

*Б.А.Чумаченко, Е.П.Власов,
В.В.Марченко*

СИСТЕМНЫЙ АНАЛИЗ

при
геологической
оценке
перспектив
рудоносности
территорий

Б. А. Чумаченко, Е. П. Власов,
В. В. Марченко

3359

Системный анализ
при геологической
оценке перспектив
рудноносности
территорий



МОСКВА «НЕДРА» 1980



УДК 553.001.2 : 519.2

Чумаченко Б. А., Власов Е. П., Марченко В. В. Системный анализ при геологической оценке перспектив рудоносности территорий. М., Недра, 1980, 246 с.

Приведен системный анализ процесса поисково-оценочных работ, дана математическая постановка задачи оптимизации управления этим процессом. Рассмотрены современное состояние разработок по созданию человеко-машинных систем геологического прогнозирования и математические методы решения задач выделения информативных признаков и классификации геологических объектов. Предложена структура и методы организации территориальных банков данных. Описана разработанная человеко-машинная информационно-прогнозирующая система АИПС/ЕС «Регион» и показаны некоторые результаты ее практического применения при оценке перспектив рудоносности территорий.

Книга предназначена для геологов, геофизиков, геохимиков, специалистов производственных и научно-исследовательских организаций, сотрудников вычислительных центров, занятых разработкой и использованием современных средств и методов обработки информации при геологических исследованиях, управлением поисково-разведочными работами и решением прогнозно-оценочных задач. Она также представляет интерес для специалистов различных отраслей, занимающихся комплексной интерпретацией и анализом данных наземных, аэро- и космических съемок.

Табл. 25, ил. 27, список лит. — 165 назв.

Рецензент кандидат геолого-минералогических наук
А. В. Канцель

ПРЕДИСЛОВИЕ

В последние годы наблюдается резкий рост спроса на различные виды минерального сырья. В соответствии с этим в сферу поисков и разведки вовлекаются новые площади и территории. Арсенал поисковых методов непрерывно обогащается, увеличивается глубина поисковых работ, повышается комплексность исследований полезных ископаемых. Начали использоваться в практике поисков материалы космических исследований. Все это приводит к резкому возрастанию объема геологической и другой информации. Одновременно растет и «цена» каждой ошибки в определении перспективности того или иного геологического объекта (рудопроявления, месторождения и т. п.). В связи с этим весьма актуальным является повышение эффективности геологических исследований на основе применения современных математических методов и вычислительных средств обработки разнородной геологической информации для увеличения оперативности и надежности прогнозных заключений.

К настоящему времени в отечественной и зарубежной литературе появилось много работ, посвященных применению современных методов обработки информации и ЭВМ для решения задач прикладной геологии.

Характерной особенностью настоящего этапа внедрения ЭВМ в практику работ является использование системного подхода к решению проблемы повышения эффективности геологических исследований. Такой подход является перспективным и открывает новые пути как для решения практических задач, так и для геологических научных исследований.

Новейшими разработками в этом плане являются интегрированные человеко-машинные системы, сочетающие логику мышления, теоретические знания, практический опыт и интуицию специалистов с большими возможностями ЭВМ по хранению, накоплению и оперативной обработке информации и современных математических методов.

С точки зрения системного подхода поисково-разведочные работы включают в себя многоступенчатую систему последовательных операций сбора и переработки разнородной геологоструктурной, минералогическо-геохимической, геофизической и другой информации, в процессе которых формируются геологические прогнозные заключения: на первых стадиях о перспективности или бесперспективности (с точки зрения потенциальной рудоносности) отдельных площадей; на последующих — о промышленной ценности рудопоявлений и, наконец, на заключительных стадиях о масштабах рудных объектов и их экономико-промышленных характеристиках.

При этом эффективность поисково-разведочных работ существенно зависит от качества процессов накопления, хранения и обработки собираемой информации и формирования на ее основе прогнозных заключений. Поэтому проблеме накопления, хранения и обработки поисково-разведочных данных и формированию прогнозных заключений в человеко-машинных системах посвящена данная работа.

Создание человеко-машинных систем, сочетающих опыт и теоретические знания специалистов с возможностями ЭВМ, определяет качественно новый этап развития геологопоисковых работ. Этот этап характеризуется формированием и многоцелевым использованием интегрированных банков данных, оперативным ретроспективным анализом накапливаемой информации, комплексным исследовательским прогнозированием в зависимости от различных факторов и геологических гипотез, решением сложных многопараметрических информационно-логических задач управления и оптимизации поисково-разведочных работ.

В настоящей работе рассматривается одно из перспективных направлений создания человеко-машинных систем, связанное с комплексной обработкой картографической информации, прогнозированием и оценкой перспектив рудоносности территорий.

В главе I проводится системный анализ геологопоискового процесса, выделяется важнейшее звено этого процесса — геологическое прогнозирование. Показывается, что геологическое прогнозирование на различных этапах поисков и разведки месторождений методологически остается неизменным, меняется лишь используемая исходная информация. Подчеркивается специфика формирования прогнозных заключений на любой стадии поисково-разведочных работ — недостаточность исходных данных и, как следствие, необходимость применения специальных математических методов, повышающих надежность принимаемых решений.

На основе аналитического обзора отечественных и зарубежных публикаций (глава II) дается обоснование того, что одним из возможных путей повышения эффективности принятия прогнозных решений является разработка и использование интегрированных человеко-машинных систем. Предлагается и описывается структура таких систем, включающая:

- блок методов принятия оптимальных решений при неполной информации;
- машинный банк данных;
- блок диалоговых процедур «специалист — ЭВМ» и технологических модулей обработки информации по заданной схеме.

Математические методы выделения информативных признаков и принятия прогнозных решений при неполной информации приводятся в главе III, в которой рассматривается ряд конкретных рекомендаций и оригинальных алгоритмов, учитывающих специфику задач классификации геологических объектов.

Современное состояние информационного обеспечения задач прогноза и оценки геологических ситуаций на основе ЭВМ описывается в главе IV. Показывается необходимость коренного улучшения информационного обеспечения, перехода от позадачного подхода к сбору и накоплению информации к интегрированному ее ведению и многоцелевому использованию на основе машинных банков данных. Описываются возможные типы машинных банков геологической информации, а также предлагается новая система управления базой данных универсального назначения. Сформулированы принципы создания территориальных машинных банков картографической информации, при этом геологическая карта рассматривается как основополагающая исходная документация.

Методика и технология решения задач при формировании прогнозных заключений приводятся в главе V, где на примере разработанной оригинальной человеко-машинной информационно-прогнозирующей системы АИПС/ЕС «Регион» рассматриваются различные аспекты взаимодействия специалиста и ЭВМ при создании территориального банка данных, комплексной обработке информации и оценке геологических ситуаций.

Некоторые результаты использования разработанной системы АИПС/ЕС «Регион» для мелко- и среднемасштабного прогнозирования приводятся в главе VI. При этом отмечается, что разработанная система ориентирована на решение задач исследования геологических гипотез и оценки перспектив рудоносности территорий на любой вид минерального сырья по картографической информации, включая данные космических съемок.

В главе VII рассматриваются перспективы развития человеко-машинных систем с позиций дальнейшего расширения и сопряжения функций специалиста и ЭВМ в едином комплексе работ по сбору, накоплению, обработке и использованию поисково-разведочных данных.

Б. А. Чумаченко написаны «Предисловие», главы I и VII, разделы III.5, IV.2, V.2; Е. П. Власовым — глава III (за исключением раздела III.5), разделы IV.1 и IV.4; В. В. Марченко — главы II и VI, разделы IV.3, V.1 и V.3. В написании отдельных разделов принимали участие: В. А. Яковлев (V.2), В. Г. Филоян и В. А. Ковалев (IV.4), Э. А. Немировский (III.9).

В разработке методов, моделей, алгоритмов решения задач и практическом опробовании системы принимали участие: В. А. Яковлев, Э. А. Немировский, Л. С. Лукьянов, М. Л. Чернов, Е. Е. Разин, Л. В. Головин, В. И. Котеликов, Г. Н. Кричевец, Б. Н. Павлов, В. П. Долгов, В. Н. Грязнов, А. А. Сапунков, Л. В. Оганесян и Г. Г. Солопов.

Авторы считают своим долгом выразить глубокую благодарность А. Н. Еремееву, В. Ф. Семенову, А. А. Дерягину, Л. П. Зеликовскому, полезные советы и предложения которых способствовали отработке основной концепции создания системы АИПС/ЕС «Регион».

Авторы признательны В. В. Попову, И. Д. Вороне, Н. П. Лаверову, М. В. Шумилину, Л. Ч. Пухальскому, В. М. Волкову и В. Н. Брюханову за помощь и поддержку в проведении исследований. Особую благодарность авторы выражают директору Международного научно-исследовательского института проблем управления члену-корреспонденту АН СССР С. В. Емельянову за ценные рекомендации по дальнейшему развитию и совершенствованию человеко-машинных систем регионального геопргноза с использованием данных космических и наземных съемок.

Авторы с благодарностью примут все отзывы и замечания читателей по книге.

I. СИСТЕМНЫЙ АНАЛИЗ ГЕОЛОГОПОИСКОВЫХ ИССЛЕДОВАНИЙ

В последние годы поисково-разведочные работы ориентируются в основном на выявление перекрытых и слепых месторождений. Это повлекло за собой развитие новых методов поисков, увеличение глубины исследований и объема собираемой информации и, как следствие, удорожание поисково-разведочного процесса. Соответственно возросла и «цена» каждой ошибки в определении перспективности того или иного месторождения. В этой связи все большее значение приобретает разработка методов повышения экономической эффективности геологоразведочных работ и, в частности, применение для этой цели математических методов и электронно-вычислительной техники.

К настоящему времени накоплен значительный опыт использования математических методов и ЭВМ при решении задач прикладной геологии, оптимизации отдельных элементов геологоразведочного процесса. Так, например, ЭВМ используются сейчас для хранения и математико-статистической обработки данных полевых наблюдений; широко применяются математические методы при выборе оптимальной плотности разведочных сетей и подсчете запасов; алгоритмы распознавания образов и методы теории статистических решений используются при решении задач прогнозирования и оценки геологических ситуаций и т. д.

Все это создает объективные предпосылки для разработки современных систем обработки информации, предназначенных для широкого использования в практической геологии. Это должна быть интегрированная человеко-машинная система, позволяющая автоматизировать основные процессы выработки и обоснования вариантов решений о направлении геологических исследований на различных этапах геологоразведочного процесса, что определяет конечную эффективность геологоразведочной деятельности в целом. Такая система должна сочетать логику мышления, опыт и интуицию специалистов с большими расчетными возможностями ЭВМ.

Построение подобной системы представляется возможным лишь на базе системного подхода. Поэтому необходимым условием оптимизации процесса геологоразведочных работ и повышения его эффективности является анализ структуры процесса, разработка его системной модели и выявление основных звеньев, определяющих его эффективность.

1. СХЕМЫ ЭТАПНОСТИ ПРОЦЕССА ГЕОЛОГОПОИСКОВЫХ И РАЗВЕДОЧНЫХ РАБОТ

В практике прогнозирования и оценки перспектив рудоносности отдельных территорий принята определенная этапность работ, заключающаяся в последовательной детализации исследований и уменьшений размеров исследуемых участков. Принципы такого построения геологоразведочного процесса были описаны и обоснованы в фундаментальном труде В. М. Крейтера [54]. В общем виде содержание отдельных этапов и стадий изложено в «Методических указаниях по проведению отдельных этапов геологоразведочных работ» [70].

Применительно к специфике поисков урановых месторождений схема этапности геологопоисковых и разведочных работ впервые опубликована Д. Я. Суражским [118]. Согласно этой схеме, весь процесс геологопоисковых и разведочных работ разделяется на пять этапов. На первом этапе («региональные металлогенические исследования») объектом оценки является металлогеническая провинция в целом. Второй этап, определяемый как «собственно поисковые работы», имеет своей задачей оценку потенциально рудоносного района. Основное назначение третьего этапа («предварительная разведка») состоит в оценке перспективности рудного поля в целом. Четвертый этап («детальная разведка») предназначается для оценки рудного месторождения. Наконец, пятый этап («эксплуатационная разведка») является завершающим и должен дать оценку промышленной ценности отдельного эксплуатационного блока (рудного тела). В данной схеме этапности рудный район рассматривается как система рудных полей, а каждое рудное поле — как система месторождений, состоящих, в свою очередь, из рудных тел.

Другим примером анализа этапности геологоразведочных работ является схема, разработанная на производственном предприятии. Первый этап по этой схеме заключается в проведении рекогносцировочно-тематических работ («выбор перспективных площадей на основе прогноза в масштабах 1 : 100 000—1 : 200 000»). На втором этапе («поисково-съёмочном») проводится комплекс геолого-геофизических и геохимических поисковых работ в масштабах 1 : 25 000—1 : 50 000. Третий этап («поисково-оценочные работы и разбраковка аномалий») состоит в проведении комплекса геолого-геофизических и геохимических работ в масштабе 1 : 10 000. Наконец, четвертый этап («разведочные работы») заключается в оценке масштабов оруденения комплексом геолого-геофизических методов.

Применительно к глубинным поискам погребенных месторождений урана аналогичная методика была разработана А. Н. Еремеевым [38]. Как отмечает автор, ...«основное ее направление состоит в увеличении детальности работ при одновременном сокращении количества площадей, на которых ведутся геологопоисковые работы».

Первым этапом указанной методики является «геологопрогнозный» этап, заключающийся в прогнозировании на основе информации съёмки масштаба 1 : 200 000 рудных полей и в выделении площадей, перспективных для проведения в их пределах следующего, «геологосъёмочного» этапа исследования. На основе полученной в результате съёмочных работ масштаба 1 : 50 000 геологической и другой информации определяются участки, перспективные для проведения работ третьего, «геологопоискового» этапа. Основным его содержанием является проведение глубинных радиометрических поисков в масштабе 1 : 10 000, обработка полученной информации и на ее основе прогнозирование месторождений (выделение перспективных аномалий и возможных рудоносных структур для проведения работ четвертого этапа). Заключительный четвертый, «оценочный» этап включает в себя детальные работы масштабов 1 : 1 000—1 : 2 000, на основе которых дается оценка перспектив отдельного месторождения.

Рассмотренные схемы этапности, отражающие основную структуру поисково-разведочных работ на уран, целесообразно сопоставить с общими методиками организации геологоразведочного процесса. В 1972 г. были опубликованы две детально разработанные схемы этапности общего процесса геологопоисковых и разведочных работ [9, 46], в которых наиболее полно рассматриваются вопросы содержания отдельных

этапов и стадий с целью повышения эффективности геологоразведочного процесса в целом. Так, в статье В. И. Бирюкова, Б. М. Косова и Н. А. Хрушова [9] отмечается: «Одним из коренных условий повышения эффективности методики поисков и разведок месторождений полезных ископаемых является выполнение их в целесообразной последовательности. Рациональное разделение процесса геологоразведочных работ на стадии позволяет исключить излишние затраты на исследования непромышленных объектов или преждевременные затраты на изучение районов, рудных полей и месторождений, освоение которых относится к далекому будущему». В цитируемой статье предлагаются разработанные в ВИЭМСе исходные данные по определению стадийности геологоразведочных работ в качестве основы для составления нового проекта методических указаний по проведению отдельных этапов геологоразведочных работ. В общем геологоразведочном процессе авторы выделяют три крупных этапа: региональная геологическая съемка, поиски месторождений и разведка месторождений. Первые два этапа разбиваются на две стадии каждый, а третий — на четыре стадии.

Первый этап: 1) познание общего геологического строения крупных регионов; 2) составление прогнозных карт с выделением площадей для более детального геологического картирования и поисков.

Второй этап: 1) выделение перспективных участков с оценкой прогнозных запасов II и III групп; 2) выделение месторождений, заслуживающих предварительной разведки. Подсчет запасов по категории C_2 и прогнозных I группы.

Третий этап: 1) расчет временных кондиций и предварительная геолого-экономическая оценка месторождения с подсчетом запасов по категориям C_1 и C_2 . Составление ТЭД и вывод о промышленном значении месторождения; 2) подсчет запасов по категориям А+В. Материалы должны быть достаточны для проектирования строительства горнодобывающего предприятия; 3) изучение слабо освещенных участков месторождения, перевод из низших категорий запасов в высшие, выявление новых тел; 4) подсчет запасов, готовых к выемке, уточнение направления выработок, определение потерь и разубоживания.

Несколько отличающаяся от рассмотренной схемы этапность предлагается в работе А. Б. Каждана [46]. Автором делается попытка четко определить целевое назначение стадий геологопрогнозного этапа путем использования категории прогнозных запасов «Д» (D_3 — рудоносные провинции и районы, D_2 — рудные поля и узлы, D_1 — глубокие горизонты и фланги месторождений). Кроме того, в отдельный этап выделяются разведочные работы на эксплуатируемых месторождениях. Всего в общем геологоразведочном процессе автор выделяет четыре этапа: «геологопрогнозный», «поисковый», «разведочный» и «геолого-разведочные работы в условиях действующего предприятия». При этом первые три этапа разделяются на две стадии каждый. Основные задачи отдельных стадий этапов определяются автором следующим образом.

Первый этап: 1) прогноз возможных масштабов проявления полезных ископаемых по всему району без их количественной оценки; 2) оценка прогнозных запасов ведущих полезных ископаемых для всей территории в целом (категория D_3), выделение локальных площадей для производства поисковых работ.

Второй этап: 1) обнаружение участков повышенной концентрации полезных ископаемых и оценка суммарных прогнозных запасов по всем перспективным участкам (категория D_2); 2) выявление объектов,

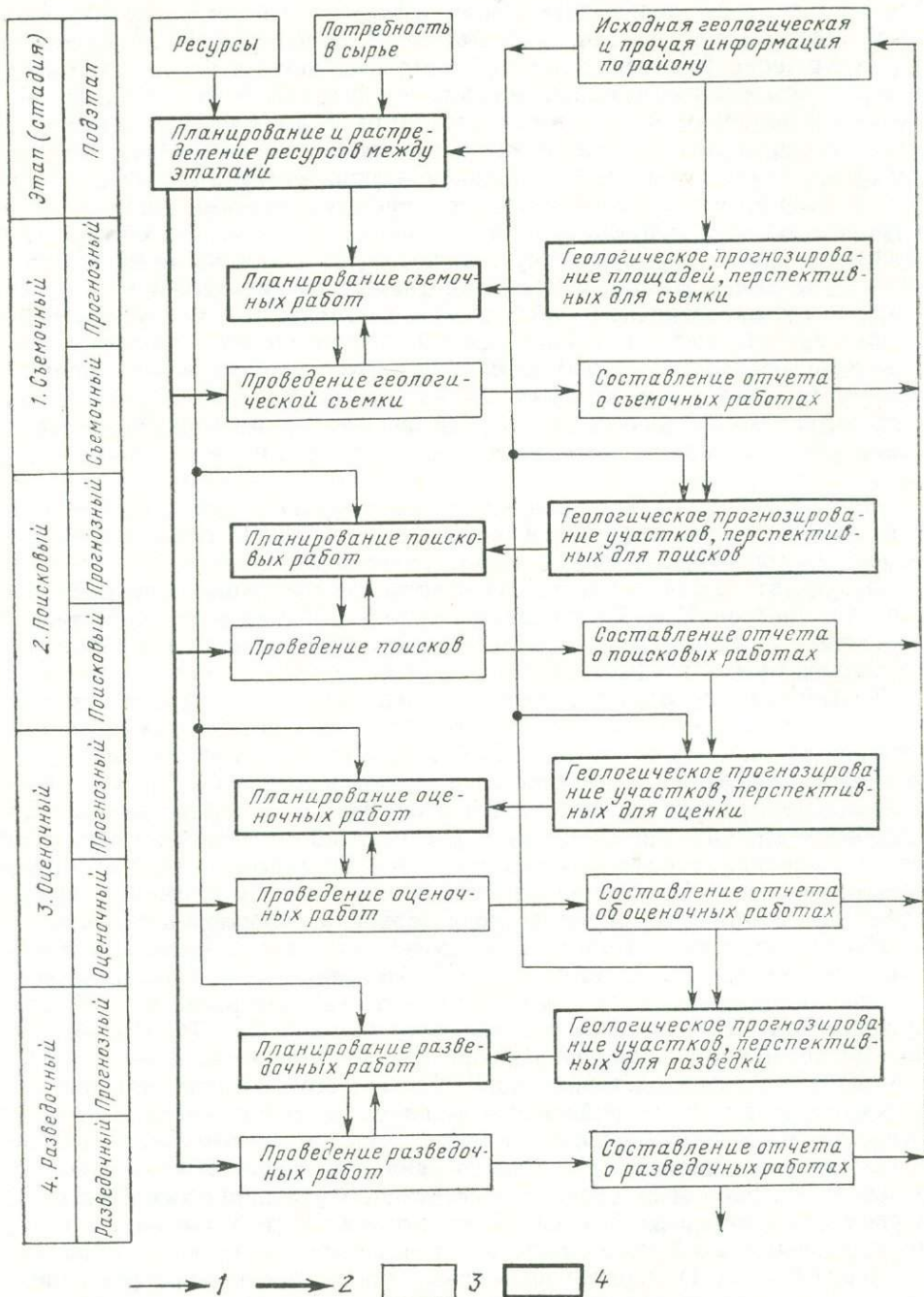


Рис. 1. Системная модель планирования и управления геологоразведочным процессом. 1 — информационные потоки; 2 — материальные, денежные и прочие ресурсы; 3 — процесс; 4 — процедуры принятия решений

заслуживающих разведочных работ, и оценка цифр прогнозных запасов участка по категории D_1 (на отдельных участках — C_2).

Третий этап: 1) подсчет запасов по категории C_2 (на отдельных участках — C_1), составление временных кондиций и ТЭД; 2) подсчет запасов по категориям А, В, C_1 ; составление проекта строительства горного предприятия.

Четвертый этап. Уточнение сведений о запасах для их планомерного вскрытия, подготовки и выемки, расширение минерально-сырьевой базы действующего предприятия.

Рассмотренные схемы этапности несколько отличаются одна от другой. Определенные различия имеются как в формулировках содержания и задач отдельных этапов и стадий, так и в числе последних.

Очевидно, среди исследователей нет единства в определении завершающих стадий геологоразведочного процесса. Так, например, пятый этап, «эксплуатационная разведка» в схеме Д. Я. Суражского, соответствует двум стадиям схемы В. И. Бирюкова и др. — «промышленной разведке» и «эксплуатационной разведке».

В то же время можно отметить единый подход исследователей к определению начальных этапов геологопоисковых работ: этапы (стадии), соответствующие по задачам «геологопрогнозному» и «геологосъемочному» этапам схемы А. Н. Еремеева и др., имеются во всех схемах.

Таким образом, анализ приведенных схем этапности показывает, что хотя особенности поисков конкретного вида минерального сырья или другие причины могут влиять на структуру геологоразведочного процесса, но общие принципы его построения остаются неизменными. Разработка методики оптимального управления, пригодной для оптимизации всего цикла геологопоисковых и разведочных работ в любом районе, должна основываться на этих принципах.

С общей системной точки зрения весь процесс геологопоисковых и разведочных работ в каком-либо районе представляет собой последовательность взаимосвязанных этапов, имеющих сходное внутреннее содержание. Действительно, на каждом этапе осуществляются, в принципе, два процесса. Во-первых, на основе анализа имеющейся геологической, геофизической, геохимической, экономической и другой информации составляется обоснованное заключение о перспективности тех или иных районов (участков, месторождений, рудных тел) и принимается решение о проведении дальнейших исследований. Во-вторых, на выделенных площадях проводятся полевые работы и осуществляется сбор новой, более детальной информации.

С учетом всего изложенного в настоящем разделе в качестве исходной для построения системной модели геологоразведочного процесса принята схема этапности, изображенная на рис. 1. Основным ее отличием от рассмотренных выше схем является то, что в каждом этапе выделено два подэтапа, соответствующих выполнению двух различных по характеру процессов: аналитического (обработка и анализ информации, принятие решений о проведении работ) и производственного (проведение полевых работ).

2. ПОСТРОЕНИЕ СИСТЕМНОЙ МОДЕЛИ УПРАВЛЕНИЯ ГЕОЛОГОРАЗВЕДОЧНЫМ ПРОЦЕССОМ

В настоящем разделе геологоразведочный процесс и управление им анализируются на основе системного подхода [86]. При таком подходе исследуемый объект или процесс (система) представляется как единая

взаимосвязанная совокупность составляющих его звеньев (подсистем). Каждая такая подсистема рассматривается как определенный процесс преобразования входных данных в выходные, а внутренняя структура элементарной подсистемы не раскрывается. Такой способ построения моделей позволяет за счет разумного выбора подсистем учитывать при анализе реальных объектов все их существенные характеристики и взаимосвязи без привлечения ненужных деталей.

Для создания оптимальной системы управления геологопоисковыми и разведочными работами необходимо построение модели, достаточно полно отражающей структуру процесса поисково-разведочных работ и управления им. В то же время разработка совершенно точной, детальной модели, исчерпывающе описывающей моделируемый процесс, привела бы к значительным трудностям ее анализа. Таким образом, возникает задача создания системной модели, с одной стороны, учитывающей многочисленные особенности геологоразведочного процесса и, с другой — достаточно простой для анализа.

Решением этой задачи является иерархическое описание процесса [68]. Весь процесс задается совокупностью моделей, описывающих его с точки зрения различных степеней детальности. При этом звено, которое являлось подсистемой в модели более высокого уровня общности, оказывается системой в более детальной модели.

Рассмотренные в предыдущем разделе схемы этапности с точки зрения иерархического подхода можно в некотором смысле считать системными моделями геологоразведочного процесса наиболее высокого уровня общности. Подсистемами в этих моделях являются отдельные этапы (стадии) поисково-разведочных работ.

С позиции теории оптимального управления геологоразведочный процесс впервые рассмотрен в работе [91], в которой предложена системная модель планирования и управления всеми стадиями геологоразведочных работ как единым целостным комплексом взаимосвязанных операций.

Принципиальной особенностью как отдельных этапов, так и всей системы поисково-разведочных работ является последовательность в проведении операций сбора, переработки и оценки информации. Решения и рекомендации, принимаемые на каждом этапе поисково-разведочных работ (на основе оценки накопленной и уже имеющейся информации), можно рассматривать в качестве входной управляющей информации для проведения работ на следующих этапах. В то же время поисковые и оценочные критерии, установленные в процессе более детальных исследований, могут быть использованы, и часто используются, в практике для переоценки рекомендаций, сформированных на ранних этапах поисково-разведочных работ. Наличие подобных контуров обратной связи в итерационном процессе поисково-разведочных работ характеризует систему управления данным процессом как достаточно сложную кибернетическую систему.

В настоящем разделе предлагается более полная системная модель управления геологоразведочными работами, позволяющая сформулировать задачу оптимального управления геологоразведочным процессом и выделить основные звенья (подсистемы) управления. В главе II одна из этих подсистем (геологическое прогнозирование) будет являться предметом более подробного анализа.

Предлагаемая системная модель изображена на рис. 1. Модель охватывает четыре этапа поисково-разведочных работ (съемочный, поисковый, оценочный и разведочный), проводимых в пределах некоторо-

го региона. Аналогичные модели могут быть построены в зависимости от необходимости для большего или меньшего числа этапов; внутренняя структура этапов от этого, очевидно, не изменится. Общим входом модели являются: потребность в сырье, денежные, материальные и людские ресурсы, а также геологическая, геофизическая, геохимическая, экономическая и другая информация, полученная в результате проведения региональных исследований и из других источников. Выходом модели являются материалы разведочных работ в виде отчета, содержащего данные о количестве разведанных запасов, утверждаемого ТКЗ или ГКЗ.

На каждом из четырех этапов в отдельные подсистемы выделены четыре различные процедуры, две из которых связаны с принятием решений. Рассмотрим последовательность работ, например, на втором поисковом этапе.

Этап начинается с выполнения процедуры «геологического прогнозирования участков, перспективных для поисков». Соответствующая подсистема на рис. 1 имеет своим входом геологическую и прочую информацию об оцениваемых площадях, полученную в результате проведения съемочных работ предыдущего этапа, а также всю имеющуюся информацию по изучаемому району. Выходом подсистемы является заключение о перспективности исследуемых участков для постановки на них поисковых работ (прогнозная карта).

Следующим звеном является принятие решения о том, на каких именно из выделенных перспективных участков следует ставить в планируемом периоде поисковые работы, и выбор методов опискования выбранных участков («планирование поисковых работ»).

Входом соответствующей подсистемы является заключение о перспективности участков (выход предыдущей подсистемы) и информация о выделенных для проведения поисков ресурсах. Выход этой подсистемы совместно с материальными, денежными и другими ресурсами поступает на вход следующей подсистемы «проведения поисков». Последним звеном рассматриваемого этапа является обработка результатов полевых работ, составление геологических карт и отчетов, которые будут являться основой для оценки на следующем этапе перспективности опискованных участков. Здесь следует указать, что между подсистемами «планирования» и «проведения работ» существует тесная обратная связь (это положение действительно для всех этапов, как это и показано на рис. 1). В самом деле, любой новый положительный результат, полученный в процессе проведения работ, может привести к перераспределению имеющихся технических средств, сил и других возможностей. Этим самым подчеркивается адаптируемость общей системы геологоразведочного процесса к изменяющейся внешней ситуации. Как справедливо отмечается в работе [72]: «Достижение конечных целей геологических исследований осуществляется путем сочетания и чередования технических и логических операций».

На остальных трех этапах осуществляются процедуры, аналогичные по содержанию описанным.

Координация работ всех четырех этапов обеспечивается подсистемой «планирования и распределения ресурсов между этапами». Эта подсистема не относится ни к одному из этапов. Ее входом являются: потребность в сырье, ресурсы, выделенные на проведение всего комплекса поисково-разведочных работ в данном районе в течение планируемого периода, вся геологическая, геофизическая, геохимическая, экономическая и прочая информация, имеющаяся по исследуемому району.

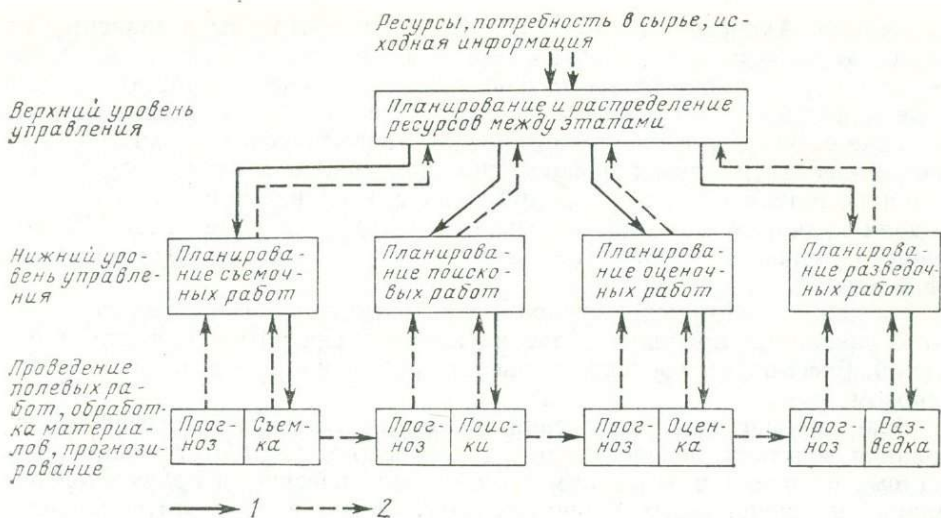


Рис. 2. Двухуровневая структура процесса управления поисково-разведочными работами.

1 — управляющее воздействие, 2 — информационные потоки

Кроме того, на ее вход поступает в режиме «обратной связи» информация с подсистем планирования съемочных поисковых, оценочных и разведочных работ — возможные варианты проведения исследований в планируемом периоде на каждом из этапов.

На основании всей имеющейся информации в рассматриваемой подсистеме осуществляется выбор наилучшего варианта планирования всего комплекса работ и производится соответствующее распределение ресурсов между этапами.

Таким образом, в построенной системной модели геологоразведочного процесса выделяется пять типов процедур, четыре из которых осуществляются на каждом этапе и одна — общая для всей системы.

Процедуры проведения полевых исследований (подсистемы «проведение геологической съемки», «проведение поисков», «проведение оценочных работ» и «проведение разведочных работ»), а также процесс обработки результатов этих исследований, составления геологических карт и отчетов на каждом этапе (подсистемы «составление отчета о съемочных, поисковых, оценочных, разведочных работах») не связаны непосредственно с принятием решений и планированием геологоразведочного процесса, а обеспечивают получение информации, необходимой для планирования, и сами являются объектом планирования.

В отличие от них подсистемы «геологического прогнозирования» на каждом этапе являются решающими подсистемами. В них формируются решения о перспективности или неперспективности тех или иных площадей или участков, и на этих решениях основывается процесс планирования поисково-разведочных работ. В самом деле, если, например, анализируя имеющуюся информацию, геолог даст ошибочную оценку перспектив рудоносности какого-либо участка и на нем будет запроектировано и запланировано проведение комплекса работ, то самые качественные полевые исследования позволят лишь установить ошибку в принятом ранее решении и не дадут эффективной отдачи (если не считать, что отрицательный результат — тоже результат).

Собственно процесс планирования поисково-разведочных работ имеет двухуровневую структуру. Верхним уровнем является подсистема «планирование и распределение ресурсов между этапами», осуществляющая управление геологоразведочным процессом в целом. Нижний уровень образован подсистемами: «планирование съемочных работ», «планирование поисковых работ», «планирование оценочных работ», «планирование разведочных работ». Блок-схема такой структуры управления показана на рис. 2.

3. К ПОСТАНОВКЕ ЗАДАЧИ ОПТИМИЗАЦИИ УПРАВЛЕНИЯ ПОИСКОВО-РАЗВЕДОЧНЫМИ РАБОТАМИ

В предыдущем разделе на основе системного анализа геологоразведочного процесса было выделено 3 типа процедур (по две на каждом этапе и одна общая для всей системы), непосредственно связанных с управлением геологоразведочным процессом (см. рис. 1): «планирование и распределение ресурсов между этапами», «геологическое прогнозирование» и «планирование съемочных (поисковых, оценочных, разведочных) работ». Наиболее важной из этих процедур, с точки зрения влияния на эффективность геологоразведочного процесса в целом, является «геологическое прогнозирование» на каждом из этапов, т. е. оценка по данным полевых исследований перспективности того или иного района, участка, месторождения. Действительно, собственно процесс управления поисково-разведочными работами («планирование и распределение ресурсов между этапами» и «планирование работ») основывается на оценке перспективности исследованных площадей, и от правильности такой оценки зависят как размеры непроизводительных затрат на изучение непромышленных участков, так и вероятность пропуска промышленных месторождений.

В силу изложенных причин большое внимание в настоящей работе было уделено вопросам оптимизации процедур прогнозирования на основе применения современных математических методов и ЭВМ.

В последующих главах процедура прогнозирования рассматривается более детально, где предлагается методика построения человеко-машинной автоматизированной информационно-прогнозирующей системы (АИПС) и описывается конкретный пример реализации такой системы для решения задач геологического прогнозирования.

Реализация информационно-прогнозирующей системы является первым шагом на пути построения автоматизированной системы управления геологоразведочным процессом. Дальнейшие шаги должны заключаться в создании методики комплексного решения задач оптимального распределения ресурсов, выбора участков и методов их исследования.

В настоящее время такой комплексной методики не существует, хотя каждая из названных задач в отдельности решалась различными исследователями. Так, например, многие авторы занимались определением оптимальной плотности поисковых и разведочных сетей, появлялись в печати работы, посвященные обоснованному распределению затрат на поисково-разведочные работы по этапам; имеется обширная литература по оптимальному в том или ином смысле (с точки зрения максимизации прибыли, минимизации «среднего риска», или вероятности пропуска месторождения и т. п.) выбору участков для постановки работ.

Всякая задача оптимизации, вообще говоря, задается множеством возможных состояний оптимизируемого процесса и определенной на этом множестве целевой функцией. Таким образом, разработка методики оптимального планирования поисково-разведочных работ должна включать в себя формальное описание множества возможных (допустимых) планов работ в данном районе, выбор критерия, характеризующего эффективность каждого конкретного плана, и разработку математического аппарата, обеспечивающего решение поставленной задачи. В настоящем разделе предлагается способ описания процесса планирования поисково-разведочных работ, дающий возможность строгой постановки задачи планирования.

Опишем формально ситуацию, возникающую при планировании. Для простоты будем предполагать, что планирование проводится на год, хотя аналогичное описание возможно и при перспективном планировании. Предположим, что в некотором районе S одновременно проводятся работы всех 4 этапов (согласно принятой здесь схеме этапности, см. рис. 1). Пусть к рассматриваемому моменту в результате проведенных исследований выделены площади S_l , перспективные для постановки работ l -го этапа ($l=1, 2, 3, 4$). Каждая из этих площадей может исследоваться различными комплексами методов M_k . Обозначим, кроме того, участки, выделенные как перспективные в результате работ 4-го, разведочного этапа S_5 . Будем считать, что каждая из площадей S_l ($l=1, 2, 3, 4$) состоит из «элементарных участков» $s_{i,l}$: $S_l = \cup V_i s_{i,l}$. Предполагается, что эти участки, выделенные из соображений геологической однородности, географических условий и пр., рассматриваются при планировании как минимальные самостоятельные единицы, т. е. работы на любом таком участке $s_{i,l}$, либо не ставятся вообще, либо исследуется весь участок одним комплексом методов.

Обозначим через $V_l(s)$ единицу количества запасов на площади S , сделанную по результатам работ l -го этапа. Будем считать, что для каждого участка $s_{i,l} \subset S_l$ имеются оценки $V_{l-1}(s_{i,l})$. На 4-м, разведочном этапе это могут быть оценки запасов по категориям C_1 и C_2 , на 3-м, оценочном — по категориям C_2 и прогнозной D_1 , на 2-м, поисковом — по категории D_2 и на 1-м, геологопрогнозом — по категории D_3 [46]. В настоящем отчете не обсуждаются способы получения оценок по прогнозной категории D . Очевидно, требуется специальная разработка таких способов и в особенности разработки методики определения точности этих оценок. Здесь необходимо лишь отметить, что для строгой постановки задачи оптимального планирования поисково-разведочных работ требуется на каждом этапе получать числовые оценки запасов полезного ископаемого (хотя, естественно, на начальных этапах эти оценки не могут быть сделаны с большой точностью).

Всякий план работ Π в данном районе состоит из выбранных участков $s_{i,l}$ для каждого l -го этапа вместе с методами $M_{k(i,l)}$ их исследования

$$\Pi = \{ \langle s_{ii}, M_{k(i,l)} \rangle s_{ii} \subset S_i^* \}, \quad l=1, 2, 3, 4. \quad (I.1)$$

Здесь символом S_i^* обозначаются площади, входящие в план работ

$$\Pi(S_i^*) \subset S_i.$$

Понятно, что не каждый выбор (I.1) участков и методов может определять реальный план. Для описания множества допустимых планов следует ввести ряд ограничений.

Один тип ограничений определяется ресурсами, выделенными на проведение комплекса поисково-разведочных работ в данном районе в планируемом периоде. Обозначим имеющееся количество ресурсов некоторого вида R , а количество ресурсов этого же вида, требующееся для исследования единицы площади участка s_{il} комплексом методов $M_k - C_{il}^k$. Тогда ограничение по этому виду ресурсов запишется в следующем виде:

$$\sum_{l=1}^4 \sum_{s_{il} \in S_l^*} C_{il}^k(i) |s_{il}| \leq R, \quad (I.2)$$

где под символом $|s_{il}|$ понимается величина площади участка s_{il} . Соотношение (I.2) должно выполняться по всем видам ресурсов, потребляемым геологоразведочным процессом.

План (I.1), удовлетворяющий ограничениям (I.2), является реальным в том смысле, что он может быть выполнен исходя из имеющихся ресурсов. Однако всякий план должен обеспечивать необходимый прирост запасов, и это положение должно быть отражено при описании множества допустимых планов. Соответствующие ограничения на площади S_l^* , входящие в план, могут быть записаны, например, в следующем виде:

$$V_3(S_4^*) + V_4(S_5) - v_{(1)} - \Delta_{(4)} = \delta_4 \geq 0,$$

$$V_{l-1}(S_l^*) + V_l(S_{l+1}) - S_{l+1}^* - \delta_{(l+1)} - v_{(l-1)} - \Delta_l = \delta_l \geq 0, \quad (I.3)$$

где $v(t)$ — необходимый по плану прирост разведанных запасов в t -ый год, начиная с момента планирования; δ_l — произвольные неотрицательные числа (соотношение (I.3) должно выполняться хотя бы при одном наборе чисел $\delta_l \geq 0$; $l = 1, 2, 3, 4$).

Смысл чисел Δ_l необходимо разъяснить особо. Рассмотрим, например, первую строку соотношения (I.3). Величина $V_3(S_4^*)$ представляет собой в общем виде ожидаемую величину запасов, которые будут разведаны на площади S_4^* . Поскольку действительное количество запасов, разведанных на этой площади после проведения работ 4-го этапа, $V_4(S_4^*)$ может оказаться меньшим, в соотношение (I.3) и введен член Δ_4 , увеличивающий надежность обеспечения необходимого прироста запасов. Вычисление значений Δ_l тесно связано с вопросом точности оценок запасов V_{l-1} и должно являться предметом специального исследования.

Таким образом, множество допустимых планов Π , выполнимых исходя из имеющихся ресурсов и обеспечивающих с заданной надежностью необходимый прирост запасов, может быть задано ограничениями (I.2) и (I.3). Понятно, что в дальнейших исследованиях эти ограничения могут быть дополнены новыми.

Гораздо более сложным вопросом по сравнению с формальным описанием множества допустимых планов является формирование целевой функции. В настоящее время, насколько известно авторам, среди исследователей нет единого мнения по поводу выбора показателя эффективности геологоразведочного процесса. В качестве такого показателя предлагаются: стоимость разведки единицы полезного ископаемого, средний риск, ожидаемая прибыль и пр. Выбор конкретного вида этого критерия опять-таки должен быть сделан в результате специального исследования. В настоящем разделе важно лишь отметить, что в принципе целевая функция любого из указанных типов может быть вычис-

лена на основе имеющейся в распоряжении специалистов информации по исследуемому району.

Таким образом, при выборе конкретного показателя эффективности геологоразведочного процесса задача оптимального планирования может быть сформулирована так: из множества планов вида (I.1), удовлетворяющих ограничениям (I.2) и (I.3), выбрать план, обеспечивающий максимум выбранному критерию. При дальнейшей разработке такая постановка может стать основой для создания методики оптимального планирования поисково-разведочных работ.

4. НЕКОТОРЫЕ МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ ПРОГНОЗИРОВАНИЯ ПОЛЕЗНЫХ ИСКОПАЕМЫХ ПРИ СОЗДАНИИ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННО-ПРОГНОЗИРУЮЩИХ СИСТЕМ

В настоящее время прогнозирование полезных ископаемых осуществляется на основе прогнозных карт различного масштаба в зависимости от детальности проводимых исследований. Любая прогнозная карта составляется в результате обобщения данных металлогенической карты [92]. Методика составления этих карт и основные требования к ним неоднократно обсуждались в печати [7, 40 и др.]. Прогнозная карта отражает степень перспективности изучаемых площадей, месторождений и рудопроявлений на основе металлогенических наблюдений и содержит основные выводы по выделению перспективных площадей и рекомендации по проведению дальнейших поисково-разведочных работ. «Металлогенические и прогнозные карты рудных районов, рудных зон и рудных узлов предназначены для выявления закономерностей размещения полезных ископаемых на их территории и служат для прогнозного выделения рудных полей, а иногда месторождений, в том числе скрытых, не выходящих на поверхность» [92].

Общепринятые методы и принципы геологического прогнозирования не позволяют избежать большого разнообразия и неоднозначности составляемых различными авторами прогнозных карт. Такое положение обусловлено следующими основными причинами: 1) огромным многообразием геологических ситуаций, объектов прогнозирования и возникающих при этом связей; 2) отсутствием четкого определения понятия «прогноз» при геологических исследованиях. Последнее является результатом особенностей естественного геологического языка, его неоднозначности и неопределенности. Четкое определение понятия «прогноз» при геологических исследованиях является наиболее важным моментом при разработке методов формализации процесса прогнозирования и однозначной интерпретации принципов составления прогнозных карт и требований к ним. Наличие таких принципов и требований в настоящее время крайне необходимо в связи с развитием работ по созданию автоматизированных информационно-прогнозирующих систем. Это обусловлено тем, что в отличие от материальных процессов, которые осуществляются в результате переноса и преобразования вещества, энергии, геологический прогноз является только информационным процессом, т. е. процессом получения, передачи, хранения*, переработки и выдачи информации.

Следовательно, управлять процессом геологического прогнозирования — означает управлять информационными процессами, позволяю-

* Хранение информации может быть рассмотрено как особый вид передачи информации, а именно ее передачи во времени.

щими анализировать существующую геологическую ситуацию, и путем синтеза разнообразных геологических факторов (в соответствии с решаемой задачей) определять вероятность обнаружения представляющего интерес рудного объекта. Это определение соответствует современному понятию геологического прогноза [64, 124 и др.].

Создание автоматизированных человеко-машинных систем геологического прогнозирования означает применение кибернетических методов в помощь специалистам-геологам, поскольку именно кибернетика занимается процессами, протекающими в системах управления.

Управление не может осуществляться без наличия критериев оценки его воздействия. Следовательно, необходимо четко определить требования к прогнозным картам и заключениям, выявить критерии, которые могут быть положены в основу эффективности управления геолого-прогнозными работами.

При формировании любого прогнозного решения геолог должен гарантировать себя от ошибок двух видов. Во-первых, он стремится не забраковать площадь, в пределах которой может иметь место прогнозируемый объект. Иначе говоря, исследователь стремится не забраковать перспективную площадь. По принятой в теории вероятностей терминологии это значит, что исследователь стремится обезопасить себя от ошибки первого рода. Во-вторых, геолог старается избежать ошибки второго рода, которая заключается в том, чтобы не выделить в качестве перспективной площадь, которая на самом деле не перспективна.

Поскольку прогноз всегда имеет вероятностный характер, то исключить ошибки первого и второго родов в принципе невозможно. А если учесть, что геологические закономерности проявляются через случайные величины, то станет очевидным, что задачи прогнозирования полезных ископаемых могут быть только вероятностными, а не детерминированными. Следовательно, необходимо найти компромиссное решение между уровнями ошибок первого и второго родов. Для нахождения такого компромисса нужно знать цену каждой из ошибок.

Цена ошибки первого рода C_1 складывается из потерь, которые возникают в результате пропуска искомого объекта вследствие ошибочной забраковки перспективной площади.

Ошибка второго рода характеризуется теми потерями C_2 , которые возникают в результате дальнейшего изучения ошибочно выделенной «перспективной» площади, которая на самом деле не имеет перспектив на обнаружение прогнозируемого объекта.

При формировании геологических прогнозов можно выделить стадии ретроспективного анализа, диагноза и собственно прогноза. Стадия ретроспективного анализа начинается с момента получения задания на прогнозирование объекта и завершается созданием гипотетической модели объекта прогнозирования и выбором метода реализации прогноза.

К диагностическим относятся задачи, решение которых преследует цель изучить причины и содержание возникаемых ситуаций. На стадии диагноза производится анализ геологической ситуации путем расчленения сложных геологических факторов на отдельные более простые единицы. В геологической практике к такому типу относятся задачи по изучению геологоструктурных условий локализации месторождений, характера геологических процессов, все задачи, связанные с выявлением генетических особенностей различных геологических образований (в том числе и месторождений, рудных тел и т. д.), закономерностей размещения полезных ископаемых и т. д. Диагностика позво-

ляет уточнить созданную при ретроспективном анализе гипотетическую модель объекта прогноза.

Стадия собственного прогноза заключается в практическом приложении выбранных методов прогнозирования. При этом определяется направление действий и условия, осуществление которых может привести к появлению прогнозируемого события. На этой стадии должны быть сформулированы критерии оценки осуществления прогнозируемого события.

Геологический прогноз можно считать прогнозом только в том случае, если в зависимости от поставленных целей предусмотрены возможные варианты действия, которые могут привести к осуществлению прогнозируемого события. Говоря более конкретно, прогнозное заключение должно содержать указания на возможные варианты и виды геологических работ (в зависимости от возможных вариантов принятых моделей объекта), осуществление которых может привести к выявлению прогнозируемого события. Такая постановка вопроса совпадает в общем виде с требованием о том, что прогнозная карта должна позволить... «определить дальнейшее направление поисковых и разведочных работ; давать возможность проведения поисково-разведочных и горно-буровых работ на конкретных точках» [92]. Прогноз — это предплановая разработка многовариантных моделей плана. План же, в отличие от прогноза, содержит однозначно определенные сроки и условия осуществления какого-либо события.

Стадии ретроспективного анализа, прогноза, планирования и оперативного управления должны находиться в постоянном взаимодействии между собой и характеризоваться наличием обратных связей.

В результате прогноза должны разрабатываться меры, обеспечивающие наиболее эффективное развитие ситуации, а также меры по предупреждению проявления нежелательных последствий. В процессе выполнения запланированных действий необходимо осуществлять сравнение возникающих ситуаций с ранее запланированными. Если возникшая ситуация не совпадает с прогнозируемой, следует возвратиться к стадии диагностики с учетом новой информации, составить новое прогнозное заключение и, как следствие, откорректировать план дальнейших действий. Этот процесс необходимо повторять до достижения желаемого результата или до получения убедительных данных о невозможности появления прогнозируемых событий.

Четкое разграничение задач различных стадий прогнозирования при геологических исследованиях необходимо по следующим соображениям. Во-первых, это позволит правильно понимать цели и задачи прогноза и разработать соответствующие методы и требования к геологическим прогнозам. Во-вторых, появится возможность разработать критерии оценки прогнозных заключений. В настоящее время многим задачам, которые по сути являются диагностическими, часто предъявляются требования прогноза и в результате делаются ошибочные заключения о «необоснованности» прогнозов, хотя на самом деле эти задачи не могли быть решены по критериям, предъявляемым к прогнозам. В итоге многие ценные диагнозы не доводятся до стадии прогноза (а тем более планов), получив отрицательную оценку при ошибочном рассмотрении их в качестве прогнозов. Из изложенного следует, что геологический прогноз, как и любой другой прогноз, является вероятностным процессом, в связи с чем необходимо иметь способ оценки вероятностей осуществления прогнозируемого события. Прогнозное заключение должно

быть получено методами, эффективность которых проверена на других объектах той же природы и сложности. Логическая строгость и непротиворечивость вывода и его соответствие основным законам природы является единственным критерием оценки правильности прогнозного заключения в случае прогнозирования событий, не имеющих аналогов среди известных и прогнозирование которых не может быть обосновано методами аналогии.

В общем виде проблема прогнозирования геологических объектов* может рассматриваться как поиск в геометрическом пространстве участков с промышленной концентрацией полезных ископаемых по совокупности геологических, геохимических и геофизических косвенных признаков рудной минерализации. Данная проблема может быть сведена к решению ряда частных задач отыскания математических моделей связи между геологическими, геохимическими и геофизическими параметрами геологических объектов и масштабностью оруденения.

Геологической предпосылкой для постановки подобных задач является представление о существовании вокруг рудных объектов анизотропного околорудного пространства, формирование отдельных свойств которого причинно связано с процессом рудообразования. Под анизотропией околорудного пространства в данном случае понимается его неоднородность, выраженная в устойчивом симметрично-асимметричном распределении (в частности, относительно рудного процесса).

Процесс рудообразования, как и любой другой геологический процесс, является многофакторным и протекает, за исключением экстремальных случаев, в условиях открытой системы, где основными управляющими моментами являются геолого-тектоническая ситуация, существующая в момент рудообразования, физико-химические условия, определяемые вмещающей средой, тип рудообразующих растворов и т. д. [21 и др.]. Реализация этих условий при становлении месторождений приводит к неоднородности околорудного и рудного пространства. Устойчивость этой неоднородности определяется сходством комплекса перечисленных выше условий, в которых протекает формирование рудных месторождений. Сходная макронеоднородность, по-видимому, сохраняется в месторождениях на уровне рудных формаций и металлогенических провинций при значительных вариациях микронеоднородностей.

Таким образом, по-видимому, можно считать, что одним из путей повышения эффективности геологических прогнозных заключений является изучение в многомерном пространстве комплексных ореолов косвенных геологических признаков на основе использования современных математических методов и ЭВМ. Такие пространственные ореолы характеризуются специфическим сочетанием косвенных геологических, геофизических и геохимических признаков; они описывают все околорудное пространство и включают в себя рудные тела месторождения. Выделение их на современном этапе представляется таким же оправданным, как выделение, например, ореолов гидротермальных изменений вмещающих горных пород. Естественно предположить, что такие ореолы косвенных признаков являются благоприятными для локализации оруденения и по существу являются ореолом месторождения в пространстве косвенных признаков, который распространяется значительно шире сосредоточения промышленных рудных тел.

* Под геологическим объектом здесь понимаются рудный район, рудный узел, рудное поле, месторождение и т. д. [130].

II. ЧЕЛОВЕКО-МАШИННЫЕ СИСТЕМЫ ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ КАК СРЕДСТВО ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ПРОГНОЗОВ

Как было показано в предыдущей главе, процедура геологического прогнозирования определяет эффективность всего комплекса геологических исследований. На любом этапе геологу приходится давать оценку перспективности изучаемых объектов по комплексу косвенной геологической информации, объем которой неуклонно возрастает с каждым годом. Поэтому возникает объективная необходимость разработки эффективного инструмента для оперативного комплексного анализа разнородной геологической информации. Это обуславливает разработки различных систем накопления и обработки информации с использованием ЭВМ. Известно, что новейшими разработками в этом плане являются интегрированные человеко-машинные системы, сочетающие логику мышления, опыт и интуицию специалистов с большими возможностями по хранению и обработке информации с помощью современных ЭВМ. Необходимым условием разработки такой системы является системный анализ традиционной схемы геологического прогнозирования.

1. СОВРЕМЕННОЕ СОСТОЯНИЕ РАЗРАБОТОК ПО СОЗДАНИЮ ЧЕЛОВЕКО-МАШИННЫХ СИСТЕМ ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

В последние 10—15 лет наблюдается широкое внедрение современных средств обработки информации и ЭВМ в практику геологических исследований; интенсивно растет количество научных публикаций по данной тематике. Анализируя эволюцию внедрения вычислительной техники в практику поисково-разведочных работ, можно отметить некоторые характерные особенности. В начале и середине 60-х годов основное внимание исследователей уделялось вопросам разработки отдельных алгоритмов, решению задач поисков статистических закономерностей; в меньшей степени рассматривались общие методологические вопросы использования математических методов и ЭВМ для анализа и переработки геологической информации, примеры решения некоторых частных задач геологического прогноза. Такой фрагментарный подход не мог оказать значительного влияния на повышение эффективности геологических исследований.

Накопленный в этот период опыт позволил выявить «узкие» места применения ЭВМ в геологической науке и практике. К ним прежде всего относилась большая трудоемкость предварительной подготовки и обработки информации, ее машинного кодирования. Затраты времени на эти операции составляли до 80—90% от всего процесса решения задачи с помощью ЭВМ. Кроме того, необходимость даже незначительного изменения или корректировки первичных данных требовала от исследователей повторения всей предварительной работы.

Другая трудность состояла в том, что информация, подготовленная для решения по конкретному алгоритму, оказывалась непригодной при использовании другого алгоритма или требовала дополнительной пере-

кодировки. После получения результатов решения использованная информация не находила дальнейшего применения. Таким образом, наблюдалось явное несоответствие между большими возможностями обработки информации с помощью ЭВМ и отставанием информационного обеспечения.

К началу 70-х годов рядом исследователей были разработаны основные принципы методики решения прогнозных задач с помощью ЭВМ [33, 36, 47, 49, 99, 126 и др.]. Стала очевидной необходимость поиска новых путей, устраняющих отмеченные выше трудности. Один из путей выхода из создавшегося положения состоял в использовании системного подхода. На основе такого подхода интенсивно разрабатывались различные системы автоматизированной обработки геологических, геофизических, геохимических, гидрогеологических данных [11, 34, 94 и др.]. Вместе с тем одна из главных трудностей широкого использования ЭВМ — несоответствие вида исходной геологической информации требованиям автоматизированной технологии обработки данных — оставалась нерешенной. В связи с этим в последние годы интенсивно развивается направление по разработке автоматизированных систем хранения, накопления и поиска информации (банков данных).

Рассмотрим некоторые разработки по обоим направлениям применительно к проблеме геологического прогнозирования. Впервые задача прогнозирования геологических объектов, как задача принятия решений в условиях неполной информации для рудных месторождений, была рассмотрена в работе Г. Н. Баласанова и др. [45]. Наиболее полно вопросы применения математических методов при прогнозировании полезных ископаемых освещены в работе А. Н. Бугайца и Л. Н. Дуденко [13]. В этой работе приводится краткое описание общей структуры автоматизированной системы прогнозирования. При этом отмечается, что общая последовательность действий состоит в постепенном переходе от более низких степеней к более высоким с использованием на каждом шаге (если это возможно) сочетания соответствующих вероятностных и логических (эвристических) алгоритмов с выработкой того или иного решающего правила. В работе достаточно подробно освещены методы описания и предварительной обработки геолого-геофизических данных, но недостаточно разработаны вопросы информационного обеспечения системы.

В. И. Ароновым, Э. Я. Невельской и В. С. Самариним [4] разработана автоматизированная система поиска и статистической обработки геологической информации. Система, реализованная на ЭВМ БЭСМ-4 и М-220, предназначена для накопления больших массивов данных на магнитных лентах, осуществления выборки из массивов согласно запросу, вычисления оценок центральных моментов, проведения многомерного корреляционного и регрессионного анализа.

Б. А. Горлицкий, Л. С. Финкель [31] приводят характеристики автоматической системы массовой обработки геохимических данных «комплекс». Основные виды операций реализованы в системе: «ввод», «описание», «сравнение», «классификация», «поиск». Система позволяет проверять геологическую однородность и «компактность образа», вычисляет необходимые статистические характеристики и классификацию объектов.

В. В. Ломтадзе [65] рассматривает автоматизированную систему обработки геофизических данных (АСОГ). В состав АСОГ входит управляющая, обслуживающие и рабочие программы. Автор обосновывает необходимость наличия в автоматизированной системе средств

для описания числовых массивов; все массивы должны иметь одинаковую структуру. Такая организация данных обеспечивает возможность стыковки программного обеспечения независимо от тематической характеристики массивов.

Некоторые результаты разработки автоматизированной информационно-поисковой системы «Инженерная геология» даны в работе И. С. Комарова, А. Р. Зайделя, Я. И. Файна и др. [74]. Объекты, описываемые системой, разделены на 3 категории: точечные, линейные и площадные. Точки наблюдения по способу привязки в пространстве разделяются на расположенные бессистемно, по линии с равным шагом, в узлах прямоугольной решетки и в узлах пространственной решетки. Каждая точка наблюдения может быть охарактеризована определенным набором как количественных, так и качественных признаков, объединяемых в вектор признаков; последовательность признаков в векторе постоянная. Из векторов одинаковой структуры формируется массив хранения; массивы разделяются также по типу объектов и способу привязки точек наблюдения. Предмашинный (промежуточный) носитель информации имеет форму таблиц; каждому типу массива соответствует свой макет таблицы. Информация вводится в виде слов (словосочетаний) или их сокращений длиной не более 12 букв и числовых обозначений. Математическое обеспечение позволяет проводить поиски и статистическую обработку информации (оценки параметров распределения, корреляционный анализ), построение карт тренда.

Л. М. Штайн [134] приводит описание диалоговой специализированной системы обработки данных. Он отмечает, что существующий в большинстве отечественных серийных ЭВМ режим пакетной обработки обладает рядом недостатков, самый существенный из которых — довольно большое время ответа. Если непосредственный контакт программиста с ЭВМ при отладке программ позволяет ускорить этот процесс, то пакетная обработка информации на основе традиционного набора алгоритмических языков не обладает достаточной гибкостью в том смысле, что не обеспечивает никаких преимуществ при решении специализированных классов задач. Описываемая система должна обеспечить решение определенного класса задач по статистической обработке информации в диалоговом режиме. Система должна также предоставить пользователю возможность на выносном терминале выполнение всех операций пакетной обработки. Язык, обслуживающий диалоговую систему обработки данных, обеспечивает программирование процедур статистической обработки информации для пользователей, не являющихся специалистами в области программирования. База данных, обслуживающая диалоговую систему обработки данных, позволяет создавать файлы с переменной иерархической структурой.

В. Н. Николенко приводит сведения об автоматизированной системе классификации и прогнозирования геологических объектов (АСКИП) [77]. Система позволяет по квадратной сети исходных полей формировать множество характеристик (признаков), проводить анализ делящих свойств полученных характеристик и с учетом их коррелируемости выбирать информативное сочетание признаков, осуществлять выбор алгоритма распознавания и выполнять диагностику объектов. Опыт эксплуатации АСКИП свидетельствует о необходимости использования более совершенного технического и математического обеспечения. Автор отмечает, что перевод системы на вычислительные машины третьего поколения позволит привлекать более сложные и надежные алгоритмы обработки данных.

Организация автоматизированной системы для количественной оценки перспектив территории при поисках месторождений полезных ископаемых приведена в работе В. В. Давиденко, В. С. Романова, Г. И. Дервянченко и Г. Г. Стихарной [90]. Эта система относится к классу систем распознавания с обучением. Система приспособлена для сбора, хранения и анализа исходных материалов, представленных в виде моделей и состоит из четырех соподчиненных блоков: управления, подготовки исходных данных и вычисления признаков, формирования решения и алгоритмов принятия решения.

В. Ф. Семеновым, А. Н. Еремеевым и др. [2] предложена автоматизированная информационно-прогнозирующая система (АИПС). Авторы, разрабатывая эту систему, исходили из того, что эффективность геологоразведочного процесса в значительной степени определяется качеством выполнения процедур прогнозирования на каждом этапе поисково-разведочных работ.

АИПС реализована как интегрированная человеко-машинная система на основе рационального сочетания возможностей ЭВМ и специалистов-геологов.

Вопросам разработки банков данных в последние годы также уделяется все возрастающее внимание. Вопросы систематизации и типизации банков данных подробно разобраны С. И. Агаджаняном [3], организации службы хранения и поиска геолого-геофизической информации посвящено исследование В. Н. Боганика и Т. Г. Шабельниковой [10]. Принципы обработки информационных массивов в автоматизированных системах управления изложены в работе В. М. Глушкова, В. П. Гладуна, Л. С. Лозинского и С. Б. Погребинского [80].

Пример практической реализации банка данных для автоматизированной системы прогнозирования на этапе мелко- и среднемасштабных геологических исследований приведен в работе Б. А. Чумаченко, Е. П. Власова, В. В. Марченко и др. [88].

При разработке банка данных авторы исходили из того, что для построения с помощью ЭВМ различных прогнозных карт необходимо обеспечить, во-первых, точную привязку информации и, во-вторых, компактную организацию в машинной памяти большого количества различной геологической информации в виде, удобном для последующей обработки. Наличие банка данных позволило осуществлять многовариантное прогнозирование рудоносности территорий на основе различных эталонных объектов обучения. Оказалось также возможным провести сравнительный анализ прогнозирования на основе различных геологических гипотез, оценить информативность геологических факторов при решении конкретных задач геологического прогноза.

Быстрыми темпами растет число публикаций по банкам данных за рубежом. М. Р. Андертон [137] предлагает гибкую систему записи информации по скважинам (банк данных). Комплекс сведений по скважине представляется в виде иерархической структуры, состоящей из 5 уровней, на которых записываются соответственно стратиграфия, литология, тип породы, состав и минерализация. Каждый из этих уровней может представлять собой набор сведений, формируемых по определенным интервалам глубин. Запись данных по скважине имеет переменную длину. Используется мнемонический четырехзначный код, структура которого отражает последовательность уровней.

П. Беккерт, Р. Вебстер, Г. Мишель [138] подробно рассматривали вопрос создания банка данных для исследования эволюции Земли.

Проблема хранения и обработки данных в геологическом картировании рассмотрена в работе Х. Бернера, Т. Экстёма и др. [140]. Разработанная авторами система «Геомап» позволяет сохранять и обрабатывать данные полевых геологических наблюдений; система рассчитана на лиц, не знакомых с программированием. Основу системы составляет комплекс программ, составленных на языке ФОРТРАН II. Текст единичного полевого описания кодируется не более чем на 10 перфокартах. Введенные данные сортируются по географическим координатам, номеру обнажения и автору описания, проверяются на наличие ошибок по нескольким логическим тестам и записываются на магнитной ленте. Для построения геологических карт используется широкоформатный графопостроитель; основные данные или их комбинации в виде цифр и знаков наносятся на прозрачную основу и сопоставляются с другими картами. Статистическая обработка предусматривает: подсчет частот с выдачей гистограмм, расчет среднего значения и среднего квадратического отклонения, одномерный дисперсионный анализ, дискриминантный, факторный, регрессионный анализы и некоторые другие операции.

В. Штайн, Хеер Ханс и др. [139] описывают используемую геологической службой Нидерландов систему автоматизированной обработки материалов геологической документации скважин. Данные из бурового журнала, содержащие информацию о глубинах залегания, литологии, ископаемых организмах, структурно-текстурных особенностях печатаются на пишущей машинке в специальном формате, пригодном для последующего чтения с помощью оптического читающего устройства. Для описаний используется специальный словарь. Технология обработки информации включает этап контроля ошибок. Система позволяет выполнять преобразование различных форматов с целью обработки геологических материалов различных стран Западной Европы (Дания, Англия, Франция).

Общие принципы автоматизированной обработки геологической информации и характеристика создаваемого банка данных в отраслевом информационном центре (геофонд, Прага) приводится в работе И. Книжковой и Х. Бужковой [158].

В ЧССР разработан и используется в практической деятельности банк геологических данных по Северо-Чешскому угольному бассейну [156]. К 1973 г. в банке сохранялись сведения по 16 тыс. скважинам. Каждая скважина содержит сведения о географической привязке, техническую информацию и геологические данные с описанием разреза. Банк данных непрерывно пополняется и применяется с целью автоматического построения разрезов по угольным пластам, подсчета запасов, построения моделей месторождений и др.

Проект системы статистической обработки данных при региональных исследованиях предложен И. Жеффруа и Т. К. Вигналлом [150]. В общем виде этот подход состоит в том, что исследуемый регион разделяется на квадратные ячейки, часть из которых (контрольные) хорошо изучена и содержит известные месторождения. Остальные ячейки являются объектами изучения. Для всех ячеек в вычислительном центре собираются все необходимые материалы, обрабатываемые в последующем с использованием моделей соответствия, тренд-анализа и характеристического анализа. В итоге обработки на ЭВМ выявляются участки, перспективные для локализации месторождений; при этом указываются вероятность обнаружения и некоторые экономические параметры.

Оценивая перспективы развития информационных систем по нефтяным и газовым скважинам, Форготсон и Стоут [148] отмечают, что к 1970 г. в США хранились в ЭВМ материалы более чем на 600 тыс. скважин. Накопленная информация позволяет осуществлять управление процессом разведки и проводить статистическую обработку данных. По каждой скважине может сохраняться до 480 параметров.

К 1972 г. в США и Канаде машинный банк охватывал уже материалы более чем по 700 тыс. скважин [149]. Обработка информации проводится на различных стадиях поисково-разведочных работ. При этом решаются, например, такие задачи, как выбор перспективных стратиграфических горизонтов, площадей под детальную разведку и т. п. Так, в районах Скалистых гор эта информация использовалась для региональной оценки продуктивности свиты Мадди (нижний мел). По районам распространения этой свиты (Вайоминг, Монтана, С. и Ю. Дакота) в результате машинной обработки были построены серии карт (структурных, мощностей, продуктивности скважин, поверхности тренда перспективности свиты). По выданным рекомендациям впоследствии был открыт ряд месторождений с суммарными запасами 33 млн. т нефти.

Интересные работы по созданию объемных моделей геологического строения района выполнены Грендером, Раппопортом и Седжерсом [154]. Территория общей площадью в 2500 км² (долина Сан-Хоакин, Калифорния) была выбрана в качестве модельного объекта для кодирования и обработки на ЭВМ больших массивов геологической информации. Исследуемый район был разбит на элементарные объемы размером по 3,2×3,2 км по площади и 150 м по глубине (всего 5976 ячеек). Геологическая информация по каждому объему содержит 60 параметров. Созданный таким образом массив, состоящий из 240 тыс. чисел, обрабатывается с помощью ЭВМ; в результате чего могут быть построены карты глубин фундамента, разрезы по произвольным направлениям, карты мощностей, изученности и т. п.

Геологической службой Канады успешно используется региональная информационно-поисковая система (проект «Кост Маунтинз») [155]. Начиная с 1962 г. все материалы поисково-картировочных работ в районе Кост Маунтинз (Британская Колумбия) по территории 110 тыс. км² заносятся с помощью специальных предмашинных носителей в ЭВМ. Информация более чем по 25 тыс. описаний обнажений горных пород подвергается различной машинной обработке (построение геологических границ и др.). Основное назначение системы — сбор и поиск всех данных, необходимых для картирования и построения предварительных геологических карт.

Как отмечает А. С. Смирнова [114, 115], по состоянию на 1973 г. в мире насчитывалось не более 23 действующих систем накопления, хранения и обработки геологических данных. Из них в США 5 систем (3 по нефти и газу), в Канаде также 5 систем (2 по нефти и газу), во Франции 4 (1 по нефти и газу), в Англии 2, в Чехословакии 2, в Австралии 2.

Вместе с тем геологическая служба Канады и Международный союз геологических наук приступили к составлению металлогенической карты Северной Америки — часть проекта по составлению металлогенической карты мира. Такие же работы проводятся во Франции. Составление металлогенических карт сопровождается накоплением информации в информационно-поисковых системах.

В настоящее время геологическая служба Канады использует массив ОПСЕП, содержащий сведения по 700 рудным месторождениям (на 1974 г.), а также массив по молибденовым месторождениям («Молифайл»), в котором приводятся сведения по 1500 объектам. В функции этой системы входит изучение размещения молибденовых месторождений в национальном масштабе; организация данных для быстрого сканирования, сортировки и индексирования; непосредственный доступ к данным для получения карт различных масштабов и составления различных таблиц, корректировка, суммирование данных и оценка ресурсов и др.

Геологической службой Канады используется также банк данных для статистического анализа [146]. Он предназначен для прогнозирования месторождений полезных ископаемых и изучения геологических процессов в региональном масштабе. Банк охватывает данные по двум крупным областям Канадского щита (Верхней и Южной) и состоит из двух массивов: геологической информации и месторождений полезных ископаемых. Массив геологической информации содержит литостратиграфические описания 8500 ячеек размером 10×10 км, что составляет 49% всей площади двух указанных областей Канадского щита. Для сбора информации было использовано 55 геологических карт масштабов 1 : 292 000 и 1 : 146 000.

Массив по месторождениям полезных ископаемых накапливался одновременно с геологическим. Он содержит данные по 492 месторождениям Верхней области и включает месторождения меди, никеля, цинка, свинца, серебра, золота. Каждое месторождение характеризуется координатами, запасами и содержанием полезного компонента по каждому металлу и др.

Эти массивы использовались для статистического изучения размещения отдельных видов полезных ископаемых в зависимости от лито-возрастных характеристик и тектоно-магматических зон. С помощью собранной информации составлены карты распространения типов пород с определенной вероятностью нахождения определенных месторождений.

Одна из последних разработок канадских геологов «Банк данных для оценки минеральных ресурсов в Аппалачском регионе» [146, 160, 164], охватывающий провинции Нью-Брансуик, Квебек, Новую Шотландию, Ньюфаундленд. В задачу банка входит количественное представление геологических параметров, которые прямо или косвенно могут контролировать размещение полезных ископаемых. Аппалачский регион отличается разнообразным геологическим строением, наличием большого числа различных типов горных пород, метаморфическими преобразованиями и интенсивной интрузивной деятельностью. Основными полезными ископаемыми этого региона являются медь, свинец, цинк.

Банк данных состоит из двух массивов: регионального геологического файла и файла месторождений полезных ископаемых. Оба массива взаимно дополняют друг друга. Региональный геологический файл рассматривается как численное выражение геологической компилятивной карты. Все характеристики, находящиеся в виде чисел в банке данных, могут быть обработаны с помощью ЭВМ и использованы для статистического анализа. ЭВМ строит всевозможные карты, на которых показана корреляция между различными геологическими параметрами. При обработке карты применяется теория множеств. Например, карты показывают вероятностные значения приуроченности отдельных полезных ископаемых к определенным типам горных пород или дру-

гим геологическим характеристикам, позволяют оценивать вероятность выявления месторождений различных типов на определенных участках региона. Сведения, содержащиеся в банке данных, все время дополняются и уточняются.

Файл месторождений полезных ископаемых содержит сведения по 180 месторождениям, 844 рудопроявлениям и 695 точкам минерализации. Для описания этих объектов используется 7 групп входных данных: идентификация, местоположение, геологическое положение (по отношению к интрузиям или несогласию), вмещающие породы, минералогия, классификация в соответствии с морфологическими и генетическими типами, запасы и содержание руды.

Для некоторых объектов кодируются расстояния от интрузивного контакта или несогласия и параметры контакта или несогласия. В записи вмещающих пород используются те же литостратиграфические термины, что и в региональном геологическом файле, снимаемые с геологической карты, но помимо них учитываются породы, непосредственно содержащие рудную массу, которые на геологической карте не всегда выделяются в самостоятельные единицы.

Минералогическая характеристика включает список из 20 минералов, присутствующих либо в руде, либо в жильной массе. Данные по запасам и содержанию (всего 12 характеристик, включая годовой прирост запасов, ежегодную добычу и др.) указываются для каждого компонента.

Одним из самых крупных автоматизированных массивов по месторождениям полезных ископаемых в мире является информационно-поисковая система МИНДЕП [161, 165]. Она содержит информацию о 8 тыс. месторождений полезных ископаемых Британской Колумбии (Канада).

Разработка этой системы была начата в 1973 г. как реализация исследовательского проекта, предназначенного для изучения месторождений полезных ископаемых на базе ЭВМ в научных и промышленных целях, а также для усовершенствования процессов ввода, редактирования, обработки и использования информации. В системе предлагалось собрать максимально возможное число сведений о месторождениях с целью применения их для решения вопросов региональной металлогении, генезиса, моделирования процессов поиска и разведки месторождений, для оценки ресурсов и других видов геологического и экономического изучения. В настоящее время система пользуется финансовой поддержкой федерального правительства, властей провинций и промышленных компаний.

В кратком обзоре невозможно дать критический анализ различных подходов к решению рассматриваемой проблемы. Для более полного ознакомления с публикациями в этой области можно адресовать читателя к работам [51, 115].

Пожалуй, одним из наиболее существенных негативных моментов рассмотренных разработок является недостаточное привлечение геологов-специалистов для постановки задач, содержательного анализа промежуточных и окончательных материалов. В некоторых случаях наблюдается даже стремление к полной автоматизации получения окончательных выводов о перспективности тех или иных геологических объектов. Однако геологическая наука оперирует с огромным количеством неформализованной информации, поэтому игнорирование опыта и теоретических знаний специалистов при решении прогнозных задач нецелесообразно. Опыт специалистов следует непременно использовать,

например, в виде экспертных оценок [85], эвристического моделирования и др., как это практически реализовано в системе АИПС [2].

В последнее время известны разработки, ориентированные в значительной степени на эвристический подход при решении задач геологического прогноза [48].

Можно в какой-то мере понять стремление отдельных специалистов (например, геофизиков, геохимиков и др.) не только обработать специальную информацию, но и дать определенные рекомендации прогнозно-геологического характера. Вместе с тем не следует забывать, что более достоверные решения могут быть получены лишь при комплексном анализе всей имеющейся геологической, геофизической, геохимической и прочей информации.

Видимо, наличие таких «частных» подходов и содержание различных узкоспециализированных систем можно объяснить тем, что отдельные разработки выполняются без общего системного анализа всего процесса поисково-разведочных работ и выделения на этой основе места и роли математических методов и ЭВМ как фактора повышения эффективности всего геологоразведочного процесса.

Наконец, практика убедительно свидетельствует о том, что самая лучшая разработка системы геологического прогнозирования не будет иметь прикладного значения без соответствующего информационного обеспечения. Отсюда со всей неизбежностью следует вывод о необходимости совмещения систем обработки информации с банками геологических данных.

Практика указывает на неперспективность подхода применительно к решению геологических задач, когда отдельно рассматриваются системы обработки данных и системы накопления информации, ее хранения и выдачи.

Стремление к объединению систем обработки и хранения информации выражено В. М. Сидоровым [112]. Он пишет: «Так сложилось в геологической практике последних пяти лет, что разработка и внедрение информационных систем (ИС) и автоматизированных систем обработки данных (АСОД) проходят, как правило, независимо друг от друга». Автор предлагает сочетать оба подхода при разработке автоматизированных информационных систем.

Аналогичный подход предложен В. А. Нагорским [73], который отмечает о целесообразности машинного формирования достаточно полного первичного признакового пространства, необходимости использования и учета априорной неформализованной информации о механизмах рудного контроля, отражающей опыт и интуицию геологов, а также о эффективном использовании графической геологической информации (карты, планы, схемы и т. п.) при решении задач геологического прогноза. Решение геологических задач в системе «человек — ЭВМ» рассмотрено в специальной работе В. М. Омелина [83].

А. Н. Бугаец и Л. Н. Дуденко в монографии [13], посвященной применению математических методов при геологическом прогнозировании, приходят также к выводу о необходимости создания класса методов «человек — машина».

Уместно заметить, что еще в 1969 г. Л. Б. Розовский [108] полагал, что «прогнозы действия современных геологических процессов на основе метода аналогии с помощью ЭВМ также будут выполняться при обязательном участии геолога-исследователя, который будет выполнять роль эксперта и критика (Н. Винер), контролирующего поступающую информацию и результаты ее обработки машиной».

Таким образом, из краткого анализа современного состояния разработок автоматизированных систем геологического прогнозирования, вытекают следующие выводы.

1. Автоматизированная система геологического прогнозирования должна представлять собой человеко-машинную систему, предназначенную для логического анализа разнородной информации. Она должна органически сочетать огромные возможности современных ЭВМ по накоплению, хранению и обработке больших массивов информации с теоретическими знаниями и практическим опытом специалистов геологов.

2. Система должна состоять, по крайней мере, из двух главных частей: территориального машинного банка данных многоцелевого назначения и блока обработки информации на основе использования различных математических и эвристических подходов, реализованных в виде наборов соответствующих алгоритмов (технологических модулей).

3. Специалисты должны, по крайней мере, выполнять в человеко-машинной системе следующие операции: формулировать постановку задачи геологического прогнозирования, давать оценку представительности исходной информации, анализировать промежуточные и окончательные результаты.

4. Система должна иметь современные устройства автоматизированного ввода-вывода информации. Реализацию системы следует осуществлять на базе ЭВМ третьего поколения (ЕС).

Разработанная с учетом изложенных выше принципов автоматизированная система геологического прогнозирования будет представлять собой современный инструмент комплексного анализа разнородной геологической информации, практическое использование которого позволит ускорить процесс получения геологических прогнозных заключений и повысить качество прогнозов.

2. СИСТЕМНЫЙ АНАЛИЗ ПРОЦЕССА ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

В главе I было показано, что основными звеньями, определяющими эффективность геологоразведочного процесса, являются процедуры геологического прогнозирования на каждом этапе поисково-разведочных работ. Рассмотрим подробнее процесс принятия специалистами геологами прогнозных решений о перспективности отдельных геологических объектов.

Геологический прогноз является реализацией определенных теоретических представлений, взглядов, гипотез на природу и процесс рудообразования. От того, насколько эти представления соответствуют реальной действительности, в конечном счете и зависит качество прогнозных оценок. В ряде случаев прогнозное построение осложняется из-за наличия нескольких, подчас взаимоисключающих взглядов на генезис. Возникает необходимость сопоставления и сравнения альтернативных генетических гипотез (с точки зрения их правильности, обоснованности и пригодности) для описания и объяснения особенностей изучаемого процесса по результатам наблюдений. Подобный анализ возможно провести методом воссоздания картины процесса рудообразования в его развитии путем конструирования моделей.

На различных этапах поисково-разведочных работ этот процесс методологически одинаков и состоит в самом общем виде из следующих стадий.

Вначале, после того как поставлена задача прогнозирования, специалисты изучают имеющуюся по данному району информацию, определяют ее достоверность и представительность, проводят отбор данных, необходимых для получения прогнозного заключения. На основе имеющейся информации специалисты восстанавливают историю геологического развития исследуемой территории, выявляют место и время возможного образования оруденения в этом процессе. Изучая металлогенические особенности района, специалисты стараются определить эмпирическую связь между отдельными геологическими факторами и масштабами оруденения и разрабатывают поисково-оценочные критерии. Прежде чем приступить к оценке какого-либо объекта или района, специалисты знакомятся с максимально возможным количеством информации по исследуемой территории, характеризующей основные черты и особенности ее геологического строения, геофизическими аномалиями и т. п.

Особое место в этом процессе занимает ознакомление с металлогенией района, с известными рудными месторождениями и рудопроявлениями; тщательно изучаются особенности расположения месторождений в геологических структурах, геофизических полях и т. д. Сравнение потенциально рудоносных площадей осуществляется в соответствующем масштабе исследований, имея при этом в виду, что прямых аналогов в природе не существует. Отсюда сравнительный анализ, например, рудоносных зон (в масштабах исследований 1 : 500 000—1 : 200 000) выполняется на основе изучения геологических и магматических формаций, вмещающих те или иные объекты, гравимагнитных аномалий, имеющих размеры сотни квадратных километров. Сравнение же площадей в пределах выделенных рудоносных зон производят по иной совокупности данных (фациальный состав толщ; гидротермально-метасоматические изменения, непосредственно предшествующие рудоотложению; первичные и вторичные ореолы рассеяния и т. д.).

Далее, используя разработанные поисково-оценочные критерии, производится содержательный геологический анализ исследуемой территории с целью выделения участков сочетания критериев, благоприятных для локализации оруденения. При этом, как правило, используется метод геологической аналогии.

На всех уровнях исследования, как правило, за основу принимается историко-геологический подход, в общем виде сформулированный Ю. А. Билибиным: «Процессы минерализации, ведущие к возникновению минеральных, в частности, рудных месторождений, представляют одну из сторон единого и сложного процесса геологического развития земной коры. В своем историческом развитии они теснейшим образом взаимосвязаны с другими сторонами того же процесса, т. е. осадконакоплением, тектоническими движениями (развитием структур), магматической деятельностью, метаморфизмом. Процессы минерализации могут и должны изучаться лишь в своем историческом развитии и в теснейшей взаимосвязи с процессом развития земной коры» [8].

В ходе решения прогнозных задач методом аналогии допускается предположение, что месторождения, относимые к одной и той же рудной формации, могут отличаться зонами в зависимости от истории геологического развития.

На заключительной стадии специалисты составляют заключение о проведении дальнейших исследований. После экспертизы и утверждения на научно-техническом совете данное заключение является основа-

нием для принятия решения о проведении полевых (съемочных, поисковых, оценочных или разведочных) работ.

Очевидно, конкретное содержание исследований на каждой из описанных стадий меняется в зависимости от того, на каком этапе поисково-разведочных работ производится прогнозирование (меняется собственно объект прогнозирования, рассматриваются различные комплексы факторов и т. д.). Тем не менее описанная общая структура процесса прогнозирования на каждом из этапов остается неизменной и может быть в самом общем виде представлена схемой, изображенной на рис. 3.

Анализ процесса геологического прогнозирования позволяет выделить в нем целый ряд процедур, выполнение которых с помощью математических методов и ЭВМ более эффективно, чем традиционными способами.

К числу таких процедур следует отнести прежде всего хранение, обновление и первичную обработку геологической, геофизической и геохимической информации, расчет различных статистических характеристик и т. п. В самом деле, арсенал методов исследования геологических объектов за последнее время многократно увеличился, и это повлекло за собой соответствующее увеличение объема информации, учитываемой при прогнозировании. Ручная обработка больших информационных массивов, выделение полезной информации и ее анализ без использования ЭВМ становятся практически невозможными.

При определении рудоконтролирующих факторов и разработке поисково-оценочных критериев традиционными способами, без применения математических методов неизбежно большую роль играют различные субъективные причины. В зависимости от геологической школы, генетического подхода, практического опыта различные специалисты могут по-разному оценивать, например, важность тех или иных геологических факторов. К тому же следует учесть, что даваемые специалистами оценки значимости факторов и поисковые критерии имеют, как правило, качественный, а не количественный вид. В связи с этим оче-

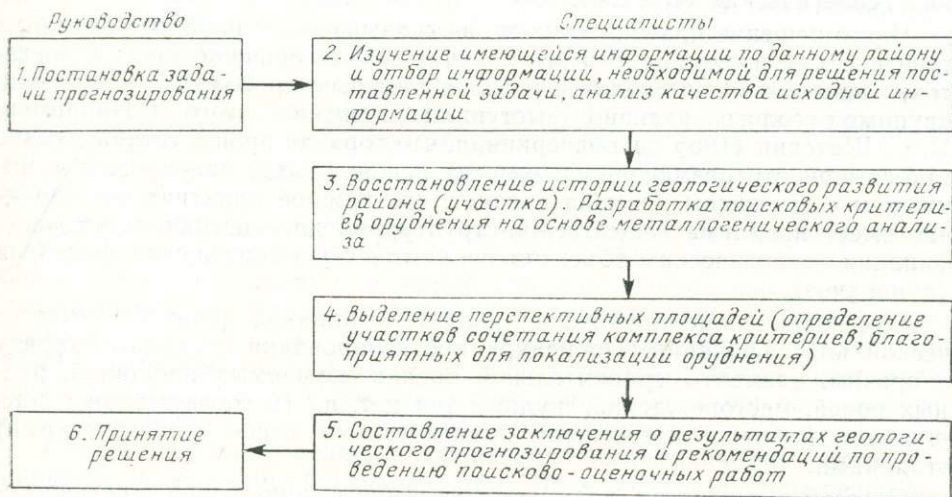


Рис. 3. Принципиальная схема последовательности решения геологических прогнозных задач.

видно, что применение соответствующих математических методов для выделения наиболее важных геологических факторов и построения поисково-оценочных критериев на количественной основе может в значительной степени уменьшить влияние субъективизма отдельных исследований на конечный результат и повысить качество прогноза.

При выделении перспективных участков на исследуемых площадях специалист анализирует отдельные геологические объекты с точки зрения наличия поисково-оценочных критериев, благоприятных для локализации оруденения конкретного типа.

Такой анализ при традиционном прогнозировании проводится, как правило, на основе качественных свойств геологических объектов, а сам процесс сравнения по методу аналогии выполняется также на качественной, а не на количественной основе. Следовательно, и на этой стадии процесса геологического прогнозирования применение математических методов диагностики и оценки геологических объектов на основе разработанных поисково-оценочных критериев может привести к повышению эффективности и обоснованности прогноза.

Таким образом, на каждой стадии процесса геологического прогнозирования имеются возможности для повышения его эффективности путем применения математических методов и ЭВМ. Поэтому в последние годы наблюдается тенденция все более широкого использования ЭВМ и методов теории статистических решений, распознавания образов, многомерного статистического анализа и т. п. в практике решения прогнозных геологических задач.

Использование математических методов в прикладной геологии является составной частью внедрения математики в геологию вообще и относится к тому направлению этого внедрения, которое имеет целью получение практических выводов из существующих теоретических выводов и представлений. В рамках существующих теоретических представлений и моделей объединяются различные подходы к изучению геологических объектов. Один из подходов — генетический, второй — основанный на исследовании пространственных закономерностей размещения геологических объектов.

Часто решение практических задач связывается с наиболее сложным генетическим подходом. В итоге получается, что решение многих, достаточно простых задач сводится к более сложным проблемам. Некоторые крупные геологи активно выступили против этого. Например, Н. С. Шатский (1965 г.) подчеркивал, что гораздо проще устанавливать закономерности размещения полезных ископаемых и оперировать с ними, чем с генезисом. Отсюда следует, что важное практическое значение имеет проблема вещественно-структурной диагностики, или классификация геологических объектов по комплексу объективных факторов (признаков).

Прогнозирование оруденения является типичной задачей геологической классификации. Процедура прогнозирования содержит, наряду с прочим, элементы сравнительной оценки изучаемых площадей, рудных полей, месторождений, рудных тел и т. п. (в соответствии с масштабами работ) с известными практическими (либо теоретическими) эталонами.

По всей вероятности, разумное сочетание и дополнение генетического и статистического подходов на основе использования современных средств обработки информации позволит повысить эффективность составления геологопрогнозных заключений.

Здесь необходимо подчеркнуть, что процесс геологического прогнозирования является творческим процессом, в котором ведущую роль играют опыт и знания геологов-специалистов. Решение таких задач, как, например, восстановление истории геологического развития того или иного района, разработка различных геологических концепций, определение на их основе закономерностей временного и объемного взаимодействия различных геологических факторов и т. п., не исчерпывается формально-логическими и арифметическими операциями, осуществляемыми ЭВМ. Поэтому, как представляется авторам, неправильно было бы ставить вопрос о замене процесса геологического прогнозирования полностью автоматическим, «машинным» прогнозированием.

Задача, следовательно, состоит в том, что необходимо «встроить» ЭВМ в процедуру геологического прогнозирования таким образом, чтобы предоставить специалистам возможность оперативно использовать разнообразный математический аппарат для формально-логического и количественного анализа имеющейся информации, сохранив за ними постановку задачи, выбор основных факторов и критериев, содержательное истолкование результатов работы ЭВМ и принятие окончательных решений.

3. СТРУКТУРА И ПРИНЦИПЫ ПОСТРОЕНИЯ ЧЕЛОВЕКО-МАШИННОЙ СИСТЕМЫ ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

В настоящем разделе предпринята попытка на основе анализа традиционной схемы геологического прогнозирования и опыта применения математических методов для решения задач прикладной геологии разработать с позиций системного подхода общую схему взаимодействия человека (специалиста геолога) с информационно-вычислительной системой, ориентированной на решение задач геологического прогнозирования. В такой системе необходимо, во-первых, в максимальной степени использовать творческие способности специалистов; во-вторых, использовать большие возможности современных ЭВМ по накоплению, хранению и обработке больших массивов информации; в третьих, увязать технологию решения задач в системе со сложившейся традиционной методикой геологического прогнозирования; в-четвертых, четко разграничить функции специалистов и ЭВМ и, наконец, в-пятых, в такой системе следует найти способ, резко сокращающий затраты ручного труда по предварительной подготовке и обработке исходной информации.

В результате исследований была разработана автоматизированная информационно-прогнозирующая система (АИПС), представляющая собой по существу инструмент выработки многовариантных решений задач геологического прогнозирования на основе рационального взаимодействия специалистов с ЭВМ [2]. Одной из важнейших особенностей АИПС является последовательное выполнение отдельных составляющих процедур процесса геологического прогнозирования на основе использования машинного банка данных многоцелевого назначения.

Постоянное расширение использования математических методов в прикладной геологии обусловлено тем, что в последние годы резко возрастает объем информации, получаемой в процессе геологических исследований. Основные трудности практического использования всей поступающей информации в значительной мере обусловлены рядом взаимосвязанных причин: недостаточным уровнем наших знаний о моделях

связей между отдельными признаками и свойствами геологических объектов; недостаточной надежностью существующих оценочных и прогнозных критериев; отсутствием единого подхода к вопросам комплексной интерпретации исходных данных; ограниченными возможностями человеческого мозга при анализе многомерного пространства признаков, характеризующего описание геологического объекта; отсутствием эффективных средств комплексной оперативной обработки поисково-разведочных данных для нужд диагностики, прогноза и оценки геологических ситуаций.

К настоящему времени начинает складываться определенная методика решения прогнозных геологических задач с использованием математических методов и ЭВМ. Достаточно детально основные принципы ее применительно к задачам рудной геологии изложены в работах [12, 39, 45, 51, 99 и др.], применительно к нефтяной геологии — в работах [4, 67, 126 и др.]. В указанных исследованиях, основанных на применении различного математического аппарата для классификации геологических объектов, определен круг вопросов, при решении которых в настоящее время применяются математические методы и подробно рассмотрены эти методы. В частности, проанализированы возможные подходы к выбору признаков для решения прогнозных задач и методы минимизации числа признаков, даны сведения о различных используемых алгоритмах и т. п.

Наибольшее распространение получает в настоящее время подход, основанный на применении алгоритмов распознавания образов. В общей постановке задачи распознавания применительно к прогнозированию геологических объектов предполагается, что в изучаемом районе имеются объекты двух или более классов (например, рудоносные и безрудные площади), описанные некоторым комплексом признаков («объекты обучения»). С помощью алгоритма распознавания на основе этих объектов обучения на ЭВМ строится решающее правило, позволяющее относить любой объект, описанный тем же комплексом признаков, к одному из рассматриваемых классов.

Процедуру нахождения решающего правила принято называть обучением классификации, а отнесение объекта, не принадлежащего к обучающей выборке, к одному из эталонных классов — собственно распознаванием.

Наибольшей трудностью, возникающей при практическом решении прогнозных задач по такой схеме, является большой объем ручного труда по предварительной подготовке данных для ввода их в ЭВМ. Для обучения и классификации объектов с помощью алгоритма распознавания образов (а также многих других алгоритмов) требуется ввести в машину матрицу из m признаков (m — суммарное число геологических объектов обучения и объектов, подлежащих классификации, n — число признаков, характеризующих каждый объект). Такая информация, подготавливаемая в настоящее время, как правило, вручную, имеет узкоцелевой характер, и при необходимости решения новой прогнозной задачи по тому же району (например, при поиске других типов месторождений) вся трудоемкая работа по подготовке исходных данных должна проделываться заново. Для преодоления указанных трудностей следует, в первую очередь, организовать хранение данных в ЭВМ на принципиально новой основе. Во-первых, информация должна вводиться в ЭВМ в таком виде, чтобы требовалась лишь минимальная ее предварительная обработка. Во-вторых, информация, хранящаяся в ЭВМ, должна быть ориентирована не на решение отдельной конкрет-

ной задачи, а на ее неоднократное использование при решении различных задач геологического прогнозирования на данной территории. В-третьих, способ хранения информации должен предусматривать возможность ее дополнения, изменения и исключения ее отдельных частей, а также возможность оперативного извлечения любых интересующих специалистов данных. И, в-четвертых, необходимо создание специального математического обеспечения, позволяющего автоматически, по запросам специалистов формировать из хранящегося в памяти ЭВМ исходного фактического материала по всему району («первичного массива») массивы данных, необходимые для решения конкретных прогнозных задач («целевые массивы»), в виде, пригодном для непосредственного применения математических методов классификации. В дальнейшем такого рода машинную систему накопления, хранения, обновления и извлечения геологической информации вместе с ее математическим обеспечением, удовлетворяющую сформулированным выше требованиям, будем называть банком данных.

При наличии такого банка данных специалист получает возможность без дополнительной подготовки и переработки исходного материала осуществлять многовариантное прогнозирование на основе различных гипотез, моделей и критериев, используя соответствующие математические методы. Иными словами, появляется возможность построить такую систему взаимодействия специалиста-геолога и ЭВМ, при которой первый почти полностью освобожден от механической обработки больших массивов данных и занимается решением содержательных геологических вопросов, используя при этом в качестве инструмента анализа разнообразный арсенал математических методов.

Функциональная схема первого модельного варианта системы изображена на рис. 4. Рассмотрим предлагаемую схему по отдельным процедурам (номера процедур соответствуют рис. 4).

1. Задание на формирование банка данных многоцелевого назначения выдается руководством территориальной геологической службы и выполняется группой специалистов — геологов, геофизиков, геохимиков и сотрудниками вычислительного центра. Основное назначение работы — выбрать наиболее качественные и представительные материалы и поместить их в память ЭВМ для последующего использования при решении различных задач геологического прогнозирования.

2. Анализ имеющейся информации и подготовка исходных данных. Подготовка исходных данных состоит в детальном анализе качества геологических, геофизических, геохимических и других карт. По существу от этой работы, предшествующей вводу информации в ЭВМ, зависят, в конечном итоге, результаты прогнозных исследований. Именно поэтому для каждого из факторов, вводимых в память ЭВМ специалистами экспертами, оценивается качество проведенных работ, т. е. пригодность конкретной информации для решения поставленных задач.

При анализе имеющейся исходной информации необходимо учитывать:

— изученность территории геологическими, геофизическими и другими съемками различных масштабов;

— качество съемок различных лет, включая их сопоставимость, сети наблюдений и др.;

— наличие трансформированных карт геофизических полей и т. д.

Эта работа выполняется группой специалистов (геологов, геофизиков, геохимиков и др.). После ознакомления со всей доступной информацией по исследуемому району (объекту) составляется заключение

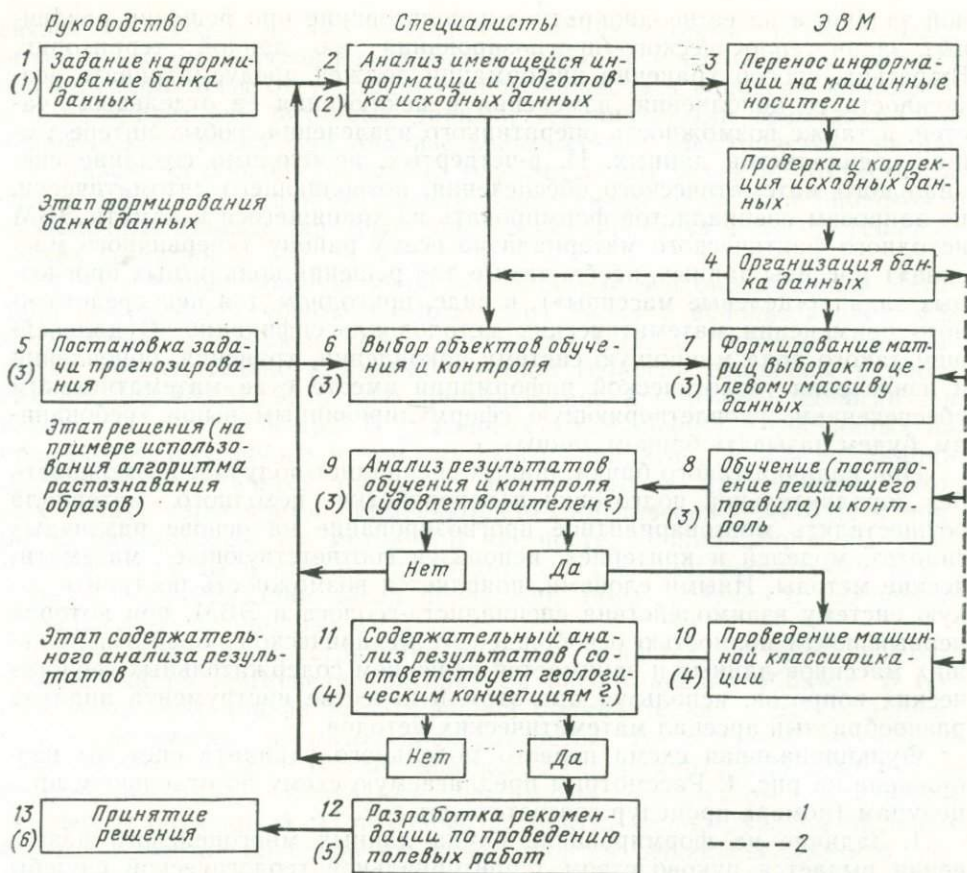


Рис. 4. Принципиальная схема автоматизированной информационно-прогнозирующей системы для отдельного этапа геологических исследований. В скобках указаны порядковые номера соответствующих процедур, выполняемых при традиционных методах геологического прогнозирования (см. рис. 3).

1 — последовательность выполнения операций, 2 — передача информации

о принципиальной возможности формирования банка данных*. На основе этого отбирается качественная и представительная информация, которая в последующем составит основу банка данных. При этом для уменьшения влияния субъективности отдельных исследователей и выработки коллективного мнения о качестве геологической информации предусматривается применение математических процедур обработки экспертных оценок.

По той роли, которую играют в прогнозировании оруденения различные геологические признаки, они могут быть подразделены на две главные группы: разделяющие и отождествляющие. Разделяющие признаки выполняют основную функцию — классификацию объектов, выбранных с учетом отождествляющих признаков, которые отражают общность геологической позиции объектов рассматриваемого типа. Та-

* При наличии технических возможностей в банк данных следует заносить всю надежную исходную информацию.

ким образом, следует выбрать комплекс тех признаков, которые имеют прямое отношение к прогнозируемому оруденению, и исключить из дальнейшего рассмотрения так называемые «шумовые», или «паразитные» [35], которые объективно не содержат полезной информации.

При выборе признаков известны два подхода. Первый состоит в учете физической сущности изучаемых явлений. Способом выбора комплекса признаков при этом являются экспертные оценки квалифицированных специалистов. Второй подход заключается в оценке информативности признаков по их роли в распознавании объектов. Наилучшим является, по-видимому, комплексный подход, когда предварительный выбор комплекса признаков производится в соответствии с мнением специалистов (по этим признакам формируется банк данных), а окончательный выбор информативных признаков осуществляется в процессе разработки решающего правила с последовательным исключением признаков, неинформативных для решения конкретной задачи геологического прогнозирования.

Важным требованием к используемой информации является одинаковый уровень изученности ко всему исследуемому району (площади, объекту и т. д.). В самом деле, для того чтобы использовать метод аналогии, очевидно, необходимо, чтобы «эталонный» объект и объект, по которому делается заключение, были бы описаны одними и теми же признаками. Следовательно, для формирования информационной базы следует использовать лишь те данные, по которым имеются сведения по всей исследуемой территории (району и т. п.).

Однако в практике исследований часто невозможно получить по тем или иным причинам одинаковую по качеству исходную информацию.

Главные трудности при оценке качества и достоверности исходных материалов возникают при решении прогнозных задач мелкого и среднего масштаба. На практике часто наблюдаются случаи, когда по отдельным частям какой-либо рудной провинции имеются несколько вариантов геологических и других карт, составленных многими авторами. При этом имеет место различная трактовка отдельных вопросов магматизма, спорные определения возраста отдельных толщ горных пород, различные уровни геофизических полей и т. п. Поэтому при решении задачи мелкомасштабного прогноза может возникнуть необходимость в составлении сводной карты по всему району, что может повлечь за собой разработку единой, приемлемой для всех авторов стратиграфии и т. п.

Легче решаются вопросы «увязки» уровня геофизических полей между отдельными частями района, для чего может быть использована ЭВМ. Следует подчеркнуть, что эта работа по существу носит тематический характер и обобщает геологические результаты многочисленных исследований различных вариантов по данному району. Выполнение такой работы позволяет создать комплекс-сводных геологических и других карт, занести их в последующем в память ЭВМ и использовать в дальнейшем не только для прогнозных заключений, но и в качестве справочного материала по запросам специалистов.

На этапе среднемасштабного прогнозирования основная трудность состоит в том, чтобы «увязать» между собой результаты геологических исследований по различным листам геологической съемки (например, в масштабе 1 : 50 000). Однако на этом этапе специалисты уже стараются «стыковать» свои данные с соседними планшетами, тем более, что такого рода работы выполняются, как правило, в одном территориаль-

ном геологическом управлении (комбинате, экспедиции), что обуславливает использование стандартной (по крайней мере, в пределах данной организации) геологической легенды и одного уровня геофизических полей. Все это значительно облегчает задачу оценки качества и представительности информации на этапе среднemasштабного прогнозирования.

В этой связи на основе либо данных экспертных оценок, либо других приемов целесообразно вводить соответствующие коэффициенты, учитывающие качество исходных данных и уровень изученности.

3. Перенос информации на машинные носители и проверка исходных данных. Чтобы ввести имеющуюся информацию в ЭВМ необходимо, во-первых, снять (измерить) нужные данные с геологических, геофизических карт и других документов; во-вторых, отперфорировать эти данные в кодах используемой ЭВМ на перфолентах или перфокартах. Именно на эти операции в настоящее время приходится подавляющая часть всех затрат труда при решении какой-либо задачи с помощью ЭВМ. В последние годы наметился значительный сдвиг в решении этой проблемы: разработаны и выпускаются различные устройства, автоматизирующие ввод информации в ЭВМ. Применение этих устройств позволяет значительно сократить время для ввода информации и, тем самым, устранить одно из основных препятствий, стоящих на пути широкого использования вычислительной техники при решении геологических задач. Процедура проверки правильности занесенных на машинные носители данных также является в настоящее время одной из самых трудоемких. Наличие специального программного обеспечения для выявления наиболее часто встречающихся ошибок, а также возможность распечатки введенных данных на графопостроителе существенно облегчают визуальный контроль картографической информации.

4. Организация банка данных. В результате данной процедуры собранная специалистами исходная информация, зафиксированная на машинных носителях, заносится в память ЭВМ (на магнитные ленты, диски) и организуется там в порядке, обеспечивающем возможность ее оперативного пополнения, исправления и извлечения (организуется «первичный массив» исходных данных). Далее, с помощью комплексов программ математического обеспечения банка данных по этой информации рассчитываются значения признаков, необходимых для решения задачи прогнозирования, и полученные результаты записываются в памяти ЭВМ в удобном для последующей работы виде («целевой массив»).

Следует иметь в виду, что в принципе первичный массив банка может содержать значительно большее количество информации, чем это необходимо для решения конкретной задачи, и использоваться также при решении других задач (при выдаче различных справок по запросам и пр.).

5. Постановка задачи. Конкретная постановка задачи геологического прогнозирования задается «руководством», в качестве которого может рассматриваться, например, научно-технический совет геологической организации (главка, управления, экспедиции). К задачам такого рода могут быть отнесены следующие: «определить площади, перспективные на обнаружение оруденения определенного типа с целью проведения геологосъемочных работ», «выбрать наиболее перспективное развитие поисково-разведочных работ», «провести классификацию аномалий», «ранжировать площади по их перспективности на конкретный тип оруденения» и т. д.

6. Выбор объектов обучения и контроля. Поскольку решение прогнозных задач основано на использовании метода аналогии, то непременным условием является наличие так называемых «эталонных объектов», т. е. хорошо изученных объектов, аналоги которых могут быть обнаружены в пределах изучаемой территории. Этот вопрос является одним из главных в прогнозировании. Успех прогноза целиком зависит от того, насколько типичными являются эталоны различных классов, насколько существенны различия геологических характеристик разномасштабных объектов и насколько общей является позиция эталонных объектов и объектов оценки. К выбору объектов исследования следует подходить с большой осторожностью ввиду неоднозначности понятий «рудная формация», «рудное поле» и т. д. [43] и, как правило, спорности вопросов генезиса месторождений. Во всех случаях пределы области эталона и объекта прогнозирования должны быть ограничены некоторыми рамками, которые зависят от масштаба или уровня исследования. В качестве таких границ могут в различных случаях выступать различные геологические единицы, как например, металлогеническая зона, рудное поле, отдельное месторождение и т. д. Наконец, как эталонные объекты, так и объекты оценки должны характеризоваться одинаковой степенью изученности.

Вопросы выбора эталонных объектов тесно связаны с выбором вероятных типов рудных месторождений, обнаружение которых принципиально возможно в пределах исследуемой территории. Решение этого вопроса происходит на всех этапах работ и предшествует постановке специальных исследований. По существу успех решения задачи геологического прогнозирования во многом зависит от правильного выбора генетических и промышленных типов месторождений, нахождение которых возможно в исследуемом районе. Выбор эталонных рудных объектов (данного генетического типа) основывается на заданных ограничениях по параметрам месторождений. К степени изученности эталонных объектов предъявлялись достаточно высокие требования.

Анализ существующих геологических гипотез (моделей) процесса рудообразования и использования одной или нескольких моделей и соответствующих им эмпирических поисковых критериев на входе прогнозирующей системы позволяет конкретизировать формирование территориального банка данных и сделать эту процедуру достаточно целенаправленной. Если таких эталонных объектов в пределах исследуемой территории не имеется, то принципиально допускается использование в качестве эталонов объектов из других районов, общность геологических позиций которых с данной ситуацией представляется специалистам бесспорной. Процедура выбора эталонных объектов имеет принципиальное значение для применения методов распознавания образов в прогнозировании и оказывает непосредственное влияние на надежность окончательных выводов о перспективности или бесперспективности рассматриваемого района (участка).

По своей сути эта процедура может выполняться лишь специалистами экспертами, однако применение математических методов и использование ЭВМ может оказать существенную помощь в принятии ими окончательных решений.

Для определения согласованности экспертов может быть использован метод обработки экспертных оценок, описанный в работе [71], а также другие способы [50, 85, 107 и др.].

Наличие банка данных существенно облегчает процедуру выбора эталонных объектов. Так, после того как эксперты определяют эталон-

ные объекты, можно быстро «вызвать» из машинного банка данных все характеризующие их признаки, построить и проанализировать гистограммы распределений объектов по различным признакам. Однако анализ распределений объектов в одномерном пространстве (по гистограммам) не дает возможности проанализировать «однородность» эталонных классов сразу по всем признакам вместе. Такой анализ в n -мерном пространстве (где n — количество геологических и других факторов, использованных для описания объектов) возможен при использовании специальных математических приемов. В качестве такого вспомогательного способа анализа «компактности» объектов может быть использован, например, алгоритм таксономии [79].

Главные требования, предъявляемые к объектам обучения, заключаются в следующем:

— среди объектов обучения и контроля должны быть представлены все разновидности объектов, поиск аналогов которых проводится в районе;

— эталонная выборка должна включать в себя возможно большее число изученных объектов одной формации.

7. Формирование матриц — выборки по целевому массиву. После того как специалисты на основе своих знаний и опыта сформулируют из хорошо изученных объектов выборку обучения и контроля, из банка данных должны быть автоматически выбраны значения всех требуемых признаков по соответствующим объектам. В зависимости от цели дальнейшей обработки эти данные могут быть выведены в виде матриц «объект — признак» на АЦПУ, перфокарты или магнитную ленту. По каждому объекту необходимо выдавать его условные координаты.

8. Обучение и контроль. На основе сформулированной обучающей выборки производится построение формального поисково-оценочного критерия (так называемого «решающего правила») с помощью соответствующего алгоритма распознавания образов. Оценка качества решающего правила осуществляется по данным классификации объектов контроля, не использовавшихся при процедуре обучения.

9. Анализ результатов обучения и контроля. Построенные при выполнении предыдущей процедуры формальные решающие правила могут быть применены для построения прогнозных карт лишь после их содержательного анализа. Специалисты анализируют результаты контроля, учитывая при этом число ошибок и отказов от классификации, и выявляют, по возможности, содержательные причины последних. В случае, если построенное решающее правило по тем или иным причинам не удовлетворяет специалистов, они могут в оперативном порядке корректировать его, изменяя выборку обучения и набор учитываемых признаков или используя другие алгоритмы. Таким образом, процедуры 6—9 могут выполняться в итерационном цикле несколько раз, пока не будут построены удовлетворительные решающие правила.

Кроме того, использование данной системы позволяет оперативно проверять правомерность различных гипотез и точек зрения (например, о генетической общности отдельных объектов), высказываемых отдельными специалистами. Например, если кем-либо из экспертов высказывается особое мнение о степени подобия некоторых объектов, то легко, приняв их за эталонные, выполнить процедуру обучения (процедура 8) и по результатам классификации быстро установить, к каким последствиям (в смысле геологического прогноза) в целом по исследуемой территории приведет данная точка зрения.

10. Проведение машинной классификации. На основе разработан-

ного решающего правила производится автоматическое разбиение всей исследуемой территории в соответствии с полученным решающим правилом, рассматриваемом в качестве комплексного поисково-оценочного критерия. Результаты такой классификации печатаются в виде варианта прогнозной карты. Дополнительно по требованию специалистов может выдаваться числовой материал, поясняющий с формальных позиций результаты отнесения тех или иных объектов (площадей) к различным классам. Наличие банка данных по исследуемой территории позволяет выполнять данную процедуру с большой скоростью (несколько минут).

11—12. Содержательный анализ результатов прогнозирования и разработка рекомендаций. Специалисты оценивают результаты машинной классификации с содержательных позиций, тщательно анализируют все необычные с их точки зрения ситуации и т. п. В результате выполнения этой процедуры, являющейся завершающей, выдается откорректированная информация в виде прогнозной карты с пояснительной запиской, содержащей обоснование выбора перспективных участков и рекомендации по очередности их изучения.

Оценка результатов при многовариантном прогнозировании приобретает особенно важное значение. Поэтому помимо общепринятого критерия оценки по числу ошибок I и II рода используется и ряд других:

— величина суммарной площади перспективных участков в процентах от исследуемой территории;

— геологические особенности пространственного размещения перспективных участков [39] и др.

Особенности геологического положения перспективных участков в пределах исследуемой территории могут быть использованы в качестве вспомогательного (более тонкого) критерия оценки качества прогнозных карт, учитывающего зональность размещения перспективных участков и соответствие их расположения существующим геологическим представлениям.

13. Принятие решения. На заключительном этапе проводится анализ полученных результатов прогноза на научно-техническом совете организации (предприятия) и их утверждение.

Первый вариант рассмотренной системы был реализован на ЭВМ типа М-222 и опробован на практических материалах. В качестве информационной основы банка данных использовались геологические, геофизические и геохимические карты.

Опытная эксплуатация системы показала, что ее применение для решения задач прикладной геологии позволяет повысить оперативность, обоснованность и качество составления геологических прогнозных заключений за счет более полного использования имеющейся информации, коллективных знаний специалистов геологов и современных математических средств обработки информации.

В то же время экспериментальные исследования эффективности данного варианта АИПС позволили установить ряд недостатков, касающихся технологического процесса обработки геологических данных в системе и связанных в основном с ограниченными техническими возможностями ЭВМ второго поколения. С целью устранения этих недостатков был разработан второй вариант данной системы на базе ЭВМ третьего поколения, обладающий улучшенными технологическими характеристиками и значительно большими возможностями для решения практических задач в режиме промышленной эксплуатации (АИПС/ЕС «регион»).

III. МАТЕМАТИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

1. ОБЩИЕ ВОПРОСЫ ПРИМЕНЕНИЯ МЕТОДОВ РАСПОЗНАВАНИЯ ОБРАЗОВ В ЗАДАЧАХ ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

В прикладной геологии многие прогнозные и интерпретационные задачи решаются методами аналогии. При этом производится сопоставление изученных геологических объектов по выбранному комплексу признаков с известными эталонными объектами. Подобный подход открывает широкие перспективы для использования математических методов распознавания образов [45] применительно к решению задач геологического прогнозирования.

Проблемы распознавания образов при геологическом прогнозировании связаны с проблемами классификации объектов по косвенным данным. В принципе задача сводится к построению (обучение распознаванию) модели связи между масштабностью оруденения и косвенными геологическими признаками и отнесению (собственно распознавание) объекта с помощью построенной модели к одному из классов объектов, например к рудоносным или бесперспективным в отношении рудоносности. Концепцию распознавания образов при геологическом прогнозировании можно выразить следующим образом.

Предположим, что всякому реальному геологическому объекту можно поставить в соответствие набор действительных чисел x_1, x_2, \dots, x_n , которые представляют собой различные измеренные геологические, геофизические, геохимические и другие характеристики объекта.

Тогда каждая такая совокупность чисел может быть интерпретирована как вектор или точка в многомерном пространстве. В этом случае множеству всех возможных объектов какого-либо одного класса (образа) в пространстве признаков соответствует некоторое множество точек.

Задача обучения распознаванию образов заключается в построении в пространстве признаков поверхности, разделяющей множества точек, соответствующие различным классам объектов. После построения такой поверхности новые объекты (точки) могут идентифицироваться в зависимости от положения относительно разделяющей поверхности.

Рассмотрим геометрическую интерпретацию процесса распознавания или классификации образов.

Пусть имеется N классов, на которые необходимо разделить объекты. Пронумеруем эти классы целыми числами $1, 2, \dots, N$ и примем в случае рассматриваемого ниже примера $N=3$.

Класс	Прогноз
1	Рудоносный объект
2	Объект бесперспективен в отношении рудоносности
3	Решение не принимается

Представим каждый объект в виде точки в n -мерном евклидовом пространстве E^n . Декартовы координаты такой точки суть действитель-

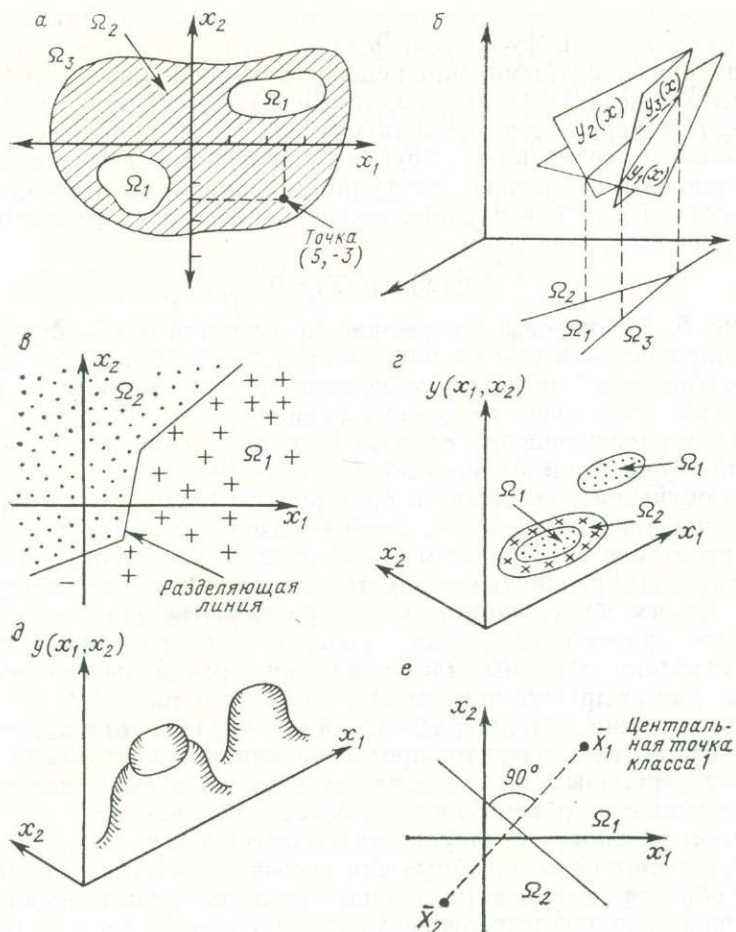


Рис. 5. Геометрическая интерпретация классификации при распознавании образов (по Е. П. Власову)

ные числа x_1, x_2, \dots, x_n . Распознаванием объектов является отображение точки из E^n в числа 1, 2, 3, которыми пронумерованы классы. Обозначим через Ω_i множество точек из E^n , которые отображаются в i . Тогда для каждого такого числа x_n имеется множество точек из E^n , обозначенное одним из символов $\Omega_1, \Omega_2, \Omega_3$. В качестве примера классификации рассмотрим множества, показанные на рис. 5, где $n=2$. Любая точка плоскости в зависимости от ее принадлежности к множествам 1, 2 или 3 отображается соответственно в одно из чисел 1, 2 или 3. Например, вновь изучаемый объект (точка) с координатами-признаками $(5, -3)$ был бы отнесен к классу 2.

На рис. 5, а множества точек трех рассматриваемых классов объектов отделены друг от друга разделяющими поверхностями (кривыми в E^2) на взаимно непересекающиеся области в пространстве признаков.

Множества точек, на которые разделяющиеся поверхности делят E^n и которые соответствуют одному из N классов, называют областями решений.

Разделяющие поверхности любого вида можно полностью определить N скалярными функциями $y_1(\bar{x}), \dots, y_N(\bar{x})$. Эти функции, называемые дискриминантными, или решающими функциями, выбирают так, чтобы для всех $\bar{X} \in \Omega_i$ выполнялось $y_i(\bar{x}) > y_j(\bar{x})$ при $i, j = 1, 2, \dots, N; i \neq j$. Так, i -ая решающая функция на области Ω_i принимает наибольшее значение по сравнению с другими решающими функциями.

Предположим, что решающие функции непрерывны на разделяющих поверхностях. Тогда поверхность, разделяющая смежные области Ω_i и Ω_j , определяется уравнением

$$y_i(\bar{x}) - y_j(\bar{x}) = 0. \quad (\text{III.1})$$

На рис. 5, б показаны три решающие функции и области решений, которые порождаются ими на плоскости x_1, x_2 . Заметим, что разделяющие поверхности в плоскости x_1, x_2 являются проекциями на эту плоскость линий пересечения решающих функций.

Обучение распознаванию образов или построение разделяющей поверхности осуществляется по выборке известных эталонных объектов (выборке обучения), которым в пространстве измеряемых признаков соответствует множество точек, рассеянных внутри некоторой области данного пространства. При этом предполагают, что совокупность точек, соответствующих различным объектам, принадлежащим к одному и тому же классу, будет компактна в пространстве признаков в том смысле, что «расстояние» между членами каждого класса будет в среднем невелико, а члены, представляющие различные классы, будут отделены один от другого некоторой искомой границей.

Математические методы распознавания образов для каждой конкретной задачи геологического прогнозирования существенным образом зависят от априорных знаний и предположений о свойствах распределения точек классов объектов в пространстве признаков.

При «односвязных» совокупностях классов объектов, как показано на рис. 5, в, применяют линейные или кусочно-линейные методы распознавания образов, которые разделяют точки классов гиперплоскостью или кусочно-линейной гиперповерхностью, состоящей из ряда гиперплоскостей.

Если классы объектов, которые должны быть отделены друг от друга, «многосвязные» (рис. 5, г) (класс 1 состоит из двух разъединенных подклассов, один из которых находится в окружении точек класса 2), то применяют методы распознавания образов, которые позволяют строить разделяющие поверхности для областей, решений произвольной формы и сложности. Для распознавания классов, изображенных на рис. 5, г, можно использовать, например, нелинейные методы, которые строят полиномиальные разделяющие функции (рис. 5, д).

Чтобы представить общие принципы построения разделяющих поверхностей, используемых в алгоритмах классификации, рассмотрим простейший пример и параметрический и непараметрический методы обучения распознаванию.

Предположим, что имеется два класса объектов. Из априорных данных известно, что классы объектов образуют два компактных множества точек, разнесенных в пространстве признаков. При этом точки, соответствующие объектам класса 1, группируются около некоторой точки \bar{X}_1 (центральной точки), а точки, соответствующие объектам класса 2, — около другой центральной точки \bar{X}_2 (рис. 5, е). Причем координаты точек \bar{X}_1, \bar{X}_2 заранее неизвестны. Если бы они были известны, то было бы разумным разделить пространство объектов на две области

Ω_1 и Ω_2 , используя для этой цели какую-либо простую разделяющую поверхность, скажем плоскость. В качестве такой плоскости можно использовать плоскость, которая проходит через середину отрезка, соединяющего точки \bar{X}_1 , \bar{X}_2 , и ортогональна к нему. Уравнение разделяющей поверхности, которая выполняла бы подобное деление, имеет вид

$$(\bar{X}_1 - \bar{X}_2) \cdot \bar{X} + \frac{1}{2} |\bar{X}_2|^2 - \frac{1}{2} |\bar{X}_1|^2 = 0, \quad (III.2)$$

где $\bar{X} \cdot \bar{X}$ — скалярное произведение векторов \bar{X} и \bar{X} , а $|\bar{X}|^2 = \bar{X} \cdot \bar{X}$ — квадрат модуля вектора \bar{X} .

На рис. 5, *e* для случая $n=2$ показана разделяющая поверхность, соответствующая уравнению (III.2).

Для рассматриваемого примера можно использовать следующие методы обучения распознаванию.

Параметрический метод основан на оценке параметров, характеризующих классы объектов, и последующем определении разделяющей поверхности как функции данных параметров. Параметры задаются до обучения распознаванию из каких-либо априорных соображений.

Возьмем в качестве таких параметров координаты точек \bar{X}_1 и \bar{X}_2 , поскольку они определяют расположение точек каждого класса в пространстве признаков.

В этом случае обучение распознаванию сводится к получению оценок координат \bar{X}_1 и \bar{X}_2 по выборке обучения. Оценки определяются как выборочные средние по каждому классу объектов. После того как получены оценки этих средних, можно использовать уравнение (III.2) для определения разделяющей поверхности. На этом собственно и заканчивается процесс обучения по параметрическому методу.

Для классификации некоторого объекта H необходимо его признаковое описание $\bar{X}^H(x_1^H, \dots, x_n^H)$ подставить в левую часть уравнения (III.2) вместо \bar{X} и определить численное значение или хотя бы знак получившегося выражения. Если оно окажется положительным, то объект относится к классу 1, если отрицательным, то к классу 2. При нулевом значении выражения решение не принимается (случай, когда точка объекта H лежит на разделяющей поверхности).

Непараметрический метод характеризуется тем, что до обучения распознаванию из априорных соображений производится выбор форм разделяющей поверхности определенного класса, например в виде линейной или какой-либо нелинейной функции относительно признаков x_1, \dots, x_n с известными коэффициентами.

При выборе линейной зависимости разделяющая поверхность будет иметь следующий вид:

$$\sum_{i=1}^n a_i x_i + a_{n+1} = 0. \quad (III.3)$$

Обучение распознаванию в непараметрическом методе сводится к процессу определения по выборке эталонных объектов таких значений коэффициентов $\bar{A}(a_1, a_2, \dots)$, при которых разделяющая поверхность обеспечивает приемлемую классификацию объектов обучения.

Для предложенного примера распознавания компактных и разнесенных в пространстве признаков классов объектов разделяющая поверхность в виде линейной функции (III.3) должна дать приемлемое решение.

Следует отметить, что априорные сведения о составе характеристических параметров классов объектов при использовании непараметрического метода иметь необязательно. В этом состоит его преимущество перед параметрическим методом обучения, так как при решении задач геологического прогнозирования, как правило, отсутствуют априорные данные о параметрах распределения классов объектов в пространстве признаков.

Рассмотренные выше примеры классификации относились к задачам распознавания образов «с учителем», т. е. к задачам, в которых о каждой точке выборки обучения известно, к какому классу объектов она принадлежит*.

Однако помимо рассмотренных, существует класс задач распознавания образов, когда отсутствует априорная информация о принадлежности точек обучения к определенным классам объектов и даже о числе классов. Задача построения разделяющей поверхности без этой информации называется задачей обучения распознаванию «без учителя». Методы обучения распознаванию в этих задачах организуют поиск компактных и разнесенных множеств точек объектов в пространстве признаков, которые и фиксируются затем как классы или однородные совокупности объектов, представленных в одной исходной выборке обучения.

Наиболее разработанной и чаще всего решаемой с помощью математических методов и ЭВМ в настоящее время при геологическом прогнозировании является задача распознавания образов «с учителем».

Проблема применения математических методов распознавания образов к решению задач геологического прогнозирования включает в себя ряд взаимосвязанных вопросов:

- выбор эталонных объектов (формирование выборки обучения);
- выбор признаков, характеризующих рудоносность геологических объектов (формирование признакового описания объектов);
- выбор метода распознавания образов для решения задачи на основе использования априорных знаний специалистов геологов и предварительного анализа выборки объектов по их признаковому описанию;
- построение с помощью выбранного метода распознавания образов по выборке эталонных объектов и их признаковому описанию решающего правила;
- классификация с помощью построенного решающего правила контрольных объектов (части эталонных объектов, не включенных в выборку обучения с целью оценки качества решающего правила), и при удовлетворительной классификации контрольных объектов — оценка вновь исследуемых объектов.

Каких-либо единых рекомендаций по решению перечисленных вопросов для различных задач геологического прогнозирования не существует и, по-видимому, существовать не может. Это связано с тем, что в каждом конкретном случае решение данных вопросов существенно зависит от уровня знаний специалистов геологов, их опыта и интуиции, которые до конца не поддаются формализации и могут по-разному учитываться при постановке и решении задач.

В тоже время, исходя из анализа накопленного опыта применения методов распознавания образов в прикладной геологии и учета свойств

* В теорию распознавания образов термин «учитель» перешел из теории обучающихся автоматов, обучение которых производится с помощью «учителя», сообщающего машине, какому множеству точек предъявляемое изображение принадлежит.

алгоритмов распознавания, можно сделать ряд общих рекомендаций по рассматриваемому кругу вопросов.

Для статистических алгоритмов распознавания образов, в которых каждый измеряемый признак объекта определяется как случайная величина, а распределение классов объектов в пространстве признаков описывается многомерной (n -мерной) функцией плотности вероятности вектора $\bar{X}(x_1, x_2, \dots, x_n)$, желательно, чтобы число объектов в выборке обучения в несколько раз превосходило количество признаков.

В большинстве известных статистических алгоритмов классификации используется параметрический метод обучения распознаванию. Здесь многомерные функции плотности вероятности векторов классов объектов задаются в виде некоторых зависимостей от неизвестных параметров, например статистик первого и второго порядка (средних значений и дисперсий классов). При этом, чем больше неизвестных параметров, которые необходимо определять в процессе обучения, тем больше должна быть выборка обучения.

Для детерминистских алгоритмов, в которых измеренные признаки объектов рассматриваются как детерминированные величины, количество членов выборки обучения не является столь критической величиной. Поэтому при решении задач геологического прогнозирования с малой выборкой обучения (число членов выборки обучения соизмеримо или меньше количества измеряемых признаков), как правило, применяют детерминистские алгоритмы классификации.

Для любого метода классификации необходимо формировать такую выборку обучения, которая представляет характерные и представительные объекты каждого класса. Увеличивать количество членов выборки обучения за счет слабо изученных объектов не рекомендуется, поскольку возможная ошибка их предварительной разбраковки может привести к снижению надежности решающего правила и классификации новых объектов.

Значительную трудность представляет выбор признакового описания объектов. Это связано, во-первых, с тем, что выбор признаков для решения задач геологического прогнозирования осуществляется часто среди большого числа факторов геолого-геофизической и геохимической информации, представленной неравномерно и не одинаково детально, и, во-вторых, трудно найти специалистов широкого профиля, которые могли бы эти признаки совместно проанализировать и сопоставить по их информативности.

Во всех случаях при решении данного вопроса необходимо учитывать как можно больше знания специалистов геологов, хотя бы по каждой группе анализируемых признаков в отдельности, чтобы исключить из них неинформативные признаки и ограничить размерность решаемой задачи. Это повышает надежность решения задачи и упрощает ее реализацию.

При выборе признакового описания объектов рекомендуется использовать математические методы экспертных оценок для обработки мнений геологов, а также методы выделения информативных признаков по выборке обучения.

После задания выборки обучения и признакового описания объектов выбирают метод распознавания образов. При выборе методов решения задач геологического прогнозирования используют априорные знания специалистов о характерных особенностях классов объектов, проводят

предварительный анализ многомерного распределения классов объектов.

Если сталкиваются со случаем, когда классы объектов имеют много характерных особенностей, члены одного и того же класса близко находятся друг от друга, а различных классов — далеко, то применяют линейные методы распознавания образов.

Если из априорных знаний и предварительного анализа следует, что классы состоят из подклассов, групп, которые имеют различные свойства, то применяют методы, позволяющие строить разделяющие поверхности сложной формы.

Когда предварительный анализ исходных данных не проводится и отсутствует априорная информация о распределении классов в пространстве признаков, используют последовательно различные методы распознавания образов до тех пор, пока какой-либо из них не обеспечит заданную надежность классификации выборки обучения. Обычно подбор метода распознавания образов начинают осуществлять с более простых и быстродействующих: линейных и кусочно-линейных. Если они удовлетворительной классификации выборки обучения не обеспечивают, начинают применять сложные и емкие по времени вычислений алгоритмы для распознавания многосвязных классов объектов.

Следует отметить, что при использовании любого метода распознавания образов желательно получать решающие функции как можно более простого вида, во-первых, потому, что их проще реализовать, а во-вторых, они дают более устойчивые решения, что особенно заметно при малых выборках обучения.

Учитывая специфику решения задач геологического прогнозирования: отсутствие одинаковых геологических объектов (каждый объект специфичен по комплексу характеризующих его признаков); наличие в классах объектов подклассов, групп объектов с различными свойствами, а иногда и уникальных объектов с особыми присущими только им характеристиками; ограниченность выборок обучения и большая размерность признакового описания объектов, можно рекомендовать для применения также методы распознавания образов, которые позволяют определять и использовать при классификации характерные свойства как классов объектов в целом, так и их отдельных изолированных групп точек в различных признаковых подпространствах. К методам распознавания образов, обладающим указанными алгоритмами обучения, относятся, например, методы классификации по правилу ближайших величин, которые основаны на сравнении расстояний в пространстве (подпространствах) признаков между оцениваемой точкой и ближайшими точками классов объектов.

В настоящее время известно большое число методов распознавания образов, используемых при решении задач геологического прогнозирования. Они достаточно широко освещены в литературе [17, 42, 96, 105, 106, 109 и др.], поэтому ниже будут рассмотрены в основном оригинальные методы классификации. Среди них особенно подробно анализируется один класс эвристических алгоритмов, предназначенный для распознавания многосвязных классов объектов и использующий правило ближайших величин. Принципиальным преимуществом этого класса алгоритмов является то, что в нем заложен последовательный принцип классификации, а также то, что пространство признаков, в котором анализируются классы объектов и их отдельные группы точек, не фиксируется.

2. РАНЖИРОВАНИЕ И ВЫДЕЛЕНИЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ * ПО РЕЗУЛЬТАТАМ ОПРОСА ЭКСПЕРТОВ

Решение задач геологического прогнозирования часто вызывает трудности из-за большой размерности исходного пространства. Кроме того, некоторые признаки, привлекаемые для разделения объектов на классы, по своей генетической природе не содержат необходимой для этих целей информации. Как уже отмечалось ранее, введение в классифицирующую функцию подобных признаков может привести к ухудшению качества классификации. В связи с этим при решении геологических задач нередко приходится переходить к новому, реконструированному пространству признаков, как правило, меньшей размерности по отношению к исходному, но позволяющему производить классификацию с заданной надежностью при наиболее простом виде решающего правила. Формирование подобного пространства может быть осуществлено либо путем отбора из n исходных признаков некоторого числа наиболее информативных признаков, либо путем конструирования системы новых признаков с определенными свойствами, либо одновременно двумя указанными методами. Следует отметить, что оценка индивидуальной информативности признака в отдельности от других признаков не всегда оказывается достоверной, поскольку может существовать статистическая связь между ними. Так, относительно независимые признаки, представляющиеся весьма ценными при индивидуальной оценке, могут иметь лишь небольшую дополнительную информацию в совокупности с другими признаками, и, наоборот, относительно малоценные признаки могут значительно увеличивать информацию системы признаков.

Для исключения неинформативных признаков и ограничения размерности задач геологического прогнозирования представляется целесообразным собрать и обработать априорные сведения об изучаемых объектах, полученные в результате опроса специалистов. Это позволяет ускорить и повысить качество процесса выделения существенных признаков, содержащих самую характерную информацию об объектах, избежать лишних экспериментальных поисков. Использование информации, обобщающей коллективный опыт специалистов, повышает достоверность исходных данных. Поскольку мнения специалистов могут расходиться, то для правильного их использования необходимо определенным образом их систематизировать и обработать.

Для решения подобной задачи могут быть применены методы ранговой корреляции [157] и дифференциального ранжирования [71].

Метод ранговой корреляции

Для систематизации анализируемых признаков можно применить так называемое ранжирование. Под ранжированием понимается расположение n признаков в порядке возрастания или убывания некоторого качественного критерия (в данном случае информативности).

Ранг x_i указывает на место, которое занимает i -ный признак среди всех n признаков, ранжированных в соответствии с их информативностью. Статистикой связи ранжированных признаков является коэффициент ранговой корреляции.

* Под информативностью признака понимается тот вклад, который вносится этим признаком в обеспечение надежности решения конкретной задачи геологического прогнозирования.

Пусть n признаков ранжированы дважды в соответствии с их информативностью:

$$x_1, x_2, \dots, x_n, \\ x'_1, x'_2, \dots, x'_n.$$

Рассмотрим признаки i и j . Для i -го признака информативности имеют ранги соответственно x_i и x'_i а для признака j — x_j и x'_j . Тогда коэффициент ранговой корреляции определится из выражения

$$\rho = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_j - x_i)(x'_j - x'_i)}{\sqrt{\sum_{i=1}^n \sum_{j=1}^n (x_j - x_i)^2 \sum_{i=1}^n \sum_{j=1}^n (x'_j - x'_i)^2}}. \quad (III.4)$$

Очевидно, что

$$\sum_{i=1}^n \sum_{j=1}^n (x_j - x_i)^2 = \sum_{i=1}^n \sum_{j=1}^n (x'_j - x'_i)^2, \quad (III.5)$$

так как $x_i = 1, 2, \dots, n$; $x'_i = 1, 2, \dots, n$,
и, следовательно,

$$\rho = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_j - x_i)(x'_j - x'_i)}{\sum_{i=1}^n \sum_{j=1}^n (x_j - x_i)^2}. \quad (III.6)$$

Проведя несложные преобразования и учитывая, что x, x' — члены натурального ряда, получим

$$\rho = 1 - \frac{6S(d^2)}{n(n^2 - 1)}, \quad (III.7)$$

где

$$S(d^2) = \sum_{i=1}^n (x_i - x'_i)^2. \quad (III.8)$$

Наиболее интересным практическим приложением ранговой корреляции является вопрос о рассмотрении корреляционной связи нескольких ранжированных рядов.

Пусть имеется ряд признаков, в разной степени информативных с точки зрения решения конкретной задачи, и группе экспертов предлагается проранжировать эти признаки в соответствии с их информативностью. В результате опроса экспертов получаем матрицу рангов.

Эксперты	Признаки					
	1	2	...	i	...	n
1	x_{11}	x_{12}	...	x_{1i}	...	x_{1n}
2	x_{21}	x_{22}	...	x_{2i}	...	x_{2n}
...
j	x_{j1}	x_{j2}	...	x_{ji}	...	x_{jn}
...
m	x_{m1}	x_{m2}	...	x_{mi}	...	x_{mn}

Определим общий коэффициент ранговой корреляции для данной группы экспертов. Суммируя ранги для каждого признака, получим суммарную ранжировку признаков

$$\sum_{j=1}^m x_{j1}; \sum_{j=1}^m x_{j2}; \dots; \sum_{j=1}^m x_{jn}. \quad (\text{III.9})$$

Рассчитаем коэффициент ранговой корреляции для приведенного ряда из n членов, каждый из которых равен среднему значению для суммарных рангов ряда (III.9).

Очевидно, что среднее значение для суммарных рангов ряда

$$a = \frac{m}{2} (n+1). \quad (\text{III.10})$$

Используя формулу (III.8), вычислим величину $S(d^2)$

$$S(d^2) = \sum_{i=1}^n \left\{ \sum_{j=1}^m (x_{ji}) - \frac{m}{2} (n+1) \right\}^2. \quad (\text{III.11})$$

Величина $S(d^2)$ принимает максимальное значение, когда все эксперты дадут одинаковые ранжировки (например, в виде натурального ряда). Тогда суммарная ранжировка ряда (III.9) будет иметь вид: $m, 2m, \dots, nm$,

$$\text{а величина } S(d^2) = \frac{1}{12} m^2 (n^3 - n). \quad (\text{III.12})$$

Коэффициент ранговой корреляции для суммарной ранжировки вычисляем по следующей формуле:

$$W = \frac{S(d^2)}{S_m(d^2)} = \frac{12S(d^2)}{m^2(n^3 - n)}, \quad (\text{III.13})$$

величина W называется коэффициентом конкордации и изменяется в пределах от 0 до 1. Значение $W=0$ означает, что связи между ранжировками исследователей не существует, а $W=1$ означает, что все исследователи одинаково ранжируют признаки по информативности.

В случае, если ранжировка содержит совпадающие ранги, формула для вычисления W принимает вид

$$W = \frac{S(d^2)}{\frac{1}{12} m^2 (n^3 - n) - m \sum_{j=1}^m T_j}, \quad (\text{III.14})$$

$$\text{где } T_j = \frac{1}{12} \sum_{t_j} (t_j^3 - t_j), \quad (\text{III.15})$$

а t — число повторений каждого ранга в j -ом ряду. Причем признакам, имеющим одинаковые ранги, приписывается ранг, равный среднему значению мест, которые эти признаки поделили между собой.

Можно рассчитать распределение коэффициента W в предположении, что между экспертами не существует согласия относительно ранжиров-

ки признаков по их информативности. Распределение частот для W приведено в работе [6] для некоторых m и n .

Для больших m и n можно воспользоваться распределением χ^2 .

Величина $m(n-1)W$ имеет χ^2 -распределение с $\nu=n-1$ степенями свободы для $n>7$.

В случае совпавших рангов χ^2 -распределение с $n-1$ степенями свободы имеет вид

$$\chi^2 = \frac{S}{\frac{1}{12} mn(n+1) - \frac{1}{n-1} \sum_{j=1}^m T_j} \quad (\text{III.16})$$

Сравнивая величину χ^2 с табличным значением, можно принять решение о значимости коэффициента W . Если в таблице значение χ^2 больше полученного, то коэффициент W не значим, а гипотеза о наличии согласия у исследователей отвергается. В противном случае гипотеза применяется.

Метод дифференциального ранжирования

Существенным недостатком рассмотренного метода ранговой корреляции является то, что специалистам предлагается всю оцениваемую группу признаков расположить в порядке убывания их информативного веса. На практике не часто возникают случаи, когда исследователю приходится одновременно анализировать большое число взаимосвязанных факторов, имеющих различную физическую природу.

Число факторов, которое вынужден рассматривать геолог, нередко превышает 100. Естественно, что в таких случаях эксперт затрудняется представить все факторы в виде натурального ряда, расположив их по информативности, поскольку не представляется возможным провести совместный анализ такого значительного числа параметров. Кроме того, применимость метода ранговой корреляции для решения ряда практических задач ограничивается тем, что с помощью него можно оценивать согласованность мнений специалистов, а значит, и достоверность результатов опроса лишь по всей совокупности изучаемых параметров. Это равносильно тому, что если согласно методу будет принято решение о несогласованности мнений экспертов, то все экспертные оценки будут аннулированы. Последнее приведет к потере ценной информации, безусловно содержащейся в данных экспертизы.

Метод дифференциального ранжирования лишен отмеченных недостатков. Сущность его состоит в вычислении дифференцированных оценок согласованности мнений экспертов по каждому из рассматриваемых факторов, выделении тех факторов, по которым мнения экспертов в статистическом смысле являются согласованными, и последующем их ранжировании в порядке убывания усредненных экспертных оценок.

Предлагаемый метод ранжирования отдельных факторов при их дискретной (балльной) экспертной оценке сводится к следующему. Производится опрос экспертов по ряду геологических факторов. В распоряжении экспертов представляется фиксированный набор оценок, например: 0, 1 и 2. Неинформативным, по мнению эксперта, факторам дается нулевая оценка. Информативным факторам приписывается оценка 1. Наиболее существенные признаки получают оценку 2. Результаты опроса экспертов сводятся следующим образом.

№ эксперта	№ фактора					
	1	2	...	i	...	n
1	x_{11}	x_{12}	...	x_{1i}	...	x_{1n}
2	x_{21}	x_{22}	...	x_{2i}	...	x_{2n}
...
j	x_{j1}	x_{j2}	...	x_{ji}	...	x_{jn}
...
m	x_{m1}	x_{m2}	...	x_{mi}	...	x_{mn}

Присвоение оценок в определенной степени зависит от субъективности мнений экспертов, от их специализации и квалификации, в результате чего оценки экспертов по одному и тому же фактору могут значительно отличаться. Поэтому представляется целесообразным экспертные оценки подвергать статистическому анализу.

Прежде всего возникает вопрос, насколько вообще можно доверять этим оценкам, как сильно влияние субъективизма в этих оценках. Для ответа на этот вопрос необходимо определить степень согласованности мнений экспертов, по величине которой можно судить о достоверности совокупности экспертных оценок по каждому фактору в отдельности и по всем факторам в целом.

Вычислим для каждого фактора среднюю оценку

$$\bar{X}_i = \frac{\sum_{j=1}^m x_{ji}}{m_i}, \quad (\text{III.17})$$

где i — номер фактора; x_{ji} — оценка i -го фактора, данная j -м экспертом; m_i — число экспертов, оценивавших i -й фактор.

Расположим факторы по возрастанию величины средней оценки. Введем величину $S_i(d^2)$, характеризующую рассеяние индивидуальных экспертных оценок относительно средней оценки следующим образом:

$$S_i(d^2) = \sum_{j=1}^m (x_{ij} - \bar{X}_i)^2. \quad (\text{III.18})$$

Очевидно, что в случае полной согласованности мнений экспертов каждая индивидуальная оценка равна средней оценке и $S_i(d^2) = 0$.

Величина $S_i(d^2)$ примет свое максимальное значение, равное $S_{\max}(d^2)$, когда половина экспертов присвоит некоторому фактору оценку 0, а другая — 2. Эта величина равна

$$S_{\max}(d^2) = \frac{m}{2} (2-1)^2 + \frac{m}{2} (0-1)^2 = m,$$

когда m — четное число,

$$\text{и } S_{\max}(d^2) = \frac{m-1}{2} \left(2 - \frac{m-1}{m}\right) + \frac{m+1}{2} \left(0 - \frac{m-1}{m}\right)^2 = \frac{m^2-1}{m},$$

если m — число нечетное.

Ясно, что чем меньше величина $S(d^2)$, тем больше согласованность между экспертами по данному фактору и тем больше можно доверять экспертным оценкам.

Введем коэффициент согласованности V_i , характеризующий степень согласованности мнений экспертов.

$$V_i = \frac{S_{\max}(d^2) - S_i(d^2)}{S_{\max}(d^2)}. \quad (\text{III.19})$$

Коэффициент V_i может принимать значения в пределах от 0 до 1. Равенство $V_i=0$ соответствует полнейшей несогласованности мнений экспертов, а равенство $V_i=1$ свидетельствует о полном единстве мнений экспертов об i -м признаке. Таким образом, по величинам $S_i(d^2)$ и V_i можно судить о достоверности оценок. Однако в силу того, что закон распределения величины $S_i(d^2)$ в отличие от распределения величины V_i описывается хорошо изученными, табулированными функциями, решение о достоверности оценок будем выносить в дальнейшем только на основании величины $S_{bi}(d^2)$, вычисленной по конкретной выборке экспертных оценок. Заметим, что если известна величина $S_{bi}(d^2)$, то легко рассчитать выборочное значение V_{bi} .

Рассмотрим вначале схему получения экспертных оценок, при которой мнения специалистов не согласованы. Предположим, что экспертам об оцениваемом i -м признаке ничего не известно. В этом случае, естественно предположить, что эксперты будут случайным образом, с равной вероятностью давать оценки 0, 1 и 2. Очевидно, что возможны m^3 различных комбинаций оценок, где m — число экспертов. Вычислив значения $S(d^2)$ для каждой комбинации и построив функцию распределения случайной величины $S_i(d^2)$, составим таблицу (табл. 1) значений S_0 , при которых вероятность $P[S_i(d^2) \geq S_0]$ равна заданной величине.

Примем S_0 за нижнюю границу распределения случайной величины $S_i(d^2)$ и будем считать мнения экспертов об i -м факторе согласованными в том и только в том случае, если выборочная оценка $S_{bi}(d^2)$ окажется меньше S_0 . Таким образом, процедуру оценки согласованности мнений специалистов по критерию $S_i(d^2)$ можно свести к следующему.

Вычисленная для каждого фактора величина $S_{bi}(d^2)$ сравнивается с табличным значением S_0 . Если окажется, что $S_{bi}(d^2) < S_0$, то с вероятностью ϵ можно утверждать, что оценки i -у фактору присваивались не случайным образом, а были подчинены некоторой объективной закономерности, выразившейся в согласованности мнений экспертов.

Если же величина $S_{bi}(d^2) \geq S_0$, то оценки i -го признака по информативности следует считать недостоверным из-за отсутствия согласованности мнений экспертов.

Таблица 1

Критические значения $S_0(m, P)$

m	ϵ						m	ϵ					
	0,99	0,93	0,95	0,90	0,85	0,80		0,99	0,98	0,95	0,90	0,85	0,80
4	—	—	0,40	0,40	0,40	0,40	12	2,80	3,35	4,45	4,80	5,30	5,85
5	—	0,40	0,40	1,0	1,0	1,0	13	3,15	4,15	4,85	5,45	6,15	6,55
6	0,40	0,40	1,10	1,10	1,40	1,75	14	3,85	4,60	5,30	6,20	6,60	6,90
7	0,40	1,10	1,10	1,60	2,40	2,40	15	4,65	5,15	5,65	6,65	7,15	7,65
8	1,20	1,20	1,70	2,40	3,20	3,20	16	5,20	5,60	6,20	7,20	7,60	8,20
9	1,80	1,80	2,10	3,20	3,80	3,80	17	5,65	6,05	7,00	7,80	8,35	8,70
10	2,05	2,25	3,25	4,05	4,25	4,45	18	6,15	6,70	7,70	8,45	9,00	9,70
11	2,35	2,65	3,80	4,35	4,65	4,80	19	6,60	7,40	8,25	9,05	9,75	10,05
							20	7,10	7,90	8,90	9,77	10,40	10,90

При достаточно большом числе экспертов ($m \geq 10$) закон распределения величины $S_i(d^2)$ приближается к распределению χ^2 с K степенями свободы, и для определения величины S_0 при больших m можно воспользоваться таблицами, содержащимися в литературе по статистике.

Число степеней свободы при этом определяется по формуле $K = m - 1$, так как одна связь расходуется на определение усредненной оценки.

Те признаки, по которым мнения специалистов признаются несогласованными, из дальнейшего рассмотрения исключаются. Оставшиеся признаки ранжируются по величине усредненной оценки \bar{X}_i , т. е. по степени информативности.

Общая последовательность операций при использовании метода дифференциального ранжирования выглядит следующим образом.

1. Подготовительный этап:

— формирование входного массива факторов

$$\{i\} = 1, 2, \dots, n;$$

— выбор шкалы оценок факторов;

— формирование группы экспертов

$$\{j\} = 1, 2, \dots, m;$$

— формирование таблиц пороговых значений согласованности экспертных оценок по каждому фактору

$$\{S_0(m_i, P)\}.$$

2. Вычислительные процедуры:

— вычисление усредненной экспертной оценки по каждому фактору

$$\bar{X}_i = \sum_{j=1}^m \frac{x_{ji}}{m},$$

где x_{ji} — оценка i -го фактора, данная j -м экспертом;

— расчет для каждого фактора характеристики рассеяния экспертных оценок относительно средней

$$S_i(d^2) = \sum_{j=1}^m (x_{ji} - \bar{X}_i)^2;$$

— определение согласованности экспертных оценок для каждого фактора с помощью условия

$$P\{S_i(d^2) \geq S_0(m_i, P)\} = P.$$

Если условие соблюдается, то далее выполняют ранжирование факторов в порядке убывания величины информативности; в случае несоблюдения условия производится отдельное рассмотрение фактора.

В дальнейшем представляется возможность использовать полученные априорные информативные веса признаков для уточнения их апостериорных информативных весов, полученных с помощью ЭВМ. Это должно повысить точность классификации прогнозируемых (оцениваемых) объектов.

Следует отметить, что наряду с выделением существенных признаков изложенные методы обработки экспертных оценок могут быть эффективно использованы для выбора эталонных объектов, что дает возможность составить достаточно представительную обучающую выборку и, следовательно, заложить более качественные основы решения задачи прогнозирования.

3. РАНЖИРОВАНИЕ ПРИЗНАКОВ ПО ДАННЫМ ВЫБОРКИ ОБУЧЕНИЯ В ПРЕДПОЛОЖЕНИИ ИХ НЕЗАВИСИМОСТИ

В настоящем разделе рассматриваются методы выбора и ранжирования по информативности геологических признаков по выборке обучения, т. е. по фактическим данным, характеризующим изучаемые классы объектов. Если ранее в основу ранжирования и выбора существенных признаков закладывались априорные сведения — экспертные оценки, то сейчас предполагается, что исследователь приступил к непосредственному изучению объектов и собрал достаточный объем апостериорной геологической информации.

Эти два подхода к ранжированию признаков не противоречат, а дополняют друг друга. Сначала геолог, приступивший к решению некоторой задачи прогнозирования, пытается собрать полезную информацию по интересующему его вопросу (по вопросу выбора существенных признаков для использования в задаче) у квалифицированных специалистов, имеющих отношение к проблеме. Отобрав и проанализировав с тем или иным уровнем доверия экспертные оценки, исследователь определяет, какие признаки необходимо измерять и учитывать при описании классов явлений, а какие не надо. Затем, приступив к эмпирическому изучению объектов, он собирает фактический материал, формирует выборку обучения (если возможно — контрольную выборку) и по результатам обработки собранной информации окончательно формирует пространство признаков, в котором целесообразно решать задачу прогнозирования. Фактические данные, содержащиеся в выборке обучения, позволяют исследователю уточнить априорные оценки, данные экспертами, выявить наиболее информативные признаки и вторично сократить размерность пространства измерений в целях эффективного решения задачи.

Ранжирование признаков по выборке обучения можно проводить либо в предположении их статистической независимости, либо, наоборот, в предположении существования между исходными признаками взаимозависимости. Если на основании некоторых априорных сведений принимается первое из названных предположений, то выбор существенных признаков производится путем их независимого рассмотрения и ранжирования по информативности согласно некоторому заданному критерию, характеризующему взаимное расположение классов и распределение объектов в них по каждому из признаков.

Оценка информативности статистически зависимых признаков требует анализа многомерных распределений объектов в пространстве измерений.

Все известные алгоритмы выбора существенных признаков до решения самой задачи классификации не могут считаться оптимальными, однако их использование способствует сокращению размерности и формированию в той или иной степени информативного пространства признаков.

Метод распределений

Рассмотрим возможные варианты плотности распределения вероятностей параметров x_i для классов A, B — $P(x_i/A)$ и $P(x_i/B)$. Есте-

* В качестве параметра, оцениваемого на информативность, может быть принята и некоторая функция от исходных параметров $\varphi(x_1, \dots, x_n)$; например в виде x_i, x_j, x_p, x_q и т. д.

венно, что качество классификации будет наилучшим по тому параметру, по которому максимально разнесены распределения отдельных классов. Следовательно, распознавание будет безошибочным в том случае, когда эти распределения полностью не совпадут, т. е.

$$P(x_i/A) \geq 0 \text{ только при } P(x_i/B) = 0 \text{ и}$$

$$P(x_i/B) \geq 0 \text{ только при } P(x_i/A) = 0.$$

Если все параметры измеряются в одном масштабе, то в качестве меры разделимости классов можно использовать величину, пропорциональную разности математических ожиданий

$$\delta x_i = a_i |m^A_{x_i} - m^B_{x_i}|, \quad (\text{III.20})$$

где $m^A_{x_i} = M\{x_i/A\}$ и $m^B_{x_i} = \{x_i/B\}$;

M — символ математического ожидания;

a_i — весовой коэффициент, $a_i > 0$.

К такой оценке можно перейти, например, при стандартизации исходной информации, т. е. вместо переменных x_i рассматривать новые переменные

$$z_i = \frac{x_i - m_{x_i}}{\sigma_{x_i}}, \quad (\text{III.21})$$

где m_{x_i} — математическое ожидание x_i .

Тогда все новые переменные z_i будут иметь математические ожидания m_{z_i} и среднеквадратичные отклонения $\sigma_{z_i} = 1$. Если же несколько параметров имеют одинаковые значения разности математических ожиданий, но измеряются в разных масштабах, то целесообразно наиболее ценным принять параметр, имеющий меньшую дисперсию по классам.

Таким образом, среди всех рассматриваемых параметров наилучшим для целей классификации следует считать тот параметр x_i , который удовлетворяет следующим двум условиям:

а) дисперсии параметров x_i по классам A и B минимальны:

$$\begin{aligned} (\sigma^A_{x_i})^2 &= M\{(x_i - m^A_{x_i})^2\}, \\ (\sigma^B_{x_i})^2 &= M\{(x_i - m^B_{x_i})^2\}, \end{aligned} \quad (\text{III.22})$$

что практически соответствует условию

$$(\sigma^A_{x_i})^2 + (\sigma^B_{x_i})^2 \rightarrow \min;$$

б) дисперсия параметра x_i максимальна

$$(\sigma_{x_i})^2 = M\{(x_i - m_{x_i})^2\} \rightarrow \max. \quad (\text{III.23})$$

Рассмотрим второе условие:

$$\begin{aligned} (\sigma_{x_i})^2 &= M\{(x_i - m_{x_i})^2\} = M\{[x_i - P(A)m^A_{x_i} - P(B)m^B_{x_i}]^2\} = \\ &= M\{[(x_i - m^A_{x_i}) + P(B)(m^A_{x_i} - m^B_{x_i})]^2\} = \\ &= M\{(x_i - m^A_{x_i})^2 + 2P(B)(m^A_{x_i} - m^B_{x_i}) \times \\ &\quad \times M\{x_i - m^A_{x_i}\} + P^2(B)(m^A_{x_i} - m^B_{x_i})^2\}. \end{aligned} \quad (\text{III.24})$$

Здесь $P(A)$ и $P(B)$ — априорные вероятности классов A и B .

$$M\{(x_i - m^A_{x_i})^2\} = P(A)M_A\{(x_i - m^A_{x_i})^2\} +$$

$$+P(B)M_B\{(x_i - m^A_{x_i})^2\} = (\sigma^A_{x_i})^2 P(A) + P(B)(\sigma^B_{x_i})^2 + P(B)(m^B_{x_i} - m^A_{x_i})^2. \quad (\text{III.25})$$

$$M\{(x_i - m^A_{x_i})\} = P(B)(m^B_{x_i} - m^A_{x_i}). \quad (\text{III.26})$$

$$(\sigma_{x_i})^2 = P(A)(\sigma^A_{x_i})^2 + P(B)(\sigma^B_{x_i})^2 + P(A) \cdot P(B)(m^A_{x_i} - m^B_{x_i})^2. \quad (\text{III.27})$$

Если $N_A = N_B = \frac{N}{2}$, то $P(A) = P(B) = 0,5$.

$$(\sigma_{x_i})^2 = \frac{(\sigma^A_{x_i})^2 + (\sigma^B_{x_i})^2}{2} + \frac{1}{4}(m^A_{x_i} - m^B_{x_i})^2. \quad (\text{III.28})$$

После преобразования

$$\left[\frac{m^A_{x_i} - m^B_{x_i}}{\sigma^A_{x_i} - \sigma^B_{x_i}} \right]^2 = 4 \frac{(\sigma_{x_i})^2 - 0,5 [(\sigma^A_{x_i})^2 + (\sigma^B_{x_i})^2]}{(\sigma^A_{x_i} - \sigma^B_{x_i})^2}. \quad (\text{III.29})$$

Обозначим

$$\frac{m^A_{x_i} - m^B_{x_i}}{\sigma^A_{x_i} - \sigma^B_{x_i}} = \Delta x_i. \quad (\text{III.30})$$

Тогда

$$\Delta x_i = 4 \frac{(\sigma_{x_i})^2 - 0,5 [(\sigma^A_{x_i})^2 + (\sigma^B_{x_i})^2]}{(\sigma^A_{x_i} - \sigma^B_{x_i})^2}. \quad (\text{III.31})$$

При выполнении указанных выше условий величина Δx_i будет максимальной. Этим определяется возможность использования критерия Δx_i для ранжировки параметров с унимодальным распределением. Однако следует отметить, что достоверная классификация по одному параметру x_i возможна (в случае распределения $P(x_i/A)$ и $P(x_i/B)$ близкого к нормальному) лишь в случае, когда $\Delta x_i \geq 3$. Если оценка оказывается меньше этой величины, нахождение классифицирующей функции требует привлечения большого числа переменных.

Рассмотренный метод оценки обладает существенным недостатком, который резко сужает области его использования: он неприменим, если распределения переменных по классам сильно отличаются от нормальных. Так, при распределении многосвязанных областей классов оценка по критерию Δx_i забракует параметр, хотя классификация по нему может быть наилучшей.

Корреляционный метод

Информативность геологических параметров может быть установлена на основе определения степени взаимосвязи отдельных параметров и их сочетаний с признаком принадлежности к классу объектов y , например путем расчета коэффициента корреляции параметра x_i с величиной y

$$r(x_i, y) = \frac{M\{(x_i - m_{x_i})(y - m_y)\}}{\sigma_{x_i} \sigma_y}, \quad (\text{III.32})$$

где m_{x_i} , m_y — значения математических ожиданий x_i и y ; σ_{x_i} и σ_y — среднеквадратичные отклонения x_i и y .

В реальных задачах коэффициент корреляции $r(x_i, y)$ оценивается по конечной выборке

$$\hat{r}(x_i, y) = \frac{\sum_{j=1}^N (x_{ji} - \bar{X}_i)(y_j - \bar{Y})}{\sqrt{\sum_{j=1}^N (x_{ji} - \bar{X}_i)^2 \sum_{j=1}^N (y_j - \bar{Y})^2}}. \quad (\text{III.33})$$

В этом равенстве

$$\bar{X}_i = \frac{1}{N} \sum_{j=1}^N x_{ji} \quad \text{и} \quad \bar{Y} = \frac{1}{N} \sum_{j=1}^N y_j, \quad (\text{III.34})$$

где N — число объектов в выборке.

Ранжировка параметров по корреляционному критерию в значительной мере должна отвечать ранжировке, полученной по методу распределений, так как величины Δx_i и $r(x_i, y)$ пропорциональны размерности математических ожиданий параметра x_i по классам A и B .

В самом деле, если

$$y_j = \begin{cases} +1 & \text{при } \bar{X}_j \in A \\ -1 & \text{при } \bar{X}_j \in B, \end{cases} \quad (\text{III.35})$$

то $m_j = P(A) - P(B)$.

$$M_{C_j} \{ (x_i - m_{x_i})(y - m_y) \} = 2P(A)P(B) [m_{x_i}^A - m_{x_i}^B],$$

$$r(x_i, y) = 2P(A)P(B) \frac{m_{x_i}^A - m_{x_i}^B}{\sigma_{x_i} \sigma_y}. \quad (\text{III.36})$$

$$\begin{aligned} (\sigma_y)^2 &= M \{ (y - m_y)^2 \} = 1 - m_y^2 = [1 - 2P(B)]^2 = \\ &= 4P(B)[1 - P(B)] = 4P(A)P(B), \end{aligned}$$

$$r(x_i, y) = \sqrt{P(A)P(B)} \cdot \frac{m_{x_i}^A - m_{x_i}^B}{\sigma_{x_i}}. \quad (\text{III.37})$$

M_{C_j} означает, что математическое ожидание берется для объектов из класса C_j ($C_j = A, B$).

Так как ранжировки параметров по методу распределений и коэффициентам корреляций являются зависимыми, что обычно подтверждается данными экспериментальных исследований, оценку информативности параметров, по-видимому, следует проводить только по одному из этих методов.

Метод информативных интервалов *

Предположим, что все элементы обучающей выборки попали в интервал OM переменной x , которая может принадлежать одному из двух классов A и B . Для удобства дальнейших рассуждений будем считать, что класс A не пересекается с классом B , а все N изображений выборки находятся в целочисленных точках $0, 1, 2, \dots, M$, примем $M \geq N$.

* Метод разработан Б. А. Чумаченко и Е. П. Власовым.

Зададим функцию $f(x)$ в целочисленных точках $0, 1, \dots, M$ следующим образом:

$$\begin{aligned} f(x) &= 0, \text{ если } P_A(x) = P_B(x) = 0, \\ f(x) &= 1, \text{ если } P_A(x) > P_B(x), \\ f(x) &= -1, \text{ если } P_A(x) < P_B(x). \end{aligned}$$

Функцию $\varphi(x)$, определенную в тех же точках и принимающую значения $0, 1$ и -1 , будем называть допустимой, если длина интервала постоянства функции $\varphi(x)$ не меньше Δx_{\min} — минимальный интервал, на котором возможным является определение $P_A(\Delta x_{\min})$ и $P_B(\Delta x_{\min})$.

Сущность метода информативных интервалов сводится к процедуре построения оптимальной допустимой функции и оценке информативности интервалов постоянства.

В классе допустимых функций $\varphi(x)$ ищется такая функция $\varphi^*(x)$, для которой величина

$$\begin{aligned} F(O, M) = \sum_{i=0}^M \left\{ & |[\varphi^*(i) - f(i)]| \frac{1}{2} f(i) \varphi^*(i) + \right. \\ & \left. + |\varphi^*(i)(1 - |f(i)|)| + |f(i)(1 - |\varphi^*(i)|)| \right\} \end{aligned} \quad (\text{III.38})$$

принимает наименьшее значение.

Величина $F(O, M)$ характеризует ошибку, возникающую при аппроксимации функции $f(x)$, допустимой функцией $\varphi^*(x)$. При этом, согласно критерию (III.38), в каждой целочисленной точке при условии $f(i) \neq \varphi^*(i)$ начисляется ошибка, равная единице. Остановимся на существе процедуры определения оптимальной в смысле (III.38) функции $\varphi^*(x)$, впервые предложенной в работе [125].

Введем необходимые определения. Точку A будем называть узлом допустимой функции $\varphi(x)$, если точка A есть первый конец интервала постоянства (скачок) $\varphi(x)$, если на интервале $(A - \Delta x_{\min} + 1, A + \Delta x_{\min})$ функция не меняет своего значения.

Легко видеть, что если точка A является узлом, то $\varphi(x)$ является допустимой как на интервале (O, A) , так и на интервале $(A + 1, M)$. Отсюда следует, что если оптимальная на (O, M) функция $\varphi^*(x)$ имеет узлом точку A , то $\varphi^*(x)$ является оптимальной на интервалах (O, A) и $(A + 1, M)$. Назовем ценой точки B величину

$$\begin{aligned} F(O, B) = \sum_{i=1}^B \left\{ & |[\varphi^*(i) - f(i)]| \frac{1}{2} f(i) \varphi^*(i) + \right. \\ & \left. + |\varphi^*(i)(1 - |f(i)|)| + |f(i)(1 - |\varphi^*(i)|)| \right\}, \end{aligned} \quad (\text{III.39})$$

где $\varphi^*(x)$ оптимальная на (O, B) .

Покажем, как найти цену точки B , если известны цены точек $0, 1, 2, \dots, B - 1$.

Пусть $\varphi^*(x)$ — функция, оптимальная на (O, B) . Покажем, что ближайший в B узел функции $\varphi^*(x)$ лежит на интервале $(B - 2\Delta x_{\min} + 1, B - \Delta x_{\min})$. Действительно, точка B — конец интервала постоянства $\varphi^*(x)$ длины не меньше Δx_{\min} . Если длина этого интервала меньше $2\Delta x_{\min}$, то конец предыдущего интервала и узел $\varphi^*(x)$ лежат на отрезке $(B - 2\Delta x_{\min}, B - \Delta x_{\min})$. Если же длина этого интервала больше $2\Delta x_{\min} - 1$, то точка $B - \Delta x_{\min}$ сама является узлом.

Пусть A — ближайший к B узел $\varphi^*(x)$. В таком случае $\varphi^*(x)$ оптимальна на интервалах (O, A) и $A+1, B)$, и цена точки равна

$$\sum_{i=1}^B |\varphi^*(i) - f(x)| = \sum_{i=0}^A |\varphi^*(i) - f(i)| + \sum_{i=0}^B |\varphi^*(i) - f(i)|. \quad (\text{III.40})$$

Первое слагаемое в этой сумме равно цене точки A , а второе легко находится. Действительно, пусть k — число точек интервала $(A+1, B)$, в которых $f(x) = -1$; l — число точек интервала $(A+1, B)$, в которых $f(x) = 1$; m — число точек интервала $(A+1, B)$, в которых $f(x) = 0$; ($k+m+l = B-A$). Так как $\varphi^*(x)$ постоянная и оптимальна на $(A+1, B)$, то $\varphi^*(x) = 1$ на этом интервале, если $l > k, m$; $\varphi^*(x) = 0$, если $m > k, l$; $\varphi^*(x) = -1$, если $k > l, m$. Следовательно, значение второго слагаемого в (III.40) равно $\min(k+l, k+m, l+m)$. Таким образом, левая часть (III.40) легко определяется, если известна точка A .

Так как для любой точки A интервала $(B-2\Delta x_{\min}+1, B-\Delta x_{\min})$ можно построить допустимую функцию, то оптимальная функция $\varphi^*(x)$ дает оценку точке B , равную минимуму цен B при условии, что ближайшим к B узлом являются точки $B-\Delta x_{\min}, B-\Delta x_{\min}-1, \dots, B-2\Delta x_{\min}+1$. Найдя этот минимум, можно получить цену точки B , а также ближайший к B узел оптимальной функции $\varphi^*(x)$.

Этим способом определяются цены и ближайшие узлы всех последующих точек, включая точку M . Дойдя до точки M , легко восстановить оптимальную функцию $\varphi^*(x)$. Она постоянна на интервале от M до ближайшего к M узла M_1 , от M_1 до ближайшего к M_1 узла M_2 и т. д.

Значения, принимаемые функцией $\varphi^*(x)$ на каждом из этих интервалов, также определяются в процессе вычислений.

Следует отметить, что изложенный алгоритм определения оптимальной $\varphi^*(x)$ можно использовать и для случая пересекающихся классов A и B .

Пусть в общем случае по обучающей выборке i -ая точка n_i^A раз попадает в класс A и n_i^B раз в класс B .

Тогда критерий (III.38) можно записать следующим образом:

$$F(O, M) = \sum_{i=0}^M \left\{ |\varphi^*(i) f(i) \frac{n_i^A - n_i^B}{2} (f(i) - \varphi^*(i))| + |\varphi^*(i) (1 - |f(i)|)| + |(n_i^A - n_i^B) (1 - |\varphi^*(i)| f(i))| \right\}, \quad (\text{III.41})$$

где

$$f(i) = \begin{cases} 1 & \text{при } n_i^A > n_i^B \\ -1 & \text{при } n_i^A < n_i^B \\ 0 & \text{при } n_i^A = n_i^B. \end{cases}$$

В критерии (III.41) ошибка аппроксимации в случае $f(i) \neq 0$ и $f(i) \neq \varphi^*(i)$ равна $n_i^A - n_i^B$, а в случае $f(i) = 0$ и $\varphi^*(i) \neq 0$ — единице.

Для всех интервалов постоянства функции $\varphi^*(i)$, кроме интервалов, на которых $\varphi^*(x) = 0$, подсчитывается величина J_j , характеризующая информативность j -го интервала

$$J_j = K_j^n \cdot K_j^k \cdot K_j^y, \quad (\text{III.42})$$

где K_j^n — коэффициент представительности j -го интервала;

K_j^k — коэффициент компактности j -го интервала;

K_j^y — коэффициент удаленности изображений, попавших в j -ый интервал одного знака, от изображений, попавших в тот из интерва-

лов противоположного знака (j' интервал), который является ближайшим к рассматриваемому j -у интервалу.

В явном виде выражение (III.42) запишется так:

$$J_j = \frac{|n_j^A - n_j^B|}{N_A - N_B} \cdot \frac{OM}{\Delta x_j} \cdot \frac{|\mu(x_j) - \mu(x_{j'})|}{OM} = \frac{|n_j^A - n_j^B| |\mu(x_j) - \mu(x_{j'})|}{(N_A + N_B) \Delta x_j}, \quad (\text{III.43})$$

где n_j^A , n_j^B — количество изображений, принадлежащих соответственно классу A и классу B и попавших в j -ый интервал;

$\mu(x_j)$, $\mu(x_{j'})$ — выборочные средние по интервалам j и j' соответственно;

N_A , N_B — количество изображений обучающей выборки, принадлежащих соответственно классу A и классу B ;

Δx_j — длина j -го интервала постоянства. Оценка информативности $J(x)$ параметра x вычисляется по формуле

$$J(x) = \frac{1}{K} \sum_{j=1}^k J_j(x), \quad (\text{III.44})$$

где $J_j(x)$ — информативность j -го интервала постоянства функции $\Phi^*(x)$;

K — количество интервалов постоянства функции $\Phi^*(x)$, не равных 0.

Параметры $x_1, \dots, x_l, \dots, x_n$ ранжируются по информативности в порядке убывания их оценок $J(x)$.

Метод информативных интервалов по существу является развитием метода распределений в том смысле, что он позволяет оценивать информативность параметра при любых законах плотности распределения вероятности классов.

Как отмечалось ранее, методы оценки информативности отдельных параметров или их комбинаций (факторов) разрабатываются и применяются при решении задач классификации для формирования наиболее информативного подпространства параметров. Нахождение такого подпространства, отсеивание неинформативных и отбор существенных факторов среди исходных — все это, с одной стороны, создает определенные предпосылки для успешного решения общей задачи классификации, с другой — способствует направленному изучению геологических объектов и минимизации их описания.

4. РАНЖИРОВАНИЕ ПРИЗНАКОВ ПО ДАННЫМ ВЫБОРКИ ОБУЧЕНИЯ В ПРЕДПОЛОЖЕНИИ СУЩЕСТВОВАНИЯ ЗАВИСИМОСТИ МЕЖДУ НИМИ

В общем случае при решении большинства задач выбор и оценку информативности отдельных признаков для формирования исходного пространства измерений нельзя производить безотносительно к другим рассматриваемым признакам. Это связано с тем, что многие признаки, характеризующие реальные физические процессы, являются зависимыми друг от друга величинами. Поэтому выбор существенных признаков по величине их индивидуальной информативности является неправомерным.

В самом деле, пусть, например, некоторые два признака характеризуются наиболее высокой индивидуальной информативностью по сравнению с другими измеряемыми признаками и имеют коэффициент взаимной корреляции, близкий к единице. Тогда при учете их статистической зависимости какой-либо один из этих признаков целесообразно исключить из формируемого пространства измерений, поскольку его

использование не увеличит количества полезной информации, содержащейся в описании явлений. В то же время этот признак будет отобран как информативный методами ранжирования, изложенными в предыдущем разделе, что приведет к необоснованному увеличению размерности решаемой задачи. Кроме того, применение методов ранжирования, не учитывающих существующую взаимосвязь между факторами, может привести к исключению существенных признаков, которые имеют невысокую индивидуальную информативность, но зато в совокупности с другими признаками обладают значительной так называемой «дополнительной» информативностью.

Наиболее эффективным подходом при решении задачи выбора информативного подпространства статистически зависимых признаков является поиск этого подпространства в процессе обучения классификации и построения оптимального, в смысле надежности распознавания, решающего правила. В этом случае задача выбора подпространства информативных признаков становится неотъемлемой составной частью задачи классификации, а естественные численные критерии, подобные вероятности ошибки классификации, используются для оценки эффективности всей системы распознавания в целом.

К работам, в которых процесс построения информативного подпространства рассматривается в единой связи с процессом обучения классификации, можно отнести публикацию Г. С. Лбова [62] и работу [162].

В работе [62] с помощью эвристического метода находится оптимальное подпространство $T^m_{\text{опт}}$ исходного n -мерного пространства признаков, в котором оценочный функционал

$$P(T^m) = \sum_{i=1}^k q_i \sum_{j=1}^k \frac{\alpha(j/i)}{N_i} \sum_{i=1}^k q_i \frac{\alpha_i}{N_i} \quad (\text{III.45})$$

принимает минимальное значение.

Здесь q_i — априорная вероятность появления i -го класса;

N_i — число объектов i -го класса в обучающей выборке;

$\alpha(j/i)$ — число точек множества M_i , попавших в некоторые области Y_i ;

K — количество классов;

M_i — множество точек, соответствующих i -у классу в T^m ;

Y_i — непересекающиеся области, на которых разбито пространство T^m по числу классов и которые определяются следующим

образом: $\bar{X} \in Y_i$,

если $\alpha_i P_j'(\bar{X}) \geq q_i P_i'(\bar{X})$, $i = 1, 2, \dots, K$; $i \neq j$,

где $P_i'(\bar{X})$ — плотность распределения вероятности по классам;

α_i — число точек множества H_i , попавших в $Y_1, \dots, Y_{i-1}, \dots, Y_{i+1}, \dots, Y_K$.

Вначале, согласно алгоритму, делается предположение, что

$$P_1 = \dots = P_n = \frac{1}{n},$$

где P_j — вероятность включения j -го признака в формируемое подпространство. Затем случайно выбираются подпространства размерности $m < n$, ($T_{11}^m, \dots, T_{21}^m$) и определяются $P_i'(\bar{X})$. Из выбранных подпространств находятся T^m_{min} и T^m_{max} , дающие $P(T^m)_{\text{min}}$ и $P(T^m)_{\text{max}}$ из $P(T_1^m), \dots, P(T_2^m)$. Далее увеличивают вероятности выбора каждого из признаков, составивших T^m_{min} , на некоторую добавочную величину h

и уменьшают на ту же величину вероятность выбора признаков, составивших T_{\max}^m . После этого производится выбор новых m -мерных подпространств $T_{12}^m, \dots, T_{22}^m$ и процедура повторяется.

В случае, если $\min\{P(T_{rp}^m)\}_{\min}$ не меняются на протяжении некоторого числа повторяющихся циклов, процедура прекращается.

Основным недостатком рассматриваемого метода поиска информативного подпространства является громоздкость вычислительной процедуры.

Метод сокращения исходного L -мерного пространства признаков в работе [162] основан на непараметрической оценке функций плотности вероятности и использовании некоторого критерия разнесенности распознаваемых классов объектов в пространстве измерений. Аппроксимация функции плотности вероятности $f(\bar{X}/A_i)$ строится в предположении, что $f(\bar{X}/A_i)$ принадлежит семейству кусочно-непрерывных функций, а само преобразование пространства $X=R^L$ в $M=R^l$, ($l < L$) осуществляется с помощью линейного преобразования, задаваемого матрицей $A(l \times L)$, т. е. для $\bar{X} \in X$ и $\bar{X} \in M$ выполняется $\bar{X}' = AX$,

где

$$A = \begin{pmatrix} a_{11}, \dots, a_{1L} \\ \cdot \quad \cdot \quad \cdot \\ a_{l1}, \dots, a_{lL} \end{pmatrix}.$$

Алгоритм сокращения размерности пространства признаков до величины l сводится к использованию градиентного метода, максимизирующего разнесенность распознаваемых классов объектов в R^l .

Этот метод менее громоздкий, чем предыдущий, однако применение его на практике связано с неопределенностью выбора значения $l < L$, при котором точность классификации объектов остается в пределах заданной величины. Этот же недостаток присущ и методу Лбова [62].

Одним из методов распознавания образов, в котором найдено оригинальное и интересное решение совмещения процессов обучения классификации и определения информативного подпространства признаков, является метод В. С. Стрижевского [116]. Не останавливаясь подробно на изложении метода, коротко поясним его сущность.

Пусть имеется в n -мерном пространстве измерений x_1, x_2, \dots, x_n два конечных множества векторов или точек $\bar{X}_p \in A$ и $\bar{X}_q \in B$, $p=1, 2, \dots, N_A$; $q=1, 2, \dots, N_B$. Тогда процесс обучения классификации можно пояснить следующим образом:

а) строятся проекции векторов \bar{X}_p и \bar{X}_q на координаты x_1, x_2, \dots, x_n ;

б) находятся интервалы $(x_{1 \min} x_{1 \max}, \dots, x_{n \min} x_{n \max})$ такие, что все проекции x_{ip} и x_{iq} ($i=1, 2, \dots, n$) векторов \bar{X}_p и \bar{X}_q принадлежат $(x_{i \min} x_{i \max})$, причем сами точки $x_{i \min}$ и $x_{i \max}$ принадлежат множеству $\{x_{ip}, x_{iq}\}$;

в) все интервалы $(x_{i \min} x_{i \max})$ разбиваются на заранее заданное число r непересекающихся и примыкающих друг к другу отрезков $(x_{i0} x_{i1}), \dots, (x_{ij} x_{ij+1}), \dots, (x_{i(r-1)} x_{ir})$, где $x_{i0} = x_{i \min}$ и $x_{ir} = x_{i \max}$;

г) из всех (rx_n) отрезков находятся такие, на которых лежат проекции векторов какого-либо одного из классов A или B ; затем из последних выбирается и запоминается такой отрезок $(x_{ij} x_{ij+1})_{\max}^1$, на котором оказалось максимальное количество проекций исходных множеств \bar{X}_p и \bar{X}_q ;

д) вектора множества \bar{X}_p и \bar{X}_q , проекции которых лежат на отрезке $(x_{ij} x_{ij+1})_{\max}^1$, из дальнейшего рассмотрения исключаются.

Затем процедуры, перечисленные в пунктах а—д, повторяются, находится следующий отрезок $(x_{ij}x_{ij+1})^2_{\max}$, и так продолжается до тех пор, пока все исходное множество векторов не будет исчерпано. В этом случае процесс обучения классификации считается законченным.

Предположим, что процесс обучения для исходной выборки векторов заканчивается на K -й итерации вместе с нахождением отрезка $(x_{ij}x_{ij+1})^K_{\max}$. Тогда некоторый новый вектор $\bar{X}(x_1, \dots, x_n)$ по данному алгоритму будет классифицирован следующим образом. Пусть проекция вектора принадлежит следующим из K найденных на этапе обучения отрезкам: $(x_{ij}x_{ij+1})^{\xi_1}_{\max}, \dots, (x_{ij}x_{ij+1})^{\xi_t}_{\max}$, причем $1 \leq \xi_1 < \xi_2 < \dots < \xi_t \leq K$. В этом случае вектор \bar{X} будет отнесен к тому из классов A или B , к которому принадлежат те векторы из множества $\{\bar{X}_p, \bar{X}_q\}$, чьи проекции лежат на отрезке $(x_{ij}x_{ij+1})^{\xi_1}_{\max}$.

Из рассмотренного алгоритма классификации легко видеть, что набор отрезков $(x_{ij}x_{ij+1})^1_{\max}, \dots, (x_{ij}x_{ij+1})^K_{\max}$ полностью определяет наиболее эффективную систему измерений, поскольку он является достаточным для получения надежного решения задачи и в то же время он не является избыточным. К этому же классу алгоритмов относятся рассмотренные в разделе III метод информативных интервалов и метод информативных гиперсфер. В названных методах процесс обучения также органически связан с направленной процедурой поиска эффективной системы признаков, причем процедура сводится к нахождению таких областей в различных подпространствах, которые в смысле некоторого заданного критерия являются наиболее информативными.

Следует особо отметить, что проблему сокращения пространства зависимых признаков нецелесообразно рассматривать отдельно от проблемы распознавания образов, ибо оценка качества решения первой проблемы может быть дана только вместе с анализом надежности системы распознавания в целом.

5. СТАТИСТИЧЕСКИЕ МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ

Оптимальное решающее правило

Пусть объекты, принадлежащие каждому из N классов, могут появляться случайно в соответствии с N функциями распределения вероятности.

Введем функцию $P(\bar{X}/A_i); i=1, 2, \dots, N$, где $P(\bar{X}/A_i)$ — вероятность появления объекта X при условии, что он принадлежит A_i классу. Пусть также $P(A_i)$ есть априорная вероятность появления объекта из i -го класса, а величина C_{ij} при $i, j=1, 2, \dots, N$ представляет потери (цену ошибки), которые возникают, когда распознающее устройство относит к j -у классу объект, в действительности принадлежащий к i -у классу. Если P_{ij} — вероятность отнесения объекта i -го класса к j -у, то суммарные потери распознавания (риск распознавания) могут быть записаны в виде

$$R = \sum_{i=1}^N \sum_{j=1}^N C_{ij} P_{ij}. \quad (\text{III.46})$$

Решающее правило будем называть оптимальным, если при его использовании величина R становится минимальной. Минимум суммарных потерь достигается, если использовать следующую решающую процедуру [57].

Вначале вычисляются величины

$$P(A_i/\bar{X}) = \frac{P(\bar{X}/A_i)P(A_i)}{P(\bar{X})},$$

где $P(A_i/\bar{X})$ — вероятность того, что данный объект \bar{X} принадлежит к A_i классу;

$P(\bar{X})$ — вероятность появления объекта \bar{X} , равная

$$P(\bar{X}) = \sum_{i=1}^N P(A_i)P(\bar{X}/A_i). \quad (\text{III.47})$$

Вычисление величины $P(A_i/\bar{X})$ производится по данным выборки обучения. Затем принимается решение (правило Байеса), что $\bar{X} \in A_i$, если

$$C_{ij}P(A_i/\bar{X}) - C_{ji}P(A_j/\bar{X}) > 0. \quad (\text{III.48})$$

$i, j = 1, 2, \dots, N; i \neq j.$

Использование критерия Байеса эквивалентно тому, что все пространство измерений разбивается на N областей. Граница между областями A_i и A_j будет проходить через точки, в которых выполняется условие

$$C_{ij}P(A_i)P(\bar{X}/A_i) = C_{ji}P(A_j)P(\bar{X}/A_j). \quad (\text{III.49})$$

Эта поверхность называется оптимальной решающей границей. В зависимости от вида функции $P(A_i)P(\bar{X}/A_i)$ решающее правило представляет собой гиперповерхность той или иной степени сложности.

Использование правила Байеса (III.48) при решении практических задач связано с большими трудностями, так как обычно условные плотности вероятностей $P(\bar{X}/A_i)$ не известны и их приходится оценивать по выборке обучения. В наиболее простом случае, если из каких-либо априорных сведений можно задаться видом законов распределения классов в пространстве измерений, то процесс обучения классификации сводится к вычислению оценок параметров этих распределений по ограниченной выборке обучения. Реализация правила Байеса осложняется, если величины $P(\bar{X}/A_i)$ не удается выразить аналитически. В этом случае необходимо по обучающей выборке вычислять и запоминать значения функций $P(\bar{X}/A_i)$ в каждой точке или некоторой достаточно малой области n -мерного пространства признаков. Но тогда объем памяти, необходимый для запоминания всех вычисленных величин, настолько велик, что такой подход нельзя реализовать, за исключением немногих простейших случаев.

При распознавании классов объектов, для которых условные плотности вероятностей $P(\bar{X}/A_i)$ заранее не заданы и не могут быть записаны в аналитической форме, используют аппроксимации неизвестных функций распределения по выборке обучения.

Разные методы аппроксимации неизвестных распределений приводят к различным алгоритмам классификации. Процесс обучения распознавания будет эффективным в том случае, если строить и проверять на исходной выборке последовательно и целенаправленно усложняющиеся правила классификации, которые учитывали бы более глубокие связи моделируемых явлений. Так последовательно усложняя вид аппроксимирующей функции путем добавления членов уравнения более высокого порядка, можно отыскать решающее правило, которое было бы наилучшим в смысле минимума ошибки классификации.

Исследование рекомендуется начинать с модели статистической независимости признаков. В этом случае функции плотности распределения можно определить следующим образом:

$$P(\bar{X}/A_j) = P(x_1/A_j) \cdot \dots \cdot P(x_i/A_j) \cdot \dots \cdot P(x_n/A_j), \quad (\text{III.50})$$

где $P(x_i/A_j)$ — вероятность появления признака x_i в классе A_j ;
 n — размерность решаемой задачи.

Если в смысле некоторого заданного критерия оптимальности аппроксимация (III.50) оказывается неудовлетворительной, то нужно учитывать статистическую зависимость между признаками x_i .

Последовательно улучшая все более глубокие статистические связи между признаками, можно построить оптимальную аппроксимацию искомых функций распределения классов объектов.

Отметим, что в одном из разработанных алгоритмов (см. ниже) успешно использовалась при прогнозной оценке аппроксимация вида

$$P(\bar{X}/A_j) = \prod_{i=1}^n P(x_m/\bar{X}_{j1(m_i)}, \dots, \bar{X}_{jK(n_i)}) A_i,$$

где K — порядок аппроксимации;

m_1, m_2, \dots, m_n — некоторая перестановка натуральных чисел от 1 до n ;

$jl(m_i)$; $l=1, \dots, K$ — некоторые целочисленные функции, удовлетворяющие для любых l и i условию: если $jl(m_i) = m_q$, то $i > q$.

Недостатком изложенного выше подхода является то, что для его реализации, как правило, требуется значительное машинное время.

Метод классификации на основе аппроксимации вероятностного распределения произведением условных вероятностей *

Пусть требуется установить принадлежность предъявленного объекта, описанного вектором измерений $\bar{X} = (x_1, x_2, \dots, x_n)$, к одному из N известных классов a_j ; ($1 \leq j \leq N$). При этом последние определены совместной плотностью распределения $P(x_1, x_2, \dots, x_n/a_j)$ и априорной вероятностью появления изображений каждого класса $P(a_j)$. Тогда в соответствии с правилом Байеса оптимальная процедура принятия решений при условии равенства цен всех ошибок сводится к определению совместных вероятностей

$$P(\bar{X}/a_j) = P_j P(x_1, x_2, \dots, x_n/a_j); \quad j=1, 2, \dots, N \quad (\text{III.51})$$

и отнесению объекта \bar{X} к тому классу, для которого значение соответствующей совместной вероятности максимально. Заметим, что если каждое измерение x_i ; ($i=1, 2, \dots, n$) принимает лишь два значения, то количество возможных значений $P(\bar{X}/a_j)$, которое необходимо определить по обучающей выборке, если совместные плотности заранее не заданы, и запомнить для целей классификации, равно $N \cdot 2^n$.

Последнее при решении большинства задач геологического прогнозирования существенно ограничивает применимость вероятностного подхода.

Рассмотрим выражение, известное из теории вероятностей,

$$P(x_1, x_2, \dots, x_n/a_j) = P(x_1/a_j) \times \\ \times P(x_2/x_1, a_j) \cdot \dots \cdot P(x_n/x_{n-1}, \dots, x_1, a_j). \quad (\text{III.52})$$

* Исследование выполнено Б. А. Чумаченко, Е. П. Власовым, Г. Н. Кричевцом.

Если предположить, что параметры $x_i; i=1, \dots, n$ статистически независимы, то формула (III.52) примет следующий вид:

$$P(x_1, x_2, \dots, x_n/a_j) = P(x_1/a_j) \cdot \dots \cdot P(x_n/a_j). \quad (\text{III.53})$$

Отсюда следует, что при выполнении условия статистической независимости параметров оптимальная классификация будет состоять в оценке Nn возможных значений $P(x_i/a_j)$ и последующем определении максимальной величины (III.51).

Принятие условия (III.53) при разбраковке геологических объектов с целью прогнозирования промышленных, как показывает опыт работ, неправомерно.

В настоящем разделе рассматривается аппроксимация дискретной совместной плотности $P(x_1, x_2, \dots, x_n/a_j)$ произведением условных вероятностей, в условии которых количество дискретных компонент ограничивается сверху заранее заданным числом K ; ($K < n$) и предлагается один алгоритм оптимальной в смысле некоторого критерия аппроксимации плотности в виде

$$P(x/a_j) = P_K(x/a_j) = \prod_{i=1}^n P(x_i/x_{j1(i)}, \dots, x_{jK(i)}, a_j). \quad (\text{III.54})$$

Параметр K выбирается в каждой отдельной задаче, исходя из априорных сведений о характере статистической зависимости параметров с учетом требуемой точности классификации.

В дальнейшем для удобства изложения будем рассматривать не совместную плотность $P(x_1, x_2, \dots, x_n/a_j)$, а плотность распределения вероятностей $P(\bar{X}) = P(x_1, \dots, x_n)$, понимая под последней распределение изображений любого из изучаемых классов объектов.

Определения и некоторые вопросы аппроксимации совместной плотности распределения

Определение 1. Аппроксимацией K -го порядка ($0 \leq K \leq n$) для $P(x_1, \dots, x_n)$ является

$$P_K(x_1, \dots, x_n) = \prod_{i=1}^n P(x_{m_i}/x_{j1(m_i)}, \dots, x_{jK(m_i)}), \quad (\text{III.55})$$

где $j_l(i)$; ($1 \leq l \leq K$) — некоторые целочисленные функции, причем для любых l и i выполняется неравенство $j_l(i) \leq i$;

m_1, m_2, \dots, m_n — некоторая перестановка из натуральных чисел от 1 до n .

Для простоты записи введем координату $x_0 \equiv 0$. Пусть $P(x)$ и $Q(x)$ — произвольные плотности распределения, и из $Q(x_0) = 0$ следует $P(x_0) = 0$.

Определение 2. Нормой аппроксимации функции $P(x)$ функцией $Q(x)$ называется число

$$I(P, Q) = \sum_{\substack{x \\ Q(x) \neq 0}} P(x) \log \frac{P(x)}{Q(x)}. \quad (\text{III.56})$$

Норму аппроксимации можно принять за критерий качества аппроксимации, так как $I(P, Q) > 0$ при $P \neq 0$ и $I(P, Q) = 0$. Для доказатель-

ства последних утверждений воспользуемся известным неравенством для выпуклой вверх непрерывной функции $f(x)$

$$\sum_{i=1}^K P_i f(x_i) \leq f\left(\sum_{i=1}^K P_i x_i\right), \quad (\text{III.57})$$

которое справедливо при любых x_i ; ($1 \leq i \leq K$) и при удовлетворяющих следующим условиям:

- 1) $P_i \geq 0$ ($0 \leq i \leq K$),
- 2) $\sum_{i=1}^K P_i = 1$.

В нашем случае, переписав (III.56) в виде

$$I(P, Q) = - \sum_{Q(x) \neq 0} P(x) \log \frac{Q(x)}{P(x)}$$

и воспользовавшись тем, что $\log x$ — выпуклая функция при $n > 0$, а $\sum_x Q(x) = 1$, имеем

$$I(P, Q) = - \sum_{Q(x) \neq 0} P(x) \log \frac{Q(x)}{P(x)} \geq - \log \sum_{Q(x) \neq 0} Q(x) = 0. \quad (\text{III.58})$$

Равенство в (III.57) возможно только в том случае, когда все x_i , соответствующие не нулевым P_i , равны между собой. В нашем случае это означает, что $I(P, Q) = 0$ только в том случае, когда для всех x , для которых $Q(x) \neq 0$, выполняется равенство

$$\frac{P(x)}{Q(x)} = \lambda. \quad (\text{III.59})$$

Так как $P(x)$ и $Q(x)$ — вероятностные распределения, то $\lambda = 1$. Наше утверждение доказано.

Заметим, что по построению функция $P_K(x)$ не может обращаться в нуль при тех x , при которых $P(x) \neq 0$; т. е. можно исключить из рассмотрения те x , для которых $P_K(x) = 0$, так как вероятность P этих значений x равна нулю.

Определение 3. Весом сомножителя $P(x_i/x_{j1}, \dots, x_{jK})$ называется число

$$I(i, j_1, \dots, j_K) = \sum_{x_i, x_{j_1}, \dots, x_{j_K}} P(x_i, x_{j_1}, \dots, x_{j_K}) \log P(x_i/x_{j_1}, \dots, x_{j_K}) \quad (\text{III.60})$$

Суммирование здесь, как и везде в аналогичных случаях, ведется по всем возможным сочетаниям значений написанных аргументов.

Определение 4. Весом аппроксимации $P_K(x)$ называется сумма всех входящих в нее сомножителей.

Теорема 1. Пусть $P_K^0(x)$ — аппроксимация K -го порядка для $P(x)$, имеющая максимальный вес среди всех аппроксимаций K -го порядка. Тогда для любой другой аппроксимации порядка выполняется следующее неравенство:

$$I(P, P_K^0) \leq T(P, P_K).$$

Теорема 1 означает, что аппроксимация с максимальным весом является оптимальной в смысле критерия (III.56).

Доказательство теоремы 1. По определению $I(P, P_K)$ имеем

$$I(P, P_K) = \sum_x P(x) \log \frac{P(x)}{P_K(x)} = \sum_x P(x) \log P(x) - \sum_x P(x) \log P_K(x).$$

$\sum P(x) \log P(x)$ не зависит от аппроксимации $P(x)$, поэтому, для того чтобы минимизировать $I(P, P_K)$, достаточно найти аппроксимацию $P_K(x)$, при которой сумма

$$\sum_x P(x) \log P_K(x) \quad (\text{III.61})$$

максимальна. Подставим в (III.61) запись $P_K(x)$ по определению

$$\begin{aligned} \sum_x P(x) \log P_K(x) &= \sum_x P(x) \log \prod_{i=1}^n P(x_{m_i}/x_{j1(m_i)}, \dots, x_{jk(m_i)}) = \\ &= \sum_x P(x) \cdot \sum_{i=1}^n \log P(x_{m_i}/x_{j1(m_i)}, \dots, x_{jk(m_i)}). \end{aligned} \quad (\text{III.62})$$

Меняя порядок суммирования, получаем

$$\begin{aligned} &\sum_{i=1}^n \sum_x P(x_1, \dots, x_n) \log P(x_{m_i}/x_{j1(m_i)}, \dots, x_{jk(m_i)}) = \\ &= \sum_{i=1}^n \sum_{x_{m_i}, x_{j1(m_i)}, \dots, x_{jk(m_i)}} P(x_{m_i}, x_{j1(m_i)}, \dots, x_{jk(m_i)}) \cdot \log P(x_{m_i}/x_{j1(m_i)}, \dots, x_{jk(m_i)}). \end{aligned} \quad (\text{III.63})$$

Последняя двойная сумма и есть вес аппроксимации. Теорема 1 доказана.

Отметим, что теорема 1 доказана в [142] для частного случая аппроксимации первого порядка.

Рассмотрим алгоритм построения оптимальной аппроксимации первого порядка, предложенный в [142].

Запишем вначале вес каждого сомножителя $P(x_i/x_j)$ в симметричной относительно x_i и x_j форме. Для этого к весу аппроксимации прибавим постоянное число, не зависящее от аппроксимации

$$M = - \sum_{i=1}^n \sum_{x_i} P(x_i) \log P(x_i). \quad (\text{III.64})$$

При этом, учитывая равенство $\sum_{x_i} P(x_i, x_j) = P(x_j)$, получим

$$\begin{aligned} &\sum_{i=1}^n \sum_{x_i, x_{j(i)}} P(x_i, x_{j(i)}) \cdot \log P(x_i/x_{j(i)}) - \sum_{i=1}^n \sum_{x_i} P(x_i) \cdot \log P(x_i) = \\ &= \sum_{i=1}^n \sum_{x_i, x_{j(i)}} P(x_i, x_{j(i)}) \cdot \log \frac{P(x_i, x_{j(i)})}{P(x_i)P(x_{j(i)})}. \end{aligned} \quad (\text{III.65})$$

Таким образом, можно рассматривать вес пары x_i и x_j как величину

$$I(x_i, x_j) = \sum_{x_i, x_j} P(x_i, x_j) \cdot \log \frac{P(x_i, x_j)}{P(x_i)P(x_j)}. \quad (\text{III.66})$$

Отсюда следует, что задача построения оптимальной аппроксимации $P_1(x)$ эквивалентна теперь следующей задаче. Требуется построить граф без циклов, состоящий из n вершин и имеющий максимальный суммарный вес ветвей. Вершины графа при этом соответствуют компонентам вектора \bar{X} , а ветви — всевозможным парам компонент

с соответствующими весами. Отсутствие циклов требуется для того, чтобы аппроксимация была сама по себе вероятностным распределением.

Алгоритм оптимальной аппроксимации

В работе [142] предложен следующий алгоритм аппроксимации. Расположим на плоскости n точек, соответствующих n компонентам вектора X . Эти точки будут вершинами строящегося графа. Запишем всевозможные пары (x_j, x_i) в последовательность в порядке убывания их весов. Пары с разными весами запишем в произвольном порядке. Соединим ветвью вершины графа, соответствующие первой паре. Вершины, соответствующие каждой последующей паре, соединяются ветвью в том случае, если при этом в графе не образуются замкнутые циклы. Процедура прекращается, когда исчерпана вся последовательность ветвей.

Выше описанная процедура, очевидно, упростится, если в алгоритм внести два изменения:

а) не будем рассматривать пары независимых компонент, так как, во-первых, вес такой пары равен нулю, и включение ее в граф не меняет веса аппроксимации, и, во-вторых, множитель $P(x_i/x_j)$ при независимости x_i от x_j все равно при любом x_j равен $P(x_i)$;

б) процедуру перебора пар будем прекращать после того, как в граф включены $n-1$ ветви. Правомомерность этого изменения объясняется тем, что связный граф без циклов с вершинами имеет в точности $n-1$ ветвь.

Для доказательства оптимальности аппроксимации, полученной по приведенному алгоритму, Хов [142] ссылается на работу Крускала, где это доказано в случае, когда веса всех пар (x_i, x_j) различны. Докажем эту же теорему для произвольных весов.

Теорема 2. Приведенный выше алгоритм строит граф с максимальным возможным для данного K весом. Другими словами, если применить алгоритм к двум последовательностям пар, расположенных по убыванию весов, и если эти последовательности различаются порядком пар с равными весами, то полученные при этом, вообще говоря, различные графы имеют один и тот же вес. Этот вес будет максимально возможным для данного порядка аппроксимации.

Доказательство теоремы 2. Зафиксируем некоторую последовательность n пар с невозрастающими весами. Поставим в соответствие каждому графу двоичное слово с n разрядами. В i -м разряде этого слова поставим 1, если i -я пара зафиксированной последовательности включена в граф, и 0 — в противном случае.

Пусть слово A соответствует графу α , который построен с помощью обсуждаемого алгоритма. Возьмем любой другой граф β с теми же вершинами, не содержащий циклов. Обозначим через B слово, соответствующее графу β .

Рассмотрим теперь двумерные векторы $\begin{pmatrix} a_i \\ b_i \end{pmatrix}$, где a_i и b_i — i -ые разряды слов A и B соответственно, и докажем, что среди первых p (p — любое целое число) векторов $\begin{pmatrix} a_i \\ b_i \end{pmatrix}$ количество векторов $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ не меньше, чем количество векторов $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Предположим противное, т. е. существует m такое, что среди первых m векторов количество векторов $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ равно t , количество векторов $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ равно l и $l > t$.

Обозначим пары нашей последовательности, соответствующие векторам $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, через $u_i (0 < i < t)$, а пары, соответствующие векторам $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, через $v_j (0 \leq j \leq t)$.

Присоединим к графу, соответствующему m первым разрядам слова A , все ветви v_j . В полученном графе G каждая ветвь v_j будет содержаться в некотором цикле, не содержащем ни одной v_i при $i \neq j$. Противное противоречило бы построению графа A . Если из графа G исключить ветви $u_i (0 < i \leq t)$, то полученный граф должен соответствовать первым m разрядам слова B , т. е. не должен содержать циклов. По индукции легко доказывается следующая лемма.

Лемма. Если в некотором графе G можно выделить набор ветвей $v_j (0 < j \leq l)$, для каждой из которых в графе найдется цикл, содержащий эту ветвь v_j и не содержащий ни одной из ветвей v_i при $i \neq j$, то отбрасыванием любых t ветвей при $t < l$ не может получиться из графа граф, не содержащий циклов.

Докажем лемму при $l=2$, что соответствует двум случаям.

1. Циклы, соответствующие ветвям v_1 и v_2 , не имеют общих ветвей. В этом случае, очевидно, путем отбрасывания любой ветви нельзя разорвать оба цикла.

2. Циклы V_1 и V_2 , соответствующие ветвям v_1 и v_2 , имеют несколько общих ветвей S_1, S_2, \dots, S_k . Очевидно, что если отбрасывать любую ветвь, отличную от S_1, S_2, \dots, S_k , можно разорвать только один из циклов V_1 и V_2 . Если же отбросить некоторую ветвь S_i , то в графе останется цикл, состоящий из ветвей $(V_1 - S_i) + (V_2 - S_i)$.

Лемма доказана.

Аналогично проводится доказательство при произвольном l . Из леммы следует, что граф, получающийся из G отбрасыванием ветвей, содержит по крайней мере один цикл, т. е. это не может быть граф, соответствующий первым m разрядам слова B .

Из доказанного следует, что j -я по порядку слева единица в слове A стоит не правее j -й единицы слова B . Отсюда в свою очередь следуют утверждения.

1. Если веса всех возможных пар (x_i, x_j) различны, то вес графа G строго больше веса любого отличного от A графа, не содержащего циклов и имеющего те же вершины (теорема, доказанная в работе [142]).

2. Если некоторые ветви имеют равные веса, то, применив алгоритм к двум последовательностям пар, отличающимся только порядком ветвей с равными весами, очевидно, получим графы, имеющие равные веса.

Описанный алгоритм применялся в задачах классификации геологических объектов. При этом по обучающей выборке для каждого класса строилась аппроксимация распределения произведением условных вероятностей, а затем осуществлялась классификация по вероятности принадлежности объекта каждому из классов при условии, что параметры объекта принимали данное значение.

6. ДЕТЕРМИНИСТСКИЕ МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ

Метод линейной классификации «Гиперпласт»

Метод предназначен для классификации объектов, каждый из которых может принадлежать к одному из двух классов.

Предположим, что обучающая выборка содержит K_1 объектов I класса и K_2 объектов II класса.

Объект описывается n -мерным вектором (x_1, x_2, \dots, x_n) . При этом координата x_i может принимать как дискретные, так и любые действительные значения.

Каждый вектор \bar{X} преобразуется в двоичный вектор \bar{Y} размерности $N > n$. Для этого область значений каждой координаты x_i разбивается на f_i интервалов. В векторе \bar{Y} для координаты x_i отводится f_i двоичных разрядов. При преобразовании \bar{X} в \bar{Y} если значение x_i попало в некоторый из f_i интервалов, то в соответствующий разряд в векторе \bar{Y} ставится 1, в остальных $(f_i - 1)$ разрядах, выделенных для координаты x_i , ставятся нули. Таким образом, получается вектор \bar{Y} размерности $N = \sum_{i=1}^n f_i$.

В пространстве размерности N проводится полоса максимальной толщины $2h$, заключенная между двумя параллельными гиперплоскостями. При этом требуется, чтобы точки обучающей выборки, принадлежащие разным классам, лежали с разных сторон от этой полосы (такая полоса существует, вообще говоря, не всегда, поэтому применение метода ограничено линейной классификацией в преобразованном пространстве признаков).

Объекты, предъявленные для классификации, разбиваются на три класса: в один из них входят объекты, попавшие внутрь полосы толщины $2\theta h$, а в два других — объекты, лежащие по разные стороны от этой полосы. Объекты, попавшие внутрь полосы, не относятся системой ни к одному из классов, а объекты, лежащие вне ее, относятся к тем же классам, что и соответствующие объекты обучающей выборки. Параметр θ выбирается в зависимости от требуемой надежности классификации.

Построение полосы

Полоса строится методом последовательных приближений. Выберем начальные значения коэффициентов нормали к полосе — \bar{C}^0 и начальное значение толщины полосы $2h_0$. Для того чтобы включить в рассмотрение полосы, не проходящие через начало координат, добавим ко всем векторам Y координату $y_0 = 1$. Аналогично нормаль к полосе также будем считать $(N+1)$ -мерным вектором (C_0, \dots, C_N) (при нормировании первый коэффициент отбрасывается).

$(N+1)$ -мерные векторы, соответствующие объектам обучающей выборки, будем обозначать через $\bar{a}_1, \dots, \bar{a}_k$, для I класса b_1, \dots и b_k для II класса.

Цикл настройки

Вычисляется скалярное произведение $\alpha = (\bar{a}_i, \bar{C}_j)$. Если выполняется неравенство

$$\alpha > h, \quad (\text{III.67})$$

то переходим к вектору \bar{a}_2 . Если неравенство (III.67) не выполняется, переходим к

$$\bar{c}^{j+1} = \frac{\bar{c}^j + d_j \bar{a}_1}{\|\bar{c}^j + d_j \bar{a}_1\|}, \quad (\text{III.68})$$

где d_j — монотонная последовательность положительных чисел, стремящаяся к нулю. Она вводится для постепенного уменьшения скорости настройки. Преобразование (III.68) повторяется до тех пор, пока не выполнится неравенство (III.67).

Эта процедура повторяется последовательно для всех векторов $\bar{a}_m (1 \leq m \leq k_1)$. Естественно, что после окончания можно гарантировать выполнение неравенства (III.67) только для a_{k_1} .

После этого аналогичная процедура выполняется для векторов $b_m (1 \leq m \leq k_2)$, при этом неравенство (III.67) заменяется на

$$\alpha < -h, \quad (\text{III.69})$$

а преобразование (III.68)

$$\bar{c}^{j+1} = \frac{\bar{c}^j - d_j b_m}{\|\bar{c}^j - d_j b_m\|}.$$

Проверка настройки

Проверяется выполнение всех неравенств для вектора \bar{c}^f , полученного при настройке

$$\begin{aligned} (\bar{a}_1, \bar{c}^f) &> h, \\ \dots \dots \dots \\ (\bar{a}_{k_1}, \bar{c}^f) &> h, \\ (\bar{b}_1, \bar{c}^f) &< -h, \\ \dots \dots \dots \\ (\bar{b}_{k_2}, \bar{c}^f) &< -h. \end{aligned} \quad (\text{III.70})$$

Если хотя бы одно из неравенств (III.70) не выполняется, цикл настройки повторяется. Одновременное выполнение всех неравенств (III.70) означает, что объекты I и II классов обучающей выборки лежат по разные стороны полосы ширины $2h$, образованной в пространстве R^{N+1} гиперплоскостями с нормалью \bar{c}^f

$$-h < (\bar{X}, \bar{c}^f) < h.$$

В этом случае ширина полосы увеличивается $h^1 = h(1+DH)$ и проводится настройка для нового h .

Если после m -кратного повторения цикла найти полосу не удастся, т. е. хотя бы одно из неравенств (III.70) не выполняется, то ширина полосы h уменьшается

$$h' = h(1-DH). \quad (\text{III.71})$$

После каждого преобразования (III.71) величина DH уменьшается вдвое. Это требуется также для постепенного уменьшения скорости настройки. Обучение прекращается, когда после преобразования (III.71) выполняется неравенство

$$DH = \varepsilon, \quad (\text{III.72})$$

где ε — заранее заданная малая величина.

По вектору \bar{X} , предъявленному для классификации, вычисляется скалярное произведение $\alpha = (\bar{C}, \bar{X})$.

Вектор \bar{C} — вектор нормали, полученный при обучении. Объект \bar{X} относится к I классу, если $\alpha > \Theta H$; ко II классу, если $\alpha < -\Theta H$; к спорным объектам, если $-\Theta H < \alpha < \Theta H$ (отказ от классификации). H — порог, полученный на последнем шаге обучения.

Следует отметить, что в исходном n -мерном пространстве признаков разделяющая поверхность, которая строится по методу «Гиперпласт», является не гиперплоскостью, а нелинейной поверхностью весьма сложного вида.

Если в качестве оценки информативности каждого интервала f_i признака x_i принять соответствующую этому интервалу компоненту C_{f_i} вектора \bar{C} , тогда относительную информативность признака x_i можно выразить в виде

$$J_{x_i} = \frac{\sum_i |C_{f_i}|}{\Sigma} \cdot 100\%,$$

где Σ — сумма информативных весов интервалов по всем признакам.

Метод нелинейной классификации с использованием процедуры случайного поиска

Рассмотрим задачу разделения двух совокупностей точек $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_{n_A} \in A$ и $\bar{X}_{n+1}, \dots, \bar{X}_N \in B$ в пространстве исходных параметров x_1, \dots, x_n с помощью нелинейных гладких поверхностей, задаваемых уравнениями полиномиального вида

$$y = (\bar{X}) = \sum_{i,k=0}^n a_{ik} x_i x_k, \quad (\text{III.73})$$

где x_i, x_k — соответственно i -я и k -я координаты вектора \bar{X} ; n — количество параметров; $k, i = 0, 1, 2, \dots, n$; $k \geq i$; $x_0 = 1$; a_{ik} — коэффициент.

Предлагается подход, позволяющий сформулированную задачу классификации свести к задаче минимизации функционала вида

$$\Phi = \frac{\sum_{j=1}^N \left[y(\bar{X}_j) - \text{Sign} \left(n_A + \frac{1}{2} - j \right) \right]^2}{N - S - 1}, \quad (\text{III.74})$$

где $S = C_{n+2}^2$ — количество коэффициентов в (III.73).

Из условия (III.74) видно, что искомые коэффициенты должны быть оптимальны в смысле среднеквадратического отклонения выражения (III.73) от заранее заданной функции $\text{Sign}(\xi)$.

Кроме того, ниже показывается, что для минимизации функционала (III.74) могут быть эффективно применены методы случайного поиска [20, 103, 104], и предлагается один алгоритм нелинейной классификации, использующий оригинальную процедуру случайного поиска с элементами обучения.

Следует отметить, что до настоящего времени для решения задач распознавания образов практически не использовались алгоритмы классификации, в основе которых лежала бы идея оптимизации с помощью методов случайного поиска.

Допустим, что в некоторой ограниченной области параметров существует такая точка $\bar{a}_{\text{опт}}(\dots, a_{ik}, \dots) = \bar{a}_{\text{опт}}(a_1, \dots, a_s)$, для которой выполнено условие

$$\Phi(\bar{a}_{\text{опт}}) = \Phi_{\text{мин.}}$$

Тогда общеизвестную и наиболее универсальную процедуру случайного поиска можно пояснить следующим образом.

Из некоторой заданной начальной точки $\bar{a}_0(a_{10}, \dots, a_{s0}) \in \Omega$ производится некоторым случайным образом перемещение в точку $\bar{a}_1(a_{11}, \dots, a_{s1})$. Если окажется, что $\Phi(a_1) > \Phi(\bar{a}_0)$, то из точки \bar{a}_0 производится снова случайным образом вторая попытка (перемещение в новую точку a_1), и так до тех пор, пока не окажется, что $\Phi(\bar{a}_1) < \Phi(\bar{a}_0)$. В этом случае попытка признается удачной, а поиск решения будет продолжен в направлении этой удачной попытки, только теперь из новой исходной точки a_1 .

Движение из точки a_1 в выбранном направлении будет проводиться до тех пор, пока при реализации некоторой попытки с номером j не окажется выполненным условие:

$$\Phi(\bar{a}_j) > \Phi(\bar{a}_{j-1}) \text{ при } \Phi(\bar{a}_0) > \Phi(\bar{a}_1) > \dots > \Phi(\bar{a}_{j-1}).$$

Тогда из точки a_j осуществляется перемещение в противоположном направлении с меньшим шагом и, если оно будет неудачным, за исходную принимается точка \bar{a}_{j-1} , и вся процедура повторяется.

В процессе поиска координаты точки, в которую производится пробное перемещение из последующей наилучшей точки \bar{a}_{j-1} , в общем случае определяются следующим образом:

$$a_{pj} = a_{p_0} + C_p \varphi_{pj} \text{ при } \begin{cases} p=1, 2, \dots, s \\ \varphi_{pj} = \varphi_{p, j-1} + M_j \Delta P_j, \end{cases} \quad (\text{III.75})$$

где s — общее количество коэффициентов; a_{p_0} — p -я координата начальной точки \bar{a}_0 ; $\varphi_{p, j-1}$ — безразмерный путь, пройденный по координате a_p из начальной точки a_{p_0} за все $(j-1)$ совершенные попытки $\varphi_{p_0} \equiv 0$; M_j — общий масштаб поиска на j -м шаге (подробнее см. ниже); M_0 — задается заранее; ΔP_j — безразмерный шаг по a_p координате, равный

$$\Delta P_j(\lambda_{j-2}^{**} \lambda_{j-1}) = \begin{cases} \xi^* \text{ при } \lambda_{j-2} = \lambda_{j-1} = 0 \\ -P_{j-1} \text{ при } \lambda_{j-2} = 1; \lambda_{j-1} = 0 \\ \Delta P_{j-1} \text{ при } \lambda_{j-2} = \lambda_{j-1} = 1 \end{cases}$$

φ_{pj} — безразмерный путь, пройденный по координате a_p из начальной точки a_{p_0} в точку a_{pj} ;

C_p — масштабный коэффициент шага по координате a_p , который выбирается из априорных соображений до начала процесса оптимизации.

* Случайное число, вырабатываемое при реализации алгоритма случайных чисел в текущий момент времени. Датчик вырабатывает случайные числа, удовлетворяющие условию

$$P(\xi) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\xi^2}{2}}.$$

** λ_j — функция успеха: $\lambda_j = 1$ при удачной j -й попытке; $\lambda_j = 0$ при неудачной j -й попытке.

В процессе поиска оптимального решения может оказаться, что определяемое по формуле (III.75) значение a_{pj} не удовлетворяет заранее накладываемому условию; тогда j -я попытка признается неудачной.

Общий масштаб поиска M_j является настраиваемой величиной. Выбор величины M_j зависит от накапливаемой предыстории и положительного опыта — параметра K_j , который поощряет рост M_j при серии удачных попыток и, наоборот, способствует его уменьшению при серии отрицательных попыток.

Масштаб поиска вычисляется по следующей формуле:

$$M_j = \begin{cases} b_3 M_{j-1} & \text{при } \lambda_{j-1} = 1 \text{ и } K_j = K_{-1} + b_1 > K_{\max} \\ b_4 M_{j-1} & \text{при } \lambda_{j-1} = 0 \text{ и } K_j = b_2 \cdot K_{j-1} < K_{\min}, \end{cases} \quad (\text{III.76})$$

где K_{\min} , K_{\max} , b_1 , b_2 , b_3 , b_4 — константы, задаваемые заранее, причем $b_1 < 1$, $b_2 < 1$, $b_3 > 1$ и $b_4 > 1$. Следует отметить, что при решении практических задач масштаб поиска ограничивается снизу некоторой постоянной величиной M_{\min} .

Недостатком рассмотренной процедуры случайного поиска с обучением является необоснованная фиксация масштабных коэффициентов C_p , $p = 1, 2, \dots, s$.

В работе [104] показано, что наилучшая сходимость процесса оптимизации достигается при выполнении в каждой из попыток условия

$$\frac{d\Phi}{dC_1} C_1 \simeq \frac{d\Phi}{dC_2} C_2 \simeq \dots \simeq \frac{d\Phi}{dC_s} C_s. \quad (\text{III.77})$$

Однако чтобы выполнить условие (III.77), необходимо вычислять не во всех, но во всяком случае в большинстве точек, от которых производится поиск лучших решений, все частные производные $\frac{d\Phi}{dC}$.

Последнее приводит к тому, что при решении многопараметрических задач резко возрастает время счета, а это в свою очередь вынуждает исследователей отказаться от использования рассмотренного метода корректировки коэффициентов C_p .

В предлагаемом ниже алгоритме классификации процедура случайного поиска с обучением существенно отличается от той, которая была описана выше. Отличается она тем, что в ней предусмотрено более эффективное использование опыта, приобретаемого в процессе последовательной оптимизации методом проб и ошибок. Поясним сказанное.

Известно, что при решении практических задач значимость и ответственность той или иной координаты a_p в ходе оптимизации может оставаться постоянной, а может изменяться и при том существенно. Если в разрабатываемой системе оптимизации не предусмотреть процесса отслеживания этих изменений и адаптации к варьированию формы поверхности отклика при переходе из одной точки поиска в другую, то это приведет к увеличению общего количества отрицательных попыток и соответственно времени решения задачи.

В большинстве случаев, встречающихся на практике, при правильно подобранном шаге варьирования форма поверхности отклика в пределах нескольких удачных попыток претерпевает регулярное изменение, которое можно вовремя отследить, если по мере совершения этих удачных шагов накапливать безусловно полезную информацию об изменении степени влияния координат a_p на оптимизируемый показатель Φ .

В предлагаемой процедуре случайного поиска адаптация к изменению формы поверхности отклика осуществляется путем взвешивания информации о степени влияния коэффициентов a_p на величину Φ за последние k удачных попыток, которая затем используется для корректировки величины текущего шага варьирования по каждой из координат. При этом если коэффициент a_p после каждой из k совершенных удачных попыток изменяется интенсивно и в одном направлении, то при корректировке соответствующей этой координате шаг варьирования увеличивается, а если изменение величины a_p было незначительным или знакопеременным, то шаг варьирования по этой переменной уменьшается.

Практически это достигается с помощью сомножителя N_{pj} , вводимого в выражение

$$\varphi_{pj} = \varphi_{p, j+1} + M_j \Delta_{pj} N_{pj} \quad (\text{III.78})$$

и равного

$$N_{pj} = CQ(\xi_{pj}), \quad N_{p0} = 1, \quad (\text{III.79})$$

где

$Q(\xi_{pj}) = \frac{\sum_{s=(l-1)k+1}^{lk} \xi_{ps}}{k}$ — математическое ожидание безразмерного случайного шага ξ_p по p -й координате за k положительных попыток с номерами $(l-1)k+1, (l-1)k+2, \dots, lk$; C — коэффициент, вводимый для варьирования масштаба поиска ($1 < C \leq 2$).

Из (III.78) и (III.79) легко заметить, что вначале, когда положительный опыт отсутствует или он незначителен, система не располагает информацией для адаптации ($N_{p0} = 1$). Затем по мере накопления в процессе поиска методом проб и ошибок успешного опыта система начинает с k -й удачной попытки приспособляться к регулярным реакциям на пробные возмущения и предугадывать дальнейший ход событий путем положительной коррекции существенных компонент a_p и отрицательной коррекции несущественных. При этом коррекция производится только по последним k успешным попыткам, что соответствует забыванию уже устаревшего и ненужного в настоящее время опыта. Априори задаваемая величина k характеризует степень изменчивости модели оптимизации.

Докажем, что если осуществляется поиск экстремума Φ_{\min} на некоторой гиперповерхности $\Phi = \Phi(\bar{a})$ из произвольной точки $\bar{a}_0(a_{10}, a_{20}, \dots, a_{s0})$ при выполнении следующих условий:

а) критерий Φ в пространстве параметров Ω является одноэкстрементальной функцией и существует траектория непрерывного спуска из начальной точки a_0 в точку экстремума $\bar{a}_{\text{опт}}$;

б) если в области Ω не существует подобласти, соизмеримой с шагом поиска, в которой производная $\frac{d\Phi}{da_p}$ была бы равна 0 или $\pm\infty$, то за конечное число шагов достигается точка экстремума с заданной точностью ϵ .

Согласно условиям на поверхности отклика $\Phi(\bar{a})$, которая может представлять собой любую функциональную зависимость от a_p , имеются некоторые непрерывные и замкнутые линии уровней $\Phi = \text{Const}$.

Ввиду того, что единичный шаг $\bar{\xi}_{p,1}$ по каждой из координат a_p из точки \bar{a}_0 формируется с помощью датчика или случайных чисел ξ , удовлетворяющих условию

$$P(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{|\xi|^2}{2}}, \quad 0 \leq |\xi| \leq D = \text{Const}, \quad (\text{III.80})$$

то вероятность смещения в любую точку пространства, заключенного в гиперкуб с вершинами, имеющими координаты $a_{p_0} \pm D$ и с центром в точке \bar{a}_0 , всегда конечна и равна

$$P(\bar{\xi}) = \prod_{p=1}^s P(\xi_p).$$

Так как движение по каждой координате в любом направлении равновероятно, то внутри гиперкуба образуются некоторые гиперсферы с центром \bar{a}_0 и радиусом.

$$\bar{\xi} = \sqrt{\sum_{p=1}^s \xi_p^2}. \quad (\text{III.81})$$

При этом точка гиперсферы равновероятна, так как согласно (III.80) и (III.81) $p(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{|\xi|^2}{2}}$ максимальная вероятность соответствует точке $\bar{a}_0(\xi=0)$, а минимальная — точкам вершин гиперкуба ($\xi_p=D$).

Так как шаг поиска в алгоритме может быть выражен формулой

$$C_p \Phi_{p-1} = C_p M_1 \xi_{p-1}^*, \quad (\text{III.82})$$

где коэффициенты M_1 и C_p конечны ($M_j > 0$, $C_p > 0$), то области равновероятного смещения при линейно (по C_p) масштабированных координатах a_p отобразятся также в виде гиперсфер при соблюдении условия (III.77) либо будут эллиптически деформированы при

$$C_p = \frac{d\Phi}{da_p} \neq \text{Const}.$$

При этом вероятность попадания в удачную область $\Phi_1 < \Phi_0$ при некотором радиусе смещения будет определяться отношением поверхностей, образуемых при пересечении соответствующей гиперсферы гиперповерхностью $\Phi_0 = \text{Const}$.

Поскольку при уменьшении радиуса смещения или M_1 вероятность выбора удачного направления при гладкой поверхности отклика стремится к 0,5, то для любой точки вероятность смещения в требуемом направлении на некоторую величину $\Delta\Phi = \Phi_0 - \Phi_1 > 0$ всегда (при $M_1 > 0$ и $C_p > 0$) конечна и больше нуля. Аналогичным образом можно доказать, что вероятность положительного приращения $\Delta\Phi = \Phi_1 - \Phi_2$, происходящего при перемещении из точки 1 в точку 2, также конечна. Продолжая вышеприведенные рассуждения, легко показать, что в точку $\bar{a}_{\text{опт}}$ можно сместиться из любой точки a_0 , отстоящей от $\bar{a}_{\text{опт}}$ на расстоянии $R \neq \infty$ за конечное число шагов с любой степенью точности $\varepsilon > 0$, если выполняются условия «а» и «б».

Рассмотренный алгоритм классификации можно отнести к классу непараметрических методов распознавания образов, так как для его

* Формула верна для первого шага $j=1$, но аналогично можно записать и для последующих j -ых шагов поиска.

исследования не требуется сведений о законе распределения исходных классов объектов. Он легко реализуем на ЭВМ, причем для этого необходим сравнительно небольшой объем машинной памяти.

К недостатку метода следует отнести то, что он не эффективен при распознавании классов объектов, плотности распределения которых не являются унимодальными функциями. Тем более он неприменим для деления многосвязных классов объектов.

Классификация по правилу одной или нескольких ближайших величин

Ниже рассматривается один класс непараметрических алгоритмов, которые могут успешно использоваться при разделении многосвязных совокупностей (часто встречающийся на практике случай) и отдельные модификации которых экономичны в смысле расходуемого машинного времени.

Если классифицируемые множества реализаций являются многосвязными в исходном пространстве признаков, то для поиска решающего правила целесообразно использовать эвристические алгоритмы, основанные на функциях принадлежности, близости, соседства и т. п. В работе [123] предлагается при наличии взаимно коррелируемых признаков отказаться от статистического подхода в распознавании и использовать следующий метод.

Каждая точка пространства измерений характеризуется совокупностью функций принадлежности к эталонным объектам. При этом функция принадлежности к представителям какого-либо класса объектов положительна, если рассматриваемая точка пространства ближе к этим представителям, чем к любой реализации другого класса. В противном случае функция принадлежности равна нулю.

Все члены выборки будут опознаны алгоритмом, основанным на принципе нахождения максимума функций принадлежности, если в точках, соответствующих реализациям некоторого одного класса, функция принадлежности хотя бы к одному представителю этого класса положительна, а в точках, соответствующих реализациям остальных классов, все функции принадлежности к представителям этого же класса равны нулю.

Если предварительно взять в качестве эталонных объектов все члены выборки обучения, а затем каким-либо образом исключить все члены, не являющиеся необходимыми для выполнения указанного выше условия, то поставленная задача будет решена.

Алгоритм построения функции принадлежности сводится к следующему.

1. Используется вся выборка, которая считается представительной.

2. Составляется матрица r_{mnq} взаимных расстояний между реализациями рассматриваемых классов объектов

$$r_{mnq} = \sum_{i=1}^N |x_{inq} - x_{imq}|, \quad (\text{III.83})$$

где m и n — номера реализаций; q — номер класса; x — значение признака; i — номер признака; N — число признаков.

3. Для каждой n -й реализации q -го класса находится расстояние до ближайшей реализации среди остальных классов и из найденного расстояния вычитается Δ -удвоенная погрешность измерения расстояния

$$R_{nq} = \min_p \sum_{i=1}^N |x_{inq} - x_{imq}| - \Delta, \quad p \neq q. \quad (III.84)$$

4. Составляется матрица функций принадлежности:

$$\eta_{mnq} = 1 - \frac{r_{mnq}}{R_{nq-\Delta}} \quad \text{при} \quad \frac{r_{mnq}}{R_{nq-\Delta}} < 1, \quad (III.85)$$

$$\eta_{mnq} = 0 \quad \text{при} \quad \frac{r_{mnq}}{R_{nq-\Delta}} \geq 1.$$

Из формул видно, что коэффициент $\eta_{mnq} > 0$ в том случае, если расстояние от n -й реализации q -го класса до m -й реализации того же класса меньше (с запасом Δ) расстояния от n -й реализации до любой реализации любого другого образа, и равен нулю в противном случае.

Для нахождения минимального количества эталонов может быть применен алгоритм нахождения минимальных покрытий [123].

В работе [41] предлагается метод распознавания, использующий в качестве решающих функций «набор гиперсфер». Для каждого из m классов объектов ищется сфера минимального радиуса, охватывающая все реализации данного класса.

Значение радиусов сфер и матрица евклидовых расстояний между центрами этих сфер позволяет выделить классы, сферы которых не пересекаются никакой другой сферой. Поиск решающих функций для таких классов заканчивается (рис. 6а).

Далее рассматривается совокупность классов, сферы которых пересекаются. Если в область пересечения не попала ни одна из точек исходной выборки, то такое пересечение считается фиктивным, а классы — пересекающимися.

Решающими функциями для них считаются сферы, полученные в самом начале процесса обучения (рис. 6б).

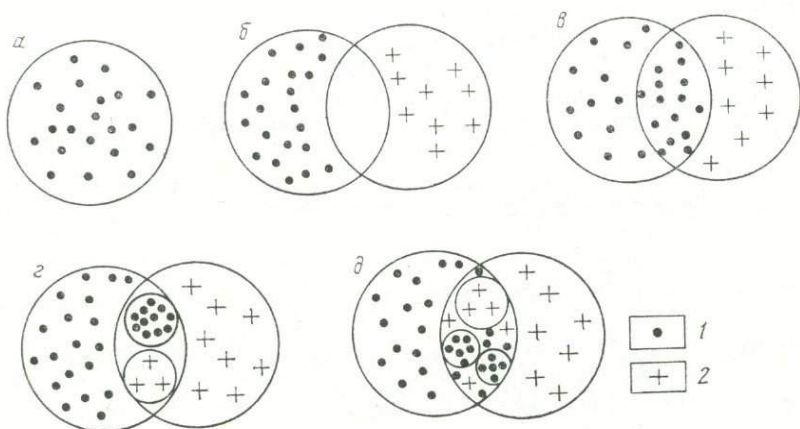


Рис. 6. Геометрическая интерпретация классификации по правилу ближайших величин.

1 — точки I класса, 2 — точки II класса

Если в области пересечения присутствуют точки только одного класса (рис. 6 в), то область этого класса считается вся сфера, а область другого — соответствующая ему сфера минус область пересечения. Наконец, могут встречаться области пересечения, в которые попадают реализации нескольких классов (рис. 6 г). Тогда проводятся сферы минимального радиуса, которые охватывают все точки одного класса, лежащие в области пересечения. Если эти сферы «второго поколения» не пересекаются друг с другом, то процедура считается завершенной, а решающее правило состоит в перечне сфер, которые охватывают точки одного класса, и указаний запретных зон внутри этих сфер, порожденных пересекающимися сферами других классов.

Может, однако, встречаться такое расположение «следующих» поколений либо совсем не улучшает разделения, либо делает это разделение недостаточно эффективным. Такой алгоритм либо не сходится, либо сходится очень медленно (рис. 6 д). Выход из этого положения может состоять в нахождении максимального числа точек одного из классов в области пересечения, которое можно охватить сферой, не содержащей точек другого класса. Оставшиеся в области пересечения точки делятся с помощью аналогичной процедуры.

Рассмотренный алгоритм экономичнее алгоритма, использующего функции принадлежности, так как при его реализации для запоминания решающего правила требуется меньший объем машинной памяти.

Е. Фикс и И. Ходжес [145] предложили простой алгоритм, для применения которого необходимо лишь существование и непрерывность функции распределения классов объектов. Он сводится к следующему.

1. Находится расстояние от испытуемой точки до всех точек обучающей выборки.

2. Выбираются из этих расстояний k минимальных. Пусть при этом окажется, что из числа k ближайших объектов обучающей выборки n_i объектов принадлежат A_i классу.

3. Находится номер i_0 , для которого

$$n_{i_0} = \max_{\bar{X}} \{n_i\} = \max \{n_1, n_2, \dots, n_m\}$$

и испытуемую точку \bar{X} причисляют к классу A_{i_0} .

Видно, что в этом алгоритме делается попытка оценить величины $P(\bar{X}/A_i)$ около точки \bar{X} ($i=1, 2, 3, \dots, m$).

Если величины $P(\bar{X}/A_i)$ аппроксимируются числами, а это справедливо при возрастании числа членов обучающей выборки, то решающее правило этого алгоритма сводится к решающему правилу Байеса. При использовании алгоритма Фикса и Ходжеса очень важным является выбор числа k , которое должно возрастать при увеличении числа объектов в обучающей выборке N , но величина k/N должна стремиться к нулю при возрастании N .

Требование большого объема памяти для реализации метода Фикса и Ходжеса делает его неприемлемым для решения многих задач прикладной геологии.

Вместо того чтобы запоминать большое число обучающих объектов, Н. Нильсоном [78] предлагается запоминать только несколько «типичных» обучающих объектов. Если обозначить местоположение локального максимума функции плотности вероятностей словом «мода», то в качестве типичного объекта для любого одного класса можно рассматривать моду функции плотности вероятности для этого класса. Зная эти моды, можно построить алгоритм распознавания. Для этого нужно

измерить расстояние от \bar{X} до каждой из этих мод и относить \bar{X} к тому классу, который имеет ближайшую к \bar{X} моду.

Так как число мод обычно намного меньше числа обучающих объектов, то для реализации этого метода не требуется большого объема машинной памяти. В тоже время результативность рассмотренного алгоритма должна быть высокой, так как велика вероятность того, что объект \bar{X} будет располагаться рядом с ближайшей локальной модой того класса объектов, к которому он принадлежит.

В работе [152] предлагается следующая модификация процедуры классификации по правилу одной или нескольких ближайших величин.

Пусть заданы M_s объектов, каждый из которых представляется в виде точки или вектора в N -мерном евклидовом пространстве. Каждому объекту $\bar{X}_j = (x_{j1}, x_{j2}, \dots, x_{jN})$; $j=1, 2, \dots, M_s$ поставлено в соответствие число θ_j , характеризующее класс, к которому относится j -й объект x_j , например $\theta=1, 2, \dots, M$.

Пусть известны признаки x_1, x_2, \dots, x_N объекта \bar{X} . Необходимо для данного \bar{X} определить θ , используя информацию об истинной принадлежности ранее опознанных объектов \bar{X}_j , индексами θ_j .

Определение. Точка \bar{X} называется ближайшей к \bar{X}_m ($m=1, \dots, M_s$) величиной, если

$$\min_j \left\{ \sqrt{\sum_{n=1}^N (x_n - x_{nj})^2} = \sqrt{\sum_{n=1}^N (x_n - x_{nm})^2} \right. \quad (\text{III.86})$$

Алгоритм на основе правила одной ближайшей величины относит \bar{X} к i -у классу ($\theta_x=i$), если ближайшей к \bar{X} величиной окажется величина \bar{X}_m из обучающей выборки с индексом $\theta_m=i$ [152].

Если объем выборки обучения велик, имеет смысл вместо одной ближайшей величины рассматривать R ближайших величин. Тогда алгоритм распознавания сводится к следующему: $\theta_x=i$, если ближайшими к \bar{X} величинами оказываются R величин из обучающей выборки, индекс которых равен i . Отметим, что если в случае правила одной ближайшей величины измеряется расстояние до ближайшей к \bar{X} величины \bar{X}_m , то в случае алгоритма R ближайших величин измеряется сумма расстояний до R ближайших величин с одинаковым индексом в каждом классе.

Значение R желательно выбирать большим, чтобы минимизировать вероятность неоптимального решения о принадлежности \bar{X} . Однако R не должно быть слишком большим, чтобы достаточно близкие к \bar{X} выборки давали точные оценки апостериорных вероятностей истинного класса объекта.

Вообще же, если распределения признаков таковы, что любое расстояние внутри класса меньше любого расстояния между классами, то лучший результат дает правило одной ближайшей величины [143].

Свойства правил ближайших величин основываются на предположении о том, что условные распределения индексов классов θ и θ_i сходятся друг к другу, когда $\bar{X}_i \rightarrow \bar{X}$.

Классификация по правилу ближайших величин предусматривает на этапе обучения нахождение такого значения $R=R_0$, при котором эффективность классификации C_{R_0} объектов из обучающей выборки максимальна. Величину C_{R_0} можно определить следующим образом:

$$C_{R_0} = \frac{M_{R_0}}{M_s} \cdot 100\%, \quad (\text{III.87})$$

где M_{R_0} — число правильно классифицируемых объектов обучающей выборки при $R=R_0$;

M_s — общее число объектов обучающей выборки.

Ниже будут рассмотрены алгоритмы, которые относятся к классу алгоритмов ближайших величин и которые по сравнению с вышеописанными алгоритмами имеют две принципиальные особенности:

— пространство признаков, в котором производится распознавание объектов, заранее не фиксируется;

— процесс классификации является последовательной процедурой, которая сводится к поочередному применению построенных на этапе обучения совокупности решающих правил, причем первое же решение, принимаемое с помощью одного из этих правил, считается истинным.

В основе этих алгоритмов лежит идея, согласно которой нецелесообразно рассматривать объекты выборки обучения в одном и том же заранее фиксируемом пространстве признаков, поскольку им свойственны как общие, так и более индивидуальные черты, присущие одновременно не всем объектам какого-либо одного класса, а отдельным группам его представителей. В силу этого более надежной процедурой обучения классификации является изучение распределений классов объектов не во всем пространстве измерений, а в различных его подпространствах с целью поиска и выявления достаточно представительных групп объектов одного класса, которые обладают близким или одинаковым набором свойств и которые существенно отличаются от представителей других классов объектов.

Кроме того, для повышения надежности распознавания в рассматриваемых ниже алгоритмах предусмотрена процедура исключения из выборки обучения однородных групп объектов по мере их нахождения.

Метод информативных интервалов

Одномерная классификация

Пусть наблюдаемая переменная x , которая может принадлежать к одному из трех классов A , B или «пустому» Ω , характеризуется плотностями вероятностей $P_A(x)$ и $P_B(x)$ и пусть определена функция $f(x)$ такая, что

$$\begin{aligned} f(x) > 0 & \text{ при } P_A(x) > P_B(x), \\ f(x) < 0 & \text{ при } P_A(x) < P_B(x), \\ f(x) = 0 & \text{ при } P_A(x) = P_B(x) = 0. \end{aligned} \quad (\text{III.88})$$

Тогда при условии равенства стоимостей ошибок классификации и априорных вероятностей оптимальное решающее правило можно записать в таком виде:

$$\begin{aligned} x \in A, & \text{ если } f(x) > 0, \\ x \in B, & \text{ если } f(x) < 0, \\ x \in \Omega, & \text{ если } f(x) = 0. \end{aligned} \quad (\text{III.89})$$

Предлагаемый алгоритм определения $f(x)$ по сходной обучающей выборке сводится к следующему.

Предположим, что все элементы обучающей выборки попали в интервал OM переменной x , причем множество изображений класса A

не пересекаются с множеством изображений класса B . Для удобства дальнейших рассуждений будем считать, что все N изображений выборки могут находиться в целочисленных точках $0, 1, 2, \dots, M$, причем $M \geq N$.

Зададим функцию $f(x)$ в целочисленных точках $0, 1, \dots, M$ следующим образом:

$$\begin{aligned} f(x) &= 0, \text{ если } x \in \Omega, \\ f(x) &= 1, \text{ если } x \in A, \\ f(x) &= -1, \text{ если } x \in B. \end{aligned} \quad (\text{III.90})$$

Функцию $\varphi(x)$, определенную в тех же точках и принимающую значения $0, 1$ и -1 , будем называть допустимой, если длина каждого интервала постоянства функции $\varphi(x)$ не меньше Δx_{\min} , где Δx_{\min} — минимальный интервал, на котором возможным является определение $P_A(\Delta x_{\min})$ и $P_B(\Delta x_{\min})$.

Требуется в классе допустимых функций $\varphi(x)$ найти такую функцию $\varphi^*(x)$, для которой величина

$$\begin{aligned} F(O, M) &= \sum_{i=0}^M \left\{ \left| [\varphi^*(i) - f(i)] \frac{1}{2} f(i) \varphi^*(i) \right| + \right. \\ &\quad \left. + |\varphi^*(i) [1 - |f(i)|]| + |f(i) [1 - |\varphi^*(i)|]| \right\} \end{aligned} \quad (\text{III.91})$$

принимает наименьшее значение.

Величина $F(O, M)$ характеризует ошибку классификации, которая связана с использованием аппроксимирующей функции $\varphi^*(x)$.

По критерию (III.91) потери в i -й точке ($i=0, 1, \dots, M$), вызванные либо отказом от классификации, либо неправильной классификацией, равны 1. Также штрафуются единицей случаи $\varphi^*(i) \neq 0$ при $f(i) = 0$.

Очевидно, что решением при Δx_{\min} является $\varphi^*(i) = f(i)$. Будем искать решение для общего случая $\Delta x_{\min} = K \leq M$, где K — целое число.

Оптимальный алгоритм A_0 построения функции $\varphi^*(x)$ из класса допустимых функций $I(x)$, которая минимизирует величину (III.91), был рассмотрен выше.

Построенная по алгоритму A_0 функция $\varphi^*(x)$ является оптимальной в смысле (III.91). И новое изображение x , не входящее в число изображений выборки обучения, будет классифицировано по следующему правилу:

$$\begin{aligned} x \in A, & \text{ если } \varphi^* = 1, \\ x \in B, & \text{ если } \varphi^* = -1, \\ x \in \Omega, & \text{ если } \varphi^* = 0. \end{aligned} \quad (\text{III.92})$$

Рассмотренный выше метод одномерной классификации можно использовать и при пересекающихся классах объектов A и B .

Так, если в общем случае по обучающей выборке i -я точка n_i^A раз принадлежала классу A и n_i^B раз принадлежала классу B , то крите-

рий (III.91) поиска оптимальной функции может быть записан следующим образом:

$$F(O, M) = \sum_{i=1}^M \left\{ \left| \varphi^*(i) f(i) \left(\frac{n_i^A - n_i^B}{2} \right) [f(i) - \varphi^*(i)] \right| + \right. \\ \left. + |\varphi^*(i) [1 - |f(i)|]| + |(n_i^A - n_i^B) [1 - |\varphi^*(i)|] + f(i)| \right\}, \quad (\text{III.93})$$

где функция $f(i) = \begin{cases} 1 & \text{при } n_i^A > n_i^B \\ -1 & \text{при } n_i^A < n_i^B \\ 0 & \text{при } n_i^A = n_i^B \geq 0. \end{cases}$

В критерии (III.93) потери в i -й точке, вызванные либо отказом от классификации, либо неправильной классификацией, учитываются с одинаковым весом и равным $(n_i^A - n_i^B)$. Кроме того, к потерям в i -й точке относится случай (штрафуется единицей), когда $f(i) = 0$, а $\varphi^*(i) \neq 0$.

Многомерная классификация

Пусть наблюдаемый вектор $\bar{X}(x_1, \dots, x_n)$ принадлежит только одному из классов — A , B или «пустому» Ω и характеризуется плотностями вероятностей $P_A(x)$ и $P_B(x)$. Тогда аналогично (III.89) решающее правило запишется:

$$\begin{aligned} x \in A, & \text{ если } F(\bar{X}) > 0, \\ x \in B, & \text{ если } F(\bar{X}) < 0, \\ x \in \Omega, & \text{ если } F(\bar{X}) = 0, \end{aligned} \quad (\text{III.94})$$

где соответственно $F(\bar{X}) > 0$ при $P_A(x) > P_B(x)$, $F(\bar{X}) < 0$ при $P_A(x) < P_B(x)$ и $F(\bar{X}) = 0$ при $P_A(x) = P_B(x) \geq 0$.

Ниже предлагается процедура многомерной классификации, в основе которой лежит вышеизложенный алгоритм одномерной классификации.

1. Для всех параметров x_s , где $s = 1, 2, \dots, n$, строятся одномерные решающие правила $\varphi^*(x_s)$:

$$\varphi^*(x_1), \varphi^*(x_2), \dots, \varphi^*(x_s), \dots, \varphi^*(x_n).$$

2. Все интервалы постоянства функций $\varphi^*(x_s)$ анализируются на информативность по критерию

$$I_{sj} = K^n_{sj} K^k_{sj} K^y_{sj}, \quad (\text{III.95})$$

где I_{sj} — информативность j -го интервала постоянства функции $\varphi^*(x_s)$ (информативность sj -го интервала пространства);

K^n_{sj} — коэффициент представительности sj -го интервала постоянства;

K^k_{sj} — коэффициент компактности sj -го интервала постоянства;

K^y_{sj} — коэффициент удаленности изображений, попавших в sj -ый интервал постоянства определенного знака (+1 или -1), от изображений, попавших в тот из интервалов противоположного знака (sj' -интервал), который является ближайшим к sj -у интервалу постоянства.

В явном виде выражение (III.95) выглядит так

$$I_{sj} = \frac{|n^A_{sj} - n^B_{sj}|}{N_A + N_B} \cdot \frac{OM_s}{\Delta x_{sj}} \cdot \frac{|\mu(x_{sj}) - \mu(x_{sj'})|}{OM_s} = \\ = \frac{|n^A_{sj} - n^B_{sj}| \cdot |\mu(x_{sj}) - \mu(x_{sj'})|}{(N_A + N_B) \cdot \Delta x_{sj}}, \quad (\text{III.96})$$

где n^A_{sj} и n^B_{sj} — количество изображений, принадлежащих соответственно классу A и классу B , и своими проекциями, попавшими в sj -й интервал постоянства;

$\mu(x_{sj})$, $\mu(x_{sj'})$ — выборочные средние по параметру x_s тех изображений, которые попали соответственно в sj и sj' интервал постоянства;

N_A , N_B — количество изображений $x \in A$ и $x \in B$;

Δx_{sj} — длина sj -го интервала постоянства.

3. Из совокупности величин I_{sj} выбирается максимальная

$$I_{a_1, b_1} = \max I_{sj}. \quad (\text{III.97})$$

4. Все изображения, проекции которых попали в интервал постоянства a_1 , b_1 , исключаются из выборки обучения и в дальнейшем не учитываются.

5. Интервал постоянства a_1 , b_1 , однозначно определяемый своими начальной x^H и конечной x^K координатами, а также индексом класса k_1 ($k_1 = A$, если интервал постоянства равен $+1$, и $k_1 = B$, если интервал постоянства равен -1), запоминается.

6. Проводится второй цикл обучения (реализация пунктов 1, ..., 5), в результате чего определяется новый, более информативный интервал постоянства a_2 , b_2 , координаты которого $x^H_{a_2}$ и $x^K_{a_2}$ запоминаются с соответствующим индексом k_2 ($1 \leq a_2 \leq n$), а величина a_2 может быть равной и не равной a_1 .

Затем производится следующий цикл обучения и так до тех пор, пока не будет исчерпана исходная обучающая выборка.

После этого процедура обучения считается законченной, а результатом ее является следующее решающее правило:

$$\left\{ \begin{array}{l} \bar{X}(x_1, \dots, x_{a_1}, \dots, x_n) \in K_1, \text{ если } x_{a_1} \in (x^H_{a_1}, x^K_{a_1}). \\ \bar{X}(\dots, x_{a_2}, \dots, x_{a_p}, \dots) \in K_2, \text{ если } x_{a_2} \in (x^H_{a_2}, x^K_{a_2}). \\ \dots \\ \bar{X}(\dots, x_{a_2}, \dots, x_{a_t}, \dots, x_{a_p}, \dots) \in K_p, \text{ если } x_{a_p} \in (x^H_{a_p}, x^K_{a_p}), \text{ но} \\ x_{a_1} \in (x^H_{a_1}, x^K_{a_1}), \dots, x_{a_t} \in (x^H_{a_t}, x^K_{a_t}), \dots, x_{a_{p-1}} \in (x^H_{a_{p-1}}, x^K_{a_{p-1}}). \\ \bar{X} \in \Omega, \text{ если } x \in \bar{K}_1, \dots, x \in \bar{K}_t, \dots, x \in \bar{K}_p, \end{array} \right. \quad (\text{III.98})$$

где p — число циклов обучения;

$(x^H_{a_t}, x^K_{a_t})$ — интервал постоянства, оказавшийся наиболее информативным при выполнении t -й итерации обучения.

Метод информативных гиперсфер

Ниже рассматривается алгоритм последовательной классификации, разработанный Б. А. Чумаченко, Е. П. Власовым и М. Л. Черновым* и являющийся развитием метода информативных интервалов.

* Практическая реализация алгоритма осуществлена с участием Е. Е. Разина.

Обучение классификации по предлагаемому алгоритму сводится к поиску такой локальной области-сферы в некотором не фиксированном, но ограниченной размерности подпространстве параметров, которая в смысле выбранного критерия оптимальности является областью повышенной плотности точек лишь одного из рассматриваемых классов.

Пусть массив обучения состоит из N объектов, каждый из которых описывается совокупностью m параметров и о которых известно, какому классу они принадлежат. Представим выборку обучения в следующем виде:

$$\begin{aligned} \bar{X}^{(1)}(x_1^{(1)}, \dots, x_m^{(1)}) &\in A, \\ &\dots \\ \bar{X}^{(N_A)}(x_1^{(N_A)}, \dots, x_m^{(N_A)}) &\in A, \\ \bar{X}^{(N_{A+1})}(x_1^{(N_{A+1})}, \dots, x_m^{(N_{A+1})}) &\in B, \\ &\dots \\ \bar{X}^{(N)}(x_1^{(N)}, \dots, x_m^{(N)}) &\in B. \end{aligned} \quad (\text{III.99})$$

Согласно алгоритму на стадии обучения производятся следующие операции.

1. Выбирается такая сфера M_1 из всех сфер M_{1i} , рассматриваемых в i -х подпространствах параметров $i=1, 2, \dots, T$ — размерности $1, 2, \dots, p \leq m^*$, которая удовлетворяет следующему критерию оптимальности:

$$k = \max K^n \cdot K^k \cdot K^y = \max \frac{|n_i^A - n_i^B|}{N} \cdot \frac{\sigma_i}{\sigma_{M_i}} \cdot \frac{\rho_{1i}}{\rho_{M_i}} \quad (\text{III.100})$$

при выполнении условия

$$n_i^A/n_i^B \geq k \text{ либо } n_i^B/n_i^A \geq k, \quad (\text{III.101})$$

где i — номер подпространства из общего числа $T = C_m^1 + C_m^2 + \dots + C_m^p$ подпространств, в которых рассматривается сфера M_{1i} ;

K^n — коэффициент представительности;

K^k — коэффициент компактности;

K^y — коэффициент удаленности;

n_i^A, n_i^B — количество точек классов A и B , принадлежащих сфере M_{1i} ;

N — длина обучающей выборки;

σ_i — стандартное отклонение точек выборки в i -м подпространстве;

σ_{M_i} — стандартное отклонение точек выборки, принадлежащих M_i ;

ρ_{1i} — суммарное расстояние ** от центра сферы M_{1i} до s ближайших точек малопредставленного в M_{1i} класса, не принадлежащих сфере M_{1i} ;

k — заданная константа, ограничивающая снизу одно из отношений (III. 101).

Поиск сфер M_{1i} производится методом последовательного перебора всех T подпространств и выбора в каждом из этих подпространств сферы M_{1i} , соответствующей максимуму критерия

$$k' \max_j \frac{|n_i^A - n_i^B|}{N} \cdot \frac{\sigma_i}{\sigma_{M_i}} \cdot \frac{\sigma_j}{\sigma_{M_j}} \quad (\text{III.102})$$

* Выбирается, исходя из технических возможностей ЭВМ.

** Расстояние ρ_{12} между точками $\bar{X}^{(1)}$ и $\bar{X}^{(2)}$ в подпространстве $(x_1, x_2, \dots, x_t, \dots, x_m)$ определяется как $\rho_{12} = \sqrt{\sum_{t=1}^m (x_t^{(1)} - x_t^{(2)})^2}$.

при выполнении условия

$$n_j^A/n_j^B \geq k \text{ либо } n_j^B/n_j^A \geq k, \quad (\text{III.103})$$

где j — номер по порядку сфер M ($j=1, 2, \dots$), удовлетворяющих условию (III. 103) и лежащих в i -м подпространстве;

n_j^A, n_j^B — число точек классов A и B , принадлежащих сфере M_j ;

σ_i — стандартное отклонение точек выборки в i -м подпространстве;

σ_m — стандартное отклонение точек выборки, принадлежащих M_j ;

ρ_j — суммарное расстояние от центра сферы M_j до s ближайших точек малопредставленного в M_j класса, не принадлежащих сфере M_j .

Для нахождения сфер M_j применяется следующая процедура. Из обучающей выборки поочередно выбираются векторы-объекты $x^{(j)}$ ($j=1, 2, \dots, N$). Для каждого $x^{(j)}$ отбирается в i -м подпространстве p ближайших векторов x^i ($p < N$) (в смысле расстояния ρ_{jl}). Расстояния ρ_{jl} ранжируются в порядке возрастания, из них определяются те, для которых выполняется условие (III. 103), а из последних выбирается такое ρ_{jl}^* , при котором K^n максимально. Сфера с центром $x^{(j)}$ и радиусом ρ_{jl}^* является искомой сферой M_j .

2. После того как в соответствии с критериями (III. 100), (III. 102) и условиями (III. 101), (III. 103) найдена сфера $M^{(1)}$, она запоминается вместе с индексом того класса точек, число которых в сфере преобладает. Все векторы, попавшие в эту сферу, исключаются из массива обучения, а на оставшейся части выборки совершенно аналогично ищется сфера $M^{(2)}$, и т. д. Эта процедура продолжается до тех пор, пока не будет исчерпана исходная выборка обучения.

После этапа обучения классификации согласно предлагаемому алгоритму получаем ряд решающих правил, по которым можно распознавать вновь предъявляемое изображение. Как это происходит, рассмотрим на следующем примере.

Допустим, на некоторой исходной выборке обучения с помощью данного алгоритма было получено 10 сфер, которым принадлежат все изображения выборки обучения.

Причем, пусть M^1, M^3, M^5, M^7, M^9 характеризуют класс A ; $M^2, M^4, M^6, M^8, M^{10}$ — класс B . Тогда решающее правило можно записать в следующем виде:

$$\bar{X} \in A, \text{ если } C^1 \vee \bar{C}^2 C^3 \vee \bar{C}^2 \bar{C}^4 C^5 \vee \bar{C}^2 \bar{C}^4 \bar{C}^6 C^7 \vee \bar{C}^2 \bar{C}^4 \bar{C}^6 \bar{C}^8 C^9 = 1,$$

$$\bar{X} \in B, \text{ если } \bar{C}^1 C^2 \vee \bar{C}^1 C^3 C^4 \vee \bar{C}^1 \bar{C}^3 \bar{C}^5 C^6 \vee \bar{C}^1 \bar{C}^3 \bar{C}^5 \bar{C}^7 C^8 \vee \bar{C}^1 \times \\ \times \bar{C}^3 \bar{C}^5 \bar{C}^7 \bar{C}^9 C^{10} = 1,$$

где

$$C_i = \begin{cases} 1, & \text{если } \bar{X} \in M^i \\ 0, & \text{если } \bar{X} \in \bar{M}^i. \end{cases}$$

Как видно, из записанного решающего правила вновь предъявляемая точка x относится к тому из классов, к которому относится некоторая сфера M^i при условии, что $\bar{X} \in M^i$ и что она попала в нее в первую очередь, если просматривать все сферы M^q ($q=1, 2, \dots, t, \dots, 10$) в порядке возрастания q .

В описанном выше методе классифицируемой точке приписывается индекс класса, к которому принадлежит большинство ближайших то-

чек обучающей выборки, т. е. используется правило ближайших величин.

Сравним вероятность ошибки классификации по правилу одной ближайшей величины с вероятностью ошибки классификации по правилу Байеса.

Пусть известны $f(\bar{X}/A_1)$ и $f(\bar{X}/A_2)$ — условные плотности вероятности для классов A_1 и A_2 и $L(i, j)$ — матрица потерь ($i, j=1, 2$). Необходимо принять решение о принадлежности объекта к одному из классов.

Условные потери в результате отнесения объекта \bar{X} к классу j — $r_j(\bar{X})$, условный байесовский риск $r^*(\bar{X})$ и риск R^* определяются соответственно следующим образом:

$$r_j(\bar{X}) = \sum_{i=1}^R P(A_i/\bar{X})L(i, j), \quad (\text{III.104})$$

$$r^*(\bar{X}) = \min_j r_j(\bar{X}), \quad (\text{III.105})$$

$$R^* = M[r^*(\bar{X})], \quad (\text{III.106})$$

где M — математическое ожидание по \bar{X} .

В работе [143] вводится понятие риска при классификации по правилу одной ближайшей величины.

Пусть $\bar{X}'_N \in \{\bar{X}_1, \dots, \bar{X}_N\}$ — ближайшая к классифицируемому вектору \bar{X} величина и A'_N — класс, к которому принадлежит объект \bar{X}'_N . Если A — истинный класс \bar{X} , то появляются потери $L(A_1, A'_N)$. Риск R и условный риск $r(\bar{X}, \bar{X}'_N)$ правила одной ближайшей величины задаются как

$$R = \lim_{R \rightarrow \infty} R_N = \lim_{R \rightarrow \infty} M[L(A_1, A'_N)], \quad (\text{III.107})$$

$$r(\bar{X}, \bar{X}'_N) = M[L(A_1, A'_N)/\bar{X}, \bar{X}'_N], \quad (\text{III.108})$$

где математическое ожидание берется по A_1 и A'_N . Согласно теореме Байеса и при $L \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ имеем

$$r(\bar{X}, \bar{X}'_N) = P(A_1/\bar{X}) \cdot P(A_2/\bar{X}'_N) + P(A_2/\bar{X}) \cdot P(A_1/\bar{X}'_N). \quad (\text{III.109})$$

Как показано в [143], $r(\bar{X}, \bar{X}'_N)$ сходится с вероятностью 1 с правой частью выражения

$$r(\bar{X}) = 2P(A_1/\bar{X})P(A_2/\bar{X}) \quad (\text{III.110})$$

при условии, что \bar{X}'_N сходится с вероятностью 1 с \bar{X} . Величина $r(\bar{X})$ является пределом условного риска классификации по правилу одной ближайшей величины.

Из (III.104), (III.105) и (III.110) нетрудно получить

$$r(\bar{X}) = 2r^*(\bar{X})(1-r^*(\bar{X})). \quad (\text{III.111})$$

Переходя от величин $r(\bar{X})$ к величинам R , получаем следующую оценку [143]:

$$R^* \leq R \leq 2R^*(1-R^*). \quad (\text{III.112})$$

Аналогичные результаты могут быть получены для правила K_N ближайших величин.

Таким образом, вероятность ошибки правила ближайших величин ограничена снизу байесовской ошибкой R^* , а сверху двойной байесовской ошибкой. Это означает, что любое другое правило, основанное на всем множестве данных, а не на информации ближайших точек, может сократить вероятность ошибки максимум наполовину. В этом смысле половина полезной информации во множестве классифицируемых объектов заключена в области ближайших величин.

Практически этот вывод понятен, так как если объекты близки друг к другу, велика вероятность того, что они будут одинаково классифицироваться или, по крайней мере, иметь близкие апостериорные вероятности.

В заключение отметим, какими преимуществами обладает предлагаемый алгоритм по сравнению с известным методом ближайших величин.

Изложенный метод позволяет организовать поиск в переменном пространстве признаков, выделяя такое подпространство, в котором имеется оптимальная в смысле критерия K сфера. Выбор подпространств производится последовательным перебором всех одномерных, двумерных и т. д. R -мерных подпространств.

Наряду с информацией о ближайших точках алгоритм учитывает информацию о всем массиве данных благодаря коэффициентам представительности K^n и компактности K^k , а также информацию об удаленности локализованной в сфере компактной группы точек какого-либо одного класса до соседней группы точек другого класса. Кроме того, число ближайших величин, т. е. число точек, попадающих в искомую сферу, не фиксируется и может меняться в довольно широких пределах.

Перечисленные преимущества безусловно повышают точность классификации разработанного метода по сравнению с методом ближайших точек. Этот алгоритм наиболее пригоден для решения задач со сложными, полимодальными и многосвязными распределениями в классах.

Тот факт, что решающее правило образуется в отдельных, наиболее информативных с точки зрения выбранного критерия подпространствах, до минимума снижает влияние на классификацию шумов, присутствующих в малоинформативных признаках.

В результате обучения алгоритм информативных гиперсфер строит решающее правило, состоящее из нескольких сфер в различных подпространствах полного пространства признаков. Очевидно, признаки, участвующие в образовании наибольшего числа таких сфер или в образовании сфер, содержащих наибольшее число точек обучения, играют значительную роль в построении решающего правила и поэтому наиболее информативны.

С учетом этого замечания в качестве показателя информативности i -го признака по алгоритму информативных гиперсфер можно выбрать следующую величину:

$$I_i = \frac{b_i}{K} + \frac{a_i}{N}, \quad (\text{III.113})$$

- где b_i — число сфер, в образовании которых участвует признак i ;
 K — общее число сфер, составляющих решающее правило;
 a_i — общее число векторов обучения, попавших в сферы, в образовании которых участвует признак i ;
 N — число векторов обучения.

7. ОЦЕНКА КАЧЕСТВА ОБУЧЕНИЯ РАСПОЗНАВАНИЮ ОБРАЗОВ С УЧИТЕЛЕМ

Используя различные алгоритмы построения решающего правила, можно получать и различные решения сформулированной задачи распознавания. Возникает, очевидно, проблема оценки качества этих решений и выбора наилучшего.

Общепринятой мерой качества распознавания является вероятность ошибочной классификации. Необходимо заметить, что иногда при решении практических задач вероятность ошибочной классификации подменяется частотой ошибок при классификации контрольной выборки. Частота ошибок на контроле является случайной величиной и потому может служить лишь более или менее точной оценкой искомой вероятности ошибки. Ниже изложены точные методы математической статистики, позволяющие оценить эту вероятность по результату классификации представительной контрольной выборки.

При небольшой контрольной выборке надежность получаемой оценки весьма невелика. В том случае, когда число изученных геологических объектов настолько мало, что невозможно ни образовать контрольную выборку, ни получить достаточно полную информацию о законах распределения признаков, вообще никакие статистические оценки качества распознавания не могут быть сделаны. В связи с этим при малой контрольной выборке оценка качества распознавания требует создания новых методов.

Оценка качества решающего правила по результату классификации контрольной выборки

Все вычисления в настоящем подразделе будут проводиться для вероятности P ошибки классификации. В принципе, сделав необходимые переформулировки, их можно было бы проводить для вероятности неправильной классификации объектов i -го класса или для вероятности отнесения объектов i -го класса к j -у классу.

Общая схема оценки вероятности ошибки P такова. Контрольная выборка, состоящая из n объектов, классифицируется с помощью исследуемого решающего правила. Предположим, что m объектов оказались классифицированы неверно, а остальные $n - m$ — правильно. Естественно после этого ожидать (если n было достаточно большим), что число $P^* = \frac{m}{n}$ (частота ошибок в эксперименте) близко к истинной вероятности ошибки P . Это число обычно и принимается за оценку вероятности ошибки.

Очевидно, что сама по себе оценка P^* без указания ее «близости» к P не имеет никакой ценности. Поэтому согласно общей теории статистических решений необходимо построить доверительный интервал (\underline{P}, \bar{P}) для истинной вероятности ошибки P с достаточно высоким коэффициентом доверия, т. е. указать такой интервал (P, \bar{P}) , в котором

$$P\{P \leq P \leq \bar{P}\} = \gamma \quad (\text{III.114})$$

(вероятность того, что P заключена в этом интервале, равна γ). Число γ выбирается исходя из практических нужд, обычно 0,90, 0,95 или 0,99.

Так как число ошибок при классификации контрольной выборки фиксированного объема является случайной величиной, распределенной по биномиальному закону, то согласно общей теории интерваль-

ных оценок границы доверительного интервала (\underline{P}, \bar{P}) с коэффициентом доверия γ вычисляются по формулам:

$$\left. \begin{aligned} \sum_{i=m}^n C_n^i \underline{P}^i (1-\underline{P})^{n-i} &= \frac{1-\gamma}{2} \\ \sum_{i=0}^m C_n^i \bar{P}^i (1-\bar{P})^{n-i} &= \frac{1-\gamma}{2} \end{aligned} \right\} \text{при } m \neq 0 \quad (\text{III.115})$$

или $\underline{P}=0$; $\bar{P}=1-(1-\gamma)^{\frac{1}{n}}$ при $m=0$.

Часто исследователя интересует только верхний доверительный предел \bar{P} для вероятности ошибки P (т. е. такое число \bar{P} , что $P\{P \leq \bar{P}\} = \gamma$). В таком случае \bar{P} находится из формулы

$$\sum_{i=0}^m C_n^i \bar{P}^i (1-\bar{P})^{n-i} = 1-\gamma \quad \text{при } m \neq 0 \quad (\text{III.116})$$

или $\bar{P}=1-(1-\gamma)^{\frac{1}{n}}$ при $m=0$.

В работе Л. В. Бельшева и Н. В. Смирнова [6] имеются таблицы доверительных интервалов для

$$\gamma = 0,90; 0,95 \text{ и } 0,99 \text{ и } m, n-m = 0(1), 20(2), 30(3), 100, 500, \infty.$$

Из этих же таблиц находится верхний доверительный предел при

$$\gamma = 0,95; 0,975; 0,995.$$

Учитывая, что $\sum_{i=m}^n C_n^i P^i (1-P)^{n-i} = I(m, n-m+1)$, где $I_x(a, b)$ есть неполная бета-функция, соотношения (III.115) и (III.116) можно переписать соответственно в виде

$$I_P(m, n-m+1) = \frac{1-\gamma}{2}$$

$$\text{и } I_{\bar{P}}(m+1, n-m) = \frac{1+\gamma}{2}, \quad (\text{III.117})$$

$$I_{\bar{P}}(m+1, n-m) = \gamma. \quad (\text{III.118})$$

В упомянутой работе помещены таблицы значений $I_x(a, b)$ для $a \leq b$, $b \geq 50$. На практике редко бывает нужна идеальная точность доверительных границ, и практически пользуются двумя следующими приближениями формул (III.117).

1. Формула Пуассона в случае $P < 0,1$:

$$\left. \begin{aligned} e^{-nP} \sum_{i=m}^n \frac{(nP)^i}{i!} &= \frac{1+\gamma}{2} \\ e^{-n\bar{P}} \sum_{i=0}^m \frac{(n\bar{P})^i}{i!} &= \frac{1-\gamma}{2} \end{aligned} \right\} \text{для } m \neq 0 \quad (\text{III.119})$$

или $P=0$; $\bar{P} = \frac{\ln(1-\gamma)}{n}$ для $m=0$.

2. Формула Муавра-Лапласа в случае $nP(1-P) > 9$:

$$\underline{P} \approx \frac{m - 0,5}{n} - t_{\frac{1+\gamma}{2}} \sqrt{\frac{P^*(1-P^*)}{n}}, \quad (\text{III.120})$$

$$\bar{P} \approx \frac{m + 0,5}{n} + t_{\frac{1-\gamma}{2}} \sqrt{\frac{P^*(1-P^*)}{n}},$$

где $t_{\frac{1+\gamma}{2}}$ есть $\frac{1+\gamma}{2} \cdot 100\%$ — квантиль нормального распределения.

Приближения Пуассона и Муавра-Лапласа для формулы (III.118) примут вид соответственно

$$e^{-n\bar{P}} \sum_{i=0}^m \frac{(n\bar{P})^i}{i!} = 1 - \gamma \quad \text{для } m \neq 0 \quad (\text{III.121})$$

или

$$\bar{P} = -\frac{\ln(1-\gamma)}{n} \quad \text{для } m = 0$$

и

$$\bar{P} \approx \frac{m + 0,5}{n} + t_{\gamma} \sqrt{\frac{P^*(1-P^*)}{n}}. \quad (\text{III.122})$$

Таблицы распределения Пуассона, доверительных пределов для P при распределении Пуассона, квантилей нормального распределения имеются, например, в работе [6].

Таблица 2

Длина доверительного интервала Δ
в зависимости от объема
контрольной выборки n
и числа ошибочно классифицированных
объектов m
(для коэффициента доверия $\gamma = 0,90$)

$n=10$	m	0	1	2	3	4
$n=10$	Δ	0,26	0,39	0,47	0,52	0,55
$n=20$	m	0	2	4	6	8
$n=20$	Δ	0,14	0,26	0,33	0,37	0,39
$n=40$	m	0	5	10	14	
$n=40$	Δ	0,07	0,19	0,25	0,26	
$n=60$	m	0	10	15	20	
$n=60$	Δ	0,05	0,16	0,20	0,22	

Таблица 3

Значения верхнего доверительного
предела \bar{P} в зависимости от объема
контрольной выборки n и числа
ошибочно классифицированных
объектов m
(для коэффициента доверия $\gamma = 0,95$)

$n=10$	m	0	1	2	3
$n=10$	\bar{P}	0,26	0,39	0,51	0,61
$n=20$	m	0	2	4	6
$n=20$	\bar{P}	0,14	0,28	0,40	0,51
$n=40$	m	0	5	10	14
$n=40$	\bar{P}	0,07	0,24	0,39	0,49
$n=60$	m	0	10	15	20
$n=60$	\bar{P}	0,05	0,27	0,36	0,45

Используя формулы (III.115), можно определить объем контрольной выборки, необходимый для достижения заданной точности оценки P . В настоящей работе приведены таблицы длины Δ доверительного интервала для P с коэффициентом доверия 0,9 (табл. 2) и значений верхнего доверительного предела \bar{P} для P с коэффициентом доверия 0,95 (табл. 3) в зависимости от объема n контрольной выборки и числа m ошибочно классифицированных контрольных объектов.

Из этих таблиц видно, что даже при контрольных выборках объема 40—60 точность оценки вероятности ошибки колеблется в пределах 0,1—0,25. Таким образом, оценку качества распознавания, сделанную на контрольной выборке объема 30—40 и меньше, следует считать недостаточно надежной.

Сравнительная оценка решающих правил в задачах с малыми выборками обучения

Рассмотрим задачу распознавания с малым числом эталонных объектов, настолько малым, что невозможно ни сделать какие-либо обоснованные заключения о законе многомерного распределения признаков, ни выделить контрольную выборку достаточного объема (к задачам такого типа заведомо относятся те, в которых число учитываемых при распознавании параметров превышает число эталонных объектов). В этом случае никакие методы определения вероятности ошибки классификации не могут привести к успеху. Возникает необходимость выбора какого-либо нового нестатистического критерия качества решающего правила. При построении этого критерия будем предполагать, что все признаки объектов имеют непрерывную область значений.

Прежде всего отметим следующее. В последнее время в теории распознавания образов получила распространение так называемая «гипотеза простоты» [41]. Согласно этой гипотезе, если два решающих правила с одинаковой точностью дают обучающую выборку, то лучшим из них будет то, которое более просто. «Простота» обычно определяется как количество степеней свободы параметров, определяющих решающую поверхность.

В обычной ситуации, с которой сталкиваются в распознавании, когда число эталонных объектов много больше размерности пространства, эта гипотеза вовсе не означает «дискриминации» сложных решающих правил, ведь поверхности с большим числом степеней свободы, вообще говоря, точнее делят обучающую выборку.

В связи со всем сказанным будет естественным под качеством решающего правила понимать некоторую геометрическую характеристику взаимного положения точек обучающей выборки и разделяющей поверхности.

Будем называть хорошим то разделение, которое удовлетворяет следующим двум эвристическим условиям (рассматриваются решающие правила, которые безошибочно классифицируют обучающую выборку):

- 1) решающая поверхность проходит возможно далеко от точек (представителей) разделяемых классов;
- 2) решающая поверхность проходит ближе к более компактному классу.

Сформулируем математическую постановку задачи. В n -мерном пространстве признаков R_n имеется обучающая выборка, состоящая из объектов двух классов: $x_1^{(1)}, \dots, x_k^{(1)}$; $x_1^{(2)}, \dots, x_m^{(2)}$. Имеется некоторый

набор D решающих поверхностей Γ , построенных различными алгоритмами, каждая из которых делит все пространство R_n на две области U_i^1 и U_i^2 такие, что $x_j^{(1)} \in U_i^1$ ($j=1, 2, \dots, k$); $x_j^{(2)} \in U_i^2$ ($j=1, 2, \dots, m$). Требуется построить функционал F , зависящий от выборки $\{x\} = \{x_1^{(1)}, \dots, x_k^{(1)}; x_1^{(2)}, \dots, x_m^{(2)}\}$ и разбиения Γ пространства и принимающий большие значения на тех разбиениях Γ , которые лучше удовлетворяют условиям 1 и 2.

Лучшей среди всех границ Γ_i из множества D будет считаться такая граница Γ , для которой «функционал качества» F принимает наибольшее значение:

$$F(\{x\}, \Gamma) = \max_{\Gamma \in D} F(\{x\}, \Gamma_i).$$

В качестве функционала F предлагается выбрать

$$F_1 = \frac{\sum_{i=1}^k \rho(x_i^{(1)}, \Gamma)}{k\sigma_1} + \frac{\sum_{i=1}^m \rho(x_i^{(2)}, \Gamma)}{m\sigma_2}, \quad (\text{III.123})$$

где $\rho(x, \Gamma)$ — расстояние от точки x до делящей границы Γ , σ_i — выборочное среднеквадратичное отклонение в i -м классе.

Функционал F_1 обладает одним существенным недостатком — его трудно вычислять. В ряде случаев при сложных решающих поверхностях, которые к тому же могут не строиться в явном виде при распознавании, вычисление $\rho(x, \Gamma)$ может оказаться вообще невозможным. Поэтому ниже мы построим функционал F_2 , достаточно близкий по геометрической сущности к F_1 и в то же время вычисляемый с разумными затратами машинного времени.

Введем несколько обозначений: $\rho_{i,j}^{(1)}$ — расстояние от $x_j^{(1)}$ до границы Γ по отрезку, соединяющему $x_i^{(1)}$ с $x_j^{(2)}$; $\rho_{i,j}^{(2)}$ — расстояние от $x_j^{(2)}$ до границы Γ по отрезку, соединяющему $x_i^{(1)}$ с $x_j^{(2)}$. Не следует, однако, думать, что обязательно $\rho_{ij}^{(1)} + \rho_{ij}^{(2)} = \rho(x_i^{(1)}, x_j^{(2)})$. Функционал

$$F_2 = \frac{\sum_{i=1}^m \min_j \rho_{ij}^{(1)}}{k\sigma_1} + \frac{\sum_{j=1}^m \min_i \rho_{ij}^{(2)}}{m\sigma_2} \quad (\text{III.124})$$

отличается от F_1 тем, что расстояние от точки до границы в выражении (III.123) заменено минимумом расстояний от точки до границы по конечному числу направлений. Выражение (III.124) содержит лишь арифметические действия над величинами $\rho_{ij}^{(1)}$ и $\rho_{ij}^{(2)}$, которые могут быть вычислены без использования явного задания решающей поверхности, а с помощью только самого алгоритма классификации.

Таким образом, функционал F_2 может быть применен при сравнительной оценке качества различных решений задач классификации с непрерывными параметрами.

Дальнейшее усовершенствование критерия качества решающего правила может заключаться в более полном учете геометрических свойств деления, в учете цен ошибок I и II рода и в обобщении критерия на случай задач распознавания, в которых часть признаков принимает дискретный ряд значений.

8. МЕТОД РАСПОЗНАВАНИЯ ОБРАЗОВ БЕЗ УЧИТЕЛЯ — «ТРАНЗИТИВ»*

Задача распознавания образов без учителя** заключается в том, что по мере изучения свойств или параметров множества исследуемых объектов (явлений) последние разбиваются на группы, классы или подмножества, причем объекты, включаемые в одно подмножество, характеризуются одинаковыми или близкими значениями некоторых параметров, но зато объекты, включаемые в разные подмножества, по этим же самым параметрам (если и не по всем, то по большинству из них) существенно отличаются друг от друга. При этом какие-либо предварительные знания о принадлежности хотя бы части рассматриваемых объектов к каким-либо качественно однородным совокупностям отсутствуют.

Задачу обычно формируют следующим образом. Имеется множество реализаций X (векторов, точек) в пространстве параметров x_1, \dots, x_n и критерий, характеризующий степень сходства или близости между любыми рассматриваемыми подмножествами точек. Требуется разделить множество реализаций на априори заданное количество классов (подмножеств) векторов таким образом, чтобы в смысле заданного критерия сходства векторы, попавшие в один класс, были расположены как можно ближе друг к другу, а векторы различных классов — как можно дальше друг от друга.

В более общем случае решения задачи количество классов заранее не задается, но зато устанавливается критерий, который позволяет оценить целесообразность продолжения процесса деления исследуемой выборки.

Следует отметить, что теория и методы распознавания образов без учителя разработаны в гораздо меньшей степени, чем теория и методы решения задач классификации априори идентифицированных реализаций. В основном в настоящее время при разработке и оценке алгоритмов таксономии приходится опираться на эмпирические соображения и использовать подход эвристического программирования.

Основная трудность решения прикладных задач, по-видимому, связана со сложностью задания объективного критерия меры подобия объектов исследования, которую необходимо использовать при группировке.

На эффективность решения задач таксономии оказывают влияние, во-первых, такие факторы, как исходное описание объектов (т. е. выбор пространства признаков) и, во-вторых, сам принцип разделения объектов в выбранном пространстве признаков (т. е. используемый критерий). Выбор пространства признаков является первым и определяющим шагом при проведении таксономии, так как в пространстве признаков, не имеющем достаточной информации о количественных особенностях интересующих классов объектов, никакой принцип разделения объектов не будет эффективен. Однако и при условии хорошо выбранного пространства признаков один и тот же принцип (критерий) разделения (таксономии) объектов не может быть одинаково эффективен во всех случаях, поэтому желательно иметь набор таких критериев.

Ниже для получения различных принципов таксономии предлагается использовать аппарат бинарных отношений.

* Предложен Л. С. Лукьяновым.

** Эта задача имеет еще другие названия: задача таксономии или распознавания образов без поощрения.

Пусть M — множество, а $\rho \subseteq M \times M$ — подмножество декартового произведения множества M на себя, тогда упорядоченная пара $\langle \rho, M \rangle$ называется отношением $\langle \rho, M \rangle$. Множество M называется областью задания отношения, а множество ρ — графиком отношения.

Обычно, если ясно, какое множество M подразумевается, то отношения обозначают одной буквой ρ . Далее будем придерживаться этого обозначения. Кроме того, если $\langle x, y \rangle \in \rho$ ($x, y \in M$), то будем писать это в виде выражения $x\rho y$.

Введем теперь некоторые операции над отношениями. Пересечением отношений $\rho \cap \sigma$ назовем отношение, определяемое пересечением соответствующих графиков отношений ρ и σ . Аналогично, объединением отношений $\rho \cup \sigma$ назовем отношение, определяемое объединением соответствующих графиков отношений ρ и σ . Для отношений можно определить понятие включения. Будем писать $\rho \subseteq \sigma$; ($\rho \subset \sigma$), если график отношения $\rho \subseteq$ графике отношения σ ($\rho \subset \sigma$).

Определим теперь переход к обратному отношению. Если ρ — отношение на множестве M , то обратное отношение ρ^{-1} определяется условием: $x\rho^{-1}y$ эквивалентно $y\rho x$. Важную роль играет операция, обозначаемая $\rho\sigma$ — произведение отношений. Эта операция определяется следующим образом: $x(\rho\sigma)y$ эквивалентно тому, что существует такое $z \in M$, для которого выполнены соотношения $x\rho z$ и $z\sigma y$. Используем еще одну операцию, которая называется транзитивным замыканием отношения ρ и будет обозначаться через $\bar{\rho}$.

Если ρ — некоторое бинарное отношение на множестве M , то его транзитивное замыкание определяется следующим образом. Соотношение $x\rho y$ считается выполненным, если существует цепочка элементов из M ($z_0 = x, z_1, \dots, z_n = y$) такая, что между соседними элементами этой цепочки выполнено соотношение ρ : $z_0\rho z_1, z_1\rho z_2, \dots, z_{n-1}\rho z_n$. Можно показать, что $\bar{\rho} = \rho \cup \rho^2 \cup \dots \cup \rho^n \cup \dots$. Отсюда следует, что $\bar{\rho} = \rho$.

Теперь можно выразить понятие таксономии в терминах бинарных отношений. Для этого понадобится понятие отношения эквивалентности.

Систему непустых подмножеств $\{M_1, M_2, \dots\}$ множества M будем называть разбиением этого множества, если

$$M = M_1 \cap M_2 \cap \dots \quad \text{и} \quad M_i \cap M_j = \emptyset \quad \text{при} \quad i \neq j.$$

Сами множества M_1, M_2, \dots называются классами данного разбиения.

Отношение ρ на множестве M называется эквивалентностью (или отношением эквивалентности), если существует разбиение $\{M_1, \dots\}$ множества M такое, что соотношение $x\rho y$ выполняется тогда и только тогда, когда x и y принадлежат некоторому общему классу M_i данного разбиения.

Пусть $\{M_1, M_2, \dots\}$ — разбиение множества M . Определим исходя из этого разбиения отношение ρ на M , т. е. $x\rho y$, если x и y принадлежат некоторому общему классу M_i данного разбиения.

Очевидно, что отношение ρ является эквивалентностью. Назовем ρ отношением эквивалентности, соответствующим исходному разбиению.

Так как таксономия — это разбиение множества на классы, то ей соответствует некоторое отношение эквивалентности. Отношение эквивалентности удобнее разбиения множества тем, что оно естественно возникает из многих других бинарных отношений. Это обстоятельство

позволяет развить некоторые эвристические соображения о возможных принципах таксономии. Проиллюстрируем это примером.

Пусть задано разбиение $\{M_1, M_2, \dots\}$ множества M . Выберем в каждом множестве M_i некоторый содержащийся в нем элемент x_i . Этот элемент будем называть эталоном для всякого элемента y , входящего в то же множество M_i . По определению считается выполненным соотношением $x_i \rho y$. ρ назовем отношением «быть эталоном». Нетрудно усмотреть, что эквивалентность $\hat{\rho}$, соответствующая исходному разбиению, может быть определена так: $y \hat{\rho} z$, если y и z имеют общий эталон: $x_i \rho y$ и $x_i \rho z$. Понятно, что любое отношение эквивалентности может быть определено с помощью отношения «быть эталоном» и, наоборот, любое отношение «быть эталоном» определяет некоторую эквивалентность. Это можно выразить формулой $\hat{\rho} = \rho^{-1} \rho$.

Отношение «быть эталоном» представляет интерес в том случае, если разбиение множества (или отношение эквивалентности) еще не задано, тогда можно получить их, используя формулу $\rho = \rho^{-1} \rho$, т. е. если имеется какой-то способ выделять «типичные» объекты, то на основе этого выделения можно провести таксономию. Таким образом, таксономия может опираться на алгоритмы типа известного алгоритма «Голотип».

Отношение эквивалентности содержательно интерпретируется как абсолютная взаимозаменяемость объектов по некоторому качеству. Однако обычно не представляется возможным непосредственно выяснить, обладает ли объект интересующим нас качеством. Поэтому прибегают к другим свойствам, формально выражаемым некоторыми бинарными отношениями, из которых затем конструируется отношение эквивалентности. Насколько адекватно построенное таким образом отношение эквивалентности отражает интересующее нас свойство, можно выяснить только из опыта.

В этом круге идей естественно рассмотреть бинарные отношения близости. Однако предварительно введем понятие эквивалентности равносильное, как известно [58], ранее введенному, но более удобное для дальнейших рассуждений.

Отношение ρ на множестве M называется эквивалентностью, если оно рефлексивно, симметрично и транзитивно. В формулах это выражается следующим образом:

$$\begin{aligned} 1) \quad & \rho \ni \Delta, \\ 2) \quad & \rho = \rho^{-1}, \\ 3) \quad & \rho \cdot \rho \in \rho, \end{aligned} \quad (\text{III.125})$$

где Δ — диагональное отношение.

Бинарное отношение близости (толерантность) на множестве M — это любое симметричное и рефлексивное бинарное отношение.

Определение естественно, так как каждый объект близок самому себе (рефлексивность), и если объект x близок объекту y , то и объект y близок объекту x (симметричность).

Обычно отношение близости в пространстве признаков задается следующим образом: два объекта (x и y) считаются близкими ($x \rho y$), если расстояние между ними $r(x, y)$ меньше некоторого параметра близости R ($r(x, y) < R$). Однако в принципе допускается любое другое (в том числе и не метрическое) определение близости; обязательны только симметричность и рефлексивность.

Вообще говоря, по сравнению с эквивалентностью толерантность более богатое в смысле содержащейся информации отношение. Однако при большом числе объектов и значительном количестве связей между ними эта информация слишком громоздка для непосредственного анализа, поэтому имеет смысл заменить отношение толерантности на отношение эквивалентности.

Это можно сделать разными способами, но если требовать единственности такой замены, то в нашем распоряжении остается только операция транзитивного замыкания.

Таким образом, бинарному отношению близости ρ ставится в соответствие его транзитивное замыкание $\bar{\rho}$, которое является эквивалентностью, т. е. некоторой таксономией. Именно этот принцип был заложен в алгоритм «Транзитив».

Пусть имеется K n -мерных векторов $\bar{X}(x_1, x_2, \dots, x_n)$. Требуется выполнить процедуру таксономии описанным выше способом. В качестве меры близости в алгоритме «Транзитив» выбрано отношение близости, инвариантное относительно евклидовых преобразований.

Таким образом, все векторы, близкие к нулевому вектору, задаются условием:

$$x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2 < R.$$

Близость к другим векторам (в силу инвариантности отношения близости относительно параллельных переносов) задается условием: $(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2 < R$. Здесь x_1, x_2, \dots, x_n — координаты одного вектора; y_1, y_2, \dots, y_n — координаты другого вектора; R — положительное число. Тогда

$$R(x, y) = \sqrt{\sum_{i=1}^n C_i (x_i - y_i)^2}, \quad (\text{III.126})$$

где $C_i = 0$ или 1.

Подразумевая естественное требование связности, определим отношение близости с точностью до выбора положительно числа R . Алгоритм осуществляет таксономию для целого ряда значений R , выбираемых из содержательных соображений. В алгоритме предусмотрена возможность игнорировать близость по некоторым координатам, что позволяет выяснить вклад этих координат в общую картину таксономии. Массив векторов можно по желанию подвергнуть нормализации.

В результате работы алгоритм «Транзитив» разбивает исходное множество векторов на таксоны (группы), обладающие следующими свойствами.

1. Для любых векторов x и y из одного таксона существует такая последовательность векторов z_1, \dots, z_t ; ($t \geq 0$) этого же таксона, что $R(x, z) \leq R^*$, $R(z_1 z_2) \leq R^*$, ..., $R(z_{t-1}, z_t) \leq R^*$, $R(z_t y) \leq R^*$.

2. Для любых векторов x и y из разных таксонов $R(x, y) > R$. Выделенные таксоны являются группами схожих в смысле указанной меры близости векторов.

Изменение коэффициентов C_i позволяет выяснить значение отдельных признаков (факторов) для группировки векторов. Изменение величины R^* позволяет регулировать меру близости векторов внутри таксонов и величину таксонов.

9. РЕЗУЛЬТАТЫ СОПОСТАВЛЕНИЯ ЭФФЕКТИВНОСТИ НЕКОТОРЫХ АЛГОРИТМОВ РАСПОЗНАВАНИЯ ОБРАЗОВ С УЧИТЕЛЕМ

С целью сопоставительной оценки эффективности ряда различных алгоритмов распознавания образов, применяемых в прикладной геологии, было проведено исследование на трех типах задач геологического прогнозирования, охватывающих разные этапы геологических исследований (всего пять задач [105]).

1. Оценка редкометалльных рудных полей по комплексу геолого-геофизических и геохимических данных — мелкомасштабное прогнозирование.

2. Оценка масштабности месторождений по комплексу косвенных геологических данных (месторождения ртути и россыпные месторождения олова) — среднемасштабное прогнозирование.

3. Определение уровня эрозионного среза редкометалльного месторождения по геохимическим данным (две задачи) — крупномасштабное прогнозирование.

Выбор конкретных задач был обусловлен прежде всего необходимостью сопоставления эффективности алгоритмов распознавания при различных типах исходной информации и объемах обучающих выборок. Исходная информация в выбранных задачах носила качественный (задачи оценки месторождений олова и ртути) и количественный характер (определение уровня эрозионного среза месторождения). При решении задачи оценки рудных полей исходная информация имела смешанный характер: рудные поля характеризовались как качественными (геологическими), так и количественными (геофизическими, геохимическими) признаками одновременно.

Обучающие выборки в решаемых задачах были как весьма небольшими по объему (например, при оценке месторождений ртути обучающая выборка включала 11 объектов), так и достаточно представительными (при решении задачи определения уровня эрозионного среза месторождения обучающая выборка состояла из 60 объектов).

Для сопоставления были взяты следующие алгоритмы: «Гиперпласт» (см. раздел III.6), тупиковые тесты [32], модификация алгоритма тупиковых тестов [120], «Потенциал-2» [109], «Перебор» [98], «Обобщенный портрет» [116].

Выбор указанных алгоритмов распознавания был обусловлен тем обстоятельством, что эти алгоритмы нашли достаточно широкое применение в прикладной геологии.

Краткие сведения о сопоставляемых алгоритмах

Алгоритм тупиковых тестов. Алгоритм основан на поиске матрицы T , являющейся описанием объектов обучения (объекты, как правило, кодируются в двоичной системе) специальных подмножеств — так называемых тупиковых тестов.

Подмножество столбцов (i_1, i_2, \dots, i_l) матрицы T называется тестом, если после удаления из T всех столбцов, за исключением (i_1, i_2, \dots, i_l) , получается матрица, все строки которой различаются между собой. Тест называется тупиковым, если после исключения любого столбца он перестает быть тестом.

Для каждого из признаков x_1, x_2, \dots, x_n , характеризующих объекты, вычисляется информационный вес признака $P(x_i) = \frac{n_i}{N}$ ($i=1, 2, \dots, n$), где n — число вхождений признака x_i в тупиковые тесты; N — общее число тупиковых тестов для матрицы T .

Информационный вес объекта, подлежащего классификации, определяется как сумма информационных весов тех признаков, которые присутствуют на данном объекте (принимают кодовое значение «1»). Значения информационных весов объектов служат основой для построения классификации объектов.

Программа, реализующая алгоритм тупиковых тестов на ЭЦВМ типа БЭСМ-6, позволяет обрабатывать матрицы с числом объектов обучения $l \leq 14$, характеризующихся числом признаков $n \leq 48$.

Модификация алгоритма тупиковых тестов. Авторами алгоритма предложено наряду с информационным весом признака $P(x_i)$; ($i=1, 2, \dots, n$) вычислять значение разделяющего веса признака x_i соответствующего значения, которое может принимать признак.

Пусть для определенности каждый признак может принимать два значения — «0» и «1» (признаки закодированы в двоичной системе). Поставим в соответствие каждому объекту обучения действительное число y_j ; ($j=1, 2, \dots, m$, где m — число объектов обучения), так называемую степень j -го эталона; например, при решении задач оценки месторождений это может быть число, пропорциональное запасам эталонного месторождения. Тогда разделяющие веса i -го признака, соответствующие значениям «0» и «1», вычисляются как

$$P_i^*(0) = \frac{\sum_{j \in I_0} y_j}{\sum_{j=1}^m y_j} \cdot P(x_i); \quad P_i^*(1) = \frac{\sum_{j \in I_1} y_j}{\sum_{j=1}^m y_j} \cdot P(x_i),$$

где I_0 и I_1 — подмножества, состоящие из объектов обучения, на которых признак x_i принимает соответственно значение «0» и «1».

Информационный вес объекта определяется как сумма разделяющих весов признаков, соответствующих значениям, принимаемым ими на объекте. Как и ранее, информационные веса объектов служат основой для построения классификации объектов.

Алгоритм «Потенциал-2» основан на вычислении в n -мерном пространстве признаков для каждого из классифицируемых объектов «потенциалов»

$$V_i^s = \frac{1}{l_s} \sum_{k=1}^{l_s} \frac{\alpha_k}{\left[\sum_{j=1}^n \beta_i (x_{kj} - x'_j)^2 \right]}; \quad i=1, 2, 3,$$

где x_{ki}, x_j^1 — координаты j -го признака k -го объекта обучения и классифицируемого объекта;

l_s — число объектов обучения s -го класса;

α_k, β_i — «веса» k -го объекта обучения и j -го признака соответственно, обусловленные совокупностью объектов разных классов; классифицируемый объект относится к тому из классов, потенциалы которого будут большими, чем для других классов.

С целью контроля эффективности работы алгоритма на самих объектах обучения предусмотрена процедура поочередного исключения

объектов из обучающей выборки и их контроль на основе остающихся объектов. Число δ ошибочно классифицированных объектов обучения при этом служит одним из показателей эффективности распознавания. Для выделения информативной совокупности признаков указанная процедура повторяется. Вначале при поочередном исключении каждого из n признаков признак, исключение которого обусловило наименьшее δ , выбрасывают из совокупности как «наихудший». Таким же путем осуществляется выбрасывание «наихудшего» признака в остающихся совокупностях из $n-1$, $n-2$ и т. д. признаков.

Программа, реализующая алгоритм «Потенциал-2» на ЭВМ типа БЭСМ-4, позволяет обрабатывать массивы, характеризуемые величиной $(n+1)l_{\Sigma} < 1886$ (l_{Σ} — число объектов обучения и контроля, n — число признаков, характеризующих объекты).

Алгоритм «Перебор» основан на той же идее перебора всех возможных сочетаний признаков по 2, по 3, по 1, что и широко известный алгоритм «Кора-3», но работает с другим критерием отбора сочетаний признаков (признаки кодируются в двоичной системе). Изменение критерия отбора сочетаний позволило устранить ряд особенностей алгоритма «Кора-3», создающих весьма серьезные затруднения при решении многих геологических задач.

Проверяемое сочетание A_i (с фиксированными значениями признаков) отбирается как характеристическое, если оно удовлетворяет следующим условиям:

$$m_1 > k_1 \text{ и } \frac{P_1}{P_1 + P_2} \geq \alpha$$

$$\text{или } m_2 \geq k_2 \text{ и } \frac{P_2}{P_1 + P_2} \geq \alpha,$$

где m_1 , m_2 — число объектов обучения I и II классов (из общего числа l_1 и l_2 соответственно), характеризуемых данным сочетанием A_i ; $P_1 = \frac{m_1}{l_1}$; $P_2 = \frac{m_2}{l_2}$; k_1 , k_2 , α — выборочные пороги.

Условия в правой части критерия используются для того, чтобы отбираемые сочетания характеризовали преимущественно один из классов (были бы достаточно «контрастными»).

Отнесение объектов к тому или иному классу, а также оценка информативности признаков производится с учетом частоты вхождения признаков в отобранные характеристические отношения.

Программа, реализующая алгоритм «Перебор» на ЭВМ типа БЭСМ-4, позволяет обрабатывать до 90 объектов обучения I и II классов, характеризуемых числом признаков $n \leq 124$.

Алгоритм «Обобщенный портрет» предназначен для построения линейного решающего правила, разделяющего пространство E_n на два класса. Для построения разделяющей плоскости в том случае, когда обучающая последовательность полностью разделима линейным правилом, находится такой единственный вектор $\bar{\varphi}$, что

$$\min_{x \in \text{I кл.}} (\bar{\varphi} \bar{X}) \geq k \cdot \max_{x \in \text{II кл.}} (\bar{\varphi} \bar{X}); \quad k < 1 \text{ и}$$

при этом достигается условный максимум $\min_{x \in \text{I кл.}} (\bar{\varphi} \bar{X})$.

Решающее правило определяется следующим образом:

$$x \in I \text{ кл., если } (\bar{\varphi} \bar{X}) \geq \frac{1+k}{2}; \quad x \in II \text{ кл., если } (\bar{\varphi} \bar{X}) < \frac{1+k}{2}.$$

Информативность i -го признака оценивается по величине $(|\bar{\varphi}_i|)$, где φ_i — i -я координата вектора $\bar{\varphi}$. Если материал обучения не делится гиперплоскостью, алгоритм определяет векторы, наиболее «мешающие» разделению, и исключает их из обучения. Для оставшихся объектов строится вектор $\bar{\varphi}$, как описано ранее. Признаками для обработки алгоритмом «обобщенный портрет» кодируются в двоичной системе. Программа, реализующая алгоритм на ЭВМ типа БЭСМ-4, позволяет обрабатывать объекты, характеризуемые числом признаков $n \leq 200$, число объектов обучения может достигать значения $\frac{40000}{n}$, но не более 700.

Решаемые геологические задачи

1. *Оценка редкометалльных рудных полей по комплексу геолого-геофизических и геохимических данных.* Для решения задачи оценки редкометалльных рудных полей в пределах одного из регионов СССР было выделено 24 рудных поля. Тринадцать из них (пять промышленных и восемь непромышленных) составили обучающую выборку. Одиннадцать рудных полей (одно промышленное и десять непромышленных) составили независимую контрольную выборку.

Каждое из рудных полей было описано 59 признаками, закодированными в двоичной системе. Этими признаками были охарактеризованы: особенности гравимагнитных полей (№ кода 1—12, 43, 44); основные элементы стратиграфии и магматизма (№ 23—35, 48, 58, 59), структурные факторы (№ 36—42, 52, 57); положение относительно сопутствующих эндогенных проявлений (№ 45—47), данные анализа подземных вод (№ 49—50), площадные аномалии гелия (№ 51).

2. *Оценка россыпных месторождений олова.* Для решения задачи оценки россыпных месторождений олова методами распознавания была сформирована обучающая выборка из семи промышленных и семи непромышленных месторождений, кроме того, 12 месторождений (6 промышленных и 6 непромышленных) были взяты в качестве независимых объектов контроля.

Каждый из объектов был представлен совокупностью 48 геологических и геоморфологических признаков, характеризующих коренные месторождения и рудопроявления в бассейне (№ 1—3); принадлежность коренных источников к той или иной рудной формации (№ 4—7); морфологию рудных тел коренных источников (№ 8—9); положение россыпи относительно интрузивных массивов (№ 10—12); относительную величину денудационного среза (№ 13—15); интрузивные породы в области питания россыпи (№ 16—19); содержание касситерита—вольфрамитов в шлиховых пробах (№ 20—23); комплексы шлихо-минералогических ассоциаций (№ 24—27); положение коренных источников относительно россыпи (№ 28—31); уклон долины и характер рельефа (№ 32—36); возрастные комплексы оловоносных рыхлых отложений (№ 37—42); крутизну и форму склонов долины (№ 43—48). Подробнее задача описана в работе [101].

3. *Оценка масштабов ртутных месторождений.* Для решения задачи оценки ртутных месторождений было отобрано 5 крупных и 6 мел-

ких месторождений, составивших обучающую выборку; кроме того, 8 ртутных месторождений (2 крупных и 6 мелких) были взяты в качестве независимых объектов экзамена (более подробно условия решения этой задачи приведены в работе [96]).

Каждое из отобранных месторождений принадлежит к одной рудной формации и описано набором значений из 70 геологических признаков, характеризующих региональные геологические структуры (№ 1—4); пространственное совмещение с месторождениями других металлов (№ 5—12); рудоконтролирующие и рудовмещающие структуры (№ 13—30); рудовмещающие (№ 31—34) и рудолокализирующие породы (№ 35—41); типы рудных тел (№ 42—51), особенности их залегания (№ 52—55); окологрудные изменения (№ 56—63) и магматические породы, ассоциированные с оруденением (№ 64—70). Все признаки кодировались в двоичной системе.

4—5. *Определение уровня эрозионного среза по геохимическим данным.*

Для решения задачи об определении уровня эрозионного среза по геохимическим данным методами распознавания образов на одном из редкометальных месторождений в верхней и нижней частях эталонного рудного столба было выделено два блока. В пределах выделенных блоков было отработано 45 и 44 пробы соответственно, каждая из которых характеризовалась значениями содержаний 13 химических элементов: Co, Mo, Pb, As, Zn, Ag, Be, Zr, Sn, Y, Yb, Mn, Ba.

С целью проверки эффективности используемых алгоритмов распознавания в зависимости от объема обучающей выборки поставленная задача разделения верхней и нижней частей рудного столба решалась в двух вариантах.

В первом варианте в обучающую выборку вошли 30 проб верхней и 30 проб нижней частей рудного столба. Оставшиеся 29 проб (15 с верхней и 14 с нижней части рудного столба) служили в качестве независимых контрольных объектов.

Во втором варианте обучающая выборка состояла из 20 проб (по 10 проб с верхней и нижней части рудного столба), а оставшиеся 59 проб (35 с верхней и 24 с нижней) составили независимую контрольную выборку. Значения содержаний химических элементов были закодированы как в баллах, так и в двоичной системе.

Результаты решения задач различными алгоритмами

Результаты решения геологических задач, описанных выше, приведены в табл. 4. Обработка контрольных материалов производилась в зашифрованном виде. За меру качества решения задач разными алгоритмами распознавания был принят процент правильной классификации независимых контрольных выборок, причем отказ от классификации принимался эквивалентным 0,5 ошибки.

По поводу результатов решения геологических задач алгоритмами распознавания можно сделать следующие выводы.

1. Оба алгоритма тупиковых тестов показали неприемлемые результаты при оценке ртутных месторождений (эффективность распознавания 62,5% против 87% в среднем по другим алгоритмам). Алгоритм тупиковых постов при классификации 12 контрольных месторождений россыпного олова дал 8 отказов от классификации (эффективность распознавания 66,7%). Модифицированный алгоритм тупиковых тестов при оценке редкометальных рудных полей показал эффективность

Эффективность распознавания
при решении геологических задач различными алгоритмами, %
(по А. Н. Еремееву и др. [105])

Задачи	„Ги-пер-пласт“	Тупико-вых тестов	Тупико-вых тестов модифицированный	„Потенциал-2“		„Перебор“	„Обобщенный портрет“
				контроль по всем признакам	контроль по информативной совокупности признаков		
1. Оценка рудных полей по геолого-геофизическим и геохимическим данным. 5 + 8 объектов обучения. 59 признаков (мелкомасштабное прогнозирование)	94,4	91,9	73,6	81,8	91,9	100,0	91,9
2. Оценка россыпных месторождений олова по геологическим данным. 7 + 7 объектов обучения. 48 признаков (среднемасштабное прогнозирование)	83,4	66,7	79,1	87,5	75,0	79,1	58,3
3. Оценка месторождений ртути по геологическим данным. 5 + 6 объектов обучения. 70 признаков (среднемасштабное прогнозирование)	87,5	62,5	62,5	100,0	87,5	87,5	87,5
4. Определение уровня эрозионного среза по геохимическим данным. Первый вариант: 30 + 30 объектов обучения. 13 признаков (крупномасштабное прогнозирование)	94,8	Не считалась	Не считалась	98,2	100,0	96,6	100,0
5. Определение уровня эрозионного среза по геохимическим данным. Второй вариант: 10 + 10 объектов обучения. 13 признаков (крупномасштабное прогнозирование)	95,6	То же	То же	79,0	94,2	94,9	97,1
Эффективность распознавания по всем задачам	91,3	73,7	71,7	89,3	93,5	91,6	86,9
Пределы отклонений (min—max = Δ)	11,3	29,4	16,6	21,0	25,0	20,9	41,7

распознавания 74% против 92—95% в среднем по остальным алгоритмам. Однако небольшое число решаемых задач, малый объем контрольных выборок не позволяют сделать вывод о непригодности этих алгоритмов.

2. Алгоритм «Гиперпласт» успешно справился с решением всех геологических задач. Кроме того, он показал наилучшую стабильность при решении всех классов задач (пределы колебаний 11,3%).

3. Алгоритмы «Потенциал-2», «Перебор» и «Обобщенный портрет» показали в целом близкие результаты. Однако у всех этих алгоритмов стабильность решения различных классов задач почти в два раза хуже, чем у алгоритма «Гиперпласт». Одним из недостатков алгоритма «Потенциал-2» является ухудшение эффективности распознавания на контроле по информативной совокупности признаков (выбранных самим алгоритмом) для задач 2 и 3.

Алгоритм «обобщенный портрет» показал неприемлемые результаты по задаче 2 (эффективность распознавания 58,3%, при средней по другим алгоритмам около 80%). Вместе с тем полученные результаты сопоставления алгоритмов не могут служить основанием для окончательных выводов ввиду ограниченного количества как объектов контроля, так и решаемых задач.

Оценка «информативности» признаков

Результаты оценки «информативности» признаков при использовании разных алгоритмов приведены в табл. 5. Следует отметить, что при

Таблица 5

Результаты выделения информативных совокупностей признаков (номера выделенных признаков указаны в порядке убывания их «информативности»)

Задачи	„Гипер- пласт“	„Потен- циал-2“	„Перебор“	„Обоб- щенный портрет“	Тупико- вых тестов	Тупиковых тестов модифици- рованный
1. Оценка рудных полей (мелкомасштабное прогнозирование)	2, 5, 15, 3, 27, 44	51, 2, 44, 15, 49, 30	51, 2, 15, 3, 18	17, 22, 38, 15, 3, 6	27, 18, 37, 49, 35, 38	18, 27, 37, 35, 38, 4
2. Оценка россыпных месторождений олова (среднемасштабное прогнозирование)	1, 36, 47, 12, 31, 13	49, 47, 1	3, 31, 45, 1, 18, 47, 43, 25	43, 31, 47, 18, 10, 3, 19, 25	28, 14, 36, 31, 1	1, 40, 43, 8, 45, 34
3. Оценка месторождений ртути (среднемасштабное прогнозирование)	7, 3, 39, 66, 29, 15, 49	3, 7, 29, 39	3, 39, 7, 24, 31	7, 5, 49, 34, 21	19, 48, 54, 65, 2, 7, 44	19, 21, 54, 24, 55, 30
4. Определение уровня эрозионного среза, первый вариант (крупномасштабное прогнозирование)	As, Be, Yb	As, Co, Sn	As, Yb	As, Be, Yb	Задача не решалась	Задача не решалась
5. Определение уровня эрозионного среза, второй вариант (крупномасштабное прогнозирование)	As, Be, Ba	As, Co, Zn	As, Yb	As, Be, Sn	То же	То же

оценке выделения информативных совокупностей признаков необходимо прежде всего учитывать эффективность решения задачи классификации объектов данным алгоритмом.

Полученные результаты позволяют сделать следующие замечания.

1. С ростом эффективности решения задачи наблюдается лучшее соответствие наиболее «информативных» признаков, выделяемых различными алгоритмами.

2. Алгоритмы, построенные на различных принципах классификации объектов, выделяют разные совокупности «информативных» признаков.

3. В число наиболее «информативных» признаков, выделенных различными алгоритмами, во многих случаях попадают одни и те же признаки. Так, при решении задачи оценки рудных полей систематически выделяются признаки № 15, 51, 2; при оценке россыпных месторождений олова — № 1, 31, 47; при оценке ртутных месторождений — № 7, 3, 39, 29; при определении уровня эрозионного среза в двух вариантах выделяются As, Be, Yb. Такое совпадение в оценке «информативности» признаков, полученное отличающимися по заложенным принципам классификации алгоритмам, свидетельствует, во-первых, о наличии несомненной связи между косвенными признаками и масштабом оруденения различных геологических объектов и, во-вторых, о принципиальной возможности практического использования апостериорного подхода к выбору информативных совокупностей геологических признаков.

IV. МЕТОДЫ ОРГАНИЗАЦИИ И ФОРМИРОВАНИЯ ГЕОЛОГИЧЕСКИХ БАНКОВ ДАННЫХ

Геологические исследования представляют собой по сути дела традиционно сложившуюся систему последовательных процедур по получению, анализу и переработке разнородной геолого-геофизической и геохимической информации, в процессе которых осуществляется локализация оцениваемой территории на основе альтернативных решений о перспективности или бесперспективности отдельных ее частей. Некоторые исследователи вообще предлагают рассматривать геологоразведочные работы «как особое информационное производство, непосредственной продукцией которого является информация» [72].

К настоящему времени известны многочисленные работы по применению современных методов обработки геологической информации с использованием ЭВМ [13, 52, 60, 88 и др.]. Вместе с тем вопросы автоматизации информационного обеспечения задач геологического прогнозирования рассматриваются, как правило, в гораздо меньшей степени, хотя существующее отставание в подготовке исходных данных является одной из главных причин, препятствующих более эффективному и широкому использованию современных ЭВМ в прикладной геологии.

В монографии А. Н. Бугайца и Л. Н. Дуденко [13], являющейся наиболее полным трудом по использованию математических методов и ЭВМ в геологическом прогнозировании, эти вопросы также не нашли достаточно полного отражения.

Ниже рассмотрены некоторые аспекты использования современных средств обработки, хранения и преобразования информации с помощью математических методов и ЭВМ применительно к проблеме геологического прогнозирования.

1. СОВРЕМЕННОЕ СОСТОЯНИЕ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ ЗАДАЧ ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ ЭВМ

В настоящее время специалисты, занимающиеся поисками и оценкой месторождений полезных ископаемых в пределах конкретного региона, располагают большими объемами геологической, геофизической, геохимической и прочей информации, требующей определенной систематизации, многоаспектного анализа, переосмысливания и периодической проверки. Вместе с тем многочисленные примеры свидетельствуют о том, что существующее информационное обеспечение не вполне соответствует требованиям, которые необходимы для использования современных средств обработки информации на основе математических методов и ЭВМ. Почти все исследователи, использующие в своей работе ЭВМ, чувствуют некий «барьер», не позволяющий широко и плодотворно применять математический аппарат в научных и производственных целях. Если несколько лет назад это был барьер недоверия к возможностям использования математических методов для описания геологических явлений и ситуаций, то сейчас одной из главных тормозящих причин, как отмечает ряд авторов [22, 23, 39, 43 и др.], является несо-

ответствие исходной геологической информации стандартным требованиям, предъявляемым современным ЭВМ. Использование для решения отдельных задач геологического прогнозирования даже простейших статистических методов требует большого объема подготовительных работ, связанных со сбором, систематизацией и кодировкой информации. При сравнении объектов исследования часто приходится сталкиваться с ситуацией, когда объекты одного характера (масштаба) описаны различными, несопоставимыми наборами признаков, что затрудняет или даже исключает их сравнительный машинный анализ. Все это приводит к тому, что при решении многих задач геологического прогнозирования учитывается лишь незначительная часть всей имеющейся информации. В общем случае как избыток, так и недостаток информации приводит к снижению эффективности решения конкретной задачи. Поэтому большую актуальность приобретает вопрос о выборе оптимального количества информации, необходимой для составления прогнозных заключений. Это положение согласуется с фундаментальной формулой Шеннона, связывающей в логарифмической функции количество информации с мерой неопределенности. Интерпретация этой формулы показывает, что увеличение количества информации не приводит к пропорциональному уменьшению неопределенности.

Внедрение новых, более производительных методов обработки информации вступает в противоречие с традиционными методами представления геологической информации. Известно, что наиболее удобным видом кодирования исходных данных для проведения последующего машинного анализа является бинарное; однако при этом часть ценной информации может быть потеряна.

Наконец, в настоящее время назрела необходимость в количественной интерпретации геологических данных, подобно тому как это делается в геофизике и геохимии; необходимо найти способы измерения информативности тех или иных сведений для повышения достоверности геологического прогнозирования.

Как уже отмечалось, существование математических методов заключается в том, что они применимы только к формальным понятиям. Из этого следует, что корректное и эффективное использование математического аппарата для решения задач геологического прогнозирования должно опираться на формализованную систему понятий. В настоящее время исследования в этом направлении развиваются как по линии разработки формализованных геологических языков применительно к конкретным разделам геологической науки, так и по линии частичной формализации в рамках отдельных задач [26, 89, 114, 135 и др.].

Рассматривая процесс геологического прогнозирования как сложное сочетание операций сбора, свертывания, хранения, поиска, передачи информации и принятия решений (т. е. как процесс информационный), представляется возможным для изучения этих процессов использовать современные методы обработки информации. В то же время последовательное внедрение ЭВМ для хранения и обработки геологических данных является благоприятным аспектом для эволюции геологической документации в смысле постепенного приспособления ее к требованиям, предъявляемым технологией машинной обработки.

Существующий процесс перевода исходной геологической информации на машинные носители и ее кодирования характеризуется большой трудоемкостью. Главное преимущество использования ЭВМ заключается в том, что при этом оказывается возможным анализировать большие массивы разнородной информации; однако чем больше исходных

данных, тем сложнее их предварительная подготовка. Например, при машинном анализе геологической информации одного планшета масштаба 1:200 000 (площадь 5200 км²) по 50 факторам требуется снять с различных карт и закодировать исходные данные общим объемом до 260 000 чисел. Если учесть, что при этом должно быть исключено наличие ошибок, то задача предварительной подготовки данных представляется весьма сложной.

Естественно, геологи, занятые на производстве, не обладают практическими возможностями подготовки таких объемов исходных материалов для анализа с использованием ЭВМ.

Следовательно, назрела необходимость в разработке какого-то принципиально иного подхода, который позволил бы резко снизить затраты ручного труда на процедуры подготовки и кодирования исходной информации без потери ценности исходных данных.

Впервые такой подход в самом общем виде был сформулирован, по-видимому, А. П. Куклиным [56]. Одним из недостатков такого подхода являлось то, что исходные материалы предлагалось кодировать и использовать для машинного анализа только в бинарном виде.

Ниже будет рассмотрен способ, сочетающий автоматизированный ввод и автоматическое кодирование исходной геологической информации и лишенный отмеченного недостатка.

Таким образом, имеющийся опыт использования математических методов и ЭВМ в практике геологического прогнозирования свидетельствует о наличии определенного несоответствия между существующим способом подготовки, предварительной обработки и кодирования исходных данных и реальными возможностями геологического производства. Именно это положение и является сейчас главным препятствием для широкого внедрения современных средств обработки информации с помощью математических методов и ЭВМ, как основы для повышения эффективности и качества составления геологических прогнозных заключений. Одним из путей устранения такого несоответствия является создание автоматизированных систем накопления, хранения и обновления исходной информации (банков данных). Поэтому такие системы получают в настоящее время все большее практическое распространение. В то же время в литературе пока не описаны такого рода системы применительно к хранению картографической информации в виде, удобном для последующего человеко-машинного анализа. Такая система должна включать автоматизированный (или автоматический) ввод исходных данных, архив, а также комплекс программ математического обеспечения предварительной обработки данных, ориентированных непосредственно на пользователей — специалистов геологов. Вопросы разработки такой системы посвящены следующие разделы главы.

2. ПРИНЦИПЫ ОРГАНИЗАЦИИ БАНКА КАРТОГРАФИЧЕСКИХ ДАННЫХ ДЛЯ РЕШЕНИЯ ЗАДАЧ ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

Применительно к проблеме геологического прогнозирования понятие банка данных было впервые конкретизировано в работах [88, 95]. Под банком данных при этом понималась система накопления, хранения и ведения геологической информации в ЭВМ, а также соответствующее математическое обеспечение.

Все задачи геологического прогнозирования на различных этапах исследований решаются на основе использования разнообразной первичной документации. По-видимому, создание единой, стандартной си-

стемы описания геологических объектов в виде, удобном как для практического использования, так и для оперативного перевода на машинные носители,— проблема чрезвычайно сложная, и вряд ли она будет решена в ближайшее время. Возникает вопрос: как в создавшейся ситуации использовать широкие возможности современных ЭВМ по обработке и оперативному применению их для составления прогнозных заключений? В настоящее время решением этого вопроса занимаются многие исследователи [60, 69, 76, 82, 88, 133 и др.].

В большей части работ исходная информация, вводимая в ЭВМ, организуется в виде матриц «объект—признак», либо основное внимание уделяется комплексу обрабатывающих программ, а вопрос подготовки исходных данных затрагивается частично или вообще не затрагивается. Выгодно отличается в этом плане система АСКИП [76], в которой заложена возможность получения признаков для прогнозирования с помощью ЭВМ при условии ввода в машину минимально обработанных данных (числовых полей). Однако при этом остается открытым вопрос использования собственно геологических (картографических) материалов. Кроме того, принятые в АСКИП процедуры получения признаков достаточно «формалистичны».

В работе [4] сформулированы совершенно правильные требования к автоматизированной системе обработки данных, но на первый план выдвигается вопрос об унификации первичных геологических документов, что придает обсуждению абстрактный характер. В. В. Немец [75] совершенно справедливо утверждает: «Опыт показывает, что путем последовательных улучшений с немедленным их внедрением в практику можно скорее и лучше подойти к автоматизированным системам управления, нежели развитием строго теоретически обоснованных научно-исследовательских программ без прямой связи с требованиями промышленности».

Как отмечалось в работе [95], начало практического решения рассматриваемой проблемы возможно уже в настоящее время на следующей основе.

1. Для построения прогнозных моделей с помощью ЭВМ необходимо, во-первых, обеспечить точную топографическую привязку исходной информации и, во-вторых, организовать в машинной памяти большие массивы информации в виде, удобном для последующей обработки. В связи с этим целесообразно условное подразделение всей геологической информации (здесь и далее под геологической подразумевается вся информация, используемая при прогнозировании, включая геофизическую, геохимическую и др.) на отдельные типы. В самом деле, организация компактного хранения в памяти ЭВМ, например, картографического материала будет существенно отличаться от организации хранения информации по буровым скважинам, а информация по горным выработкам какого-либо месторождения должна сохраняться в виде, обеспечивающем воспроизведение их пространственного положения в объемной модели месторождения. Таким образом, при создании универсальной информационной основы для решения задач геологического прогнозирования необходимо разработать принципы построения на базе ЭВМ архивов, по крайней мере, для этих трех типов данных.

2. На машинные носители, по-видимому, следует заносить информацию не по всей первичной геологической документации, а лишь по документам, уже синтезировавшим исходные данные (например, геологические и другие карты, погоризонтные планы, разрезы и т. п.). Это положение вытекает из того, что синтез первичных материалов в

геологическую карту (разрез) не всегда под силу современным ЭВМ. Кроме того, составление таких «синтезированных» документов специалисты выполняют с учетом многих неформализуемых процедур.

3. На машинные носители следует переносить информацию, наиболее часто применяемую для практической работы и решения различных задач геологического прогнозирования, имея в виду возможность постепенного ее расширения в пределах конкретных территорий.

4. Необходимо автоматизировать процедуру ввода исходной информации в ЭВМ и осуществление контроля введенных данных.

5. Следует применить такие способы кодировки первичных геологических материалов, использование которых не привело бы к потере исходной информации при ее последующем анализе с помощью ЭВМ.

Можно считать, что различная информация, используемая для решения прогнозных задач, разделяется на горизонтальные слои (по масштабам исследования, см., например, [72]). Вместе с тем на каждом из этих уровней будет находиться информация с одинаковой организацией хранения в машинной памяти (например, во всех слоях имеется картографическая информация). Иными словами, исходная информация также естественно распадается и на вертикальные слои (по типам организации хранения исходных данных). Каждый из вертикальных слоев требует для работы со своими данными специализированного программного обеспечения, и, следовательно, по характеру информации, сохраняемой в машинной памяти, и ее организации используемые в настоящее время банки данных могут быть условно подразделены на «территориальные», «линейные» и «объемные», как это показано в табл. 6. Поясним это положение подробнее.

Таблица 6

Возможные типы банков данных, используемых при решении задач геологического прогнозирования

Тип банка данных (по виду ячеек носителей информации)	Возможные размеры элементарной ячейки хранения информации	Характеристика информации, сохраняемой в памяти ЭВМ
„Территориальный“ (картографические материалы)	Любые в зависимости от исходного материала и типа решаемых задач (1 см ² , 0,25 см ² и др.)	Всевозможные виды геологических, геофизических, геохимических, структурных и других карт различных масштабов, погоризонтные планы и т. п.
„Линейный“ (материалы по скважинам колонкового бурения)	Поинтервальное хранение различной информации (1 м, 0,1 м и т. д.) либо характеристика скважины в целом	Геологическая документация, геофизические данные, данные различных анализов, описания шрифтов и т. д.
„Объемный“ (материалы, характеризующие объемные модели месторождений)	Элементарные ячейки заданного объема (1 м ³ , 10 м ³ , 1000 м ³ и т. д.)	Геологические характеристики, результаты геофизического, бороздowego и другого опробования, сведения о рудоносности и др.

Банк картографических данных

При выполнении региональных геологических прогнозов и проведении различных статистических расчетов с помощью ЭВМ многие исследователи предлагают сохранять исходные геологические карты в машинной памяти в виде отдельных ячеек. Так, Н. А. Кухаренко [60] считает, что для информационных систем по металлогении следует использовать два типа объектов: объекты, относящиеся к геологической среде — площадные, и объекты, относящиеся к полезным ископаемым — месторождения, рудопроявления, ореолы, точки минерализации. Для площадных объектов он предлагает использовать международную разграфку листов масштаба 1:1 000 000, что обеспечивает достаточную формализацию объекта «территория». За ячейку сбора информации предлагается брать участок земной коры, ограниченный рамками планшета масштаба 1:200 000, что соответствует масштабности металлогенических работ, проводимых на территории СССР. Он отмечает, что количество информации в пределах территории обуславливается не размерами площади, а сложностью геологического строения. Однако при этом не учитываются возможности прогнозирования в более крупных масштабах, т. е. подход не может являться универсальным. Подобный подход описан М. С. Лахно и А. Н. Шуевым [61]. Общие требования к обоснованию размеров площадных объектов, выбранных в качестве «элементарных ячеек» для сбора информации при автоматизированном прогнозировании месторождений полезных ископаемых, сформулированы Л. Н. Дуденко [37]. Он указывает, что размер элементарных ячеек должен превышать максимально возможные размеры месторождения прогнозируемого типа и быть меньше минимального расстояния между месторождениями. Число элементарных объектов (ячеек) должно соответствовать техническим возможностям получения и хранения геологического описания каждого из них. Размер элементарных объектов должен меняться с изменением масштаба прогнозирования. Указывается, что с учетом изложенных требований оптимальным размером элементарного объекта для рудных месторождений, по-видимому, является близкий к 1 см² в масштабе карты.

А. П. Куклин [55], осуществляя машинное прогнозирование, использовал в качестве элементарных ячеек площади, равные планшету масштаба 1:50 000, а в другом случае — площадки в 1 см² масштаба карты. В работах Б. А. Чумаченко, Е. П. Власова, В. В. Марченко и др. [88, 95] в качестве элементарной ячейки прогнозирования применялась площадь размером в 1 см² в масштабе карты. Этот подход был опробован и показал положительные результаты при исследованиях различных масштабов (1:50 000, 1:100 000; 1:200 000, 1:500 000), при этом, однако, исходные геологические данные вводились по сети 0,5×0,5 см в масштабе карты.

Аналогичный подход успешно применяется и другими исследователями [146, 151, 160 и др.]. Таким образом, при формировании универсального территориального машинного банка картографических данных в качестве элементарной ячейки хранения первичной информации целесообразно брать площадь 1 см² в масштабе исходной карты. Вместе с тем для наиболее эффективного использования информации (например, не только для целей прогнозирования, но и для составления различного рода справочных данных) следует предусмотреть возможность выдачи исходных данных в виде, удобном для пользователей. Отсюда вытекает условие о необходимости сохранения в памяти ЭВМ

геологических карт, с одной стороны, в исходном виде, а с другой — в виде, удобном для машинной обработки.

Применение способа хранения информации исходных карт по элементарным ячейкам имеет и свои недостатки. Главный из них заключается в том, что некоторые геологические обозначения, имеющие линейный характер на исходных картах (например, контакты пород, тектонические нарушения и т. п.), будут запоминаться ЭВМ в виде полос шириной, по крайней мере, в одну элементарную ячейку.

Следует отметить, что территориальный банк картографического материала может содержать и сведения «объемного» характера, например потенциальные геофизические поля, их различные трансформации. В этом случае данные, приписываемые какой-либо отдельной элементарной площадке, характеризуют не только ее саму по себе, но и всю сложность ее взаимоотношений с соседними площадками и одновременно отражают в косвенном виде информацию об особенностях глубинного строения территории. Заметим, что и сама геологическая карта по своей природе отражает в косвенном виде элементы глубинного строения, т. е. несет «глубинную» информацию.

Общие размеры территорий (карт), которые могут быть помещены в память ЭВМ в виде характеристик отдельных элементарных площадок, практически неограничены.

Рассмотрим более подробно организацию территориального банка картографической информации, созданного для среднемасштабного прогнозирования и реализованного на ЭВМ типа М-222. Не затрагивая здесь вопросов ввода исходной картографической информации, отметим, что вся территория изучаемого региона представляется в виде поля, разбитого на квадратные ячейки размером $0,5 \times 0,5$ см² масштаба карты. Кроме того, для удобства хранения данных в ЭВМ весь регион подразделяется на одинаковые квадратные листы по $72 \times 72 = 5184$ ячейки в каждом. Информация по всем факторам была перенесена с исходных карт на перфокарты, причем пробивка на перфокарте означала наличие фактора в соответствующей ячейке. При этом сведения по одному фактору в пределах одного листа размещались на шести перфокартах. Полученная, таким образом, на машинных носителях модель карты региона представляет собой набор бинарных матриц (таблиц карт факторов), каждая из которых несет информацию об области распространения одного фактора.

Однако при решении задач прогнозирования с помощью ЭВМ используется обычно не сама исходная карта, а некоторый набор признаков, полученный путем ее обработки. В этот набор могут входить различные характеристики пространственного положения изучаемых геологических факторов относительно тех или иных геологических контуров, результаты всевозможных статистических параметров распределений и т. п. В дальнейшем изложении модель исходной карты, сохраняемую в памяти ЭВМ, будем называть первичным массивом, а результаты его обработки и преобразования с использованием ЭВМ — целевым массивом.

Территориальный банк картографических данных состоит из трех массивов (первичного, целевого и каталога) и комплекса программ, обеспечивающего выполнение следующих функций:

- 1) ввод исходных данных в память ЭВМ;
- 2) вывод на АЦПУ или графопостроитель материала в форме, адекватной входным формам;

- 3) коррекцию данных по всем показателям;
- 4) пополнение массивов новыми данными;
- 5) исключение данных из массивов;
- 6) целевую обработку массивов данных для решения задач геологического прогнозирования;
- 7) составление различных выборок.

Первичный массив содержит информацию о регионе в виде карт распространения геологических, геофизических и прочих факторов. Массив сохраняется на магнитной ленте и организован следующим образом.

1. Минимальной единицей информации является «страница» размером 72×72 элементарных ячеек, содержащая сведения об области распространения одного фактора на данном листе карты.

2. «Страницы» объединяются в «листы» (сведения о распространении одного фактора по всему региону).

3. Листы записаны на магнитной ленте последовательно, причем каждый лист занимает две смежные зоны.

4. Листы объединяются в «книгу» (сведения о распространении всех факторов по всему исследуемому региону).

Целевой массив содержит информацию, необходимую для решения задач геологического прогнозирования, полученную путем машинной обработки первичного массива.

Целевой массив сохраняется на магнитной ленте и организован следующим образом:

1) минимальной единицей информации является «описание точки», содержащей значения всех признаков для каждого узла сети;

2) описания точек объединяются в «описание листа», которое занимает две смежные зоны;

3) описания листов объединяются в «описание региона».

Каталог содержит информацию о структуре первичного и целевого массивов в виде схемы распространения листов в районе, схемы изученности района по факторам и списка значений некоторых параметров; включающих:

- числа сохраняемых в территориальном банке факторов;
- числа листов в районе;
- размера области, учитываемой при вычислении признаков;
- допустимой точности расчета признаков;
- списка наименований факторов и др.

Банк данных по скважинам

Банк этого типа предназначен для хранения и использования различной информации по скважинам колонкового бурения. Естественно, что координаты скважин могут быть «привязаны» к территориальному банку картографических данных. В настоящее время уже имеются разработки по хранению отдельных видов информации по скважинам в памяти ЭВМ (нефтяная геология, промысловая геофизика и т. п.).

В качестве примера сошлемся на системы хранения информации, приведенные в работах Андертона М. Р. [137], В. Штайна и др. [139], И. Еленковой [156] и др.

Известны устройства по непосредственной записи результатов каротажа на машинные носители. Основная трудность в этом случае состоит в занесении в машинную память геологических данных, характе-

ризующих литологические, стратиграфические, минералогические и прочие характеристики.

Очевидно, при формировании банков данных по скважинам следует исходить в первую очередь из потребностей в такой информации, которая наиболее часто требуется при решении задач геологического прогнозирования. Такие банки данных должны содержать комплексную информацию, так как невозможно составлять прогнозные заключения, основываясь лишь на геофизических или геохимических материалах. В качестве элементарной ячейки хранения исходной информации может быть принята либо скважина в целом, либо определенный интервал ее длины — это зависит от масштаба карты и характера решаемых задач геологического прогнозирования. Следует оговориться, что понятие «линейная информация» является относительным. Например, результаты гамма-каротажных исследований, отнесенные к единице длины скважины, характеризуют и некоторый объем окоლოსкважинного пространства.

Современные ЭВМ позволяют сохранять практически неограниченные объемы информации. Так, например, Д. Форготсон и Ф. Штарк в своей работе [149] отмечают, что к 1972 г. в США и Канаде машинный банк содержал информацию более чем по 700 тыс. скважин и использовался, в частности, для прогноза нефтяных залежей.

В качестве примера возможной реализации структуры рассмотрим в общих чертах принципы организации скважинного банка данных, как составной части информационного обеспечения задач крупно- и среднемасштабного геологического прогнозирования.

Такой банк данных может состоять из трех основных массивов (первичного, целевого и структурного) и комплекса программ, обеспечивающих ввод и вывод информации, ее обработку, выдачу ответов на запросы и т. п. Первичный массив содержит исходные данные, скважины объединяются в нем по территориальному признаку. Информация записывается последовательно по скважинам, и все сведения о месте хранения первичных данных заносятся в каталог (структурный массив), сохраняемый также на магнитной ленте. В отличие от банка картографической информации, в первичном массиве банка данных по скважинам сохраняются числовые значения признаков, поэтому формирование целевого массива, непосредственно используемого при решении задач прогноза, не связано с большими трудностями. Для хранения информации по скважинам, видимо, целесообразно использовать структуру разветвленного дерева. Такая структура позволяет достаточно оперативно осуществлять различные выборки, необходимые для анализа и составления характеристик геологических объектов различных классов.

Банк объемных данных

В литературе описаны многочисленные примеры реализации информационно-поисковых систем и банков данных по различным месторождениям полезных ископаемых, используемых для статистического анализа и геологического прогнозирования (см. главу II). Однако в таких системах обычно каждое месторождение описывается определенным, заранее заданным набором признаков и характеристик. В отличие от этого, банк «объемных» данных предполагает сохранение в машинной памяти непосредственно объемной модели месторождения. Поэтому в

таким образом в качестве элементарной ячейки должны приниматься определенные геометрические объемы пространства. Поясним это положение следующим примером. В практике широко известны объемные модели месторождений, состоящие из погоризонтных геологических планов на прозрачной основе. Подобная модель может быть «помещена» в память ЭВМ в виде набора погоризонтных банков картографических данных. Можно создать и более совершенную модель, построенную на принципе хранения информации, характеризующей элементарные объемы. Естественно, что создание таких объемных банков данных не является самоцелью, а служит современной информационной базой для решения геологопрогнозных задач на стадиях эксплуатационной разведки.

Так, если в памяти ЭВМ будет сохраняться информация, характеризующая геологическую позицию данного элементарного объема по отношению к основным рудоконтролирующим структурам, интрузивным образованиям и т. п. (определение этих положений может производиться автоматически), то это создаст предпосылки для проведения объемного геологического анализа и определения перспектив рудоносности различных частей месторождения. Кроме того, наличие такого банка данных открывает возможности использования геологической информации для технико-экономического моделирования, планирования и выбора оптимальных вариантов систем обработки и бортовых содержаний полезного ископаемого.

При формировании объемного банка данных информация общегеологического характера может вводиться с погоризонтных планов, разрезов и т. п. Сведения о подземных скважинах и горных выработках могут сохраняться в памяти ЭВМ в соответствии с принципами, изложенными в предыдущем разделе.

Ввиду того, что в процессе эксплуатационной разведки месторождения количество горизонтов и выработок непрерывно увеличивается, банк объемных данных должен представлять собой «открытую» систему, которая может пополняться новой информацией.

В общем виде объемный банк данных применительно к решению задач геологического прогнозирования должен состоять, по крайней мере, из трех массивов:

- 1) первичного массива, содержащего три подмассива:
 - информацию об общем геологическом строении месторождения (по интерполяции данных погоризонтных планов и разрезов);
 - информацию по скважинам колонкового и подземного бурения;
 - информацию по подземным горным выработкам;
- 2) целевого массива, представляющего собой объемную модель исследуемого месторождения в заданном пространстве признаков;
- 3) структурного массива, содержащего сведения о сохраняемой информации и взаимосвязях всех массивов и подмассивов.

В принципе, банк объемных данных является универсальным и может использоваться для решения различных задач статистического анализа и геологического прогнозирования. По-видимому, основная трудность реализации банка объемных данных заключается в разработке достаточно сложной структуры хранения информации. В качестве основы такой структуры может быть рекомендована структура банка данных общего назначения на дисках, описание которой приведено ниже (см. раздел 4 настоящей главы).

На различных этапах и стадиях геологопоисковых и разведочно-эксплуатационных работ для решения наиболее важных задач геоло-

**Относительные оценки использования информации
при решении задач геологического прогнозирования
(для рудных месторождений)**

Этапы геологических исследований	Наиболее часто употребляемые масштабы работ ^а	Наиболее важные задачи геологического прогнозирования	Относительная доля информации при решении прогнозных задач, %		
			Картографические данные	Данные по скважинам	Данные по подземным горным выработкам
Прогнозотематические исследования	1: 1 000 000, 1: 500 000, 1: 200 000	Выбор рудных районов, площадей для проведения съемочных работ. Решение задач регионального (мелкомасштабного) прогнозирования	80—100	0—20	—
Геологосъемочный	1: 200 000, 1: 100 000, 1: 50 000	Выбор перспективных участков для поисков и оценки. Решение задач среднемасштабного прогнозирования	70—90	10—30	—
Поисковооценочный	1: 25 000, 1: 10 000, 1: 5 000	Выбор перспективных участков для проведения детальных поисков и предварительной оценки. Решение задач средне- и крупномасштабного прогнозирования	50—70	20—40	0—10
Разведочный	1: 10 000, 1: 5 000, 1: 2 000, 1: 1 000	Предварительная оценка масштабы оруднения, выбор перспективных направлений разведочных работ. Решение задач крупномасштабного прогнозирования	20—50	30—50	0—30
Эксплуатационный	1: 5 000, 1: 2 000, 1: 1 000, 1: 500	Классификация запасов, планирование развития подготовительных и разведочных работ. Решение задач крупномасштабного прогнозирования. Оценка флангов и глубоких горизонтов месторождения	20—30	30—40	30—50

^а Масштабы исследований и соответственно доля участия различных видов информации могут изменяться в зависимости от генетических типов объектов прогнозирования.

гического прогнозирования информация различного типа (картографическая, по скважинам, по горным выработкам) используется неодинаково. Так, если на этапе металлогенического прогнозирования окончательные решения о перспективности отдельных территорий могут быть приняты на основе только картографического материала, аккумулировавшего все необходимые и достаточные данные, то на этапе поисков и оценки существенное значение приобретают сведения, получаемые по

результатам бурения скважин. На этапе эксплуатационной разведки определяющее значение приобретают данные документации и опробования подземных горных выработок. В табл. 7 приведены относительные оценки использования геологической информации при принятии решений на различных этапах геологопоисковых и разведочных работ.

3. КАРТОГРАФИЧЕСКАЯ ИНФОРМАЦИЯ — ОСНОВА ФОРМИРОВАНИЯ ТЕРРИТОРИАЛЬНЫХ БАНКОВ ДАННЫХ

Анализ развития естественных наук свидетельствует, что их совершенствование на определенных этапах как в теоретическом, так и в прикладном аспектах осуществлялось на основе привлечения математического аппарата для построения фундаментальных и частных моделей, описания исследуемых объектов, их свойств и признаков, а также для анализа существующих между ними связей. Видимо, справедливо с этих позиций в современных условиях научно-технической революции оценить значение геологической картографической информации для составления прогнозных заключений и оценки геологических объектов и ситуаций с помощью ЭВМ.

Многие исследователи признают, что использование картографических данных при моделировании геологических объектов и явлений позволяет наряду с генетическим подходом широко применять и статистические модели. Это положение ни в коей мере не умаляет исторически сложившегося генетического подхода к изучению геологических объектов, но, напротив, дополняет и обогащает его.

Применение современных средств обработки информации на базе математических методов и ЭВМ открывает широкие перспективы комплексного использования различной картографической информации (в том числе данных космических съемок) при решении конкретных задач прикладной геологии.

Значение картографической информации при геологическом прогнозировании

Различная картографическая информация широко используется на всех этапах и стадиях геологоразведочных работ. Значение этой информации для принятия решений о перспективности объектов исследований колеблется от 30—40 до 100%. Оценивая общее значение картографической информации в современных условиях, Е. Е. Ширяев [132] справедливо указывает, что традиционная карта продолжает оставаться лучшим средством передачи информации человеку. В то же время в своем исходном виде карта является неудобным носителем геонформации для ее анализа с помощью ЭВМ. «В настоящее время в области кибернетики стоит первоочередная задача — разработать такие формы представления данных и ввода в ЭВМ, которые были бы доступны и человеку и ЭВМ. От этого зависит эффективность диалога человека с машиной, его взаимодействие с ЭВМ в процессе совместной работы. В этом отношении ведущая роль принадлежит именно картографической форме отображения информации. Такова точка зрения ведущих советских и зарубежных ученых в области технической кибернетики» [132].

В прикладной геологии в настоящее время усиленно развивается направление разработки стандартных, унифицированных форм описания геологических объектов [43, 69, 84, 115, 159 и др.]. Вместе с тем

в геологической науке и практике имеется такой стандартный документ адекватного отображения геологических ситуаций — геологическая карта. Карта понятна и читаема всеми специалистами, имеет стандартные обозначения и международную разграфку; без применения геологических карт геология немыслима как наука. Видимо, тот факт, что в своем исходном виде геологическая карта не может анализироваться ЭВМ с такой же детальностью, как это делают специалисты, в какой-то мере оказал свое влияние и на формирование мнения о трудности использования ЭВМ в этой области. Но именно в том и состоит проблема, чтобы каким-либо образом трансформировать геологическую карту и перевести ее в вид, приемлемый для машинного анализа.

В настоящее время создавалась своеобразная ситуация, когда вспомогательная по своему назначению информация, например геофизические и геохимические карты, обрабатывается без особых сложностей с помощью ЭВМ, в то время как геологическая карта — основной носитель сведений об особенностях геологического строения территории — не поддается в необходимой степени такой обработке. Другой причиной недостаточного использования ЭВМ для анализа геологических карт является, по-видимому, то обстоятельство, что на практике довольно часто имеют место случаи, когда соседние листы карт, составленные различными исполнителями, не «стыкуются» между собой, что вообще делает невозможным их совместный анализ. Скорее всего именно более широкое внедрение машинных методов хранения картографической информации позволит исключить подобные факты.

Как носитель информации, карта выгодно отличается от известных машинных носителей в следующих отношениях: первичностью и наглядностью восприятия информации, доступностью широкому кругу специалистов, возможностью упрощения взаимодействия человека с машиной при обработке данных, возможностью произвольной выборки заданной информации, своей высокой информативной емкостью [132].

Нельзя не согласиться с Е. Е. Ширяевым о том, что «поскольку карта одновременно отвечает принципам графического и математического моделирования, ее можно назвать графоматематической моделью, использующей для передачи информации свой, присущий только картографии язык. Это делает ее самостоятельной, выдвигая в разряд особых моделей» [132].

В самом деле, картографическая геологическая информация, даже в упрощенном виде, при машинном прогнозировании применяется многими исследователями, и на этом пути имеются определенные успехи, не вызывающие уже скептического отношения, по крайней мере, у большинства специалистов [1, 11, 39, 136, 160 и др.]. Ю. Г. Леонов и В. Н. Шалпо [63] справедливо указывают, что существующее представление о качественном характере геологической информации не является препятствием для использования математических методов. Они считают, что формализацию понятий в геологии следует проводить в первую очередь путем совершенствования «традиционных» понятий геологии [63]; геологическая картографическая информация в этом плане представляет собой благоприятный объект для такой формализации.

Отрицание ценности понятия «геологическая карта» сторонниками чисто формального определения геологических понятий получает резкую критику специалистов [15].

Все большее значение в последнее время приобретает информация о геологическом строении земной поверхности, получаемая с искусст-

венных спутников, являющаяся по своей сути также картографической. Обработка и анализ этой информации уже немислимы без использования ЭВМ.

Изучая глобальные закономерности размещения крупных рудных месторождений М. А. Фаворская вводит понятие «геологической аномалии», предполагающей наличие «определенного разрыва между аномалией и окружающим ее фоном» [27]. Под такой аномалией понимается «статистически значимое отклонение на данном участке тех или иных геологических признаков однородных геологических объектов от господствующего фона соответствующих признаков подобных объектов на данном этапе развития структуры в целом» [27]. М. А. Фаворская считает, что применительно к интерпретации аномалий подобного типа в ближайшее время будет разработан не только качественный, но и количественный подход.

Наибольшее значение картографическая информация приобретает на этапе регионального прогнозирования. Оценивая значение математических методов в этом аспекте, Р. М. Константинов и А. Д. Щеглов отмечают особую роль этих методов для «выявления районов поисков скрытого оруденения и решения других кардинальных проблем геологии рудных месторождений» [52].

Наконец, на основе анализа картографической информации с использованием ЭВМ открывается возможность проведения картографического моделирования, представляющего собой совершенно новую область моделирования, которая требует неперменного участия специалиста в проведении исследования. Таким образом, становится ясным огромное значение геологической картографической информации, как исходной базы данных для выполнения и автоматизированного прогнозирования с использованием современных средств обработки данных, предоставляемых специалистам исследователям научно-технической революции.

Способы преобразования картографической информации в дискретный вид

Любая автоматизированная система сбора, хранения и обработки информации должна опираться на массивы стандартных документов. Это требование обуславливает необходимость наличия стандартных первичных документов для описания геологических объектов, позволяющих без всякой перекодировки или составления промежуточных документов вводить необходимые сведения в ЭВМ. Неоднократно предпринимались попытки использовать в качестве основного первичного документа для хранения картографических данных перфокарты ручного обращения. Однако у такого способа имеется много недостатков: ограниченное кодировочное поле, использование различных ключей кодирования, необходимость перекодировки информации для ввода в ЭВМ, ограниченный круг решаемых задач и др. Имеющийся опыт хранения больших объемов картографических данных на машинных перфокартах также свидетельствует о существенных недостатках такого способа [95].

В настоящее время существует аппаратное и программное обеспечение, позволяющее полностью автоматизировать регистрацию и обработку данных о физических полях, физических характеристиках и вещественном составе горных пород и руд. В то же время способы автоматической обработки геологической картографической информации

еще весьма несовершенны, требуют больших затрат труда и времени на подготовку информации к вводу в ЭВМ. Этапу автоматизированной обработки данных предшествуют трудоемкие ручные операции по формированию массивов для решения отдельных задач геологического прогноза. В этом свете понятны усилия, предпринимаемые различными исследователями по преобразованию первичных картографических данных [29, 81, 132, 141 и др.].

Опыт математической обработки геологической информации свидетельствует о том, что фактически для каждой вновь сформулированной задачи необходимо заново вручную создавать информационные массивы. Дальнейшее совершенствование автоматизированных систем обработки геологических данных, видимо, немыслимо без разработки и использования новых подходов к преобразованию и машинному хранению картографической информации. Такие подходы должны позволить уже при сегодняшнем программном обеспечении автоматизировать и ускорить процессы предварительной подготовки информации для ее последующей машинной обработки, при этом, однако, следует принять определенные меры для того, чтобы избежать возможного снижения информативной ценности исходных картографических данных.

Автоматизация процесса поиска и обработки информации предполагает ее предварительное кодирование. Коды должны удовлетворять по крайней мере требованиям однозначности, кратности, аддитивности, программируемости. В то же время процесс кодирования не должен быть весьма сложным и трудоемким делом, т. е. кодирование исходных данных, предпринятое для последующего машинного анализа при решении конкретной задачи, должно компенсировать затраты на предварительную подготовку данных.

Проблема определения списка фиксируемых на различных стадиях геологоразведочного процесса данных для последующего их кодирования, сводится в настоящее время к проблеме предварительной оценки информативности отдельных факторов, необходимых для решения конкретных задач геологического прогнозирования. По-видимому, кардинальным решением вопроса явилось бы полное сохранение исходных карт в машинной памяти с последующим выбором любых факторов (составляющих карту), требуемых специалисту для решения конкретной прогнозной задачи.

Известны способы [111] представления контуров изображений в натуральных координатах, в которых поверхность изображения, например карты, с имеющимися областями распространения некоторого фактора разбивается на элементарные ячейки правильной сетью, и затем ячейки, попавшие в области, где присутствует данный фактор, помечаются числом «1», а ячейки, лежащие в областях, где данный фактор отсутствует — числом «0». Под фактором при этом понимается какое-либо конкретное геологическое образование, имеющееся на карте (например, стратиграфо-литологический комплекс, магматическое образование, область гидротермального изменения пород, геологическая формация и т. п.). Между тем как из практики традиционного геологического прогнозирования, так из геологической теории следует, что влияние отдельного фактора не ограничивается областью его локализации, а распространяется на смежные области, причем особое значение имеет область вблизи контакта (границы области распространения фактора). Ниже рассматривается возможный подход к преобразованию геологических и других карт в дискретный вид, лишенный рассмотренных недостатков. Способ практически опробован на фактиче-

ских данных и показал достаточно устойчивые результаты при составлении прогнозных заключений с помощью ЭВМ, сопоставимые с прогнозами, полученными на основе традиционной методики.

Суть способа состоит в следующем.

1. Исходная карта делится на стандартные листы квадратной формы.

2. Каждый лист разбивается на одинаковое, заранее заданное число элементарных ячеек, размер которых определяется необходимой точностью сохранения исходного изображения.

3. В пределах каждого стандартного листа исходная геологическая карта последовательно разделяется на отдельные, составляющие ее картографические факторы (подобно цветовому типографскому клише). Под термином «картографический фактор» будем понимать конкретное геологическое образование, определяющее геологическое строение территории и отображенное на легенде карты в виде самостоятельной единицы. Факторы могут иметь площадной, линейный или точечный характер.

4. Площадь распространения каждого фактора в пределах стандартного листа кодируется в бинарном виде («1» — наличие фактора в элементарной ячейке, «0» — отсутствие). Каждый фактор сохраняется на машинных носителях независимо от других. Информация этого типа по всем факторам будет представлять собой первичный массив. При наличии специальных устройств вывода информации первичный массив позволяет восстановить и графически изобразить карту в виде, адекватном исходному оригиналу.

5. Каждый стандартный лист по одному картографическому фактору преобразуется в некую «потенциальную функцию». Преобразованное значение картографического фактора в каждой элементарной ячейке будем называть «картографическим признаком». Под этим термином, например, может пониматься некоторая числовая характеристика элементарной ячейки карты, функционально отражающая пространственное положение этой ячейки относительно области распространения данного фактора. Эта характеристика может быть получена путем автоматического измерения или расчета, т. е. картографический признак является в некотором смысле количественной мерой картографического фактора.

Рассмотрим в качестве наиболее простого примера характеристики элементарной ячейки карты по отдельному фактору — наикратчайшее расстояние от центра этой ячейки до ближайшей границы области распространения фактора. Строго говоря, в этом случае в качестве характеристики элементарной площади p по некоторому фактору можно принять значение функции

$$F = f\{(r(p))\},$$

$$\text{где } r(p) = -\inf_{q \in \Omega} \rho(p, q) + \inf_{q \in \bar{\Omega}} \rho(p, q),$$

Ω — область распространения фактора;

$\bar{\Omega}$ — область, в пределах которой фактор отсутствует;

$\rho(p, q)$ — расстояние от центра элементарной площадки p до центра элементарной площадки q ;

$$f(r) = \begin{cases} -R & \text{при } r < -R \\ r & \text{при } -R \leq r \leq R \\ R & \text{при } r > R, \end{cases}$$

где R — радиус влияния данного фактора, определяемый специалистами.

Таким образом, в данном случае каждая элементарная ячейка массива признаков геологической (либо другой) карты будет характеризоваться набором чисел, определяющим ее точное пространственное положение относительно ближайших границ всех факторов, имеющих на данной карте. Количество этих чисел будет равно количеству составляющих карту отдельных факторов. Будем считать эти числа положительными в случае отчета по нормали от эндоконтакта и отрицательными при отчете от экзоконтакта, т. е. при удалении от границы фактора.

Расчеты пространственного положения каждой элементарной ячейки относительно контура геологического фактора могут рассчитываться автоматически на ЭВМ с помощью специального комплекса программ. Информация подобного типа, сохраняемая в памяти ЭВМ, будет соответствовать «массиву признаков».

Исходя из содержательных геологических соображений, числовые значения картографического признака могут быть сгруппированы специалистами в отдельные интервалы (например, «зона эндоконтакта», «зона экзоконтакта», «неизменные породы» и т. п.). С помощью такого и других преобразований могут быть закодированы все обозначения, имеющиеся на геологических и прочих картах.

Применение такого подхода позволяет выполнять процедуру машинного кодирования и описания геологических объектов на более качественной основе, весьма удобной для использования ЭВМ с целью машинного хранения и всевозможной обработки картографической информации. Вместе с тем при этом машинный анализ картографических данных приближается к логике мышления специалиста геолога, который при анализе геологических ситуаций учитывает не только наличие или отсутствие какого-либо отдельного фактора, влияющего на локализацию конкретного типа оруденения, но и оценивает обстановку по совокупности взаимоотношений всех факторов, представляющих интерес для решения конкретной задачи геологического прогнозирования.

Видимо, при таком двойном способе описания элементарной ячейки (в виде первичного массива и массива признаков, как это показано на рис. 7) достигается, с одной стороны, сохранение исходного картографического материала в оригинальном виде, а с другой — возможность выполнения формально-логического машинного анализа. Описание элементарной ячейки в виде картографических признаков позволяет сохранить и учитывать при машинном анализе «объемность» геологической карты, выражающуюся в специфическом сочетании составляющих ее разновозрастных образований. Структура хранения таких массивов картографических данных может быть, например, аналогичной, описанной в работе [95]. При использовании рассмотренного подхода оказывается возможным получать статистические характеристики отдельных геологических факторов в пределах исследуемой территории, такие, как математическое ожидание, дисперсия и др. Это обстоятельство открывает широкие перспективы для привлечения современных средств обработки данных с использованием математических методов и ЭВМ для различного рода статистических исследований в прикладной геологии. Оказывается, например, возможным оперативно составлять выборки по геологическим объектам различных классов (всей территории, месторождениям, рудопроявлениям и т. п.), анализировать их статистические параметры по отдельным факторам и признакам (см.

главу VI). В этих случаях представляется также возможным выделение аномальных значений отдельных признаков и их комбинаций в пределах рассматриваемой территории. Открывается принципиальная возможность выявления статистически значимых региональных «геологических аномалий» [14, 27], выделения «неблагоприятных для оруденения геологических обстановок», характеризующихся достаточно простым геологическим строением [55], прослеживания «литологической анизотропии», как фактора, благоприятного для локализации эндогенного оруденения [129], «безэталонного районирования» и анализа «компактности выборок» [79] и т. п.

Более того, принципиально оказывается возможным осуществление различных трансформаций массивов картографических геологических данных подобно тому, как это имеет место в геофизике.

В заключение отметим, что наличие современных устройств автоматического и автоматизированного ввода обеспечивает быстрое внедрение указанного подхода в практику машинного анализа геологической картографической информации.



Рис. 7. Схема хранения картографической информации в территориальном машинном банке данные (по В. В. Марченко)

Следует подчеркнуть, что рассмотренный подход, вероятно, может быть распространен и на другие виды картографической информации (схемы интерпретации геофизических данных, экономико-географические, лесные, почвенные и другие карты).

Важное значение при этом приобретают способы автоматизированного ввода картографической информации в ЭВМ и ее визуализация в процессе машинной обработки данных.

4. СТРУКТУРА И ОРГАНИЗАЦИЯ УНИВЕРСАЛЬНОГО БАНКА ДАННЫХ ОБЩЕГО НАЗНАЧЕНИЯ (БОНД)

Общие сведения

Рассмотренные ранее специализированные системы для хранения картографической геологической информации эффективны с точки зрения решаемых на их основе задач, поскольку как структура данных, так и способ их размещения на физических носителях в наибольшей степени соответствуют этим решаемым задачам. С другой стороны, такие системы хранения не всегда оказываются удобными при решении новых задач на тех же информационных массивах, поскольку включение новых задач связано, как правило, с очень трудоемким процессом их реализации. Кроме того, трудоемкость реализации новых задач существенно возрастает при необходимости вводить в систему качественно новую и разнородную информацию. Отсутствие необходимости включения принципиально новых задач, нацеленность на решение определенного класса задач и их специфика, требования к скорости обработки информации и специальные способы ее ввода и первичной обработки — все это определяет необходимость существования рассмотренных систем информации, а именно эти причины приводят к реализации относительно замкнутых систем, в которых информация и некий класс решаемых на ее основе задач образуют единое целое, зависимое друг от друга.

Формирование территориальных машинных банков данных, реализация задач экономического и учетного характера, оперативного управления, возможность быстрой реализации незапланированных справочных функций и новых задач немыслимо без мощной и развитой системы хранения первичной информации, максимально независимой от решаемых на основе ее задач. В качестве возможной основы для этой цели может использоваться универсальный банк данных общего назначения на дисках (БОНД)*.

БОНД представляет собой программный комплекс, предназначенный для обеспечения хранения, накопления и ведения информационной базы данных практически произвольной структуры, а также для осуществления гибкого интерфейса между пользователем и данными, хранящимися в системе.

Вся информация, хранящаяся в системе, представляется в виде некоторой логической структуры, с помощью которой в наглядной графической форме отображаются основные типовые элементы информации и типовые связи между этими элементами. Логическая структура строится пользователем системы в рамках принятых в системе основных понятий и терминологии на основе анализа задач, выявленных к моменту их реализации с помощью системы БОНД. Таким образом, конструирование информационной системы сводится к выбору и конструи-

* В разработке БОНД принимали участие В. А. Ковалев, В. Г. Филоян и др.

рованию логической структуры данных (описание логической структуры будет дано ниже). Система также осуществляет размещение на физической среде хранения конкретной информации в соответствии с выбранной для ее представления логической структурой. Таким образом осуществляется физическая независимость данных, под которой подразумевается изменение физического расположения и организации данных в процессе эксплуатации системы при неизменности как общей логической структуры данных, так и конкретных программ пользователей, использующих эту информацию. Основным средством, обеспечивающим физическую независимость данных, является язык манипуляции данными (ЯМД), для использования которого достаточно лишь представления о логической структуре данных в системе.

Система, кроме того, допускает и логическую независимость данных, которая означает возможность изменения логической структуры данных в процессе эксплуатации системы без изменения прикладных программ пользователя, реализующих те или иные задачи в конкретной сфере применения системы БОНД.

Логическая и физическая независимость данных в наибольшей степени обеспечивает соблюдение принципа развития — одного из основных требований, предъявляемых к современным системам управления базами данных. Реализация этого принципа позволяет без больших затрат на ведение информационной системы включать в ее состав качественно новые задачи и качественно новую информацию с возможным расширением логической структуры данных.

Совокупность всех данных, хранимых в системе в соответствии со спецификой конкретной сферы применения, в части ведения и использования информации может быть расчленена между рядом независимых пользователей, ответственных и заинтересованных в решении своих задач. Каждый пользователь может иметь локальное представление о составе и логической структуре данных, хранимых в системе. Однако все локальные системы ведения информации, сведенные вместе для управления единой системой управления данными, а также отобразенные на совокупной логической структуре данных в рамках системы БОНД, образуют единую информационную систему, в которой без дополнительных затрат могут реально решаться качественно новые задачи, использующие совокупную информацию всех локальных систем. Этим достигается соблюдение следующего основного требования, предъявляемого к информационным системам, — реализация принципа интеграции данных.

Использовавшиеся ранее методы и средства конструирования информационных систем приводили к созданию информационных фондов с высоким уровнем избыточности информации, которая возникает из-за необходимости дублирования данных. Дублирование данных в существующих информационных системах неизбежно возникает из-за необходимости обеспечения связей между данными, а также для осуществления приемлемого времени обработки информации и увеличения производительности вычислительной техники. Отрицательные последствия наличия избыточной информации заключаются не столько в перерасходе объема памяти внешних запоминающих устройств и времени, необходимого для неоднократного изменения одних и тех же данных, сколько в опасности разрушения информационной системы. Это может произойти в результате одновременного обновления одних и тех же копий информации, что приводит к выдаче противоречивых данных, полученных в результате решения различных за-

дач при использовании ими пересекающихся информационных множеств.

Возможность построения развитой и сложной логической структуры данных обеспечивает в большинстве случаев исключение избыточности данных. Если же, по предположению пользователя, избыточность информации необходима в целях получения дополнительного эффекта, то эта избыточность данных, по крайней мере, может контролироваться при рациональном использовании средств системы управления базой данных.

Все программное обеспечение системы БОНД может быть разделено на три основные составные части, каждая из которых предназначается для выполнения относительно автономных функций по конструированию и использованию информационной системы.

К первой функциональной части системы принадлежит программный комплекс, осуществляющий отображение логической структуры данных, определенной пользователем в конкретной сфере применения системы, на физическую среду хранения. Она же обеспечивает разметку и выделение участков памяти на физических носителях информации. Ядром первой функциональной части системы является язык описания структур (ЯОС), с помощью которого пользователь описывает выбранную им логическую структуру данных. Использование языка описания структур определяет первый этап в применении системы БОНД: первичное определение структуры данных и передачу этой структуры под контроль системы управления базой данных.

В частности, графически определенная структура данных описывается с помощью ЯОС и вводится в систему, которая на этапе первичного отображения структуры данных на среду хранения осуществляет синтаксический контроль правильности использования ЯОС. Кроме того, логическая структура испытывается системой на корректность с точки зрения соблюдения регламентов при ее конструировании, накладываемых системой БОНД. Простота ЯОС обеспечивается отсутствием необходимости определения каких-либо физических характеристик будущего файла хранимой совокупной информации. По специальному запросу пользователя, при условии отсутствия синтаксических и семантических ошибок в тексте описания, логическая структура воспринимается системой БОНД и запоминается для обеспечения нормальной работы двух других функциональных частей программного комплекса.

Основной составной частью системы управления базой данных является программный комплекс, обеспечивающий связь проблемных программ пользователя с информационной базой данных. Обращение к базе данных для занесения или извлечения информации осуществляется при помощи языка манипуляции данными. Он характеризуется достаточно высоким логическим уровнем построения, что позволяет просто и эффективно формулировать в прикладных программах сложные информационные запросы независимо от сложности расположения конкретной информации и адресации данных. С точки зрения пользователя системы, каковым в случае использования ЯМД является программист, этот программный комплекс выступает как единая программа, преобразующая директивы языка в конкретные действия системы по обеспечению прикладной программы необходимыми данными. Каждая директива передается системе в виде фиксированного набора параметров, так же как и в случае обращения программы к подпрограмме. Таким образом, связь с системой обеспечивается традиционными, понятными для программиста средствами.

Система управления базой данных реализована на машинах серии ЕС ЭВМ. Соблюдение стандартных соглашений, принятых в операционной системе, обеспечивает взаимосвязь системы с программами, реализованными на различных языках программирования: ASSEMBLER, COBOL, FORTRAN, PL-1.

На основе ЯМД могут быть построены не только проблемные программы конкретной сферы применения, но и универсальные программы, например для обеспечения случайных, незапланированных запросов на информацию. Более того, на основе ЯМД может быть реализован высокоуровневый язык запросов или генератор отчетов, обеспечивающий возможность получения разнородных данных без прикладных программ. В частности, в составе системы БОНД предусматривается наличие подсистемы выдачи информации и справок на экран дисплея, в которой информационный запрос и схема поиска информации может быть задана для системы в результате диалога пользователя с системой БОНД. Полученная на экране в некоторый момент информация может быть задокументирована (распечатана), а результат диалога может запоминаться системой для возобновления заданной схемы поиска и выдачи информации в другое время, начиная с любого места. Например, можно возобновить просмотр информации сначала (опуская этап диалога для определения схемы поиска) после того, как независимые пользователи обновят или изменят свою локальную информацию. Эта подсистема полностью реализована на основе ЯМД и не зависит от конкретной сферы применения системы БОНД.

Кроме того, система содержит сервисные программы, предназначенные для ведения информационной системы. К перечню этих программ можно причислить стандартные средства и программы, представляемые в составе операционной системы. Эти стандартные средства обеспечивают выполнение организационно-технологических функций, связанных, например, с образованием и ведением библиотек программ системы БОНД, личных библиотек пользователей, создания резервных копий базы данных, образование поколений данных и т. п. Основной сервисной функцией, которая должна выполняться в процессе ведения информационной системы, является периодическая чистка базы данных и переоптимизация физического размещения информации в базе. Необходимость в этой функции объясняется тем, что в процессе добавления и исключения данных нарушается оптимальное физическое размещение данных, а также наряду с потребностью в дополнительной памяти среды хранения образуются неиспользуемые участки памяти. Моменты использования программы «чистки и переразмещения» определяются пользователем либо по получении сообщения от системы о достижении предельного количества обновлений информации, либо в результате анализа времени, необходимого для решения тех или иных задач, так как временные характеристики при частом обновлении базы данных могут ухудшаться.

В состав сервисных программ входят также программы восстановления базы данных, использование которых необходимо в случаях аварийного разрушения в результате аппаратных сбоев, некорректного обращения к базе данных в процессе отладки и т. п. Существует несколько способов восстановления базы данных. Одним из таких способов является запуск программы, которая восстанавливает базу данных на основе последней резервной копии и системного журнала, в котором в процессе использования системы регистрировались все изменения информации от момента образования резервной копии.

Структура данных

Отправным этапом при создании БД и работе с ним является определение структуры данных, которая может быть наглядно представлена в виде графа и которая является единственным внешним (пользовательским) представлением о структуре хранимых данных.

Система позволяет создавать и работать с такими типами структур, как деревья, простые сети, сети с циклами (в частности с петлями), сети с множеством дуг между двумя вершинами. Кроме того, могут быть использованы структуры, в которых с любой дугой графа связан вектор данных.

Введем основные понятия о структуре данных в системе, используя пример, изображенный на рис. 8.

Основной единицей обмена между системой БОНД и проблемной программой пользователя является сегмент данных. Сегмент — совокупность непересекающихся полей данных. Особым полем сегмента является поле ключа сегмента.

В поле содержится неделимая единица информации — данное, хранимое в заданном пользователем формате. Формат поля определяет способ хранения данного в поле; например, число с фиксированной точкой, символьное данное, упакованное десятичное число и др. Кроме формата, поле характеризуется именем и длиной в байтах. Имя поля, формат и длина составляют описание поля. Поля в сегменте расположены плотно; таким образом, длина сегмента равняется сумме длин составляющих его полей.

Каждому сегменту соответствует некоторый формат, представляющий собой совокупность описаний полей сегмента, порядок следования которых соответствует порядку следования данных в сегменте.

Тип сегментов определяет некоторое множество сегментов одинаковой природы, имеющих равные форматы. Каждый тип именуется восьмисимвольным набором.

В БОНД допускается наличие не более 255 типов сегментов. Каждому типу ставится в соответствие формат, описывающий все сегменты, принадлежащие данному типу.

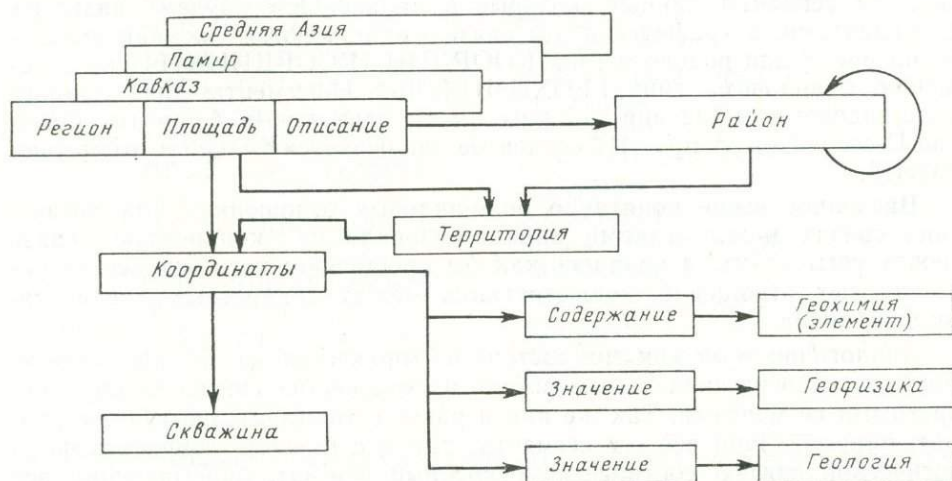


Рис. 8. Пример логической структуры банка данных

Все сегменты одного типа сортируются в порядке возрастания содержимого ключевого поля. Полным аналогом хранения сегментов одного типа является последовательный файл отсортированных записей. Так, логическая структура, изображенная на рис. 8, состоит из 8 типов сегментов с именами: РЕГИОН, ТЕРРИТОР (ТЕРРИТОРИЯ), РАЙОН, ЭЛЕМЕНТ и т. д. В типе РЕГИОН кроме поля ключа содержится поле ПЛОЩАДЬ региона в формате числа с плавающей точкой длиной в 4 байта (E4) и поле, содержащее словесное ОПИСАНИЕ региона в символьном формате длиной в 122 байта (C122). Тип РЕГИОН, в свою очередь, состоит из сегментов с ключами КАВКАЗ, ПАМИР, СРЕДНЯЯ АЗИЯ и т. д. Любой тип в БД может вступить в родо-видовое отношение с любым другим типом или же с самим собой.

Логическое родо-видовое отношение графически изображается стрелкой, соединяющей эти два типа или один и тот же тип (петля). В дальнейшем тип, из которого исходит стрелка, будем называть родовым типом по отношению к тому типу, на которой эта стрелка приходит. Второй же тип по отношению к первому будем называть видовым типом. Очевидно, что один и тот же тип в различных связях может выступать как в качестве родового, так и видового.

Между любыми двумя типами может быть определено до тридцати родо-видовых отношений (до 15 «стрелок» одного направления и до 15 «стрелок» противоположного направления).

Формально одно родо-видовое отношение от другого отличается номером связующей стрелки и ее направлением. Смысловое значение родо-видовых отношений определяется пользователем системы БОНД.

Система БОНД позволяет устанавливать на более 15 родо-видовых отношений типа петля для одного и того же типа сегментов.

Система БОНД позволяет хранить и работать не только с информацией, связанной с типом, но и с информацией, принадлежащей одновременно двум типам, если только эти два типа включены в родо-видовое отношение.

Таким образом, с любым родо-видовым отношением могут быть связаны сегменты данных, которые в дальнейшем будем называть U-сегментами, а графически эти связи изображаются так, как показано на рис. 8 для родового типа КООРДИН (КООРДИНАТЫ) по отношению к видовому типу ГЕОХИМИЯ. Все U-сегменты для данного родо-видового отношения должны иметь одинаковые форматы. Отличие U-сегментов от простых сегментов заключается только в отсутствии ключей.

Введенное выше понятие о родо-видовых отношениях или логических связях между типами характеризовало не конкретную связь между сегментами, а являлось как бы прообразом того множества родо-видовых отношений, существующих между конкретными сегментами этих типов.

Аналогично тому как тип состоит из множества конкретных сегментов, так и логическая связь состоит из множества связей между конкретными сегментами. Так же как и ранее с типом мы связывали формат, описывающий все его сегменты, так и с каждой информативной логической связью соотносится некоторый формат, описывающий все U-сегменты для каждой конкретной связи, входящей в множество данной логической связи.

Язык описания структур (ЯОС)

Как уже отмечалось, выбранная пользователем логическая и физическая структура любого банка данных (БД) описывается посредством языка описания структуры (ЯОС). Основное назначение такого описания состоит в определении типов сегментов (их форматов и имен) и логических связей между типами.

Одновременно посредством ЯОС могут описываться конкретные сегменты введенных типов и их данные, а также физические связи между введенными сегментами.

Оформленное соответствующим образом описание БД обрабатывается служебными программами БОНД, реализующими чтение описания БД и заполнение физической среды хранения (дисков) формируемыми служебными таблицами описания структуры БД и конкретными данными с оптимальным их размещением на физическом носителе информации. Пользователь не обременяется необходимостью знать физическое расположение информации на носителе. Для реализации БД и работы с ним достаточно знать его структуру.

Не вдаваясь в подробное описание синтаксических правил ЯОС, приведем в качестве примера описание логической структуры, изображенной на рис. 8.

ОПИСАНИЕ

ИМЯ—БОНД=ГЕОКАРТЫ.	СВЗ—ФОРМ.
ПОР—СОРТ=0123456789.	КОЛИЧ—СВ=1.
АБВГДЕЖ ЭЮЯ., %.	СВЗ—СЕТЬ=РЕГИОН,
СТРУК—БД.	ТЕРРИТОР, 1.
КЛ—ТИПОВ=8.	СВЗ—СЕТЬ=РЕГИОН,
ПРЕД—ТИП=РЕГИОН, 30*	КООРДИН, 1.
ПЛОЩАДЬ, Е, 4.	СВЗ—СЕТЬ=РЕГИОН,
ОПИСАНИЕ, С, 122.	РАЙОН, 1.
ПРЕД—ТИП=ТЕРРИТОР,	СВЗ—СЕТЬ=ТЕРРИТОР,
30*	КООРДИН, 1.
ПЛОЩАДЬ, Е, 4.	СВЗ—СЕТЬ=КООРДИН,
ЦЕНТР, С, 20.	ГЕОХИМИЯ, 1.
ПРЕД—ТИП=КООРДИН, 8.	*ЗНАЧЕНИЕ, Е, 4*
ПРЕД—ТИП=РАЙОН, 30*	СВЗ—СЕТЬ=КООРДИН,
ПЛОЩАДЬ, Е, 4.	ГЕОФИЗИК, 1.
ОПИСАНИЕ, С, 100.	*ЗНАЧЕНИЕ, Е, 4*.
ПРЕД—ТИП=ГЕОХИМИЯ,	СВЗ—СЕТЬ=КООРДИН,
10.	ГЕОЛОГИЯ, 1.
ПРЕД—ТИП=ГЕОФИЗИК,	*СОДЕРЖ, Е, 4*.
10.	СВЗ—СЕТЬ=КООРДИН,
ПРЕД—ТИП=ГЕОЛОГИЯ,	СКВАЖИНА, 1.
10.	СВЗ—СЕТЬ=РАЙОН,
ПРЕД—ТИП=СКВАЖИНА,	ТЕРРИТОР, 1.
20.	СВЗ—СЕТЬ=РАЙОН, 1.
	КОНЕЦ—БД.

Из приведенного описания видно, что сначала определяются типы сегментов, а затем связи между ними. В предложениях, определяющих тип сегментов, число, стоящее за именем вводимого типа, означает длину ключевого поля в сегменте в свободном формате. Система

БОНД всегда предполагает символьный формат ключевого поля. Число, фигурирующее в предложениях описания связей, определяет номер связи между двумя конкретными типами. Номер связи позволяет отличать одну логическую связь от другой для случаев, когда между двумя конкретными типами устанавливается не одна логическая связь. Предложение ЯОС — ПОР — СОРТ = позволяет устанавливать порядок сортировки сегментов, отличный от того, который возникает в результате использования стандартных кодов символов.

Связь программ пользователя с БОНД

Управление данными в системе, под которым подразумевается осуществление операций ввода/вывода из проблемных программ пользователя, происходит с помощью набора элементарных функций оперирования данными.

Каждая из функций ввода/вывода, реализованных в БОНД, определяется четырехсимвольным мнемоническим кодом. При необходимости совершения некоторой функции в конкретном месте проблемной программы мнемонический код этой функции должен быть передан в систему. Это осуществляется через стандартное средство обращения к подпрограмме — оператор CALL (или макрокоманде CALL для языка ASSEMBLER). В качестве подпрограмм в этом случае выступают программы системы.

Имя подпрограммы, следующее за оператором CALL, определяет точку входа в программы системы в зависимости от языка, на котором написана проблемная программа пользователя.

В качестве имен точек входа выбраны следующие:

ASBOND — если программа написана на языке ASSEMBLER;

CLBOND — если программа написана на языке COBOL;

PLBOND — если программа написана на языке PL-1;

FNBOND — если программа написана на языке FORTRAN.

В качестве аргументов в операторе CALL в установленном ниже порядке фигурируют идентификаторы полей, определяющие

— имя функции ввода/вывода (FUNK);

— имя типа сегмента (TYPE);

— ключ сегмента (KEY);

— номер логической связи (N);

— код связи с системой БОНД (KSS);

— код сохранения адресов (KSAD);

— рабочее поле (FLD).

Имя функции ввода/вывода определяет вид обмена между БД и проблемной программой пользователя и содержит четыре символа. Функции обмена делятся на два класса: функции, предназначенные для считывания информации из БД, и функции изменения данных в БД.

Имена функций считывания:

SEGM — читать конкретный сегмент или следующий сегмент данного типа (segment);

SPEC — читать видовой или следующий видовой сегмент указанного типа от данного родового (species);

GENS — читать родовой или следующий родовой сегмент указанного типа от данного видowego (genus);

ISEG — читать U-сегмент (information segment);

Имена функций изменения БД:

- INST — вставить в указанный тип новый сегмент (insert);
DLET — удалить сегмент указанного типа из БД (delete);
CONN — установить связь от данного родового к указанному видовому сегменту (connect);
DCON — уничтожить связь между данным родовым и указанным видовым сегментами (delete connection);
REPL — изменить данные сегмента (replace).

Идентификатор ключа сегмента в операторе CALL должен определять поле, длина которого не меньше длины ключа сегмента запрашиваемого типа.

Идентификатор номера логической связи должен определять поле длиной в два байта. В это поле в проблемной программе перед выполнением оператора CALL должно помещаться целое число без знака в формате с фиксированной точкой, лежащее в пределах от 0 до 15. Этот номер логической связи используется системой БОНД только при исполнении тех функций, в которых происходит переход по видовым или родовым связям. Для других функций это данное системой БОНД не используется.

Идентификатор кода связи с системой БОНД должен определять поле длиной в два байта. Посредством этого поля проблемная программа может указывать системе БОНД, каким образом ей трактовать ту или иную функцию. В свою очередь система БОНД по исполнении функции запишет в это поле код завершения функции, который может использоваться в проблемной программе как для анализа правильности исполнения запрошенной функции, так и для организации условных переходов в программе пользователя. Эти коды завершения характеризуют различные ситуации завершения функции; например: функция завершилась нормально, сегмент с запрашиваемым ключом отсутствует в БД, сегмент не загружен данными и т. д.

Основное назначение KSAD состоит в обеспечении быстрой реакции БОНД на выполнение требуемой функции. Правильное использование KSAD проблемным программистом обеспечивает значительное сокращение обращений к внешней памяти (в худшем случае не более одного раза для исполнения одной функции). KSAD формируется системой БОНД при выполнении функций обмена. Поскольку область, где содержится KSAD, доступна проблемной программе, программист должен обеспечить сохранность KSAD по крайней мере до момента его использования следующей функцией. В противном случае проблемная программа закончится аварийно, так как в системе БОНД реализованы средства защиты.

Длина рабочей области должна быть не меньше длины вызываемого сегмента. В рабочую область помещается копия данных сегмента либо системой БОНД при использовании функций чтения, либо проблемной программой при выполнении функции изменения БД.

Вся совокупность функций ввода/вывода образует ЯМД, с помощью которого реализуется произвольный доступ к данным, какова бы ни была сложна схема их обработки.

У. ПРИНЦИПЫ ФУНКЦИОНИРОВАНИЯ ЧЕЛОВЕКО-МАШИННОЙ СИСТЕМЫ ГЕОЛОГИЧЕСКОГО ПРОГНОЗИРОВАНИЯ АИПС/ЕС «РЕГИОН»

1. МЕТОДИКА РЕШЕНИЯ ЗАДАЧ ПРОГНОЗА И ОЦЕНКИ ГЕОЛОГИЧЕСКИХ СИТУАЦИЙ С ПРИМЕНЕНИЕМ ЭВМ

Как было показано в главе I, определяющими звеньями поисково-разведочных работ, с точки зрения влияния на его конечную эффективность, являются процедуры геологического прогнозирования на каждом этапе исследований. Поэтому главные усилия должны направляться на совершенствование процедуры решения задач геологического прогноза на основе использования математических методов и ЭВМ [91].

К настоящему времени прогнозирование месторождений полезных ископаемых стало крайне сложной задачей. Сложность прогнозирования обусловлена, с одной стороны, обилием накопленного фактического материала, который необходимо подвергнуть анализу, с другой стороны тем, что приходится использовать большое количество поисковых признаков, при этом каждый отдельный признак не имеет решающего значения, а важен лишь в сочетании с другими. Причем в разных ситуациях, т. е. в разных сочетаниях с другими признаками, данный конкретный признак имеет различное поисковое значение.

В этих условиях выполнение прогнозных работ традиционными методами наталкивается на трудности, связанные с ограниченными способностями человека к переработке многофакторной информации.

Выход из создавшейся ситуации возможен только путем привлечения в помощь геологу современных математических методов и средств обработки данных. Попытки привлечения математики и вычислительных машин предпринимаются давно. При этом были решены частные задачи, но существенных успехов достигнуто не было. Причины трудностей заключаются, с одной стороны, в том, что переработке на ЭВМ подлежит огромное количество исходных данных, а принципов их организации и ввода в ЭВМ не разработано, с другой стороны, нет достаточно приемлемой модели самого процесса прогнозных работ.

Полностью автоматическое решение на ЭВМ всех вопросов прогнозирования геологических объектов и ситуаций невозможно по крайней мере в течение ближайшего обозримого будущего. Поэтому интерес представляют автоматизированные человеко-машинные системы, в которых на ЭВМ производится перебор различных вариантов сочетания признаков для больших площадей, представляя геологу возможность анализа, выбора и комбинации вариантов. Нередко в руках специалистов, занимающихся поисками и разведкой минерального сырья в пределах конкретного региона, собирается большой объем разнородной геологической, геофизической, геохимической, гидрогеологической и другой информации, требующей своей систематизации, многоаспектного анализа, переосмысливания и периодической проверки.

Решение перечисленных задач может быть осуществлено, в частности, путем создания автоматизированных человеко-машинных диало-

говых систем геологического прогнозирования, в основе которых должны быть территориальные банки картографической и прочей информации.

Рассматривая вопросы организации систем «человек — ЭВМ» В. М. Глушков и др. [128] отмечают, что она может в принципе состоять из «заказчика» (пользователя или поставщика задачи), «программиста», «оператора по устройству связи с ЭВМ», «оператора по вводу и подготовке данных» и собственно ЭВМ.

В начальный период внедрения ЭВМ в практику работ, когда они использовались главным образом для решения уникальных задач, которые ставились математиками, эффективность систем «Человек — ЭВМ» определялась главным образом техническими возможностями машин. По мере развития технических средств обработки информации и увеличения парка ЭВМ стал расширяться и круг решаемых задач прикладного характера. При этом происходит «перераспределение» соотношения между количеством задач, требующим участия человека в процессе решения («диалоговые задачи») и не требующим его участия («недиалоговые задачи»), причем такое перераспределение идет в сторону увеличения задач диалогового типа.

Для удовлетворения возрастающего спроса на машинное время можно было пойти двумя путями [128]:

— создавать крупные вычислительные центры с «поточными» методами обслуживания пользователей-заказчиков;

— автоматизировать некоторые функции программистов и операторов и передавать остальные их функции поставщику задачи, обеспечив последнему возможность прямой связи с ЭВМ.

В соответствии с этим наметились две основные тенденции в развитии традиционной системы «человек — ЭВМ».

Одна тенденция заключалась в стремлении отдалить заказчика от машины, «включив» между ним и программистом консультанта, на которого возлагалось выполнение основных действий, входящих в этапы ознакомления с задачей и составление плана ее решения с помощью ЭВМ. В качестве таких консультантов могут выступать алгоритмисты или системные программисты, хорошо знающие особенности решения различных задач на основе ЭВМ. Другая тенденция сводилась к тому, чтобы приблизить заказчика к ЭВМ, причем в качестве консультанта должен выступать сам специалист-заказчик (пользователь системы).

Вторая тенденция получает в настоящее время все большее признание, особенно при использовании ЭВМ для научных исследований.

Системы «человек — ЭВМ» подразделяются в зависимости от характера связи пользователя с ЭВМ на две большие группы:

— системы прямой (on-line) связи, т. е. без посредников;

— системы непрямой (off-line) связи, т. е. через посредников.

Системы прямой связи подразделяются на диалоговые и недиалоговые. В первом случае пользователь, находясь за пультом, имеет возможность непосредственно вмешиваться в процесс обработки машиной его задания, получить и учесть результаты промежуточного решения, контрольного счета и т. п. Пользователь может немедленно устранить ошибки, ввести в ЭВМ исправленные данные и продолжить решение. В этом случае система прямой связи используется для решения диалоговых задач. В недиалоговых системах прямой связи нет и пользователь лишен возможности непосредственно вмешиваться в процесс решения. Ввод исправленного сообщения можно выполнить в таких си-

стемах только после его доработки посредниками между пользователем и ЭВМ.

В последние годы в отечественной и зарубежной литературе появилось много работ, посвященных применению современных методов обработки информации для решения некоторых задач прикладной геологии [18, 39, 90 и др.]. Характерной особенностью настоящего этапа внедрения ЭВМ в практику работ является использование системного подхода к решению проблемы повышения эффективности геологических исследований. Как следует из публикаций, такой подход, видимо, является весьма перспективным и открывает новые пути как для решения некоторых практических задач, так и для геологических научных исследований.

Известно, что новейшими разработками в этом плане являются интегрированные человеко-машинные системы, сочетающие логику мышления, теоретические знания, практический опыт и интуицию специалистов с большими возможностями ЭВМ по хранению, накоплению и оперативной обработке информации [128]. В принципе, такого рода человеко-машинные системы состоят из трех основных частей:

- машинного банка геологической, геофизической, геохимической и другой информации;

- блока диалоговых процедур «специалист — ЭВМ»;

- блока разработки вариантов решений.

Вопросы структуры и организации банков данных рассматриваются многочисленными авторами [19, 73, 95, 122 и др.]. Не затрагивая отдельных сторон некоторых особых подходов, можно считать, что все исследователи, занятые вопросами внедрения математических методов в геологии, единодушно признают необходимость использования ЭВМ для хранения различной геологоразведочной информации.

Существенным препятствием для широкого внедрения банков данных является отсутствие формализованного геологического языка и единой формы геологической документации [39, 44]. Использование ЭВМ ЕС открывает более широкие возможности для создания машинных банков данных. Так, например, реальные перспективы приобретает решение вопроса создания территориального машинного фонда картографической и числовой информации по определенному конкретному региону (территориальному геологическому управлению).

Весьма важной составной частью человеко-машинных систем является блок диалоговых процедур, на основе которого вырабатываются варианты решения поставленной задачи геологической оценки.

На необходимость построения автоматизированных систем диалогового типа указывалось, например, в работах [24, 134 и др.]. Одна из первых реализаций этого блока приведена в работе [2]. Известно, что к настоящему времени накоплен значительный опыт решения задач прикладной геологии на основе методов корреляционного и регрессионного анализа, классификации на основе алгоритмов распознавания образов и т. п. [25, 30, 105 и др.]. Вместе с тем, как показано в работе [105], даже при использовании алгоритмов распознавания окончательные результаты классификации существенно зависят от особенности работы алгоритма.

Очевидно, невозможно свести все многообразие задач практической геологии к одной стандартной форме решения задач на основе математических методов. Следовательно, использование какого-либо математического подхода к решению конкретной задачи геологической оценки

требует соответственно своей технологии диалоговых процедур «специалист — ЭВМ». Под термином «технология» мы понимаем определенную методическую схему — последовательность операций взаимодействия специалиста с ЭВМ при решении какой-либо задачи геологической оценки по имеющейся в машинном банке всевозможной геологической, геофизической, геохимической и прочей информации на основе конкретного математического метода. Так, например, при решении задачи с использованием алгоритмов распознавания образов требуется определить необходимые признаки, выбрать эталонные объекты обучения, составить матрицу характеристик этих объектов, проверить «однородность» выборки, получить решающее правило, проверить качество распознавания на контрольных объектах, выполнить процедуру классификации, проанализировать полученные результаты.

Применение методов таксономии для безэталонного районирования территории (на основе принципа общности и исключительности и т. п.) требует другой технологии взаимодействия специалиста с ЭВМ в процессе решения поставленной задачи.

Таким образом, становится очевидным необходимость разработки стандартных технологических модулей, позволяющих на основе машинного банка данных решать различные содержательные задачи оценки геологических ситуаций, используя всевозможные методические подходы.

В рассмотренной ранее первой версии АИПС [2, 39] взаимодействие между специалистом и ЭВМ осуществлялось на относительно невысоком уровне, однако наличие современных устройств автоматизированного ввода и вывода информации открывает широкие перспективы для повышения качества диалоговых процедур в процессе решения содержательных задач прикладной геологии.

Особо следует подчеркнуть, что различные технологические схемы обработки информации с помощью ЭВМ не могут в настоящее время полностью заменить традиционные методы составления геологических заключений. Поэтому важной задачей повышения эффективности применения математических методов в геологии является проблема более полного использования теоретических знаний и опыта специалистов геологов на основе эвристических методов. В то же время разрабатываемые интегрированные человеко-машинные системы можно рассматривать как качественно новый и эффективный инструмент анализа геологической информации, позволяющей существенно ускорить процесс получения окончательных вариантов решений. Например, на основе таких систем появляется возможность осуществлять имитационное моделирование, т. е. оценивать конечные результаты различных гипотез (моделей) и тем самым проверять прикладное значение той или иной геологической концепции. Такого рода «проигрывание гипотез» напоминает штабные игры, т. е. получение определенных конечных результатов в зависимости от изменяющейся исходной ситуации.

Это, в свою очередь, открывает возможность определения с помощью ЭВМ качества подготовки студентов и специалистов геологов.

Рассматривая пути и способы построения автоматизированной системы металлогенического прогнозирования, В. А. Нагорский [73] отмечает, что такая система должна

— достаточно полно охватывать определяющие стороны проблемы прогноза, в частности формирование первичного признакового пространства;

— обязательно учитывать априорную неформализованную информацию о механизмах рудного контроля, отражающую опыт и интуицию геологов;

— использовать первичную информацию не только количественного, но и качественного вида, а главное графическую информацию (геологические карты, тектонические схемы и т. п.);

— учитывать такие свойства реального геологического пространства, как нестационарность и автокоррелированность полей признаков;

— выдавать прогнозную карту «структурного» вида в изолиниях степени перспективности;

— иметь возможность оценивать прогнозные запасы полезных ископаемых в пределах площадей произвольной формы и размеров;

— получать дополнительную информацию о характере влияния рудоконтролирующих факторов на локализацию оруденения, иметь возможность определяющей проверки исходных геологических представлений;

— иметь возможность сравнения отдельных поисковых методов и их комплексов, а также различных вариантов прогнозных карт по их геологической и экономической эффективности.

Естественно, что наиболее полно такого рода требования могут быть удовлетворены лишь при реализации человеко-машинной системы диалогового типа.

К настоящему времени известно большое количество исследований по решению задач выделения и оценки рудных объектов по геолого-геофизическим данным с использованием математических методов и ЭВМ.

Анализируя методики обработки гравимагнитных данных при оценке рудоносности площадей, Б. Н. Павлов и др. [97] отмечают, что в настоящее время наметилось два крайних подхода:

— использование только геофизических данных с установлением непосредственных связей между оруденением и некоторыми особенностями физических полей, минуя этап интерпретации геофизических данных;

— использование комплекса геолого-геофизических данных с максимально полным учетом связей между геологическим строением и физическими полями (имеется в виду случай, когда рудные объекты непосредственно не выражаются в физических полях).

Различие в этих подходах носит принципиальный характер и касается методологии исследований.

Сторонники первого направления, не рассматривая содержательные аспекты связей между рудоносностью территории и ее геологическим строением, пытаются определить прямые связи между оруденением и геофизическими полями. Так, И. В. Головин и Э. И. Супруненко [30] обосновывают этот подход «неясностью структурного и магматического контроля рудных полей, процессов регионального метаморфизма и других факторов».

Второе направление, использующее максимальный перевод геофизической информации на геологический язык, основано на установлении теоретически обоснованных связей между рудоносностью и геологическим строением исследуемой территории. На основе второго направления в рамках системного анализа [2] разрабатываются автоматизированные информационно-логические системы. Опытно-производственная эксплуатация одной из них [39, 97] показала, что такая система в руках специалистов, обладающая большим опытом интерпре-

тации и хорошо знающих район исследований, позволяет достаточно эффективно решать как задачи выделения рудных объектов, так и выполнять методические исследования, направленные на выбор методики решения прогнозной задачи. Однако использование такой системы требует от специалистов не только определенного опыта, но и полноты знаний геолого-геофизических материалов по значительным территориям.

Вместе с тем использование при решении задач оценки рудоносности площадей в качестве исходных данных только геофизических данных в цифровом виде по исследуемой территории, не считая минимального объема данных по эталонным объектам при некоторых априорных допущениях, позволяет полностью автоматизировать процедуру прогнозирования и выполнять решение задач в автоматическом режиме, исключая участие специалистов. Примером такой автоматической системы обработки геофизических данных «от журналов до перспективных площадей» может служить, например, программа «Днепр».

Правда, сами авторы методики, называя ее автоматизированной системой поиска, подчеркивают, что она «служит помощником инженера-геофизика или геолога, ведущего поиска полезных ископаемых» [117]. Такого рода легкость получения результатов в виде перспективных площадей и участков фактически без усилий специалистов сильно подкупает возможностью своей реализации, и поэтому существует возможность развития подобных автоматизированных систем. Б. Н. Павлов и др. [97] достаточно убедительно показали с различных позиций, что наиболее перспективным направлением в настоящее время является путь комплексного использования всех имеющихся данных планирования и управления процессом обработки, последовательной формализации процедур по мере накопления знаний и опыта решения практических задач с использованием математических методов и ЭВМ.

Реализовать в подобном виде решение прогнозных задач возможно в настоящее время лишь в автоматизированном режиме. Пример такого рационального разделения функций, выполняемых специалистами и реализованных на ЭВМ, приведен в работе [99]. Перед решением конкретной задачи специалисты предварительно планируют все процедуры обработки данных и часть из них (не подлежащих формализации) выполняют непосредственно сами. В процессе решения задачи специалисты анализируют и оценивают ряд промежуточных результатов и при необходимости изменяют (корректируют) выборку эталонных рудных объектов, комплекс геологических факторов и др.

В автоматизированных системах, направленных на выделение перспективных участков, должна предусматриваться возможность планирования и управления процессом обработки данных. Положительным примером такого подхода может служить система, разрабатываемая В. В. Давыденко, В. С. Романовым и др. [90]. В этой системе предусматривается, что при решении каждой конкретной задачи специалист может выбирать наиболее подходящие для заданных условий способы и конструировать по своему усмотрению процесс обработки данных.

В автоматизированной информационно-прогнозирующей системе [2, 39, 99] в основу предварительной обработки информации положена традиционная методика, принимаемая при геологическом картировании и изучении глубинного строения исследуемой территории. Основное внимание при разработке, в соответствии с конкретными условиями ре-

шаемых задач, здесь уделено преобразованиям, основанным на непосредственном использовании существующей геологической информации и приводящим к построению комплексных геолого-геофизических материалов (геологических карт, схем геологического строения верхней части земной коры), используемых в последующем для комплексного анализа с помощью ЭВМ с целью оценки перспектив рудоносности исследуемой территории.

В программах, где процесс обработки гравимагнитных данных реализован в автоматическом режиме и заканчивается выделением перспективных участков [117], возможность управления процессом обработки отсутствует. В этих программах заложен фиксированный набор преобразований с установленными постоянными параметрами (табл. 8).

Следует подчеркнуть, что в общем случае автоматический режим предпочтительнее, но в настоящее время он не может быть реализован без дополнительных серьезных теоретических разработок и значительного объема экспериментальных исследований. Экспериментальные исследования следует ориентировать на проверку надежности предлагаемых критериев оценки связей, что позволит последовательно переводить выполнение все большего числа процедур в автоматический режим обработки [97].

В настоящее время, очевидно, целесообразно направить основные усилия на разработку человеко-машинных систем с такой методикой решения прогнозных задач, которая была бы лишена недостатков формального и «машинного» подхода и в полной мере использовала бы такие важные аспекты, как знания специалистов и возможности ЭВМ по хранению и оперативной обработке информации. Всем этим условиям отвечают интегрированные человеко-машинные системы.

Как указывает С. П. Никаноров в предисловии к русскому переводу книги Стенфорда Оптиера [86]: «Дисциплина мышления коллектива людей, диктуемая достаточно мощным и ясным концептуальным аппаратом, возможно, является более важной вещью, чем математический аппарат, позволяющий производить те или иные расчеты. Новейшие си-

Таблица 8

Сопоставление возможностей управления процессом обработки информации при решении прогнозных задач в автоматизированных и автоматических системах (по Б. Н. Павлову и др. [97])

Элементы системы	В автоматизированной системе АИПС [2]	В программе автоматической обработки данных «Днепр-4» [117]
Выборка эталонных рудных объектов	Возможно изменение выборки	Возможно изменение выборки
Комплекс геологических факторов	Формируется в соответствии с принятой моделью процессов рудообразования	Заменен некоторыми фиксированными априорными допущениями
Процедуры обработки гравимагнитных данных	Изменяются в соответствии с комплексом геологических факторов	Зафиксированы, изменения не предусмотрены
Параметры процедур обработки	Предусмотрено изменение параметров	То же

системы представляют собой человеко-машинные интегрированные системы, имеющие синергический характер и строго ориентированные на решаемые проблемы». Системный анализ, как методология решения проблемы, используется в таких системах в роли каркаса, объединяющего все необходимые методы, знания и действия для решения проблемы.

В принципе, современная человеко-машинная система диалогового типа применительно к решению задач геологического прогнозирования может состоять из совокупности четырех связанных между собой подсистем, как это показано на рис. 9.

Подсистема диалоговых процедур (1) содержит блок описания языка пользователя (1.1), блок диалоговых процедур и управления заданиями (1.2) и блок документирования выходной информации (1.3). Подсистема предназначена для общения специалиста с системой, ввода задачи и схемы ее решения, управления процессом решения задачи, контроля, анализа и выдачи промежуточных и окончательных результатов.

Подсистема ввода и хранения геологической информации (2) состоит из блока управления (2.1), блока автоматизированного ввода информации (2.2) и банка данных (2.3). Подсистема обеспечивает ввод, хранение и выбор первичной информации для справок и обработки в других подсистемах.

Подсистема формирования и ведения целевых информационных массивов (3) включает в себя блок управления (3.1) и комплект блоков, обеспечивающих формирование целевых массивов и их каталогизацию в виде, пригодном для обработки по различным технологическим схемам (3.2, ..., 3. n) решения задачи.

Наконец, подсистема разработки вариантов решений (4), имеющая также блок управления (4.1) и блоки реализации по различным технологиям (4.2, ..., 4. n), предназначена для исследования и разработки вариантов прогнозных оценок по рассматриваемым геологическим гипотезам.

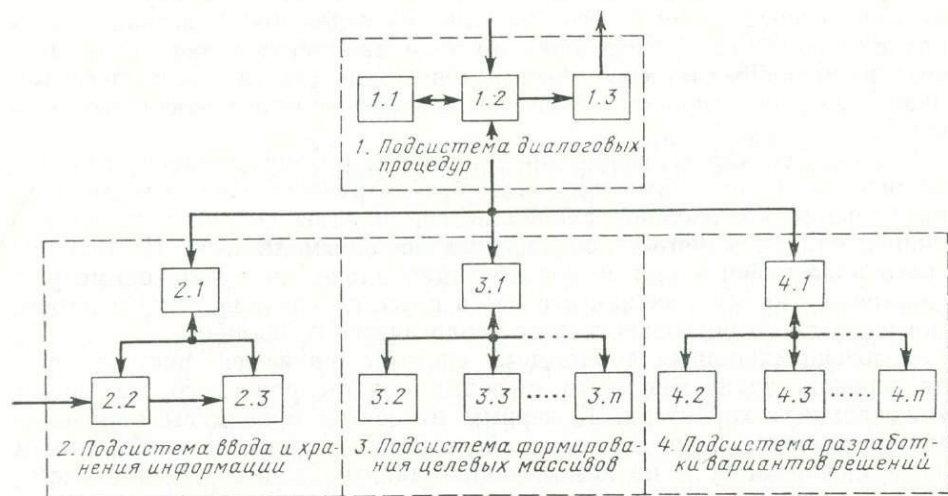


Рис. 9. Блок-схема человеко-машинной диалоговой системы для решения задач геологического прогнозирования

Решающие процедуры подсистемы включают различные алгоритмы статистического анализа, распознавания образов с поощрением, таксономии, методы экспертных оценок и т. д.

Следует отметить, что каждая из рассматриваемых подсистем может работать независимо или во взаимосвязи с другими подсистемами; это определяется задачей и технологией ее решения, задаваемой через блок диалоговых процедур (1.2).

Данная схема, естественно, не является единственно возможной, однако она содержит все основные блоки, характеризующие человеко-машинную диалоговую систему в современном ее понимании.

На основе такой схемы можно конструировать специализированные человеко-машинные системы геологического прогнозирования, позволяющие автоматизировать отдельные составляющие части общего процесса геологического прогнозирования, как иерархической системы принятия оптимальных решений при неполной информации об объекте исследования.

При разработке таких систем следует ориентироваться на существующие формы геологической информации (в том числе картографической), имея в виду, что геологическая карта является документом, достаточно понятным всем специалистам. Рассмотренный ранее способ преобразования геологических карт в цифровой вид позволяет осуществлять накопление такой информации в машинном банке данных с минимальными затратами.

Практическая реализация такого территориального банка данных является одним из неперенных условий создания автоматизированных человеко-машинных систем.

Разработка необходимого комплекса математических и прогнозных средств для переработки информации, хранящейся в банке данных, позволяет геологу провести быструю оценку информативности различных наборов признаков для возможных типов месторождений и выделить на изучаемой территории участки, имеющие сходство с эталонами по тем или иным наборам признаков. Такой материал избавляет геолога от необходимости пересматривать частные данные по всей площади, сосредотачивает его внимание на отдельных участках, подготавливает необходимые сведения по всем намеченным вариантам прогноза и высвобождает ему максимум времени для творческого осмысления задачи и принятия решений по выделению перспективных площадей.

Имеющиеся картографические материалы различных масштабов, на составление и систематизацию которых затрачены значительные усилия больших коллективов специалистов, должны сохраняться на машинных банках в формах, адекватных исходным данным. Накопление такого рода информации не только рационально, но и необходимо уже в настоящее время, так как все это в итоге обеспечивает возможность многоаспектного ретроспективного геологического анализа.

Человеко-машинные диалоговые системы являются инструментом для решения задач как чисто информационно-справочного, так и содержательного характера. К первым относятся процедуры получения справок и сводок по различным районам и площадям, по необходимым видам информации и по комплексным данным в виде таблиц, гистограмм, графиков, схем, планов и т. д., включая статистическую обработку данных. Примером задач содержательного характера являются задачи типизации и районирования территории, отдельных площадей, выделения однородных участков по отдельным видам информации и

комплексным данным на основе применения методов многомерного статистического анализа и таксономии, методов распознавания образов и т. п. Для реализации изложенных выше принципов по созданию автоматизированной человеко-машинной информационно-прогнозирующей системы требуется решить ряд задач.

1. На основе системного анализа выявить главные звенья геолого-разведочного процесса, определяющие в целом его эффективность.

2. Проанализировать методологию решения с помощью математических методов и ЭВМ задач прогнозирования геологических объектов на различных этапах поисков и разведки месторождений.

3. Разработать принципы построения автоматизированной информационно-прогнозирующей человеко-машинной диалоговой системы, характеризующейся рациональным распределением функций между специалистом и ЭВМ в процессе выработки прогнозных решений, начиная от подготовки и ввода информации в машину и организации машинного банка данных и кончая подготовкой и принятием решений.

4. Разработать способ хранения геологической, геофизической и геохимической информации в ЭВМ, удовлетворяющей, по крайней мере, следующим требованиям:

— информация должна вводиться в ЭВМ в виде, адекватном первичному;

— хранящаяся в ЭВМ информация должна быть ориентирована не на решение отдельных частных задач или задачи геологического прогнозирования, а на неоднократное и многоцелевое ее использование;

— способ хранения должен предусматривать возможность пополнения информации, ее корректуры, а также возможность привлечения и обработки любых данных.

5. Разработать алгоритмы извлечения и анализа необходимой для прогнозирования информации, а также алгоритмы формирования количественных поисково-оценочных критериев, учитывающих взаимосвязи между различными комплексами факторов и масштабностью оруденения.

6. Реализовать все необходимые алгоритмы обработки геологических данных в виде программного обеспечения на ЭВМ.

При реализации математического обеспечения системы следует в максимальной степени использовать возможности оперирования данными, имеющиеся у современных ЭВМ ЕС.

2. СТРУКТУРА И СХЕМА ФУНКЦИОНИРОВАНИЯ АВТОМАТИЗИРОВАННОЙ ЧЕЛОВЕКО-МАШИННОЙ СИСТЕМЫ АИПС/ЕС «РЕГИОН»

Описание макроструктуры системы

Информационно-прогнозирующая система АИПС/ЕС «Регион» разработана на основе рационального сопряжения традиционной методики геологического прогнозирования и методов решения прогнозных задач с помощью ЭВМ. Важнейшей составной частью системы является территориальный машинный фонд картографической информации.

Решение содержательных геолого-прогнозных задач исследования и использования геологических гипотез производится в АИПС на основе применения логических и вычислительных процедур, выполняемых ЭВМ, в сочетании с неформальными методами специалиста в процессе принятия решений при поиске, прогнозировании и оценке геологических объектов. При этом используются математические методы

оптимизации решений при неполной информации об объекте исследования.

АИПС/ЕС «Регион» состоит из трех основных частей, обеспечивающих решение содержательных задач геологического прогнозирования:

— территориального машинного фонда картографической информации, состоящего из банка исходных данных и постоянного архива преобразованной трансформированной и синтезированной информации;

— блока разработки вариантов решений, включающего блок формирования целевых массивов и блок технологических модулей решения задач на основе различных методических подходов;

— комплекса внешних и внутренних интерфейсов. К внешним интерфейсам относятся блок кодирования исходной информации, язык пользователей системы и блок оформления выходной информации. К внутренним — транслятор языка пользователей, блоки преобразования исходной информации и формирования личных целевых массивов пользователей.

На рис. 10 приведена принципиальная блок-схема системы АИПС/ЕС «Регион», т. е. ее макроструктура.

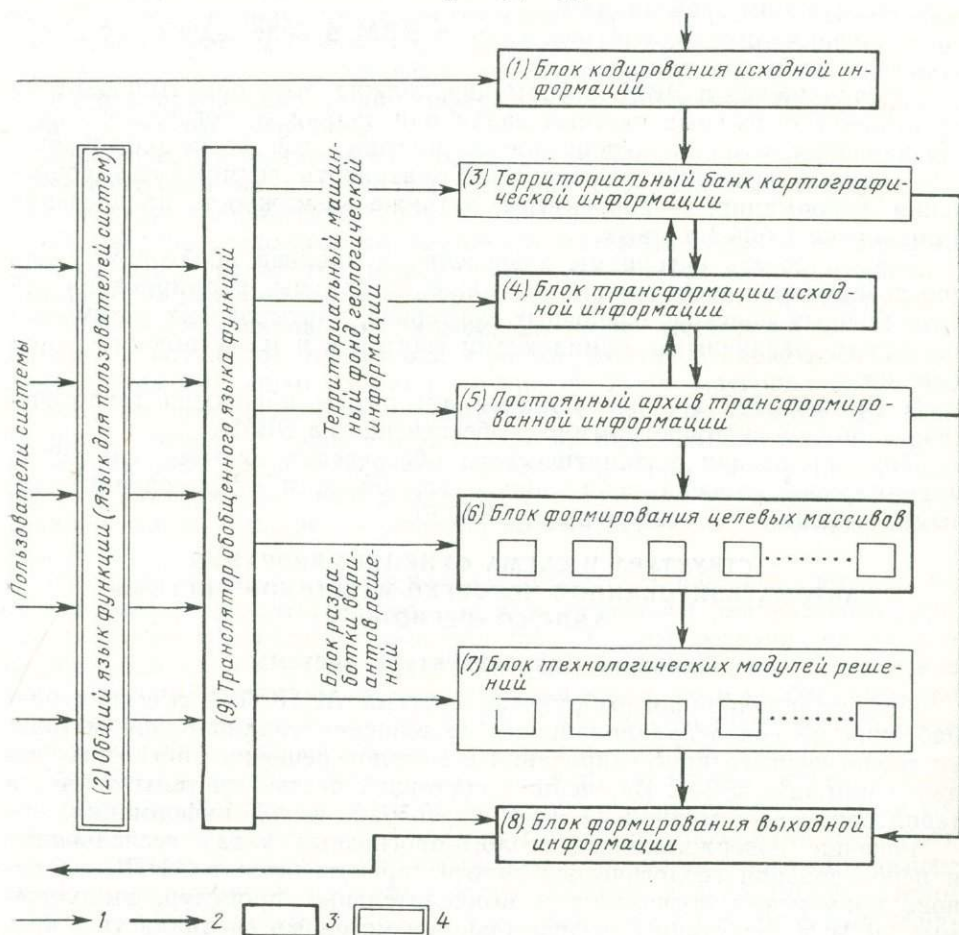


Рис. 10. Макроструктура АИПС/ЕС «Регион».

1 — управляющая информация, 2 — содержательная информация, 3 — блоки внутреннего интерфейса, 4 — блоки внешнего интерфейса

Первичные исходные данные (картографического или числового типа, а также текстовые описания геологических объектов) через блок (1) вводятся в территориальный банк данных (3). При этом осуществляется их контроль и привязка к условной системе координат.

Банк картографической информации является документированным хранилищем исходной информации и обладает всеми необходимыми для этого свойствами. Далее исходная информация подвергается предварительной обработке в блоке (4) с целью ускорения последующего процесса решения задач в подсистеме разработки вариантов решений. Исходная информация может подвергаться как логическим преобразованиям (объединению факторов, пересечению и т. д.), так и различным математическим преобразованиям. Преобразованная информация записывается и сохраняется в блоке (5). Совокупность блоков (3, 4 и 5) составляет так называемый «территориальный машинный фонд геологической информации». Последующий процесс решения аналитических и классификационных задач происходит автономно от описанного процесса накопления информации и в индивидуальном порядке. С этой целью пользователи — геологи-интерпретаторы — с помощью блоков (6 и 7) осуществляют отбор информации и решение по ней задач различных типов в соответствии с конкретными постановками и своими собственными концепциями. Совокупность блоков (6 и 7) составляет «подсистему разработки вариантов решений». Выходная информация, поступающая пользователям, в том числе справочная, оформляется в блоке (8). Запросы пользователей к системе на все виды работ формулируются на языке пользователей (блок 2), на так называемом «обобщенном языке функций». Преобразование запросов пользователей в задания системе, работающей под управлением операционной системы ЭВМ, производится в блоке (9) — «трансляторе обобщенного языка функций».

Получаемые в процессе человеко-машинного взаимодействия классификационные и прогнозные карты анализируются геологом-интерпретатором — пользователем системы, который решает вопрос о возможности и целесообразности их использования или о постановке новых экспериментов с измененными исходными данными и параметрами решения.

В качестве алгоритмов анализа и классификации при этом могут быть использованы алгоритмы распознавания образов с обучением, безэталонная классификация на основе метода таксономии, эвристическое моделирование и т. п.

Такое построение системы позволяет сочетать логику мышления, теоретические знания, практический опыт и интуицию специалистов с большими возможностями ЭВМ по хранению, накоплению и оперативной обработке различной информации.

Машинное преобразование информации в системе АИПС/ЕС

В процессе работы системы исходная информация проходит целый ряд различных преобразований. Вид и последовательность этих преобразований зависят от характеристик исходных данных (картографические, цифровые), характера требуемой обработки и форм выдачи результатов, необходимых пользователю.

Вся совокупность преобразований информации в системе разделяется на два этапа: этап накопления информации и этап решения содержательных задач прогнозирования и оценки геологических ситуаций.

Преобразование информации в процессе формирования машинного геологического фонда

Первый этап практически соответствует формированию территориального машинного фонда картографических данных. Такой машинный фонд состоит из двух основных массивов: первичного массива исходных данных и массива трансформированной информации. Первичный массив, выполняющий функции банка исходной информации, состоит из двух библиотек факторов — картографических и числовых,

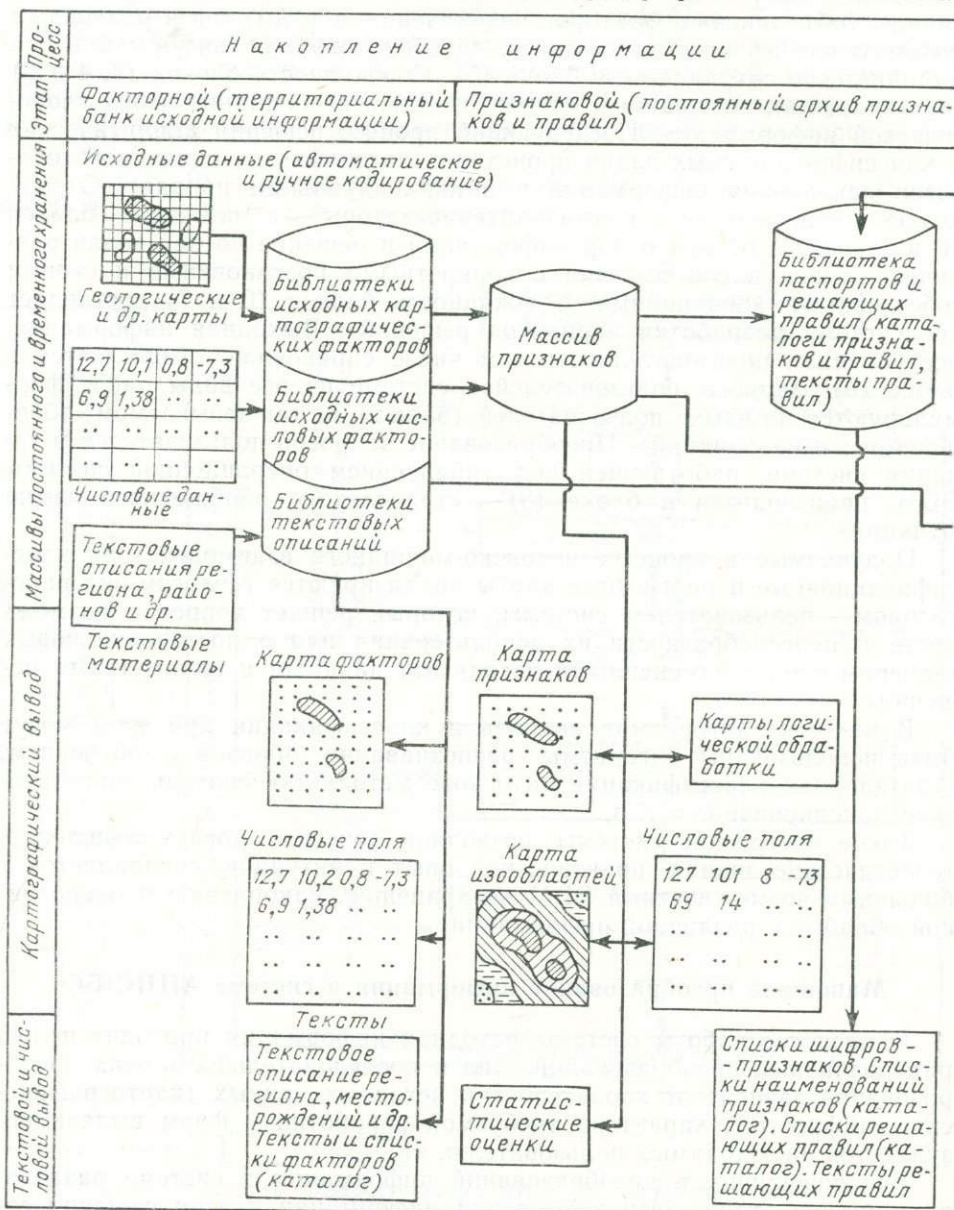
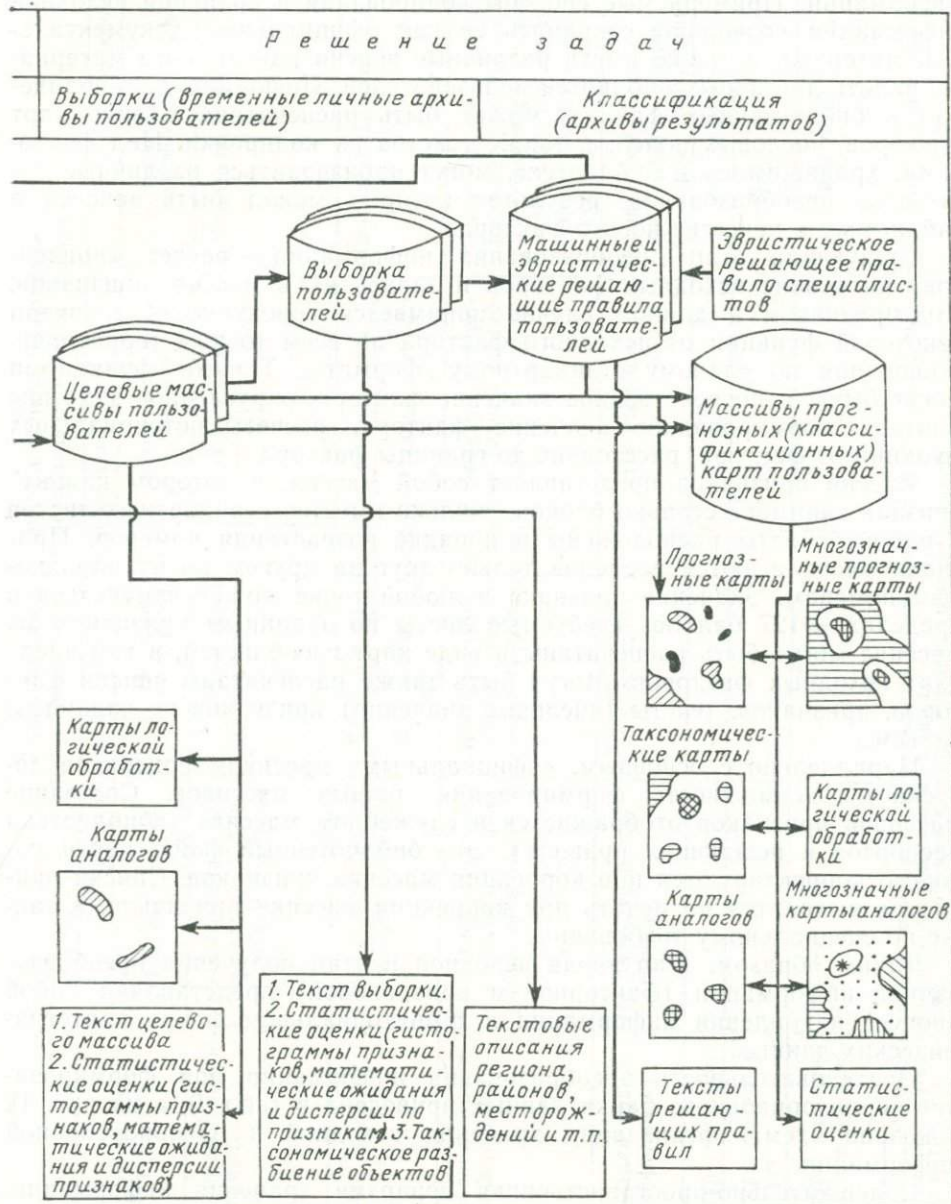


Рис. 11. Общая схема преобразования информации в АИПС/ЕС «Регион»

содержащих также и текстовые описания факторов. Он является документированным хранилищем исходной информации по исследуемой территории и служит основой для последующего преобразования исходной информации, а также для выдачи различных справок по запросам пользователей.

Второй массив является постоянным архивом преобразованной информации. Его основное назначение — ускорение последующего процесса решения задач за счет предварительной обработки информации.

На рис. 11 приведена общая схема преобразования информации в АИПС/ЕС. Исходная картографическая и числовая информация, зако-



дированная автоматизированным способом (например, с помощью прибора типа «Шифратор-2») или вручную, запоминается в ЭВМ на внешних магнитных носителях в виде библиотечных файлов.

Особо отметим, что программное обеспечение системы в принципе позволяет подключение автоматического устройства ввода картографической информации (например, оптического считывающего устройства на основе метода растровой дискретизации).

Выбранный тип файлов для хранения исходной информации весьма удобен для пополнения и выполнения коррекций факторов, осуществления различных справочных функций и последующего использования информации. Применяемые способы кодирования и хранения исходной информации позволяют сохранять ее как официальный документальный материал, а также иметь различные версии однотипного материала, вплоть до личных библиотек пользователей. Информация, сохраняемая в библиотечных файлах, может быть распечатана в виде карт факторов, числовых полей, а также текстов их кодировки. Над факторами, хранящимися в библиотеке, могут производиться различные логические преобразования, результат которых может быть занесен в библиотеку в качестве нового фактора.

Следующий этап преобразования информации — расчет «признаков» на основе исходных факторов и запись их в массив признаков. Под признаком в данном случае понимается совокупность значений некоторой функции от исходного фактора по всем точкам территории, записанная по единому стандартному формату. Такими функциями могут быть: непосредственное значение фактора, округленное значение фактора, нормированное значение фактора, расчеты потенциальных функций от фактора, расстояние до границы фактора и т. п.

Массив признаков представляет собой массив, в котором каждый признак занимает столько блоков, сколько имеется стандартных листов в районе. Листы расположены в порядке возрастания номеров. Признаки располагаются последовательно друг за другом по их порядковым номерам. Значение признака в любой точке может изменяться в пределах ± 127 баллов. Требуемые листы по заданным признакам из массива могут быть распечатаны в виде карт изообластей, в том числе карт исходных факторов. Могут быть также распечатаны списки факторов, признаков, тексты (числовые значения) признаков по заданным листам.

Параллельно с основным, «официальным» массивом признаков допускается возможность формирования личных массивов. Состояние массивов признаков отображается в служебном массиве («библиотека паспортов и решающих правил»). Это библиотечный файл, книги которого корректируются при коррекции массива признаков. Списки признаков выводятся на печать при коррекции массива признаков, а также по специальному требованию.

Таким образом, этап ввода исходной и этап получения преобразованной информации (факторной и признаковой) представляют собой процесс накопления информации в территориальном фонде картографических данных.

Последовательность преобразования информации при формировании территориального банка данных приведена в табл. 9, а на рис. 12 показана схема организации хранения исходной и преобразованной информации.

Содержательно-пространственная иерархия хранения информации, изображенная на рис. 12, требует некоторых пояснений. Банк карто-

Последовательность преобразования информации
при формировании территориального банка данных

Характеристики и процедуры	Предварительная обработка исходной информации			Формирование банка данных	
	Исходный материал	Машинное кодирование	Перенос на машинные носители	Формирование первичного массива	Формирование массива признаков
Форма представления информации	Геологическая и другие карты, геофизические и другие поля	Символьная	Символьная	Библиотечный файл	Файл прямого доступа блоками по 1032 байта Последние 8 байт шифр
Носитель информации	Традиционная карта (оцифрованная карта)	Банк ассемблера При автоматизированном вводе — перфолента	Перфокарта	Магнитный диск*	Магнитный диск
Минимальная единица информации	Качественная (количественная) характеристика точки	Строка = целому числу строк листа (числовое значение в точке)	Перфокарта = строке бланка ассемблера	„Книга“ (числовое значение фактора). Запись = целому числу строк карты = целому числу (числовые значения фактора)	Числовое значение признака в каждой точке территории
Организация хранения информации	На плане всей территории	Построчно	По столбцам	Совокупность строк, лист, фактор, группа факторов, территория	Точка, строка, лист, признак, территория
Процедуры создания	—	—	—	Ф-1, Ф-2**	П-1, Ф-8А, (П-3)
Процедуры печати и контроля	—	—	—	Ф-3 — Ф-8	Н-5 — Н-5 Н-7
Процедуры корреляции и преобразований	—	—	—	Ф-9, Ф-10	Н-8

* Хранение больших объемов информации осуществляется в виде копий дисков на магнитных лентах, которые при необходимости переводятся в рабочее состояние путем восстановления информации на дисках.

** Ф-1, Н-8 и т. д. — шифр используемых процедур преобразования информации.

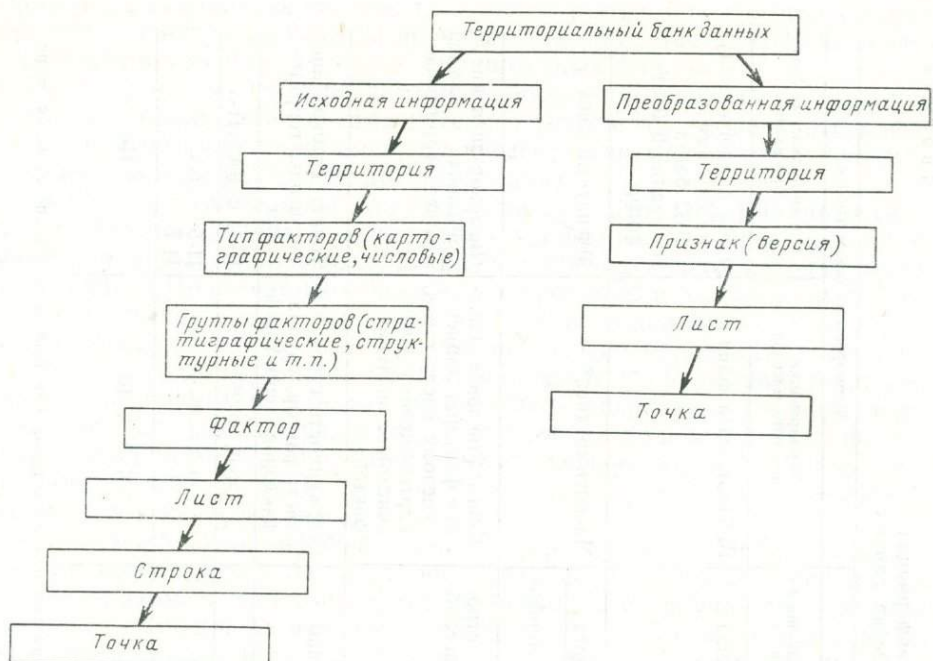


Рис. 12. Принципиальная схема организации хранения информации в территориальном машинном банке данных

графической информации системы может содержать информацию по одному или нескольким регионам. Информация по каждому из регионов состоит из двух типов факторов: картографических (двузначных) и числовых (многозначных). Оба типа факторов состоят из групп факторов, например литолого-стратиграфических, структурно-тектонических и т. п. Каждая из этих групп, в свою очередь, состоит из отдельных факторов, а информация каждого фактора хранится разделенной на части в соответствии с принятой листовой схемой деления региона. Хранимая информация одного фактора называется книгой.

Стандартным листом, принятым в системе «Регион», является лист размером 32×32 см или кратный ему. Информация отдельного фактора по одному листу называется страницей. При этом исходная информация кодируется и хранится по отдельным ячейкам по сетке 64×64 для картографических и 32×32 — для числовых факторов. Кроме того, при кодировании и хранении информация ячеек, находящихся в листе на одной горизонтали, объединяется в так называемые «строки».

Отметим также, что помимо качественной и числовой информации самих факторов в банке хранится также описательная (текстовая) информация регионов, групп факторов и отдельных геологических объектов.

Преобразование информации на этапе решения задач

Использование накопленной информации в процессе решения содержательных задач происходит параллельно и независимо от первого этапа преобразования информации в том смысле, что используется тот

объем информации, который накоплен на текущий момент. В процессе решения содержательных задач информация также претерпевает ряд преобразований, существенно зависящих от типа поставленных задач.

На рис. 11 показаны основные типы массивов, записанных на внешних магнитных носителях (дисках) в процессе решения задач. Здесь же показан вызов информации из различных массивов при использовании отдельных технологий решения.

Практически при всех технологиях используется первый тип массивов — целевые массивы. Целевой массив — это массив векторов, образуемый из массива признаков по заданным листам территории и заданному пространству признаков. Каждой точке (элементарной ячейке) листа карты соответствует вектор, координатами которого являются значения заданных признаков в данной точке.

На основе лишь одного целевого массива могут решаться как минимум задачи трех классов.

1. Расчеты статистических оценок различных признаков (например, построение гистограмм, распределение значений признаков, вычисление математического ожидания и дисперсии по отдельным признакам и т. п.).

2. Таксономический анализ однородности выборок в полном или сокращенном пространстве признаков по заданным листам территории.

3. Машинный поиск аналогов конкретной точки (с некоторой заданной мерой «близости» в полном или сокращенном пространстве признаков).

Второй тип массива, который используется в процессе решения, — массив выборок. Этот массив образуется путем выбора из целевого массива векторов в заданных точках и расположения их друг за другом в порядке выбора. Формат массива такой же, как и у целевого. Текст массива (числовые значения) может быть также выведен на печать. На основе только одного массива выборок могут решаться по крайней мере также три класса задач: 1) статистические, 2) таксономической классификации, 3) распознавание образов с обучением.

В последнем случае именно из массива выборок составляется выборка для обучения распознавания образов и контрольная выборка для оценки качества построенного решающего правила.

Третий тип массива — массив решающих правил. Этот массив образуется в результате работы процедур разработки решающего правила на основе использования алгоритмов распознавания образов или алгоритмов эвристического моделирования. Массив имеет последовательную запись. Структура массива сложная, имеет несколько типов записей и типов форматов. Как следует из изложенного, массив используется при решении двух типов задач: 1) распознавание с обучением и 2) классификация по алгоритму эвристического моделирования.

Содержание массива (решающее правило) выводится на печать как при записи, так и по специальному запросу.

В процессе решения задач также используется служебный массив, общий для обоих процессов: «библиотека паспортов и правил». Здесь при необходимости запоминаются выработанные решающие правила, а также решающие правила, составленные специалистами геологами по алгоритму эвристического моделирования. Список правил и их описание могут быть также выведены на печать.

Наконец, пятый тип массивов, используемый при решении задач, — массивы прогнозных и классификационных карт — образуются в результате обработки целевого массива каким-либо конкретным решаю-

щим правилом классификации либо методом таксономии. Формат массива тот же, что и у массива признаков. Массив может быть распечатан в виде карт изообластей; граничные значения изообластей и знаки печати при этом задаются непосредственно пользователем.

Окончательные результаты решения задач могут выдаваться в виде следующих типов печати:

- 1) гистограмм значений признаков из целевого массива или массива выборки;
- 2) математических ожиданий и дисперсии признаков в целевом массиве или массиве выборки;
- 3) карты результатов таксономической классификации целевого массива;
- 4) карты результатов поиска заданного аналога в целевом массиве;
- 5) прогнозной карты по результатам обучения распознаванию образов (с одновременным указанием апостериорных информационных вкладов отдельных признаков);
- 6) имитационной прогнозной карты, построенной на основе алгоритма эвристического моделирования (с выдачей априорных информационных весов отдельных признаков).

Последовательность преобразования информации при решении содержательных задач приведена в табл. 10.

Математическое обеспечение системы

Проблема широкого практического внедрения современной вычислительной техники в прикладную геологию осложняется в настоящее время рядом серьезных трудностей. К ним относятся, в первую очередь, высокая трудоемкость ввода исходной информации в ЭВМ и отсутствие, как правило, непосредственного общения конечных пользователей (специалистов геологов) с ЭВМ.

Основные причины такого положения объясняются сложностью существующих систем кодирования исходной информации, весьма ограниченным распространением аппаратуры автоматического кодирования, отсутствием универсальных систем хранения исходной информации, значительным смысловым разрывом между языками конечных пользователей и входным языком программного обеспечения, сложностью освоения имеющегося специализированного математического обеспечения для других проблемно-ориентированных систем, значительной сложностью разработки математического обеспечения и, наконец, сравнительно быстрым моральным старением последнего.

В связи с этим при разработке математического обеспечения АИПС/ЕС «Регион», как проблемно-ориентированной системы, наряду с требованиями, обеспечивающими выполнение целевой функции, учитывался целый ряд дополнительных требований, выполнение которых позволило бы по возможности избежать указанных выше недостатков.

В целом при разработке структуры и элементов математического обеспечения (МО) системы учитывались следующие требования.

1. Полнота обеспечения целевой функции. Иными словами, система должна обеспечить накопление и сохранение всей исходной и преобразованной информации различных типов, при этом не должно быть ограничений на объем накапливаемой информации. Вся информация постоянного хранения (в первую очередь исходная) должна сохраняться как документальный материал, т. е. она должна быть доступной для контроля, детальной корректировки, получения справок, отображения,

**Последовательность преобразования информации
при решении задач прогноза и оценки**

Характеристики и процедуры	Формирование массивов		Решение задач прогноза и оценки			
	Формирование целевого массива	Формирование массива выборок	Анализ структуры выборки	Получение решающего правила (или ввод эвристического правила)	Классификация на основе решающего правила и распечатка карты	Классификация на основе таксономии и распечатка карты
Носитель информации	Магнитный диск	Магнитный диск, печать	Магнитный диск, печать	Магнитный диск, печать	Магнитный диск, печать, карты	Магнитный диск, печать, карты
Минимальная единица информации	Значение вектора = совокупности значений признаков в каждой точке					
Организация хранения	Значение признаков в точке — векторе — листе района	Совокупность векторов в точках с заданными координатами = выборка		Библиотечный массив данных	Значение признака в одной точке — описание вектора одной точки — лист — район	
Процедуры создания	В-1	В-4	К-1	К-3, К-5	К-6	К-2
Процедуры печати и контроля	К-7, П-6, В-6	М-7, П-7 В-6	К-1	К-3, К-5	К-6	К-2

копирования и т. п. Кроме того, должна быть предусмотрена возможность хранения информации ограниченного доступа.

При использовании информации должна быть обеспечена возможность обработки различных типов информации, обработка широким спектром различных методов и способов (различные преобразования и трансформации, статистический анализ, классификация по различным алгоритмам, в том числе эвристического моделирования и др.). Вместе с тем должна быть предусмотрена возможность анализа хода решения и содержательного анализа промежуточных и окончательных результатов.

2. Удобство работы для пользователей системы, т. е. должны обеспечиваться автономность процессов поиска и использования информации; непосредственность общения пользователей с системой на естественном языке; оперативность работы с системой, в том числе минимизация времени на выполнение большинства функций системы; возможность динамического воздействия на ход решения (имитационный режим); диагностика различных ошибок в процессе решения; возможность работы через индивидуальные консольные устройства (интерактивный режим). Кроме того, необходимо обеспечить коллективную работу значительного числа пользователей при сохранении индивидуальности и независимости их работы друг от друга и простоту обучения пользованию системой.

3. Минимальная трудоемкость работы с исходной информацией: простота сбора, минимальная предварительная обработка, автоматизация кодирования, возможность использования автоматического кодирования, автоматизированный контроль и диагностика вводимой информации, однократность ввода.

4. Высокие эксплуатационные характеристики: уровень надежности, необходимый для использования системы в реальном контуре выработки управляющих решений, простота обслуживания, высокий коэффициент использования включенного программного обеспечения.

5. Общие требования: устойчивость к моральному старению, минимальная трудоемкость разработки, совместимость с другими информационными и обрабатываемыми системами.

Принципы разработки математического обеспечения системы

С учетом изложенных выше требований и имеющихся возможностей в основу разработки математического обеспечения системы были положены следующие принципы.

1. Ориентация на существующие форматы исходной информации (геологические карты, карты геофизических и геохимических числовых полей, текстовые описания отдельных факторов и целых регионов и др.).

2. Автоматизация кодирования исходной картографической информации, компактность ввода, возможность детального контроля и коррекций, совместимость различных способов кодирования и ввода данных.

3. Иерархический принцип хранения информации: по содержанию (от отдельных факторов и исходных карт до прогнозных карт); по территориальному положению (от описания отдельных элементарных площадок до описания совокупностей регионов); по длительности хранения (от временных массивов до территориального картографического машинного фонда постоянного хранения); по категориям пользова-

телей (от личных массивов пользователей до общесистемного банка и архива).

4. Наличие детального и многообразного сервиса (в части коррекции, отображения, справок, копирования) и развитой системы программ по работе с зашифрованной информацией, наличие средств защиты информации.

5. Многоаспектная гибкость, достигаемая путем использования системы открытых интерфейсов между основными элементами системы: между различными типами исходной информации и системой (способы кодирования информации), между пользователями и системой (обобщенный язык пользователей); между языком пользователей и входными языками программного обеспечения различных типов (транслятор языка функций системы); между исходной и преобразованной информацией (блок пакетов программ трансформации); между наличным составом информации постоянного хранения и текущими требованиями отдельных пользователей, а также между внутренними структурами входной информации, принятыми в используемых пакетах прикладных программ (пакет программ образования личных целевых массивов пользователей); между системой и вычислительной установкой и между составом системы и требованиями пользователей (средства генерации и настройки).

6. Единая организация языка общения с системой — типа усеченного естественного, с разделами в соответствии с типами пользователей.

7. Максимальное использование готового программного обеспечения, особенно промышленной разработки; возможность автономного использования программного обеспечения, включенного в систему; выделение пользователям индивидуальных средств общения и ресурсов системы; включение в систему сервисных средств эксплуатации (развитие, обслуживание, обучение, демонстрация и др.).

Структура математического обеспечения системы

Общая структура математического обеспечения системы сложилась в итоге попытки удовлетворить выдвинутые требования путем применения изложенных выше принципов. В ходе разработки было установлено, что выполнение всего комплекса поставленных требований практически невозможно в рамках известных структур математического обеспечения — библиотек программ, пакетов программ, систем программирования [119]. В связи с этим при реализации системы АИПС/ЕС «Регион» была разработана структура математического обеспечения, изображенная на рис. 13. Такая структура математического обеспечения была определена, как «проблемно-ориентированная мультипакетная система» или «метапакет проблемных программ» (МПП). Такой метапакет представляет собой проблемно-ориентированную совокупность пакетов программ (в широком смысле слова «пакет программ») различного назначения, интегрированных в следующих аспектах: а) единого языка обращения; б) единой технологической совокупности массивов информации; в) единой системы интерфейсов.

Основными элементами структуры метапакета являются, таким образом, проблемно-ориентированная совокупность информационных и обрабатывающих программ, технологически интегрированная совокупность информационных массивов, язык обращения к системе (язык пользователей), система интерфейсов.

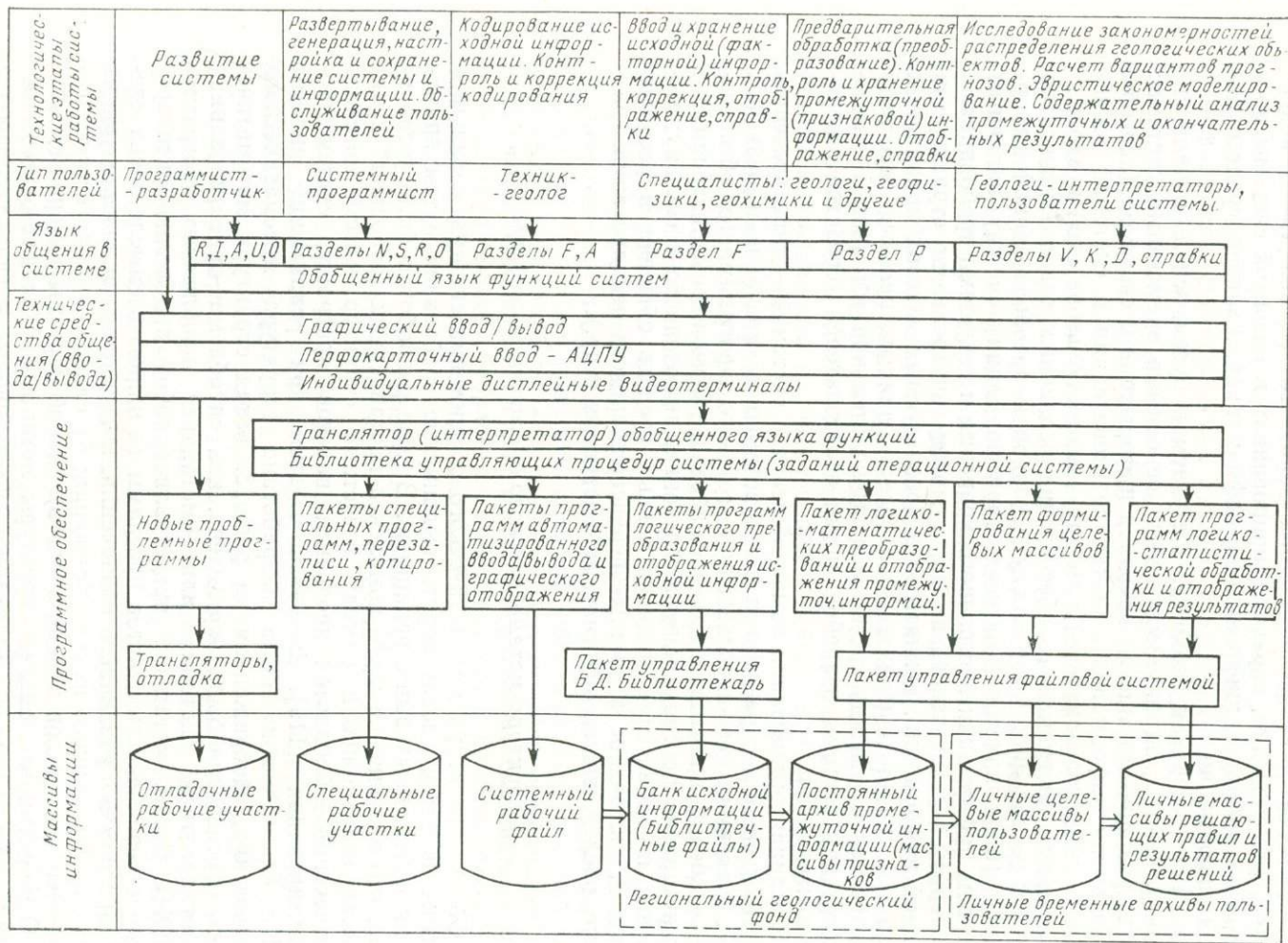


Рис. 13. Принципиальная схема структуры программного обеспечения АИПС/ЕС «Регион» (мультипакетная система коллективного пользования с обобщенным языком функций и интегрированной структурой информации)

В число программ, используемых системой «Регион», в настоящее время (включая как пакеты и программы, разработанные специально для нее, так и готовые пакеты программ, включенные в ее состав, а также пакеты и программы, входящие в состав операционной системы) входят следующие пакеты: программ копирования, программ перезаписи, программ автоматизированного кодирования и графического изображения исходных карт, программ логических преобразований и печати вариантов прогнозных и классификационных карт, программ библиотечного метода доступа, трансформации исходной информации и печати карт синтезированной информации, программ образования целевых массивов, программ логико-статистической обработки и анализа, распознавания образов, классификации, таксономического анализа, эвристического моделирования и др.

Технологический набор основных информационных массивов приведен на рис. 13. Набор таких массивов состоит из двух частей: общесистемных массивов постоянного хранения (машинного геологического фонда) и индивидуальных массивов временного хранения (временные личные массивы пользователей). В состав машинного геологического фонда входят банк исходной картографической, числовой и текстовой информации, реализованный в виде совокупности библиотечных файлов, и постоянный архив трансформированной информации, состоящий из массива признаков и библиотеки паспортов и правил.

Банк исходной информации сочетает в себе свойства документальной и фактографической систем. Структура хранения фактографического материала в банке и архиве сочетает в себе иерархический и матричный принципы.

В состав временных архивов пользователей входят: целевые массивы (массив признаков и массив выборок), массив решающих правил и массив результатов обработки (массив классификационных и прогнозных карт).

Система интерфейсов состоит из двух типов: внешних и внутренних. К внешним интерфейсам относятся интерфейс между структурами представления различных типов исходной информации вне и внутри системы, интерфейс между специальными языками пользователей и системой и интерфейс между составом программного обеспечения (ПО) системы и составом оборудования ЭВМ и требованиями пользователей. Интерфейс между структурами исходной информации осуществляется посредством способов и программ кодирования исходной информации. Способы кодирования картографической информации и числовых полей, принятые в системе АИПС/ЕС «Регион», разработаны специально для нее. Оба способа отличаются минимумом необходимой предварительной обработки исходной информации, безыбыточностью и компактностью ввода и хранения, ориентацией на применение средств автоматического (или автоматизированного) кодирования при полной совместимости и простоте ручного кодирования, возможностью детального автоматического контроля и диагностики вводимой информации.

Что касается сути разработанных способов кодирования, то они сводятся к следующему: картографические факторы записываются координатами пересечений контуров фактора при построчной развертке по каждому листу. Числовые факторы записываются целыми или десятичными числами произвольной размерности по каждой ячейке принятого размера. Кроме того, каждый фактор снабжается управляющей информацией, так называемым «паспортом», содержащим шифр, наименование, масштаб, схему изученности и другие данные.

Особо отметим, что способ кодирования картографической информации рассчитан на применение автоматического кодирования типа сортирующего барабана или оптического устройства на основе метода rasterной дискретизации.

Кодирование вводимой текстовой информации практически не накладывает ограничений, за исключением ограничения знаков в строке вводимой информации, принятого в библиотечных файлах.

Язык общения с системой (язык пользователей) реализован в виде так называемого «обобщенного языка функций» (ОЯФ), т. е. единого языка, организованного разделами в соответствии с типами пользователей. Язык неалгоритмического типа, функционально-параметрический. Параметры описываются, как правило, в ключевом формате. Разделы языка близки к усеченным естественным языкам соответствующих типов пользователей. Структура разделов усеченного языка пользователей приведена в табл. 11.

Близость ОЯФ к естественному позволяет специалистам пользователям работать непосредственно с системой; простота и компактность языка позволяют пользователям осуществлять работу с индивидуальными консольными устройствами (интерактивный режим). Наличие этих свойств подчеркивает человеко-машинный характер системы, т. е. функционирование системы в качестве инструмента обработки информации, постоянно используемого специалистами в своей работе. Системы подобного типа также иногда называют «советующими», т. е. в процессе взаимодействия человека и ЭВМ вырабатываются серии вариантов готовых решений, из которых специалистами выбираются наилучшие.

Еще одним внешним интерфейсом системы являются средства генерации и настройки. Указанные средства представляют собой программы и функции системы, позволяющие согласовывать состав системы и особенности входящего ПО с составом оборудования и составом и требованиями пользователей. К внутренним интерфейсам системы относятся: транслятор обобщенного языка функций, блок пакетов программ трансформации, пакет программ образования целевых массивов, средства генерации и настройки. Главной особенностью и смыслом наличия в системе таких интерфейсов является возможность дальнейшего развития и наращивания системы. Структура ОЯФ позволяет легко увеличивать его элементы от новых параметров функций до новых разделов языка (групп функций). Структура построения транслятора ОЯФ позволяет достаточно легко реагировать как на изменения самого языка, так и на изменения, связанные с включением в состав системы нового программного обеспечения. Это во многом определяет возможность использования нового, в том числе готового, программного обеспечения сложных структур. Блок пакетов программ трансформации удовлетворяет индивидуальные подходы пользователей к проблеме использования исходной информации в процессе геологического анализа информации, ускоряет получение конечных результатов.

Пакет программ формирования целевых массивов позволяет удовлетворять требования к структуре входной информации различных обрабатывающих программ с требованиями к структуре хранения информации одновременного пользования. Это обстоятельство является весьма важным, учитывая возможность использования нового программного обеспечения. Кроме того, введение механизма целевых массивов допускает выполнение двух существенных требований к удобству поль-

Структура обобщенного языка функций (ОЯФ) системы АИПС/ЕС «Регион»

Группы функций системы				Шифр раздела	Тип пользователей
Развитие (52)	Служеб- ные (66)	Развитие (52)	Работа с исход- ной и объект- ной библиоте- ками программ (33)	I, J, A	Программист- разработчик
			Развитие (13)	R, U	
			Отладка (6)	O	
	Обслуживание (14)	Настройка (11)	N	Системный програм- мист	
		Сохранение (3)	S		
	Эксплуата- ция (50)	Основной техноло- гии (36)	Накопление ин- формации, справки (фор- мирование бан- ков данных) (19)	Работа с исход- ной информа- цией (10)	F
Преобразование информации (9)				P	Специалисты (геологи, геофизики, геохимики и др.)
Решение задач геологическо- го прогнозиро- вания (17)			Составление выборок (3)	V	Геолог-интер- претатор, пользо- ватель систе- мы
			Обработка и классификация, анализ (14)	K	

Примечание. Числа в скобках обозначают количество функций, имеющихся в системе на 1978 г.

зования системой — автономности процессов накопления и использования информации и независимости работы пользователей друг от друга. Наличие преобразованной информации позволяет существенно сократить время решения содержательных задач, т. е. повысить оперативность работы системы с конечными пользователями.

Использование готового программного обеспечения за счет применения языкового и информационного интерфейсов улучшает характеристики системы в части надежности, простоты обслуживания, повышения коэффициента использования программного обеспечения, повышает устойчивость системы к моральному старению, уменьшает трудоемкость дальнейшего развития и облегчает совместимость ее с другими системами.

*Общие характеристики
математического обеспечения системы*

Существующая версия программного обеспечения системы АИПС/ЕС «Регион» работает под управлением и с использованием дисковой операционной системы ДОС/ЕС, начиная с версии 2.2.

Математическое обеспечение АИПС/ЕС «Регион» состоит из общесистемного и специального МО.

Специальное программное обеспечение используется при выполнении более половины всех функций системы. Остальные функции осуществляются с помощью готового программного обеспечения, в том числе входящего в состав операционной системы ЕС. Структура организации специального программного обеспечения соответствует организации типа конструктора, т. е. библиотеке наборов специализированных модулей, удовлетворяющих стандартным соглашениям о стыковке. Структура сборки программ простая, без перекрытия. Специальное программное обеспечение системы АИПС/ЕС «Регион» состоит из 5 библиотек программ, контрольного примера (3 библиотеки и 1 файл данных) и примера для демонстрации и обучения работе с системой (1 библиотека и 4 файла данных). Языки программирования — ассемблер и фортран-IV. Объемные характеристики программ приведены в табл. 12.

Физически все программное обеспечение системы находится на так называемом «субрезидентном» диске, т. е. диске, находящемся на вычислительной установке во время работы с системой. Кроме резидентного диска в работе системы используются также так называемые «информационные» диски и ленты для копирования обоих типов дисков.

Вместе с тем высокий уровень использования готового ПО влечет за собой и один из наиболее заметных недостатков, присущих разработанной структуре — некоторую избыточность ПО, включенного в систему и связанное с этим некоторое увеличение времени обработки данных. Представляется, однако, что этот недостаток в значительной степени компенсируется возможностью использования включенного ПО в других системах, а также полученными при этом преимуществами, более существенными для практического применения ЭВМ при решении задач геологического прогнозирования.

Таблица 12

**Объемные характеристики программ
системы АИПС/ЕС «Регион»**

Тип библиотеки	Число модулей	Размер участка (дорожек диска)
Абсолютная	13	90
Объектная	29	90
Исходная	33	150
Управляющих процедур	102	50
Сервисная	11	150
Всего	188	530

Выполнение функций системы инициируется посредством операторов обобщенного языка функций (ОЯФ): служебных, управляющих и функциональных.

В табл. 13 в качестве примера приводятся перечень одной группы функций, используемых в системе при решении задач классификации, статистической обработки, а также имена соответствующих функциональных аппаратов, входящих в один из разделов ОЯФ.

Выполнение функций системы инициируется посредством операторов ОЯФ служебных, управляющих и функциональных.

Операторы ОЯФ могут передаваться в систему либо непосредственно через индивидуальное терминальное устройство, либо посредством так называемых «управляющих процедур» — через перфокарточное считывающее устройство. Управляющая процедура представляет собой сравнительно небольшой набор перфокарт, фактический или записанный в библиотеку управляющих процедур (БУП), состоящий из операторов управляющей программы операционной системы и операторов ОЯФ. Управляющая процедура является заданием операционной системе и работающей под ее управлением системы АИПС/ЕС на выполнение некоторой работы. Результатом выполнения такой работы является выполнение какой-либо функции АИПС или последовательности функций.

В АИПС существуют управляющие процедуры двух типов: универсальные (многофункциональные) и функциональные одноименные, соответствующие функциональным операторам. Посредством универсальной процедуры можно передавать любые операторы и соответственно выполнять любые функции, аналогично работе с терминала. Функциональные процедуры могут передавать только некоторые конкретные виды операторов и выполнять только соответствующие функции. Применение функциональных управляющих процедур предусмотрено, как правило, только для обслуживающего персонала системы в случаях ввода и коррекции информации непосредственно через перфокарточное считывающее устройство при отладочной работе.

Наличие в системе широкого спектра функций позволяет пользователям и обслуживающему персоналу проводить обработку исходной информации по различным технологическим схемам. Решение любой задачи, в том числе информационно-справочной или содержательной, состоит в последовательном выполнении ряда функций. Технологические программные модули решения существуют на всех этапах работы системы, в том числе и на этапе содержательного решения задач, где набор таких модулей весьма велик.

В табл. 14 в качестве примера приведена одна из простейших схем обработки информации — таксономическая классификация выборки.

Ниже приводится также пример записи некоторого контрольного варианта выполнения такой схемы решения с помощью функциональных операторов ОЯФ.

§§ CELEVOJ	PRIZNAKI=3, 7—12, LISTY: 1—4, 6
§§ VYBORKA	L1, S 3, T 7, 8, 10—12, S 4, T 7—9, 11* T 20—24, 26—29; L 1, S 3, T 9—13
§§ TAXONV	RADMIN=1,0, SH4G=0,1 * RADMAX=2,0 * CHTAXMIN=2

Перечень функций классификации и статистической обработки

№ п/п	Функции	Операторы ОЯФ (и управляющих процедур)
Основные функции		
1	Построить гистограмму выборки (целевого массива)	GTGRAMMA
2	Построить решающее правило классификации по выборке обучения (обучение распознаванию „с поощрением“)	GPPLAST
3	Классифицировать целевой массив по заданному решающему правилу и распечатать классификационную карту	OCENKA
4	Ввести (записать) в систему эвристическое (имитационное) решающее правило специалиста для последующей классификации и распечатать его текст	SAPISAPR
5	Рассчитать дисперсию и математическое значение признаков целевого массива и провести его таксономию	TAXON
6	Распечатать карту таксономической классификации по заданному целевому массиву	KARTATAX
7	Провести поиск аналогов по заданным точкам (распознавание „без поощрения“) в целевом массиве и распечатать карту результатов	TAXONE
8	Провести поиск аналогов по заданным точкам (распознавание „без поощрения“) по квантованному целевому массиву	TAXONK
9	Рассчитать дисперсию и математическое ожидание выборки и провести ее таксономическую классификацию	TAXONV
10	То же, что п. 7, с „внешним“ эталоном	TAXONA
11	То же, по квантованному массиву	TAXONC
Сервисные функции		
1	Записать (каталогизировать) в системе эвристическое решающее правило	KATALZP
2	Распечатать список каталогизируемых решающих правил	OGLAVP
3	Распечатать текст каталогизируемого решающего правила	TEXTRP
4	Распечатать текст решающего правила, подготовленного для классификации	TEXTRPR
5	Записать (каталогизировать) решающее правило с рабочего участка диска	KATALZRP
6	Выбрать решающее правило из библиотеки, записать на рабочий участок диска (подготовить к работе) и распечатать его текст	VYBORKA

Таблица 14

Технологический программный модуль обработки информации при таксономической классификации произвольной выборки

Функции	Операторы функций
Образовать целевой массив Образовать выборку Выполнить процедуру таксономии выборки	CELEVOJ VYBORKA TAXONV

где §§ — символы операторов ОЯФ; * — знаки продолжения операторов; L, S, T — символы листов, строк, точек целевого массива. Остальные параметры являются латинской транскрипцией русских слов и сокращений (RADMIN — «минимальный радиус» и т. д.).

*Временные характеристики работы
основных модулей АИПС/ЕС*

Автоматизированная информационно-прогнозирующая система АИПС/ЕС позволяет выполнять расчеты и обеспечивает получение результативных карт в оперативном режиме. Это достигается отработан-

Таблица 15

Нормы машинного времени на работу отдельных программ обработки данных в АИПС/ЕС

Виды работ	Единица измерения	Минимальное время обработки, мин	Максимальное время обработки, мин	Среднее время обработки, мин
Обработка листов карты, закодированных на приборе „Шифратор-2“	1 лист-фактор	0,1 (пустой лист)	2,0 (сильно загруженный лист)	0,5
Расчет массива признаков	1 признак по 1 листу карты	1	1	1
Образование целевого массива	По 30 признакам на 20 листах карты района	9	11	10
Образование выборки	По 30 признакам для 1000 точек	2	4	3
Получение гистограмм распределений признаков	То же	14	25	20
Получение решающего правила	„ „	15	25	20
Классификация	По 30 признакам на 20 листах карты района	5	9	7

ной технологией преобразования информации на основе различных методических модулей, а также хорошими временными характеристиками обработки данных. Как и у всякой системы, технологичность определяется комплексом взаимосвязанных факторов, в частности объемом обрабатываемой информации и временными параметрами работы отдельных программ системы. Знание последних позволяет оптимально планировать процесс прохождения информации в системе.

В табл. 15 приведены нормы времени основных обрабатываемых программ АИПС/ЕС.

В ней не указаны затраты времени на подготовку первичных данных (кодирование карт факторов на приборе «Шифратор-2») и ввод полученной перфоленты программой GS/VVDTP.

Необходимо иметь в виду, что данные, приведенные в табл. 15, получены по результатам решения практической задачи на примере конкретного региона.

3. ТЕХНОЛОГИЯ ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ В СИСТЕМЕ АИПС/ЕС «РЕГИОН»

Система АИПС/ЕС «Регион» представляет собой комплекс современных методов, позволяющих реализовать автоматизированную обработку различной геологической, геофизической, геохимической и прочей информации и комплексный ее анализ при составлении прогнозных заключений и оценке перспективности геологических объектов. Система предусматривает формирование территориального банка данных, в которой вводится исходная информация, снимаемая с различных карт, планов схем и числовых полей исследуемой территории и характеризующая особенности геологического строения исследуемой территории (в том числе физических полей).

Решение содержательных задач производится по информации, сохраняемой в территориальном банке данных на основе использования различных методических схем обработки исходного материала. Характерной особенностью процесса решения является непосредственное участие в нем специалиста-пользователя системы, который в процессе решения конкретной задачи геологического прогнозирования анализирует промежуточные результаты и в «диалоговом» режиме вносит необходимые коррективы, не допуская, например, «абстрактных» или «абсурдных» результатов. Смысл такого диалога состоит также и в том, что специалист, постепенно осваивая возможности ЭВМ, получает инструмент, резко увеличивающий возможности формально-логического анализа исходных данных, т. е. повышающий «производительность» творческого труда исследователя.

Диалоговые процедуры в АИПС/ЕС реализованы на достаточно простом уровне, вполне доступном для применения на любой ЕС ЭВМ. В то же время используемые принципы при дальнейшем развитии позволяют использовать различные терминальные устройства, в том числе и аппаратуру визуального контроля.

Как следует из принципов построения АИПС/ЕС, технология человеко-машинного взаимодействия разделяется в ней на два этапа:

- 1) этап формирования территориального банка данных;
- 2) этап решения задач.

Ниже приводится описание последовательности выполнения различных процедур при формировании территориального банка данных и

решении задач геологического прогнозирования на основе АИПС/ЕС и раскрывается содержание каждой процедуры.

Реализация всех процедур рассчитана на следующий состав исполнителей:

— ведущий специалист — пользователь системы (геолог-интерпретатор);

— группа экспертов (геологов, геофизиков и др.), к которым ведущий специалист обращается при необходимости для получения экспертных оценок по какому-либо вопросу. При этом для составления согласованного заключения может быть использован, например, метод дифференциального ранжирования [85];

— группа техников (коллекторов) во главе со старшим техником (2—4 чел.), занятых на вспомогательных процедурах обработки информации и подчиненная ведущему специалисту;

— обслуживающий персонал вычислительного центра, выполняющий необходимые операции по машинной обработке данных на основе имеющегося математического обеспечения системы.

Ниже будут также приведены примерные затраты труда на реализацию отдельных операций. На рис. 14 показана схема человеко-машинного взаимодействия в процессе формирования территориального машинного фонда картографической информации (банка данных) и решения задач прогноза и оценки геологических ситуаций в АИПС/ЕС «Регион». Решение задачи (II этап) показано на примере использования алгоритма распознавания образов, для чего применяется своя технологическая схема (процедуры 4—11 на рис. 14). При использовании другого методического подхода (например, эвристического моделирования и т. п.) технология взаимодействия будет иной.

Технология человеко-машинного взаимодействия при формировании территориального банка картографической информации

Процедуры формирования территориального машинного фонда выполняются группой сотрудников геологической организации (экспедиции, управления). В задачу этой группы входят создание и оперативное ведение территориального машинного фонда, доступного для специалистов геологов — пользователей системы, и составление каталогов. Все материалы располагаются в памяти ЭВМ в соответствии с картографической схемой деления территории на стандартные листы. Заключение о качестве и представительности исходных картографических материалов, а также рекомендации о внесении их в территориальный машинный фонд могут выдаваться научно-техническим советом, рассматривавшим и утверждавшим конкретный отчет по проведенным полевым и камеральным работам.

Предварительная обработка исходной информации включает процедуры переноса данных на машинные носители.

Процедуры формирования банка данных завершаются занесением в память ЭВМ в определенной последовательности исходной и преобразованной информации и составлением каталога, который является исходным материалом для пользователей — специалистов геологов. Каталог содержит необходимые сведения о всей информации имеющейся в банке данных (масштаб, площадь, точность проведенных исследований, непосредственно результаты этих исследований и т. д.). На основе каталога могут быть сформированы сведения о изученности территории, района, отдельного листа.

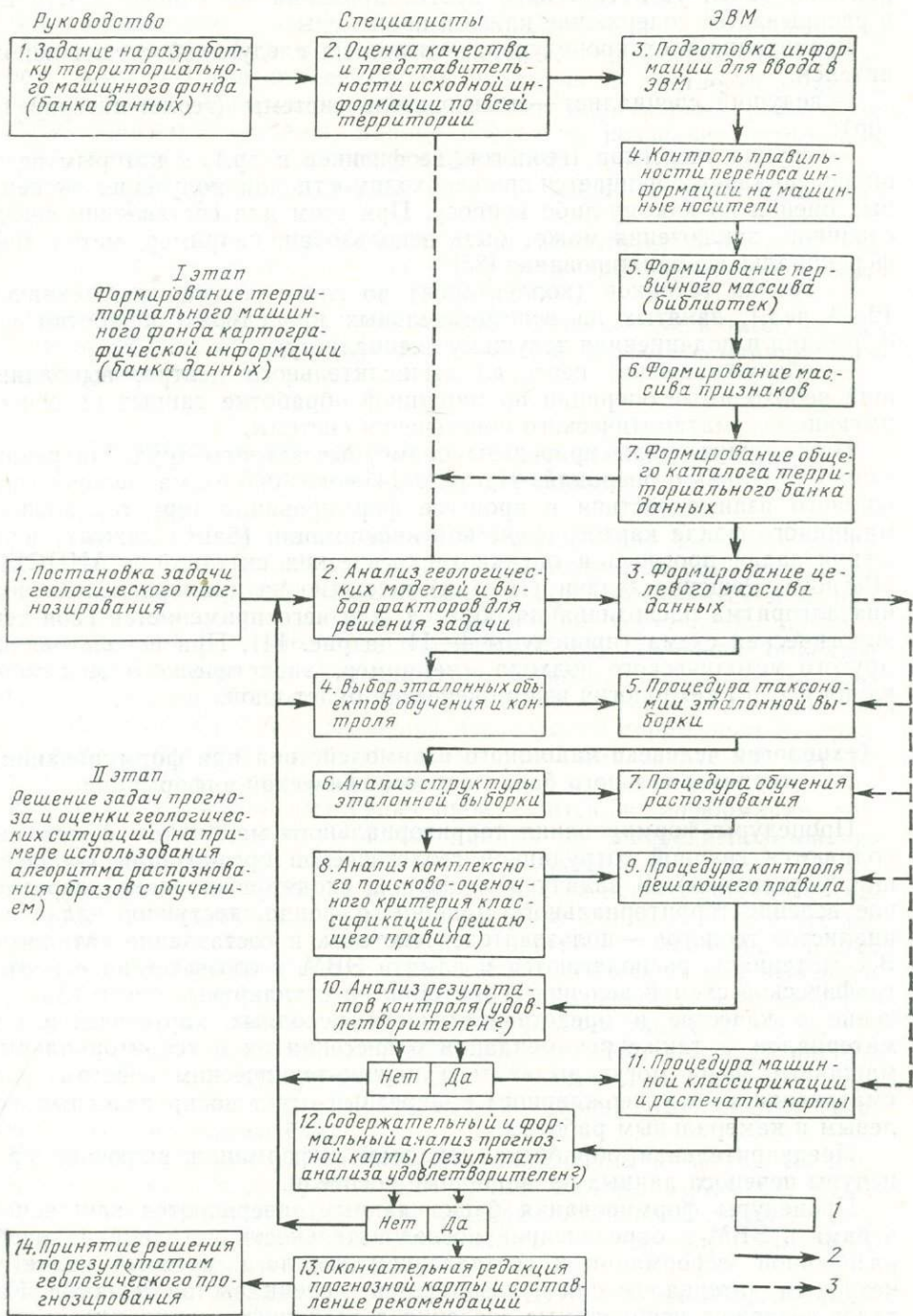


Рис. 14. Принципиальная схема процедур человеко-машинного взаимодействия в процессе формирования банка данных и решения прогнозных задач в АИПС/ЕС «Регион». 1 — процедуры и их номера, 2 — последовательность выполнения процедур, 3 — передача информации

Эти процедуры включают подготовку и экспертный анализ исходной геологической, геофизической, геохимической и прочей картографической информации.

Подготовка и анализ исходной информации. Геологическое прогнозирование осуществляется на основе имеющейся исходной информации, поэтому первым этапом разработки прогнозных заключений является анализ ведущим специалистом геологической изученности исследуемой территории. Под геологической изученностью в данном случае понимается наличие выполненных геологических, геофизических, геохимических, гидрогеологических и других исследований. Такой анализ предполагает составление сводки фактически проведенных в разные годы работ по отдельным районам, методам, масштабам.

Известно, что любые прогнозные заключения должны основываться на надежном и представительном исходном материале. Главное назначение описываемой процедуры — последовательный анализ имеющейся информации с целью составления ведущим специалистом заключения о ее надежности и представительности. Оценка качества геологических исследований осуществляется путем ознакомления с фактическим материалом и заключениями экспертов. Материалы геофизических исследований могут быть охарактеризованы качеством полевых работ (ошибки наблюдений, точность привязки и т. д.). Качество геохимических работ можно оценить точностью анализов и т. д. Следует подчеркнуть, что в машинный банк могут заноситься не только, например, исходные геофизические поля, но и схемы комплексной интерпретации геофизических данных, отражающие особенности глубинного строения территории.

Другим показателем достоверности исходной информации является ее представительность (равномерность проведенных исследований по всей изучаемой территории). Характеристику представительности трудно дать путем ознакомления с картами фактического материала.

В случае необходимости могут быть введены математические оценки тем или иным методом степени изученности отдельных частей рассматриваемой территории. В качестве аппарата для такой оценки можно, например, использовать метод статистических испытаний (Монте-Карло) [53].

В результате выполнения данной процедуры для дальнейших построений должна быть оставлена только надежная и достаточно представительная информация. Таким образом, процедура оценки качества и достоверности имеющейся информации является своего рода экспертным фильтром, с помощью которого отбраковывается недостаточно надежная информация.

Предварительная обработка информации заключается в составлении картографической схемы деления исследуемого региона на условные листы и некоторые процедуры, обеспечивающие последующий перевод информации на магнитные носители.

Базовый размер условного листа картографической информации, сохраняемого в машинной памяти АИПС/ЕС, равен 32×32 см. Такой размер обеспечивает компактное использование памяти при хранении картографических данных. Выбираемый пользователем размер листа должен быть кратен базовому либо укладываться в него целое число раз.

В связи с этим требованием вся территория, по которой формиру-

ется банк данных, условно подразделяется на несколько листов квадратной формы. Эта процедура выполняется техником-топографом (картографом) под руководством ведущего специалиста.

Эта работа необходима для того, чтобы «привязать» фактические картографические данные в памяти ЭВМ к реальной топографической основе. В памяти ЭВМ листы сохраняются в виде «страниц» квадратной формы, в отличие от трапециевидальной картографической основы. Поэтому схема деления района на отдельные листы является основой для последующего переноса результатов расчетов с помощью ЭВМ на фактическую топооснову.

Каждый условный лист сохраняется в памяти ЭВМ в виде отдельных ячеек (клеток). Совокупность исходной информации по одному показателю на всех листах исследуемой территории называется фактором. Исходная информация вводится и хранится в машинном банке данных по факторам, внутри которых сохраняется разделение на листы, строки и ячейки (точки). Информацию, сохраняемую в ЭВМ по одному фактору в пределах одного листа, будем называть «страницей».

Размер и число ячеек (точек съема информации) определяются размером листа и видом фактора. В системе различается два вида факторов; картографические (двузначные) и числовые (многозначные). К картографическим относятся факторы, которые могут принимать на всей совокупности точек исследуемой территории только два значения (т. е. имеющие только два состояния или две градации измерения). Для кодирования таких факторов достаточно двух чисел (например, 0 и 1). В содержательном плане к этим факторам относят такие, которые отражают лишь наличие или отсутствие проявлений фактора в пределах данной точки исследуемой территории (например, распространения геологических формаций, интрузивных пород, линии разломов, тектонических нарушений, элементов схем интерпретации геофизических данных и т. п.).

К числовым (многозначным) факторам относятся такие, которые могут принимать три и более значений (имеют три или более состояний или градаций измерения). Кодирование таких факторов не может быть осуществлено двумя элементарными двоичными числами и производится десятичными числами (целыми или дробными) без ограничения разрядности. В содержательном плане эти факторы отражают значение проявлений фактора (в том числе потенциальные геофизические поля, геохимические измерения и др.). К таким числовым факторам относится также, например, количество каких-либо проявлений в пределах единицы площади (количество обнажений, скважин и т. п.).

Картографические (двузначные) факторы снимаются по сети 64×64 точки в листе; числовые — по сети 32×32 точки в листе. При размере листа 32×32 см размеры ячеек будут 5×5 и 10×10 мм соответственно.

В случае отсутствия информации единого масштаба в пределах исследуемой территории, а также для повышения точности съема данных снятие информации может производиться на картах иного масштаба, отличного от принятого по данной территории с пропорциональным изменением размера листов и ячеек.

Предварительная обработка информации заканчивается составлением первичного каталога, т. е. перечня картографической информации (карт, схем интерпретации и т. д.) с указанием состава факторов, отобранных для перевода на машинные носители и последующего хранения в банке данных.

Процедуры формирования территориального банка данных

Эти процедуры включают перевод факторов на машинные носители, формирование первичных массивов исходной информации в ЭВМ, формирование признаков и каталога машинного банка данных.

Перевод картографической информации на машинные носители может осуществляться как с использованием полуавтомата «Шифратор-2», с помощью которого обводятся площади распространения факторов, так и вручную, построчным кодированием площадей распространения факторов на отдельных листах на основе бланков ассемблера.

Принципиально возможен также ввод информации с помощью автоматических устройств. Главную трудность при этом представляет разделение карты на отдельные факторы.

Не останавливаясь подробно на особенностях ввода картографической информации в ЭВМ, заметим, что конечным итогом этой процедуры является перевод исходных данных на машинный носитель (перфоленга, перфокарта, магнитная лента).

Прежде чем приступить к формированию первичного массива, необходимо проверить правильность перенесения информации. В АИПС/ЕС для этой цели используется визуальный контроль. Он состоит в том, что после некоторого машинного преобразования результаты кодирования исходных данных распечатываются (пофакторно и полистно) в виде, который может быть сопоставлен с оригиналом.

Визуальное сопоставление материалов исходной карты и машинных распечаток позволяет выявить ошибки кодирования и быстро устранить их. Откорректированная информация с помощью специальных программ записывается на магнитный диск, при этом происходит формирование библиотек исходных факторов или первичного массива.

Прежде чем приступить к процедуре формирования массива признаков, ведущему специалисту необходимо определить так называемый «радиус влияния факторов». Данная процедура является вспомогательной и предназначается для рационального использования машинной памяти. Суть ее заключается в том, что ведущий специалист (возможно с привлечением экспертов) определяет для каждого из картографических факторов область его «влияния» (R), выраженную в расстояниях от его экзоконтакта. Поясним это примером. Допустим, исследуется район размером 300×300 км в масштабе $1 : 100\,000$. В качестве одного из важных факторов для решения конкретной прогнозной задачи определены гранитные интрузии конкретного типа и определенного возраста. Максимальные размеры таких интрузий в пределах исследуемой территории 10×10 км. Требуется установить, исходя из содержательных соображений на основе геологического анализа, максимальное расстояние, на котором интрузии этого типа могут влиять на процессы рудообразования. Допустим, что такое расстояние определено в $15-20$ км. При этом лучше использовать максимальную цифру. Эта цифра является ограничением для ЭВМ при расчете в дальнейшем целевого массива банка данных. Кроме того, задается необходимая точность привязки исходных данных на карте (линий, контактов, контуров) в сантиметрах (миллиметрах).

После получения этих сведений оказывается возможным формирование массива признаков и библиотеки паспортов (машинного ката-

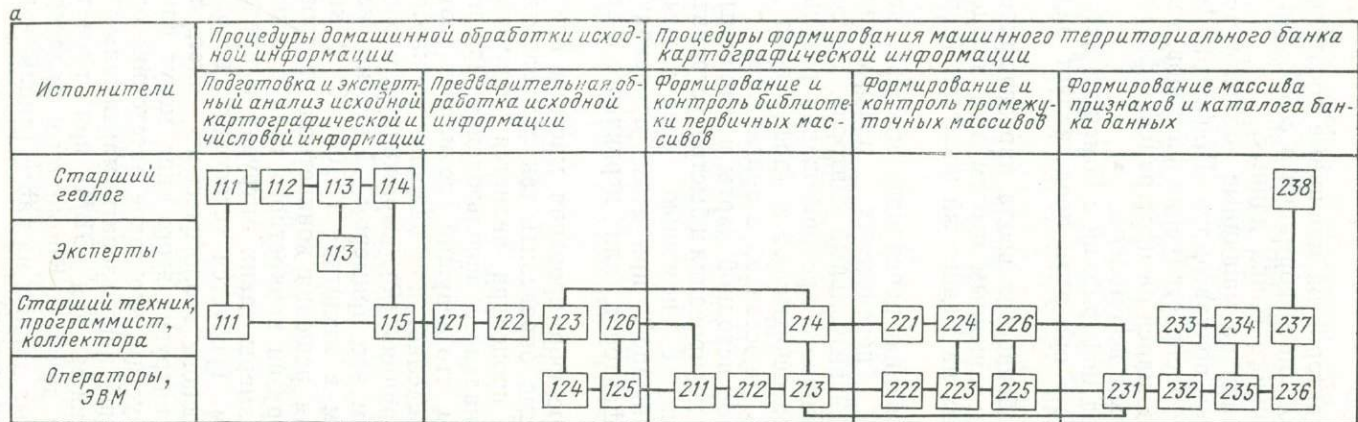


Рис. 15. Технологические схемы человеко-машинного взаимодействия в АИПС/ЕС «Регион».

а — формирование территориального машинного банка данных: 111 — составление картографической схемы деления территории на листы; 112 — анализ геологической, геофизической и прочей изученности территории; 113 — анализ качества и представительности имеющейся исходной информации; 114 — составление списка факторов и таблиц областей их распространения на листах картографической схемы территории; 115 — расчет участков для записи исходной информации на магнитных дисках ЕС ЭВМ; 121 — составление карт-факторов; 122 — составление фактор-листов (перенесение контуров на бланки — графареты); 123 — перенос информации с фактор-листов; 124 — набивка информации на перфокарты или перфоленты; 125 — контроль качества перфорации; 126 — составление задания для записи факторов в библиотеки (первичные массивы); 211 — подготовка системы к работе (восстановление дисков с лент и их разметка); 212 — формирование библиотеки (запись факторов в первичный массив); 213 — вывод на печать картографических факторов по территории, а также другой необходимой информации для последующего контроля и справок; 214 — контроль введенных в библиотеку первичных материалов, исправление ошибок и составление задания на формирование массивов признаков; 221 — составление задания на запись факторов в промежуточный массив и формирование каталога; 222 — запись факторов на диски (формирование промежуточного массива); 223 — вывод на печать записанной информации для последующего контроля и справок; 224 — контроль фактор-листов и исправление ошибок; 225 — коррекция введенной в промежуточный массив информации; 226 — составление задания на формирование массива признаков и каталога банка данных; 231 — формирование массивов признаков (образование и запись признаков на магнитный диск ЕС ЭВМ); 232 — вывод на печать значений признаков по листам и другой необходимой информации для последующего контроля и справок; 233 — контроль признак-листов; 234 — составление задания на печать каталога первичной и преобразованной информации территориального банка данных; 235 и 236 — соответственно формирование и вывод на печать каталога банка данных; 237 — контроль каталога территориального банка данных; 238 — передача каталога в территориальный геологический фонд для практического использования;

б — решение задачи геологического прогнозирования (на основе алгоритма распознавания): 111 — постановка за-

дачи геологического прогнозирования; 112 — ознакомление с исходной информацией по каталогу территориального банка данных; 113 — анализ качества и представительности имеющейся информации; 114 — оценка достаточности имеющейся информации для решения поставленной задачи; 121 — анализ имеющихся геологических гипотез (моделей); 131 — формирование модели прогнозируемого объекта и выбор факторов; 132 — определение радиуса влияния факторов; 211 — составление заявки на формирование целевого массива по требуемому району; 212 — составление задания на формирование целевого массива по требуемому району; 221 — выбор эталонных объектов для обучения распознаванию и контрольной классификации; 222 — составление заявки на формирование выборок обучения и контроля; 223 — составление задания на формирование выборок обучения и контроля; 224 — формирование выборок обучения и контроля; 225 — процедура таксономии по выборкам обучения; 226 — содержательный анализ структуры (компактности) выборки обучения; 227 — построение гистограмм одномерных распределений по выборке обучения (для каждого признака); 228 — содержательный анализ гистограмм одномерных распределений и выбор границ интервалов значений признаков; 229 — задание границ интервалов значений признаков; 2210 — построение решающего правила (комплексного критерия классификации); 2211 — содержательный анализ решающего правила (оценка информационной значимости отдельных интервалов признаков, определенной на ЭВМ); 2212 — процедура машинного контроля решающего правила на контрольной выборке; 2213 — содержательный анализ результатов контроля; 2214 — процедура машинной классификации целевого массива и построение прогнозной карты по исследуемому району; 311 — предварительный анализ результатов классификации и составление задания на выбор справочной информации по новым перспективным участкам из целевого массива; 312 — составление задания на выбор требуемой информации; 313 — выбор из целевого массива требуемой информации; 314 — содержательный анализ результатов классификации новых перспективных участков на основе оценки информативных вкладов признаков; 315 — перенос результатов классификации на геологическую основу; 321 — содержательный анализ геологической позиции перспективных площадей; 322 — составление отчета о результатах прогнозирования и рекомендаций по проведению поисково-оценочных работ; 323 и 324 — соответственно рассмотрение и утверждение отчета

лога) территориального банка данных. Этими процедурами завершается технология формирования банка. Сохраняемая в нем информация может использоваться для решения различных задач. В то же время служба ведения территориального машинного банка картографической информации должна оперативно пополнять его новыми данными, вносить уточнения в соответствии с изменениями исходных материалов и оперативно вести каталог.

На рис. 15 приведена полная технологическая схема человеко-машинного взаимодействия при формировании территориального банка картографических данных.

Таким образом, выполнение всех процедур в указанной технологической последовательности на основе взаимодействия специалистов с ЭВМ позволяет сформировать и оперативно вести территориальный банк картографических данных.

Видимо, целесообразно осуществлять такое ведение банка данных специальной группой сотрудников, входящей в состав территориальных фондов геологических организаций. Это, с одной стороны, открывает перспективу постепенного перехода хранения картографической информации на машинных носителях, а с другой, — дает возможность специалистам (геологам, геофизикам и др.) осуществлять исследования и оперативный анализ этой информации с помощью ЭВМ при решении различных задач прикладной геологии.

Технология человеко-машинного взаимодействия при решении задач

Технология человеко-машинного взаимодействия при решении задач прогноза и анализа рудоносности с использованием АИПС/ЕС разделяется на три подэтапа:

1) подготовительные процедуры домашней обработки информации;

2) процедуры взаимодействия с ЭВМ в процессе решения задачи;

3) процедуры содержательного анализа полученных результатов.

Все эти процедуры выполняются непосредственно геологом — пользователем системы.

Подготовительные процедуры домашней обработки информации

Эти процедуры включают постановку (формулирование) прогнозной или исследовательской задачи, оценку достаточности информации для ее решения, анализ имеющихся геологических моделей рудообразования (генетических концепций) и формирование рабочей гипотезы.

Под постановкой или формулированием задачи понимается постановка задания на выполнение работы по геологическому прогнозированию либо методическому исследованию. Такая постановка дается руководителем предприятия или коллективным органом — научно-техническим советом и может иметь, например, следующий вид: «выделить в данном районе площади, наиболее перспективные для обнаружения промышленных месторождений конкретного типа, с целью проведения на них поисково-съёмочных работ» или «определить участки, наиболее перспективные для выявления промышленного оруденения конкретного типа с целью проведения оценочных работ» и т. п. Точно так же, как и при традиционном методе геологического прогнозирования, пос-

ле получения задания ведущий специалист проводит его содержательный анализ. При этом определяются вероятные типы прогнозируемых объектов, сравниваются имеющиеся гипотезы (модели) формирования рудных объектов, устанавливаются реальные ограничения геологического прогноза (по размерам объектов, по территории и т. п.).

Естественно, что для успешного решения поставленной задачи прежде всего следует ознакомиться с каталогом машинного банка данных и оценить принципиальную возможность решения на основе имеющейся в банке информации. Эта процедура также выполняется ведущим специалистом (геологом-интерпретатором), при этом он может использовать аппарат экспертных оценок.

Следует сразу же оговориться, что понятие «оценка достаточности имеющейся информации для решения поставленной задачи геологического прогнозирования» является в значительной мере условным. В самом деле, прогноз представляет собой разработку рекомендаций в условиях информационной недостаточности. Поэтому при реализации данной процедуры важно установить необходимость использования дополнительных материалов, применение которых наряду с имеющимися данными позволит успешно решить поставленную задачу. Если, по мнению пользователя, имеющихся материалов недостаточно, то это отодвигает сроки выдачи прогнозного заключения. В случае, если ведущий специалист (или группа экспертов) признает, что имеющихся материалов достаточно для составления требуемого прогнозного заключения, можно переходить к следующей процедуре.

Анализ имеющихся геологических гипотез. Геологический прогноз является реализацией определенных теоретических представлений, взглядов, гипотез на природу и процесс рудообразования. От того, насколько эти представления соответствуют реальной действительности, в конечном счете и зависит качество прогнозных оценок. Во многих случаях прогнозные построения осложняются из-за наличия нескольких подчас взаимоисключающих взглядов на генезис. Возникает необходимость сопоставления и сравнения альтернативных генетических гипотез с точки зрения их правильности, обоснованности и пригодности для описания и объяснения особенностей изучаемого явления по результатам наблюдений.

Процесс прогнозирования с помощью АИПС строится на содержательной основе, поэтому во избежание сугубо формального подхода ведущий специалист должен тщательно проанализировать имеющиеся подходы, с тем чтобы в максимальной степени учесть опыт предыдущих исследователей (естественно, если рассматриваемая проблема изучалась ранее).

Использование АИПС позволяет осуществить многовариантное прогнозирование на основе различных гипотез, групп факторов, признаков, эталонов; с этих позиций изучение имеющихся гипотез ведущим специалистом является крайне необходимым. Творческий критический анализ имеющихся гипотез обеспечивает надежный отбор наиболее важной информации, необходимой и достаточной для решения поставленной задачи геологического прогнозирования.

Формирование рабочей гипотезы (модели). Процедура построения информационной модели заключается в изучении ведущим специалистом геологической истории формирования месторождения или другого геологического объекта, в зависимости от поставленной задачи и масштаба исследования. Главная задача такого моделирования — выбор комплекса факторов, имеющих важное значение для формирования

рудного объекта и эталонных объектов* — аналогов для последующего изучения количественных характеристик фактической значимости отдельных факторов и последующей классификации на этой основе всей изучаемой территории [59, 121, 130 и др.].

Построение таких информационных моделей является необходимым с точки зрения получения целостного представления о взаимосвязанной системе геологических процессов, приведших к формированию месторождения (или другого геологического объекта), выявления места в этой системе процессов накопления, миграции и переотложения рудного материала, непосредственно обусловивших возникновение его высоких локальных концентраций.

Информационная модель процесса формирования геологического объекта должна строиться на надежном фактическом материале. Таким фактическим материалом могут быть качественные и количественные результаты различных геологических, геофизических, геохимических, гидрогеологических и других наблюдений, а также экспертные оценки роли отдельных геологических процессов и факторов в рудообразовании. На данном этапе возможно привлечение большого количества факторов, выделяемых на основе априорных оценок. В дальнейшем будет определена апостериорная информационная значимость каждого фактора при решении конкретной задачи прогнозирования путем машинной обработки информации, и, тем самым, количество факторов может быть сокращено за счет исключения малоинформативных. Методические приемы построения таких геологических моделей описаны в работах В. И. Кусевича и Л. П. Зеликовского [59], Л. П. Зеликовского [44]. Особенности обработки геофизических данных приведены в работах Б. Н. Павлова, А. И. Корниенко и др. [93, 97].

Масштаб исследования определяет область рассмотрения и степень детальности изучения внутренних связей, зависимостей и количество учитываемых факторов. Например, модель формирования месторождений в масштабе 1:500 000 и 1:200 000 должна разрабатываться на фоне геологического развития региона или отдельных его районов.

При масштабе 1:100 000 — 1:50 000 область рассмотрения сужается до границ структурно-фациальных и рудоносных зон, рудных узлов или рудных полей; а при масштабе 1:10 000 и крупнее — до границ месторождений, участков, рудных тел.

При этом в качестве основных рудоконтролирующих факторов при масштабе исследований 1:200 000 выступают, как правило, наиболее общие причины рудоотложения — основные геологические процессы (осадкообразование, складкообразование, магматизм и др.), фиксируемые в реальной геологической обстановке своими продуктами (геологическими формациями, тектоническими структурами, магматическими комплексами и др.).

Модели масштабов 1:25 000 — 1:10 000 дают возможность выявить более частные причины рудоотложения, связанные с некоторыми количественными и качественными показателями геологических процессов.

Таким образом, с помощью таких моделей можно с высокой степенью обоснованности выделить группы факторов, оказывающих (в совокупности) влияние на рудообразование, что особенно важно при выборе комплекса поисковых предпосылок и признаков для решения многовариантных прогнозных задач с помощью ЭВМ.

* При использовании алгоритмов распознавания образов с обучением.

В силу того, что в АИПС/ЕС осуществляется автоматизированная обработка различного рода картографических материалов, основная сложность действий ведущего специалиста при реализации данной процедуры состоит в том, чтобы адекватно отобразить все важные для решения конкретной задачи факторы с помощью имеющейся картографической информации.

Другим важным моментом при реализации рассматриваемой процедуры является выбор эталонных объектов — аналогов прогнозируемых месторождений (рудных полей, участков и т. п.). Методика прогнозирования с помощью АИПС принципиально позволяет использовать в качестве эталонов не только объекты, имеющиеся в пределах изучаемой территории, но также объекты, расположенные и в других регионах. Однако для этого требуется, чтобы информационное описание эталонов из «чужих» территорий было идентично факторному описанию исследуемой площади.

Итак, в результате выполнения рассматриваемой процедуры ведущий специалист (при необходимости с привлечением экспертов) устанавливает комплекс геологических и других факторов, определяющих позицию прогнозируемых объектов для последующего занесения их в память ЭВМ и создания личного целевого массива по исследуемой территории.

При этом геолог-интерпретатор составляет список этих факторов и присваивает им порядковые номера.

Процедуры взаимодействия с ЭВМ в процессе решения задач

Формирование личного целевого массива является первой технологической процедурой, выполняемой по заявке пользователя системы. На основании перечня факторов, выбранных пользователем из общего каталога банка данных, в вычислительном центре (ВЦ) формируется с помощью специальных программ целевой массив, которому присваивается собственный номер. Все последующие процедуры обработки информации будут осуществляться на основе этого целевого массива, независимо от первичного массива территориального банка данных.

На рис. 14 процедура формирования целевого массива показана под №3 на II этапе.

Как отмечалось выше, решение задач прогноза и оценки рудоносности может выполняться в АИПС/ЕС с помощью следующих методических схем:

— на основе использования алгоритмов распознавания образов с обучением;

— на основе использования знаний специалиста о закономерностях локализации оруденения, формально представленных с помощью алгоритма эвристического моделирования;

— на основе безэталонной классификации с использованием алгоритма таксономии и т. д.

Ниже будет описана технология человеко-машинного взаимодействия при решении задачи геологического прогнозирования с помощью алгоритма распознавания образов с обучением.

Методику прогнозирования с использованием алгоритма распознавания образов следует применять в первую очередь, когда оказывается целесообразным выявление в пределах исследуемой территории аналогов известных рудных объектов или необходимо оценить информацион-

ную значимость отдельных факторов с точки зрения изучения закономерностей пространственной локализации конкретного типа орудения.

Первой операцией при использовании методики распознавания образов является выбор эталонных объектов обучения и контроля — процедура 4 (здесь и далее номера процедур приведены по рис. 14, этап II).

Процедура выполняется оператором ВЦ по заявке ведущего специалиста с помощью программы «Выбор».

После того как геолог-интерпретатор сформировал модель для прогнозирования и выбрал эталонные объекты-аналоги (т. е. составил реальный перечень объектов), необходимо извлечь из целевого массива соответствующую информацию, описывающую эталонные объекты в заданном пространстве признаков.

Для этого в ВЦ должны быть переданы сведения о номерах объектов обучения и контроля и об их координатах в районе (лист, строка, столбец). Результатом процедуры является машинное формирование массива значений всех признаков для каждого из объектов, или так называемого массива выборок. Одновременно эта информация распечатывается на АЦПУ. При наличии целевого массива формирование любой выборки производится весьма быстро (1—3 мин). Каждому массиву выборок присваивается свой номер, под которым он сохраняется для дальнейшего использования.

После того как сформирован массив выборки для обучения и контроля, выполняется проверка ее на «однородность», для чего используется алгоритм таксономии (процедура 5). Эта процедура выполняется при помощи программы «Транзитив», ее назначение — анализ взаимного положения объектов в принятом признаковом пространстве. В ВЦ должны быть переданы следующие данные: номер выборки, перечень признаков, которые должны быть учтены при таксономии, начальное значение параметров близости и его приращение. Результатом процедуры является несколько вариантов разложения выборки на группы «близких» между собой объектов. Первый вариант соответствует начальному значению параметра близости, второй — параметру, измененному на величину приращения, и т. д.

Содержательный анализ структуры выборок (процедура 6) выполняется непосредственно ведущим специалистом. При необходимости также могут привлекаться эксперты. Изучая разбиение выборки на таксоны при различных параметрах близости в заданном пространстве признаков, геолог-интерпретатор выясняет, насколько адекватно выбранное пространство признаков отражает известные ему геологические закономерности. Меняя шкалу исходных признаков, можно обнаружить, за счет каких факторов происходит разделение или объединение отдельных объектов выборки. Если окажется, что выборка, рекомендованная специалистами для обучения классификации и априорно считавшаяся однородной, в результате таксономического исследования оказалась неоднородной, следует исключить некоторые группы признаков или отдельные признаки, повторить процедуру таксономии, попытаться установить причины неоднородности выборки обучения и принять решение либо о том, что имеющаяся ее неоднородность несущественна с содержательных геологических позиций, либо сформировать по результатам таксономии другую выборку, которая будет однородной.

Поясним это примером. Допустим, специалистом составлена выбор-

ка из 30 объектов одного класса, каждый из которых описан n_0 признаками.

В результате процедуры таксономии выборка разделилась на 4 группы объектов:

1, 2, 3, 16, 18, 24, 30	(таксон 1);
4, 5, 6, 7, 9	(таксон 3);
7, 10, 11, 12, 13, 14, 23	(таксон 3);
15, 17, 19, 20, 21, 22, 25, 26, 27, 28, 29	(таксон 4).

Таким образом, в данном случае выборка, рекомендованная специалистом, оказывается неоднородной. Необходимо установить причины такой неоднородности путем последовательного исключения отдельных признаков и их групп с повторением процедуры таксономии до тех пор, пока все объекты войдут в один общий таксон. Это будет свидетельствовать об однородности выборки в сокращенном пространстве признаков n_1 ($n_1 < n_0$).

Кроме того, процедура таксономии может использоваться в качестве самостоятельного методического подхода для анализа различных геологических ситуаций. Например, предположим, что если в рассмотренную ранее выборку входили 7 месторождений конкретного типа (№ 1—7) и различные рудопроявления и аномальные точки (№ 8—30), то результаты таксономической классификации, приведенные выше, могут свидетельствовать, во-первых, об относительной общности объектов в каждом из четырех таксонов (в принятом пространстве признаков) и, во-вторых, о том, что все объекты разделяются на четыре типа, последний из которых не содержит известных месторождений.

На такой основе принципиально возможно осуществлять группирование всех месторождений, рудопроявлений, аномалий, точек минерализации, что позволит рекомендовать для последующей проверки полевыми работами не все аномалии, а в первую очередь лишь отдельные представители из различных «таксонов». Кроме того, появляется возможность подвергнуть процедуре таксономии часть исследуемой территории или всю территорию. Эта процедура будет соответствовать своего рода «формальному районированию», оперативно выполняемому ЭВМ на строго логической основе. В этом случае отдельные точки площади, входящие в один таксон, будут соответствовать некоторым участкам земной поверхности с особыми, присущими только им общими свойствами. Результаты такого формального районирования представляют определенный интерес с точки зрения содержательного анализа геолога-интерпретатора.

Процедура анализа структуры эталонной выборки предусматривает также построение гистограмм плотностей распределения объектов по отдельным признакам. Содержательный анализ таких гистограмм распределений также выполняется ведущим специалистом. Назначение процедуры — предварительная оценка информационной значимости отдельных факторов (в качественной форме) и определение отдельных интервалов их значений, характеризующихся преобладанием точек одного класса.

В случае использования большого числа геологических и других характеристик, часть из которых может не иметь прямого отношения к конкретной задаче геологического прогнозирования, для сокращения исходного пространства признаков следует выполнить рассматриваемую процедуру промежуточного анализа исходной информации.

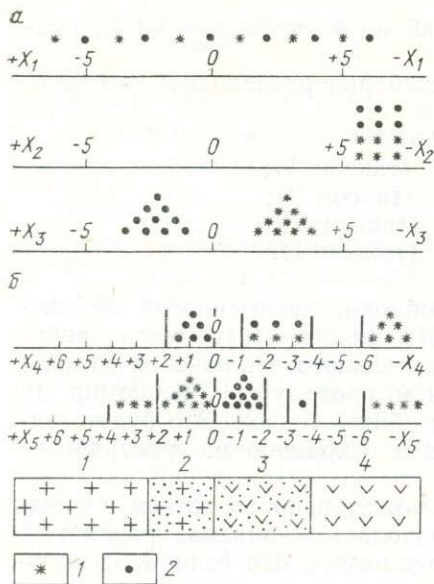


Рис. 16. Пример распределения значений фактора у отдельных объектов (а) и выделения интервалов значений фактора (б).

1 — объекты I класса, 2 — объекты II класса

двух классов и области отсутствия объектов любого класса). С другой стороны, возможно определение границ интервалов исходя из содержательных геологических соображений [93]. Во всех случаях неперенным условием является учет реальной геологической ситуации.

Поясним сказанное примером. Рассмотрим первое назначение содержательного анализа гистограмм распределений — выделение «делящих» факторов, важных для решения конкретной задачи геологического прогнозирования. Пусть имеются три гистограммы распределений объектов выборки двух классов по факторам X_1 , X_2 и X_3 (рис. 16, а).

В этом случае очевидно, что факторы X_1 и X_2 являются «неделящими» и использование их в дальнейшем при решении конкретной задачи прогнозирования нецелесообразно.

Напротив, фактор X_3 является хорошо «делящим» объекты двух классов, и использование его при решении данной задачи будет полезным.

Рассмотрим второе назначение содержательного анализа данной процедуры — определение границ интервалов факторов. В качестве примера приведем распределение объектов выборки по факторам X_4 и X_5 (рис. 16, б).

Рассматривая распределения объектов по фактору X_4 , целесообразно установить границы +2; 0; -1; -4; -6; то же необходимо сделать по фактору X_5 , т. е. установить границы +4; +2; 0; -2; -3; -4; -6.

Неперенным условием при этом является содержательный геологический анализ разделения значений фактора на отдельные интервалы. Поясним это на примере фактора X_5 . Допустим, что этот фактор характеризует «положение объектов относительно гранитной интрузии

С этой целью построенные ЭВМ гистограммы распределений отдельных факторов оцениваются специалистом и для дальнейшей работы отбираются лишь те из них, которые в той или иной мере являются «делящими» для объектов различных классов. Таким образом, одномерные распределения используются геологом-интерпретатором для выявления статистических закономерностей пространственного положения месторождений (или других геологических объектов) относительно геологических, геофизических, геохимических и других факторов, являющихся поисковыми критериями.

Другое назначение рассматриваемой процедуры состоит в том, чтобы область изменения каждого фактора, оставленного для дальнейшей обработки, разбить на отдельные интервалы. Определение границ интервалов осуществляется, с одной стороны, по результатам анализа гистограмм распределений объектов выборки (ограничиваются области наиболее плотно распределения точек каждого из

X-го комплекса». В качестве иллюстрации под фактором X_5 приведен геологический разрез, в котором интервалы 1 и 2 соответствуют области развития гранитной интрузии (1 — центральная ее часть, 2 — область эндоконтакта), интервал 3 характеризует область экзоконтакта — контактово-измененных вмещающих пород, а интервал 4 — области развития неизменных вмещающих пород. Контакт между интрузией и вмещающими породами проходит по нулевой линии. В данном случае оказывается, что объекты I класса располагаются в основном в области контактово-измененных пород, это положение согласуется с фактическими данными и имеет теоретическое геологическое обоснование.

Аналогичным образом определяются интервалы значений у всех остальных факторов. Принципиально допускается определение границ интервалов лишь на основе чисто геологических представлений.

Следующая процедура — построение решающего правила классификации на основе алгоритма распознавания образов с обучением (процедура 7 на рис. 14, этап II). Процедура выполняется в ВЦ; ее назначение — обучение ЭВМ распознаванию объектов, аналогичных представленным в исходной выборке.

При использовании алгоритма «Гиперпласт» такое обучение осуществляется с так называемым «поощрением», т. е. ЭВМ подбирает такие информационные веса для каждого из признаков, при которых происходит наилучшее разделение объектов двух классов.

Суть алгоритма «Гиперпласт» состоит в том, что, приписывая каждому интервалу геологического признака некоторый числовой вес и варьируя этими весами, алгоритм методом последовательных приближений выбирает такие их значения, при которых наилучшим образом разделяются объекты обучающей выборки.

«Информационные веса» определяются алгоритмом по обучающей выборке и соответствуют оптимальным величинам, при которых происходит наилучшее разделение объектов двух классов (например, промышленных месторождений и непромышленных рудопоявлений).

Таким образом, каждый признак получает количественную оценку его «значимости» при выделении новых объектов. Такого рода решающее правило может быть определено как комплексный критерий классификации.

Полученное решающее правило каталогизируется и сохраняется в специальной библиотеке АИПС/ЕС «Регион».

На основе полученного правила ЭВМ классифицирует контрольные объекты, не участвовавшие в обучении (процедура 9 на рис. 14, этап II). Следующая процедура — анализ комплексного критерия классификации (процедуры 8 и 10). Процедура выполняется геологом-интерпретатором. Основное назначение процедуры: оценка качества комплексного критерия (решающего правила) по результатам классификации объектов контрольной выборки и содержательный геологический анализ комплексного критерия путем оценки апостериорной информационной значимости отдельных признаков.

Получив из ВЦ решающее правило в виде распечатки на АЦПУ, геолог-интерпретатор оценивает качество распознавания контрольных объектов. Сверяя результаты машинного контроля с имеющимся фактическим материалом о принадлежности контрольных объектов к классам, геолог-интерпретатор составляет заключение о качестве построенного машинного критерия классификации. Такую оценку можно дать, например, в процентах правильной классификации.

Анализ результатов контроля также выполняется геологом-интерпретатором, который составляет окончательное заключение о результатах контроля. При этом подсчитывается количество ошибок I и II рода (неправильная классификация), а также число «отказов» от классификации. В случае, если результаты классификации удовлетворительны и отвечают существующим геологическим представлениям, геолог-интерпретатор разрешает выполнение процедуры классификации по всей исследуемой территории. Если результаты контроля не удовлетворяют экспертов, имеются возможности корректировки их путем возвращения к процедуре обучения, изменения выборки объектов обучения или уточнения геологических и прочих факторов, принятых для решения данной задачи (возвращение к процедурам 2 и 4).

Кроме того, регулируя «порог» классификации решающего правила, можно достичь оптимальных результатов распознавания путем уменьшения количества «отказов». Пример выбора оптимального порога решающего правила классификации приведен в табл. 16. В качестве критерия оптимальности при этом было принято число правильно классифицированных контрольных объектов. По этому показателю оптимальным является вариант с порогом 0,910, хотя число правильно классифицированных объектов у него несколько меньше, чем в варианте с порогом 0,880. Дело в том, что в последнем случае число неправиль-

Таблица 16
Пример выбора оптимального порога классификации решающего правила по контрольной выборке

Классы объектов контроля и их количество (N)	Результат классификации	Величина порога классификации h							
		1,000	0,980	0,960	0,940	0,920	0,900	0,880	0,910
Класс I N=94	Правильно классифицировано	29	48	61	63	65	75	84	69
	Отказ от классификации	65	46	33	31	28	17	—	25
	Неправильно классифицировано	—	—	—	—	1	2	10	—
Класс II N=53	Правильно классифицировано	19	22	27	29	35	35	45	31
	Отказ от классификации	34	31	26	24	18	17	—	12
	Неправильно классифицировано	—	—	—	—	—	1	8	—
По обоим классам N=147	Правильно классифицировано	48	70	83	92	100	110	129	100
	Отказ от классификации	99	77	59	55	46	34	—	47
	Неправильно классифицировано	—	—	—	—	1	3	18	—

но классифицированных объектов составило 18, что снижает надежность прогноза. Очевидно, что «отказ от классификации» оказывает меньшее отрицательное влияние на надежность конечных результатов, нежели «неправильная классификация».

Другой важной процедурой является анализ информативности признаков* при решении задачи геологического прогноза. В АИПС/ЕС разработан и применяется метод, использующий как формальную, так и содержательную стороны задачи. Этот метод анализирует вклад каждого признака в классификацию эталонных объектов, т. е. каждой паре признак — эталонный объект ставится в соответствие число, показывающее влияние данного признака на принятие решения при классификации эталонного объекта.

Метод основан на решающем правиле, которое строит алгоритм «Гиперпласт». В этом алгоритме область возможных значений каждого фактора разбивается на несколько интервалов. Рассмотрим, например, упоминавшийся выше фактор «положение объекта по отношению к гранитной интрузии» (см. рис. 16, б). Допустим, что для интервалов, на которые были подразделены значения фактора, обучение с помощью алгоритма «Гиперпласт» на ЭВМ дало следующие коэффициенты информативной значимости C_i : для центральной части интрузии (1) $C_1 = -20$; для области эндоконтакта (2) $C_2 = +5$; для области экзоконтакта (3) $C_3 = +93$; для области неизменных вмещающих пород (4) $C_4 = +2$.

Если объектами обучения I класса были промышленные месторождения, а объектами II класса — непромышленные рудопроявления, то оказывается, что объекты, лежащие в массиве интрузии, чаще относятся к непромышленным, а объекты, расположенные в области экзоконтакта с большой вероятностью, — к промышленным. Эти соображения могут быть объяснены с содержательных геологических позиций и согласуются с фактическим расположением объектов, использованных в качестве эталонных в выборке обучения. Естественно, это не означает, что принятие решения можно осуществлять лишь по одному наиболее информативному (или «весовому») признаку, т. е. утверждать, что в области экзоконтакта не бывает непромышленных рудопоявлений. Такие рудопоявления, конечно, могут быть, но они будут «отбракованы» другими признаками. В реальных задачах прогнозирования на основе АИПС классификация осуществляется по всем признакам по процедуре, напоминающей голосование.

Тот факт, например, что экзаменуемый объект находится в области эндоконтакта в приведенном примере прибавляет 5 «голосов» за отнесение его к I классу. Если бы объект находился в массиве интрузии, это соответствовало бы получению 20 «голосов» за его отнесение ко II классу (непромышленные объекты). Эти «голоса» суммируются по всем признакам, и если голосов «за перспективность» оказывается на величину h^{**} больше, чем голосов «за бесперспективность», объект относится к I классу.

Таким образом, решающее правило, построенное алгоритмом «Гиперпласт» по всем признакам при решении конкретной задачи геологического прогнозирования, является, в принципе, комплексным количественным поисково-оценочным критерием. Анализ составляющих решаю-

* Информативность отдельного признака — величина относительная и характерная только для конкретного сочетания признаков при решении конкретной задачи геологического прогнозирования.

** h — порог классификации, определяемый алгоритмом «Гиперпласт» в процессе обучения.

щее правило коэффициентов (информативных весов) отдельных признаков дает возможность определить, за счет каких признаков данный объект отнесен к тому или иному классу. Это позволяет установить, какие признаки существенны при классификации объектов данного конкретного типа. Кроме того, на основе анализа величины относительного информативного веса признаков для разных масштабов исследований и вариантов прогноза можно определить, какие факторы или их группы отражают наиболее существенные черты описания геологических объектов.

Рассмотрим на одном примере последовательность анализа информативных вкладов (весов) отдельных групп факторов. За 100% в табл. 17 принята суммарная величина положительных информационных весов (вкладов) отдельных факторов по каждому из эталонных объектов (гидротермальных редкометалльных месторождений 1—5), по которым получено решающее правило классификации.

Анализируя эти данные, ведущий специалист и эксперты оценивают полученное решающее правило с содержательных геологических позиций. Так, очевидно, из табл. 17 следует, что решающее правило отводит наибольшее значение структурно-тектоническим факторам, которые в общем обеспечивают приблизительно половину вклада от всех принятых для решения данной задачи признаков в величину классифицирующего критерия, обуславливающего отнесение эталонных объектов к I классу. В то же время можно заметить, что литолого-стратиграфические факторы занимают второе по значимости положение для объекта 1 (относительный положительный вклад информации 32,8%), а магматические — третье (17,1%). Для объекта 5 наблюдается обратная картина: на долю литолого-стратиграфических признаков приходится всего лишь 11,5% общего положительного вклада информации, а на долю магматических — 32,0%. Заключение экспертов, проанализировавших эти данные, свелось к тому, что сведения, показанные в табл. 17, в общем виде правильно отражают геологическую позицию и относительное значение основных рудоконтролирующих критериев каждого эталонного объекта.

Подобный анализ может быть продолжен в более дифференцированном виде: отдельно для каждого признака и его интервала. В этом случае становится возможным установить признаки, одинаково важные для всех эталонных объектов («объединяющие признаки»), т. е. отно-

Таблица 17

Таблица относительных информационных весов различных групп факторов для эталонных объектов I класса (получена по результатам обучения на эталонных объектах алгоритмом «Гиперпласт»), %

Группы признаков	Эталонные объекты				
	1	2	3	4	5
Структурно-тектонические	50,1	53,5	54,6	52,6	56,5
Литолого-стратиграфические	32,8	34,5	27,0	22,0	11,5
Магматические	17,1	12,0	18,4	25,4	32,0
Всего	100,0	100,0	100,0	100,0	100,0

сительный вклад которых в классифицирующий критерий принадлежности эталонных объектов к I классу одинаков. В самом общем виде в табл. 17 к таким «объединяющим признакам» могут быть отнесены признаки структурно-тектонического характера. Подобным образом могут быть выделены признаки, подчеркивающие «индивидуальности» каждого эталонного объекта («разъединяющие» признаки). К таким признакам следует отнести группу литолого-стратиграфических и магматических признаков, величины относительных вкладов по которым имеют более переменный характер по сравнению со структурно-тектоническими.

Продолжая анализ решающего правила, можно установить и такие интервалы признаков, которые имеют весьма малые величины положительных вкладов информации либо даже отрицательные значения. Очевидно, такие признаки оказывают незначительное влияние на результаты обучения и классификации либо даже, напротив, могут служить возможными источниками ошибок классификации. По-видимому, такие признаки отвечают понятию «шумовых», или «паразитных»; их исключение из процедуры обучения и классификации позволит упростить решающее правило и повысить надежность классификации. Однако после их исключения процедура обучения должна быть выполнена заново.

Результаты содержательного анализа решающего правила должны использоваться также на заключительном этапе решения прогнозной задачи. При этом путем сравнения информационного вклада отдельных признаков с участков, отнесенных в результате классификации к I классу, с соответствующими данными по эталонным объектам геологу-интерпретатору представляется возможность установить причины такой классификации и проанализировать их из содержательных геологических соображений.

Последовательно анализируя количественные оценки информативности отдельных интервалов, геолог-интерпретатор может установить их соответствие теоретическим представлениям, реальной геологической обстановке, процессам рудогенеза и т. п.

Таким образом, анализ решающего правила позволяет геологу-интерпретатору получить количественные оценки значимости факторов, отражающие различия между эталонными объектами I и II классов. Проверка качества решающего правила по контрольной выборке и анализ соответствия количественных оценок информативности признаков известным теоретическим представлениям являются по существу экспертизой процедур, определяющих принадлежность новых объектов к одному из классов.

После анализа решающего правила (в случае положительного заключения) геолог-интерпретатор дает разрешение на проведение процедуры машинной классификации по всей исследуемой территории (процедура II на рис. 11). Классификация выполняется в ВЦ, геолог-интерпретатор при этом указывает порог, с которым будет осуществляться классификация (машинный поиск объектов-аналогов).

Процедура классификации выполняется специальной программой и сводится к «вызову» из памяти ЭВМ значений признаков последовательно по каждой элементарной площадке в пределах всей исследуемой территории и сравнению их с полученным решающим правилом (отношением к одному из классов). Такая классификация выполняется «построчно», т. е. вначале классифицируются все точки в пределах одной

«страницы» условного листа исследуемой территории. Вся территория при этом подразделяется на площади, перспективные для обнаружения месторождений и рудопоявлений определенных типов (I класс), и площади, не перспективные для обнаружения таких объектов (II класс).

Процедуры содержательного анализа результатов

После того как результаты классификации получены в виде распечатанных карт по всей исследуемой территории, геолог-интерпретатор осуществляет предварительный анализ такой карты. При этом он проверяет соответствие результатов имеющимся геологическим концепциям. Очевидно, что при этом следует проконтролировать результаты машинной классификации исходя из содержательных геологических соображений. Для этого старшему технику сообщаются координаты участков, выделенных ЭВМ в качестве перспективных, а также тех участков, результаты классификации по которым представляются сомнительными (например, отнесение некоторых участков к неперспективным). Из целевого массива банка данных будут получены исходные данные, сравнение которых с комплексным критерием классификации (решающим правилом) послужит основой для содержательного геологического анализа.

Используя изложенный выше метод оценки «информативных вкладов признаков», геолог-интерпретатор устанавливает основные причины, послужившие основанием для формально-логической машинной классификации. Сравнивая имеющиеся исходные данные с материалами машинного обучения по эталонным объектам, можно составить заключение об относительной «близости» выделенных участков к тем или иным эталонам, а также определить, за счет каких факторов (признаков) данная конкретная точка отнесена к тому или иному классу. В целом процедура предусматривает содержательный контроль и анализ результатов машинной классификации.

Затем ведущий специалист выполняет геологическую интерпретацию полученных результатов. Эта процедура осуществляется путем геологического структурного анализа положения выделенных перспективных участков. На данном этапе проводится «генерализация» выделенных перспективных площадей в соответствии с конкретной геологической обстановкой. Например, если в пределах какой-либо единой структуры располагается несколько перспективных площадей, целесообразно объединение их в одном контуре для удобства проведения полевых работ. Решающее мнение при этом должно принадлежать экспертам, хорошо знакомым с конкретным районом.

Геолог-интерпретатор при анализе результатов находит содержательное объяснение выделенным перспективным площадям, устанавливает соответствие их конкретным концепциям и моделям формирования оруденения.

Работа завершается составлением отчета о результатах прогнозирования. Отметим, что получение окончательных результатов, в отличие от традиционных методов исследования, достигается значительно быстрее.

На рис. 15, б приведена подробная технологическая схема человеко-машинного взаимодействия в АИПС/ЕС при решении задач геологического прогнозирования на основе алгоритма распознавания образцов.

АИПС/ЕС «Регион» имеет также технологические схемы человеко-машинного взаимодействия при безэталонном районировании террито-

рии с использованием алгоритма таксономии и на основе алгоритма эвристического моделирования, дающего возможность специалисту построить поисково-оценочный критерий на основе теоретических представлений, имеющегося практического опыта и интуиции (так называемое «эвристическое правило классификации»).

Кроме того, алгоритм эвристического моделирования позволяет оперативно выделять области сочетания различных факторов, благоприятных для оруденения в пределах всей исследуемой территории, избавляя тем самым специалистов от необходимости копирования таких факторов на кальку (или другую прозрачную основу) для последовательного их совмещения. Известно, что такой способ анализа применяется специалистами при металлогеническом анализе [121].

VI. РЕШЕНИЕ ЗАДАЧ ПРОГНОЗА И ОЦЕНКИ ГЕОЛОГИЧЕСКИХ СИТУАЦИЙ НА ОСНОВЕ ЧЕЛОВЕКО-МАШИННОЙ СИСТЕМЫ

В процессе опытной эксплуатации система АИПС использовалась при решении широкого круга задач прикладной геологии в различных регионах Советского Союза со сложным геологическим строением. Масштабы исследований от 1 : 25 000 до 1 : 200 000.

Часть задач была связана с решением вопросов прикладной геологии — выделением и оценкой отдельных геологических объектов*, остальные — с выполнением научно-методических исследований [39]. В результате решения задач первого направления с помощью системы удалось выделить перспективные участки, рекомендованные для проведения поисково-съёмочных и оценочных работ.

В процессе методических исследований решались, например, задачи сравнительной оценки перспектив территории на основе конкурирующих геологических моделей, задачи оценки эффективности комплекса геологических и геофизических методов изучения территории, выявления перспективных площадей по данным экспертов и др. Ниже приводятся некоторые результаты, рассматриваемые последовательно на примере двух регионов.

Каждый из регионов характеризуется сложным строением; оруденение располагается в обоих случаях в породах фундамента, перекрытых чехлом молодых осадочных образований.

Описание результатов по каждому из регионов производится в следующей последовательности: характеристика машинного банка данных, результаты решения практических задач, результаты научно-методических исследований.

1. ПРИМЕНЕНИЕ АИПС/ЕС ДЛЯ ПРОГНОЗА МЕДНОГО ОРУДЕНЕНИЯ

Основной задачей опробования системы в районе являлось выделение площадей, перспективных на обнаружение медного оруденения. Общая площадь исследуемой территории — около 5400 км², масштаб исходных геологических, геофизических и геохимических материалов 1 : 50 000.

Краткая геологическая характеристика исследуемого района

Район исследования располагается в пределах синклинория, входящего в состав крупной геосинклинальной области. Синклинорий имеет протяженность в сотни километров при ширине от 10 до 90 км. Наиболее древними породами в его пределах являются крупнозернистые лейкократовые граниты верхнедевонского комплекса.

* Под геологическим объектом подразумевается рудный район, рудный узел, рудные поля, месторождение [130].

Для исследуемого района характерна интенсивная вулканогенно-интрузивная деятельность в каменноугольное — нижнепермское время. В отдельных структурах и тектонических блоках незначительное распространение имеют осадочные породы (песчаники, алевролиты, сланцы) с подчиненными туфовыми образованиями карбона.

По геолого-геофизическим данным, на глубине под каменноугольными вулканогенными породами расположен обширный батолит, имеющий сложное строение. В пределах синклинали широкое распространение имеют образования многофазного нижнекаменноугольного интрузивного комплекса. По составу это крупно- и среднезернистые биотит-роговообманковые граниты, гранодиориты и реже кварцевые диориты, аляскитовые граниты, а также мелкие тела средне- и мелкозернистых биотитовых порфиroidных гранитов. Активная магматическая деятельность в карбоне привела к образованию вулканических и жерловых тел, сложенных комаматами. Субвулканические и жерловые фации представлены гранит-порфирами, фельзит-порфирами, липаритовыми и дацитовыми порфирами, андезитовыми и трахидацитовыми порфиритами, трахилипаритовыми порфирами, гранит- и гранодиорит-порфирами и андезито-базальтами.

Благодаря активной интрузивно-вулканической деятельности в районе развиты значительные по размерам зоны контактово-метаморфического, гидротермального и метасоматического изменения пород, проявившегося в образовании скарновых тел, зон кварц-турмалиновых метасоматитов, площадей аргиллитизации и пропилитизации пород, многочисленных массивов вторичных кварцитов.

Район характеризуется разнообразным по составу и возрасту оруденением. Все рудные формации генетически или парагенетически связаны с различными магматическими комплексами пород и магматическими циклами. С докаледонским и раннекаледонским магматизмом, приведшим к образованию линейных гипербазитовых интрузий, связаны проявления хрома, никеля и железа с подчиненным значением минерализации меди и кобальта. Позднекаледонский и раннегерцинский магматизм привел к образованию ряда полиметаллических (иногда с молибденом и медью) месторождений и рудопроявлений. Позднегерцинские магматические циклы обусловили формирование молибденомедных месторождений, иногда с вольфрамом и рядом других элементов. Широко развито в районе меднопорфировое оруденение, для которого характерна связь с верхнепалеозойским вулканизмом. Меднопорфировое оруденение связывается с общим процессом формирования вулканогенно-плутонических комплексов, в силу чего известные месторождения различаются как по глубине формирования, так и по структурно-морфологическим признакам, характеру гидротермального изменения пород и сопутствующим элементам. Эти месторождения по положению в вулканогенно-плутонических комплексах подразделяются на следующие подтипы.

1. Месторождения небольших глубин, локализующиеся в вулканогенных сериях вулканогенно-плутонических комплексов. Для них характерно широкое проявление продуктов кислотной стадии выщелачивания (вторичных кварцитов и др.).

2. Месторождения глубин сочленения эффузивных и интрузивных серий. Для них также характерно широкое развитие продуктов кислотного выщелачивания (серцитовые кварциты, пропилиты и др.).

3. Месторождения, локализующиеся в интрузиях (калишпатизация, хлоритизация, биотитизация и др.). Эти подтипы характеризуются также различием главных парагенезисов вторичных минералов.

Характеристика территориального машинного банка картографической информации

Выбор информации для формирования машинного банка проводился на основе экспертного опроса геологов-производственников и специалистов из научно-исследовательской организации. В результате проведенного опроса было выбрано 27 основных факторов, контролирующих медное оруденение.

Сформированный по методике АИПС/ЕС банк картографической информации масштаба 1 : 50 000 позволял охарактеризовать положение любой элементарной площади (0,25 км²) в пределах территории 16 планшетов масштаба 1 : 50 000 относительно следующих факторов.

I. Литолого-стратиграфические — 5 факторов:

1) верхнекарбоновые отложения, представленные липаритодацитами собственно эффузивной и жерловой фаций. Туфы и туфопесчаники;

2) среднекарбоновые отложения — андезит-диоритовые порфириты и туфолавы липаритодацитового состава;

3) нижнекарбоновые отложения (визе — намюр), представленные лавами и кристаллокластическими туфами, липаритами дацитового состава, реже песчаниками алевролитами;

4) нижнекарбоновые отложения (турне) — песчаники, алевролиты, андезитовые и дацитовые порфириты, туфы;

5) силурийские отложения — песчаники, алевролиты, аргиллиты, конгломераты.

II. Магматические — 9 факторов:

6) средне-мелкозернистые биотитовые и лейкократовые порфировидные граниты комплекса А;

7) крупно-неравнозернистые биотитовые и аляскитовые граниты комплекса А;

8) разномелкозернистые биотитовые граниты основной фазы комплекса А;

9) дайки кислого и основного состава комплекса А;

10) гранит-порфиры и гранодиорит-порфиры (комплекс малых интрузий);

11) нерасчлененные дайки комплекса Б;

12) крупно-среднезернистые биотит-роговообманковые граниты, гранодиориты и кварцевые диориты комплекса В;

13) нерасчлененные дайки комплекса В;

14) верхнедевонские интрузивные тела крупно-среднезернистых лейкократовых и биотитовых гранитов.

III. Структурно-тектонические — 3 фактора:

15) основные разрывные нарушения предположительно глубинного заложения;

16) комплексы малых интрузий, выделяемых по локальным геофизическим аномалиям (гравиразведка и магниторазведка);

17) вулканические аппараты, выделяемые магниторазведкой.

IV. Геохимические — 4 фактора:

18) участки с повышенным содержанием меди, выделяемые по результатам металлометрической съемки;

19) то же, молибдена;

20) то же, свинца;

21) то же, цинка.

V. Метасоматические — 6 факторов:

22) биотит-амфиболовые и кварц-эпидот-полевошпатовые роговики (метасоматиты щелочной стадии);

23) калишпатизированные и альбитизированные горные породы (метасоматиты щелочной стадии);

24) биотит-амфиболовые и эпидот-хлорит-кальцитовые пропилиты (метасоматиты щелочной стадии);

25) вторичные кварциты алунит-серицит-каолинового состава;

26) прожилково-жильное и штокверковое окварцевание (метасоматиты выполнения);

27) интенсивно измененные горные породы (участки с совмещенными стадиями метасоматоза).

По этим факторам проводилось решение задач выделения площадей, перспективных на обнаружение медного оруденения. Объем сформированного территориального банка данных в последующем может быть значительно увеличен как за счет ввода в ЭВМ новых факторов, так и за счет увеличения общей площади исследования. Постоянное пополнение машинного банка новой информацией расширяет круг содержательных геологических задач, в том числе выполнение прогноза на различные полезные ископаемые.

Результаты работ

Рассматриваемые ниже результаты иллюстрируют возможности решения с помощью АИПС/ЕС некоторых задач прогнозирования и оценки геологических объектов, а также выполнение научно-методических исследований.

Решение задач осуществлялось по четырем методическим схемам («технологическим модулям»):

— выделение перспективных площадей на основе алгоритма распознавания образов с обучением на известных объектах («Гиперпласт»);

— районирование территории на основе метода таксономии («безэталонное районирование»);

— классификация месторождений, рудопроявлений и точек минерализации методом таксономии;

— выделение перспективных площадей на основе теоретических представлений и опыта специалистов (моделирование с помощью эвристического алгоритма).

Решение практических задач

Выделение площадей, перспективных на медное оруденение. Решение задачи осуществлялось по результатам многовариантного прогнозирования с использованием алгоритма «Гиперпласт». Перспективы территории оценивались на основе нескольких генетических подходов, каждый из которых характеризовался заданным набором факторов, контролирующих медное оруденение в районе (генетическая связь оруденения с интрузиями комплексов А, Б, В и в целом с нижне-верхне-карбонным магматизмом).

В качестве эталонных объектов обучения первого класса были приняты одно крупное месторождение медно-порфирикового типа и четыре мелких месторождения этого же типа. Одно мелкое месторождение, не вошедшее в выборку обучения, использовалось для контроля качества

распознавания. Ввиду того что исходное признаковое описание в машинном банке данных дается для каждой элементарной ячейки размером 0,25 км², общее количество обучающей выборки составило 84 точки, контрольной — 27 точек.

Всего было рассчитано пять различных вариантов прогноза, каждый из которых характеризовал одну из генетических концепций медного оруденения в районе. В качестве основы для прогнозной карты был принят вариант № 4, показавший наименьшее количество ошибок на контроле. Он дополнен участками, уверенно выделяющимися по всем другим вариантам. Всего на прогнозной карте было выделено 325 км² перспективных площадей, включая эталонные объекты (на прогнозной карте, построенной коллективом специалистов традиционным способом, общая площадь перспективных участков I и II очереди равна 460 км²).

На рис. 17 приведена схема сопоставления результатов прогнозирования традиционным и автоматизированным способами. В целом наблюдается достаточная сходимость результатов. Особо следует отметить, что все перспективные площади, рекомендованные коллективом специалистов для постановки поисково-оценочных работ в I очередь, уверенно выделяются при автоматизированном прогнозировании.

Анализ геолого-структурного положения выделенных площадей с точки зрения существующих поисковых критериев медного оруденения не исключает возможности наличия в их пределах промышленного оруденения. Таким образом, итоги опробования АИПС/ЕС в данном случае свидетельствуют о достаточной сходимости выделения перспективных участков. В то же время применение системы позволяет в несколько раз сократить время составления варианта прогноза.

Классификация месторождений, рудопроявлений и точек минерализации. В пределах исследуемой территории известно 43 рудных объекта (месторождений, рудопроявлений и точек медной минерализации). Под классификацией в данном случае подразумевается группирование этих объектов с помощью метода таксономии. По условным координатам из банка данных была составлена выборка по всем этим объектам. Каждый объект при этом был охарактеризован тринадцатью факторами.

1. Пространственное положение относительно средне-мелкозернистых гранитов комплекса А.
2. Пространственное положение относительно крупно-неравномернoзернистых биотитовых и аляскитовых гранитов комплекса А.
3. Пространственное положение относительно гранит-порфиров и гранодиорит-порфиров комплекса Б.
4. Пространственное положение относительно даек комплекса Б.
5. Пространственное положение относительно крупно-среднезернистых биотит-роговообманковых гранитов, гранодиоритов и кварцевых диоритов комплекса В.
6. Наличие разрывных нарушений предположительно глубинного заложения.
7. Наличие малых интрузий, выделяемых локальными геофизическими аномалиями (гравиразведка и магниторазведка).
8. Наличие вторичных ореолов рассеяния меди.
9. Наличие вторичных ореолов рассеяния молибдена.
10. Метасоматиты стадии кислого выщелачивания (калишпатизированные и альбитизированные горные породы).
11. Метасоматиты стадии кислого выщелачивания (биотит-амфиболовые и эпидот-хлорит-кальцитовые пропилиты).

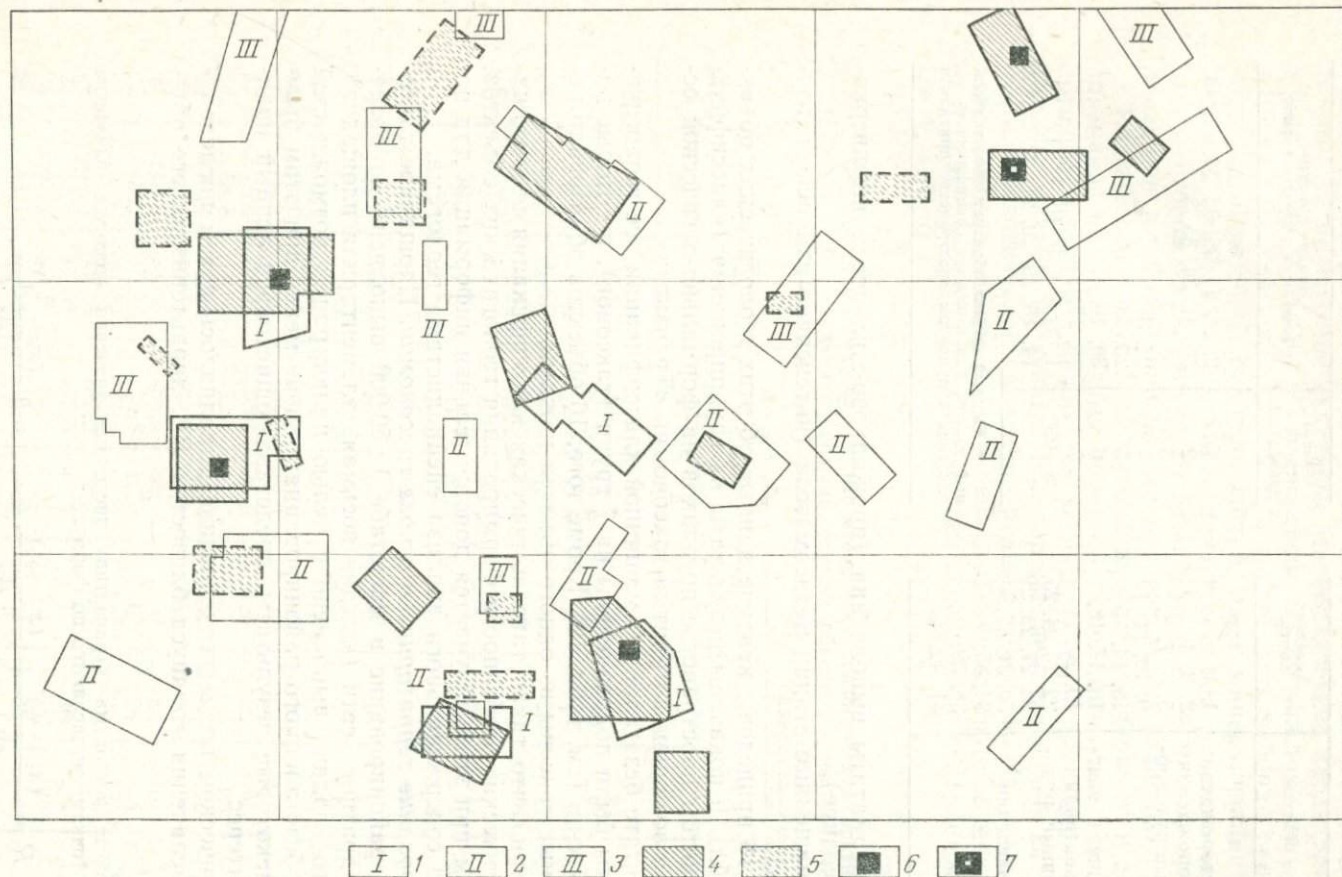


Рис. 17. Результаты геологического прогнозирования, полученные на основе традиционной методики и с использованием АИПС/ЕС «Регион».

1-3 — перспективные площади, выделенные на основе традиционной методики металлогенического прогнозирования коллективом специалистов научно-исследовательской и производственной организации: 1 — площади, рекомендуемые для постановки поисковых работ в I очередь, 2 — то же, во II очередь, 3 — то же, в III очередь; 4-7 — перспективные площади, выделенные на основе использования АИПС/ЕС «Регион»: 4 — площади, выделенные уверенно по всем вариантам, 5 — предполагаемые перспективные площади, выделяемые не по всем вариантам, 6 — эталонные объекты обучения распознаванию, 7 — контрольный объект

Таблица 18

Краткая характеристика выборки геологических объектов

Геологические объекты	Условные номера
Крупное месторождение	1
Мелкие месторождения	2, 3, 4
Перспективные рудопроявления	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Рудопроявления с неясными перспективами	16, 17, 18, 19, 20, 21
Бесперспективные рудопроявления	22, 23, 24, 25, 26, 27, 28, 29
Точки минерализации	30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43

Таблица 19

Результаты таксономической классификации

Номера групп (таксонов)	Номера объектов, вошедших в данную группу
1	2, 24, 25, 26, 32, 33, 42, 43
2	30, 31, 34, 35, 36
3	5, 6, 7
4	27, 37
5	23, 38
6	39, 40, 41
7	16, 17, 18
8—23*	1, 3, 4, 8, 9, 10, 11, 12, 13, 14, 15, 19, 20, 21, 22, 28, 29

* Каждый объект образует самостоятельный таксон, что отражает специфичность его описания в принятом пространстве признаков.

12. Метасоматиты выполнения (прожилково-жильное и штокверковое окварцевание).

13. Совмещенные стадии метасоматоза (интенсивно измененные горные породы).

В табл. 18 приведены краткие данные об этих геологических объектах, а в табл. 19 показаны результаты их группирования (классификации) в принятом пространстве признаков на формально-логической основе с использованием алгоритма таксономии «Транзитив».

В результате безэталонной машинной классификации все геологические объекты были подразделены на 7 групп (таксонов), в которые вошли 27 объектов (см. табл. 19). Кроме того, 16 объектов образовали самостоятельные таксоны из одного объекта каждый.

Естественно, что такая таксономическая классификация не отражает все генетические особенности формирования геологических объектов, вместе с тем при этом получается дополнительная информация для последующего содержательного анализа специалистами геологами.

Районирование территории на основе таксономии. Площадное таксономическое районирование в масштабе 1 : 50 000 выполнено по разреженной квадратной сети (каждая восьмая элементарная площадка). Решение этой задачи выполнено с целью иллюстрации возможностей технологии безэталонного районирования; тем не менее при этом были получены некоторые результаты, представляющие определенный практический интерес.

В методическом плане исследование выполнялось путем последовательного увеличения «радиуса близости» R^* , характеризующего общ-

* При работе алгоритма «Транзитив» численные значения «радиуса близости» между двумя точками вычисляются по формуле

$$R = \frac{(x_1 - y_1)^2}{D_1} + \frac{(x_2 - y_2)^2}{D_2} + \dots + \frac{(x_n - y_n)^2}{D_n},$$

где x_1, \dots, x_n — значения признаков первой точки;

y_1, \dots, y_n — значения признаков второй точки;

D_i — дисперсии значений соответствующих признаков.

ность сравниваемых точек выборки в принятом пространстве признаков (по 27 признакам).

Фактической основой для формального машинного анализа явилась выборка, сформированная из массива признаков банка данных. В табл. 20 приводятся некоторые математические характеристики (математическое ожидание M и дисперсия D) по каждому признаку банка данных для всей исследуемой территории, для выборки рудных объектов (месторождений, рудопроявлений, точек минерализации), а также значения признаков (x_i), характеризующих центральные части известных месторождений (крупного № 1 и мелких № 2, 3, 4, 5). Значения признаков в таблице приведены в относительных величинах.

Из табл. 20 следует, что математическое ожидание M и дисперсия D признаков выборки рудных объектов существенно отличаются от аналогичных величин в целом для всей территории исследования. Это положение служит косвенным подтверждением тезиса об «исключительности» геологической ситуации рудных объектов. Из таблицы также видно, что по своему признаковому описанию крупное месторождение № 1 практически по всем признакам существенно отличается от мелких (№ 2, 3, 4, 5).

Таблица 20

Некоторые математические характеристики геологических признаков для всей исследуемой площади, рудных объектов и месторождений

№ признаков	Для всей исследуемой территории		Для выборки всех рудных объектов		Для месторождений				
					крупное (№ 1)	№ 2	№ 3	№ 4	№ 5
	M	D	M	D	x	x	x	x	x
1	-86,4	42,92	-69,8	45,9	-30	-95	-97	-65	-110
2	-66,9	49,85	-51,5	39,9	-23	-79	-84	-82	-59
3	-52,0	48,64	-24,7	37,8	-33	15	0	-12	+17
4	-85,7	46,54	-56,0	52,16	0	28	-17	8	-48
5	-69,1	45,44	-52,5	49,1	0	-52	-34	-40	-92
6	-75,1	46,59	-49,0	41,9	-101	-15	0	-48	-17
7	-78,0	43,36	-55,0	42,1	-32	-41	-12	-126	-91
8	-68,4	44,58	-41,8	39,8	-23	-20	-41	-40	-4
9	-52,5	41,64	-26,0	33,3	-28	0	-4	-4	-32
10	-86,1	42,16	-69,2	43,7	-41	-48	-67	-126	-29
11	-82,9	54,17	-77,7	47,5	-94	-55	-88	-40	-126
12	-83,9	46,12	-89,8	47,8	-126	-52	-77	-126	-126
13	-86,4	52,42	-60,0	50,1	-87	-28	-17	4	-39
14	-110,2	32,52	-114,0	26,7	0	-93	-61	-126	-126
15	-101,0	38,85	-110,3	29,6	-126	-126	-126	-126	-65
16	-96,4	42,53	-92,9	48,4	-28	-126	-126	-126	-109
17	-95,4	40,15	-95,0	36,9	-126	-20	-48	-26	-116
18	-63,8	44,66	-52,4	41,7	-57	-36	-34	-27	-25
19	-74,4	47,91	-77,8	41,4	-31	-107	-126	-115	-59
20	-100,7	40,53	-72,4	46,7	-56	-17	0	-25	-62
21	-112,9	28,08	-86,4	47,6	0	-84	-107	-90	-54
22	-45,4	47,97	-23,8	42,8	12	28	4	-8	13
23	-103,8	39,72	-79,4	57,9	107	-126	-126	-126	-52
24	-47,6	44,13	-22,2	31,3	98	0	6	-20	-12
25	-53,8	47,09	-47,3	40,9	31	-36	-34	-27	-25
26	-102,2	36,54	-84,5	43,7	-4	-52	-45	-48	-120
27	-44,1	31,98	-15,0	18,9	4	-36	-4	-4	-32

M — математическое ожидание, D — дисперсия, x — значение признака.

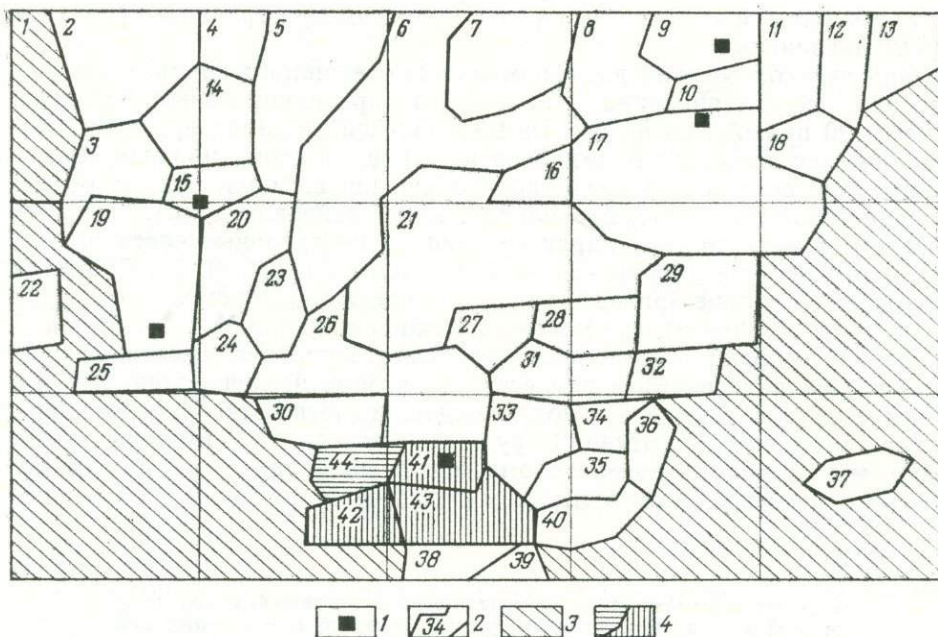


Рис. 18. Схема результатов безэталонного районирования территории на основе метода таксономии.

1 — известные медные месторождения, 2 — элементы площадного таксономического районирования и их номера, 3 — площадь «общего» таксона при минимальной величине «радиуса близости» R , 4 — «уникальные» площади, сохраняющие индивидуальность при максимальной величине «радиуса близости»

На рис. 18 изображены результаты площадного районирования, выполненные на ЭВМ с использованием алгоритма таксономии «Транзитив».

При минимальной величине «радиуса близости» R вся исследуемая территория «разбивается» на 44 различных по размерам площадных участка, образующих самостоятельные таксоны.

Самая крупная по размерам площадь, охватывающая периферийную часть исследуемой территории, образует единый «общий» таксон № 1. Эта площадь характеризуется относительно несложной (по сравнению с другими) геологической обстановкой; в ее пределах практически отсутствуют рудные месторождения.

При последующем увеличении «радиуса близости» в «общий таксон» (№ 1) постепенно входят все площади в пределах исследуемой территории, за исключением участков № 41, 42, 43 и 44. При максимальном «радиусе близости» остается лишь три таксона: № 1 (включающий участки № 1—40), таксон № 41 (включающий участки № 41, 42, 43) и площадь таксона № 44.

Интересно отметить, что в пределах таксона № 41 располагается крупное месторождение меди. Тем самым косвенным образом подтверждается своеобразная «уникальность» этого участка.

Определенный интерес представляет с этих позиций площадь, находящаяся в пределах таксона № 44. Она расположена непосредственно к западу от крупного месторождения на пересечении крупных разломов северо-западного и северо-восточного простирания. В ее пределах известны полиметаллические рудопоявления. В данном случае, естест-

венно, нельзя делать вывод, что в пределах этой площади имеется месторождение меди. Однако тот факт, что эта площадь обладает в некотором смысле «исключительными свойствами», служит основанием для более детального изучения особенностей ее строения. Разумеется, что все эти рассуждения верны лишь при описании района в принятом пространстве признаков.

Методические исследования

Сравнительная оценка перспективности территории на основе различных генетических концепций. К настоящему времени известно несколько взглядов на генезис медного оруденения в пределах исследуемого района.

Естественно, что прогнозирование перспективных площадей на основе различных концепций приводит к неоднозначным оценкам. Применение АИПС/ЕС открывает возможность оперативного получения конечных результатов прогнозирования при исследовании различных геологических моделей.

В качестве примера использования АИПС/ЕС для этих целей было осуществлено многовариантное прогнозирование по пяти геологическим концепциям, отражающим связь медного оруденения с различными интрузивными комплексами.

1. Генетическая связь оруденения с интрузиями комплекса В.
2. Генетическая связь оруденения с малыми интрузиями комплекса Б.
3. Генетическая связь оруденения с интрузивным комплексом А.
4. Генетическая связь оруденения с карбоновым магматизмом.
5. Преимущественный контроль оруденения литолого-стратиграфическими, структурно-тектоническими и другими факторами.

Решение прогнозных задач проведено с использованием алгоритма распознавания образов с обучением («Гиперпласт»).

В качестве объектов обучения I класса были использованы промышленные и мелкие месторождения медных руд. II класс был представлен заводом безрудными площадями.

Методическая схема формирования целевых массивов и выборки приведена на рис. 19. Результаты прогнозирования по всем вариантам показаны на рис. 20.

Анализ результатов показывает, что отдельные перспективные площади достаточно уверенно выделяются по всем вариантам. Это, в первую очередь, фланги известных рудных полей, а также некоторые новые площади. В то же время по различным вариантам выделяются существенно «индивидуальные» участки.

Использование разных генетических концепций при прогнозировании определяет различную информационную значимость отдельных факторов. Это положение иллюстрируется данными табл. 21.

Как следует из табл. 21, наиболее информативными факторами по варианту 1 являются прожилково-жильное и штокверковое окварцевание, геохимические ореолы свинца и меди.

В варианте 2 информативны порфировые граниты и гранодиориты комплекса малых интрузий, а также факторы, отмеченные в варианте 1.

Для варианта 3 информативными оказываются биотитовые граниты комплекса А, зоны прожилково-жильного и штокверкового окварцевания, зоны интенсивно гидротермально-измененных пород (с наложенными стадиями метасоматоза), геохимические ореолы меди и свинца.

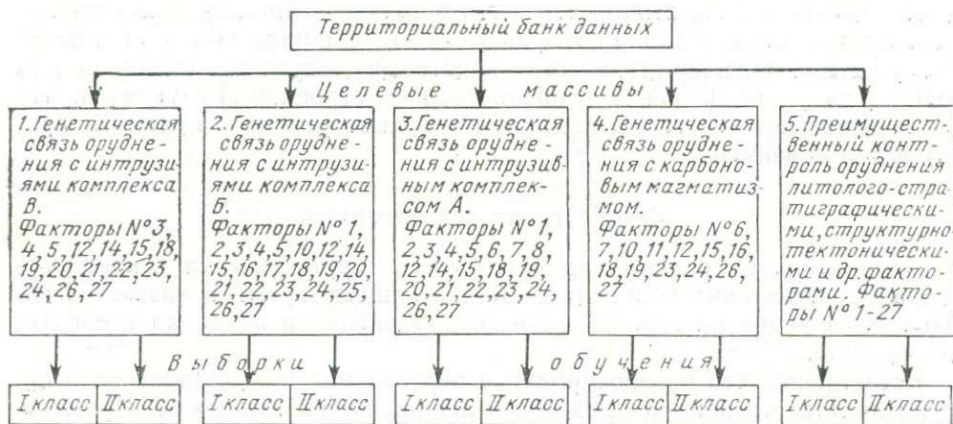


Рис. 19. Методическая схема решения задачи многовариантного прогнозирования

В варианте 4 высокой информативностью характеризуются породы комплекса малых интрузий, граниты комплекса А, зоны окварцевания и гидротермального изменения пород, а также вторичные ореолы меди.

Наконец, в варианте 5 информативны средне-мелкозернистые биотитовые и лейкократовые граниты основной фазы комплекса А, порфиоровые граниты и гранодиориты комплекса Б, а также зоны окварцевания и гидротермального изменения пород.

Соответствие каждой из названных выше генетических концепций реальной ситуации может быть охарактеризовано качеством распознавания объектов I и II классов на контрольной выборке (табл. 22).

Из табл. 22 следует, что наилучшее распознавание объектов контрольной выборки наблюдается для варианта 4.

Выделение перспективных площадей на основе теоретических представлений и опыта специалистов геологов (эвристическое моделирование). Использование АИПС/ЕС открывает широкие возможности для автоматизированного оперативного построения прогнозных карт на основе теоретических знаний и практического опыта работы специалистов геологов. Такого рода эвристическое моделирование может выполняться на основе различных эвристических алгоритмов.

Рассмотрим алгоритм эвристического моделирования (АЭМ), предложенный В. В. Марченко.

Пусть какой-либо фактор X_j , описывающий некоторое свойство объекта, предмета, явления и т. п., разделяется на K отдельных интервалов или градаций. Обозначим через $x_j^1, x_j^2, \dots, x_j^k$ — интервалы, на которые разбита область значения j -го фактора. Пусть на основе теоретических представлений, имеющегося опыта и интуиции специалист в данной области присвоит каждому из указанных интервалов некоторый информативный вес (значимость) или вклад относительно информационного описания всего объекта в целом

$$i_j^1, i_j^2, \dots, i_j^k.$$

Таким образом, последовательно различным интервалам всех факторов, описывающих конкретный объект, будут даны информационные веса, и специалистом будет сформирована информационная модель объекта.

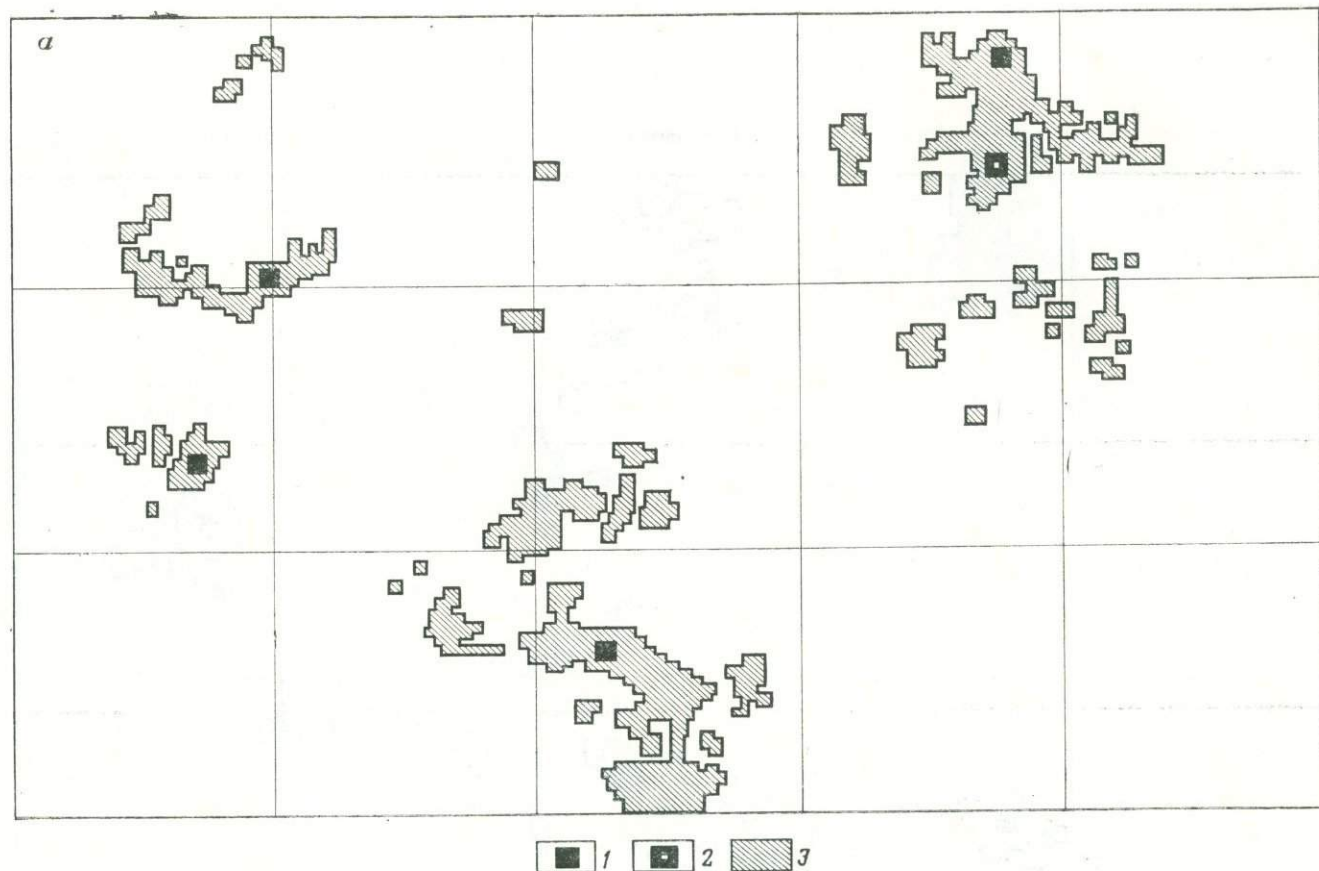
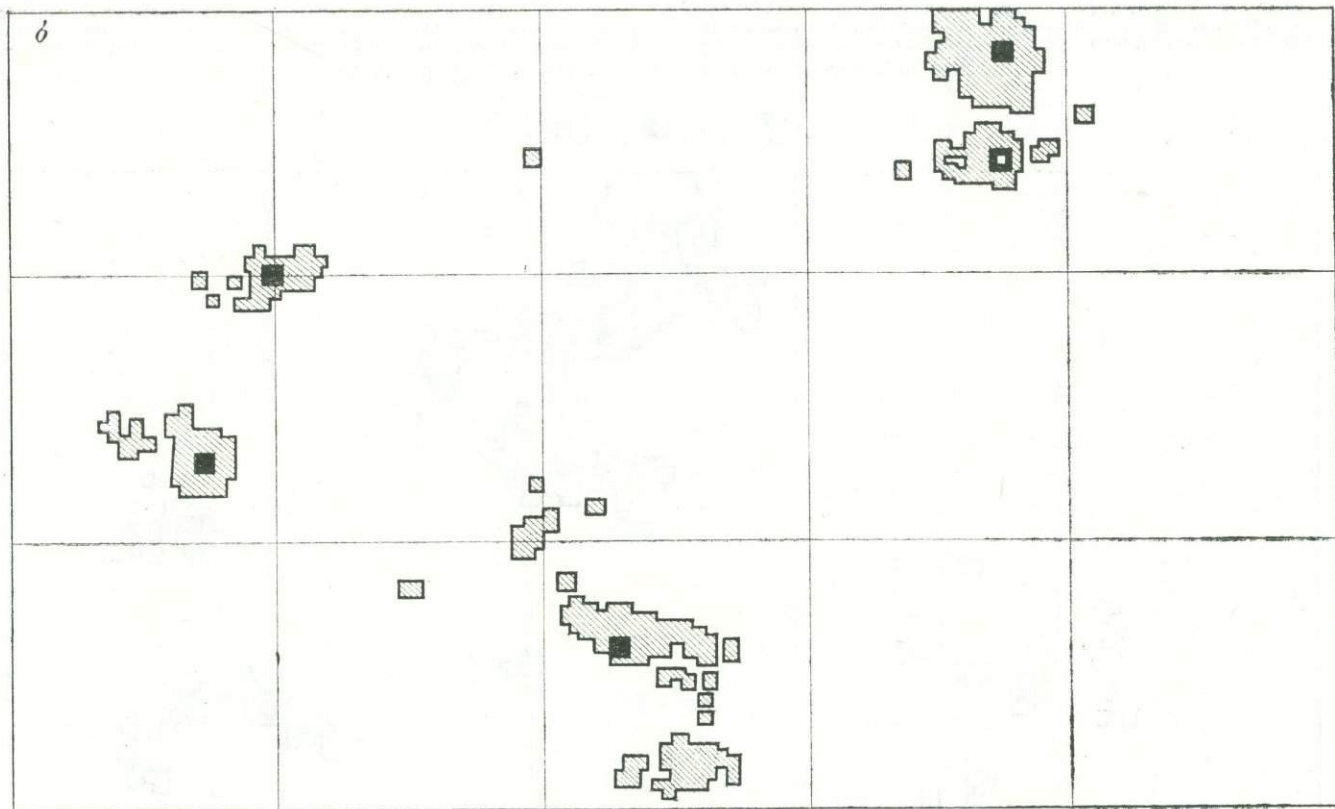
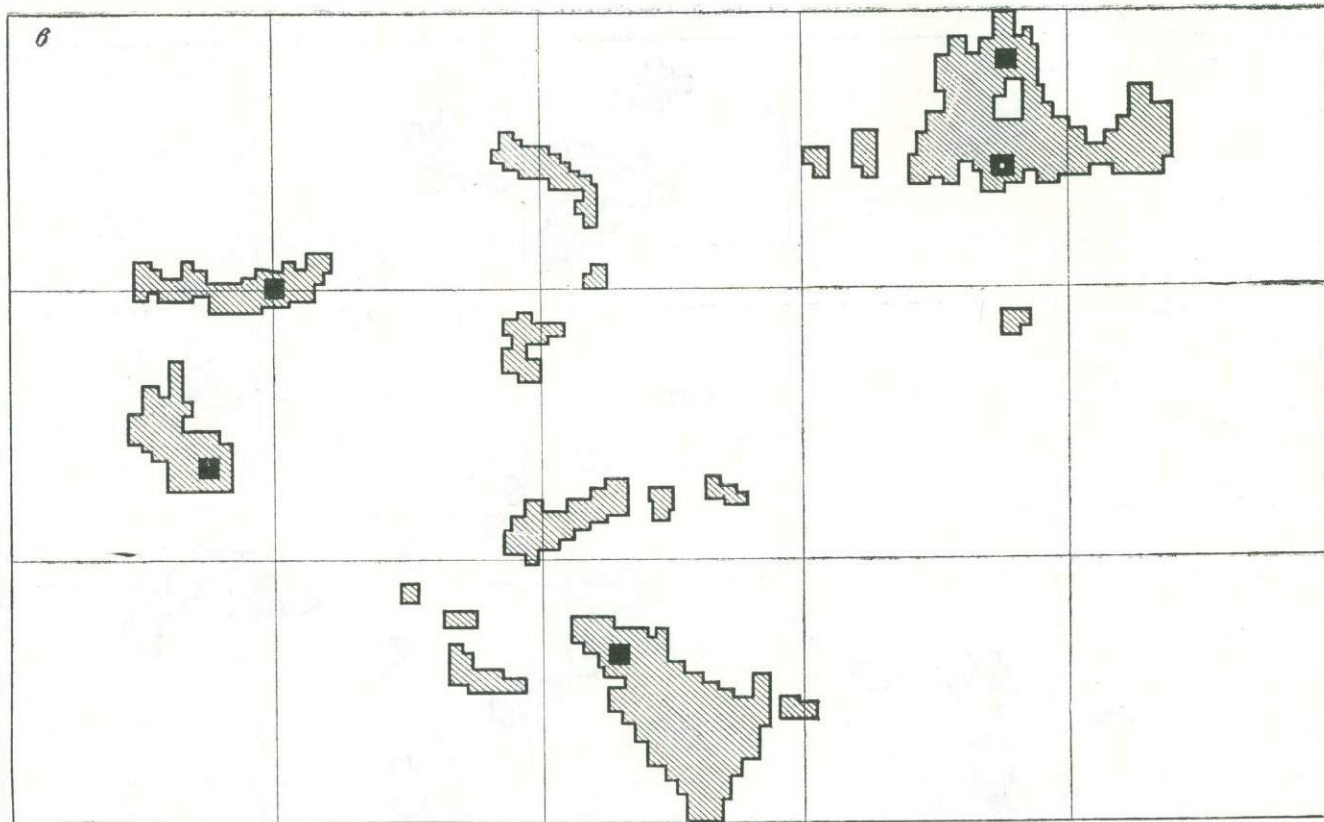


Рис. 20. Результаты многовариантного прогнозирования.

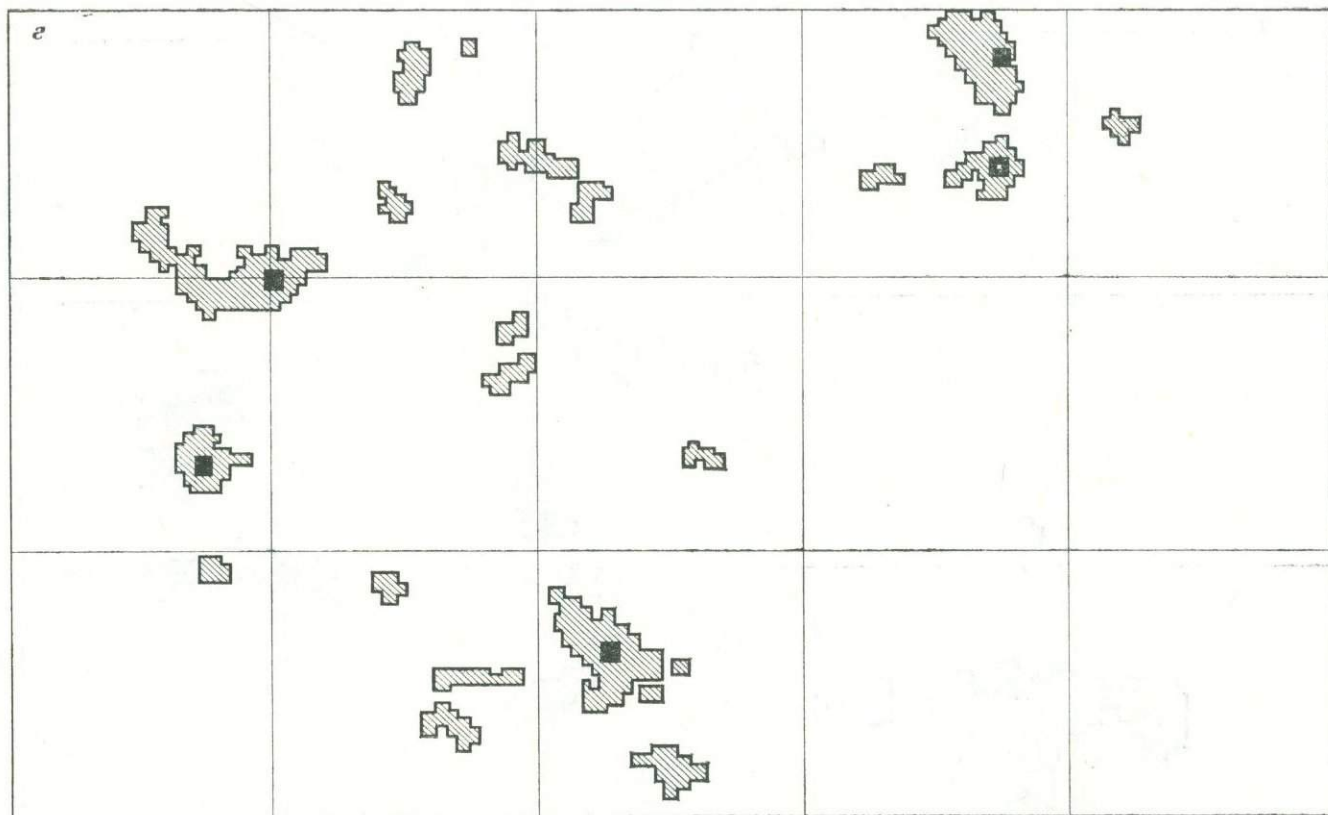
a — вариант 1 — генетическая связь оруденения с интрузиями комплекса В; *б* — вариант 2 — генетическая связь оруденения с малыми интрузиями комплекса В; *в* — вариант 3 — генетическая связь оруденения с интрузивным комплексом А; *г* — вариант 4 — генетическая связь оруденения с карбоновым магматизмом; *д* — вариант 5 — преимущественный контроль оруденения литолого-стратиграфическими, структурно-тектоническими и другими факторами.
 1 — эталонные объекты обучения распознаванию, 2 — контрольный объект, 3 — перспективные площади



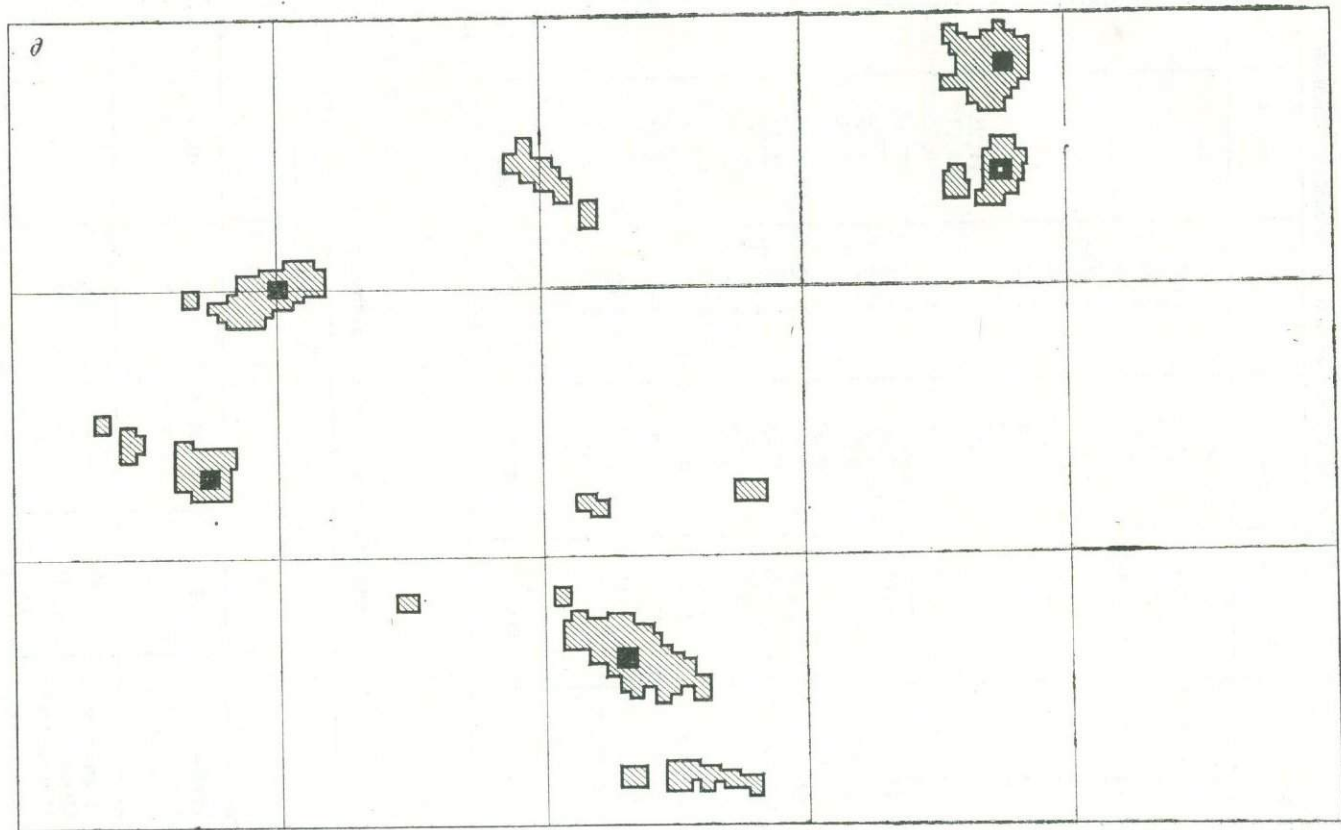
Продолжение рис. 20.



Продолжение рис. 20.



Продолжение рис. 20.



Продолжение рис. 20.

**Информационная значимость факторов по выборкам обучения,
в % относительных**

№ факторов	Информационная значимость по различным концепциям				
	Вариант 1	Вариант 2	Вариант 3	Вариант 4	Вариант 5
1	—	—	6,04	—	4,07
2	—	6,65	6,29	—	4,77
3	3,34	2,08	2,19	—	1,55
4	7,20	3,72	3,73	—	2,28
5	5,58	3,45	2,74	—	2,19
6	—	—	7,62	11,28	8,57
7	—	—	4,08	6,38	3,12
8	—	5,01	7,62	—	5,67
9	—	—	—	—	3,74
10	—	7,97	—	11,28	5,61
11	—	—	—	6,71	3,70
12	5,43	5,80	4,93	6,76	4,22
13	—	—	—	—	1,93
14	4,73	3,31	2,74	—	2,19
15	3,75	2,23	1,87	4,82	1,44
16	—	5,01	—	7,95	3,72
17	—	4,02	—	—	2,65
18	9,74	7,91	7,62	9,51	5,67
19	2,88	2,92	1,77	2,23	1,95
20	11,06	7,91	7,62	—	5,67
21	7,68	2,92	5,06	—	3,74
22	7,15	4,71	4,08	—	3,12
23	6,07	4,71	4,08	6,38	3,11
24	5,47	4,46	4,87	5,93	3,47
25	—	2,92	—	—	1,93
26	11,06	7,91	7,02	11,28	5,67
27	8,87	7,39	7,43	9,49	5,42
Итого	100,0	100,0	100,0	100,0	100,0

Таблица 22

Характеристика качества распознавания

Класс	Результаты классификации	Генетические концепции (варианты)				
		1	2	3	4	5
I	Правильно	104	111	111	108	111
	Отказ	7	—	—	3	0
	Неправильно	—	—	—	—	—
II	Правильно	84	84	84	94	84
	Отказ	27	27	27	17	27
	Неправильно	—	—	—	—	—
I и II	Правильно	188	195	195	202	195
	Отказ	34	27	27	20	27
	Неправильно	—	—	—	—	—

Для последующего поиска аналогов такого объекта на основе последовательного анализа массивов векторов введем процедуру дискретизации фактора.

Поставим в соответствие значению X_j отдельного фактора целое число m_j по следующему правилу:

$$m_j = l, \text{ если } X_j \in X_j^l.$$

Тогда классификация испытуемых векторов (поиск аналогов, отвечающих эвристически смоделированному информационному описанию объекта) сведется к суммированию по всем признакам информационных весов интервалов, в которые попадают значения соответствующих факторов классифицируемого объекта

$$M = \sum_{j=1}^n i_j^{m_j},$$

где n — количество факторов, описывающих искомый объект.

Комплексный поисково-оценочный критерий, построенный по этому алгоритму, представляет собой последовательность различных факторов с данными о важности отдельных их интервалов (с точки зрения специалиста геолога или группы экспертов) для локализации оруденения конкретного типа. Общий вид такого критерия приведен в табл. 23.

В таком виде комплексный поисково-оценочный критерий является концентрированным выражением знаний специалиста геолога и представляет собой своего рода модель объекта поиска. В результате классификации исследуемой территории (по информации банка данных) с помощью такого критерия на ЭВМ оперативно может быть распечатана результативная карта с участками максимального совпадения «важных» для решения поставленной задачи факторов локализации оруденения искомого типа. Такая карта анализируется специалистами, при этом можно легко установить, за счет сочетания каких именно факторов выделяется каждая конкретная площадь.

Здесь необходимо ввести термин «порог» алгоритма эвристического моделирования. Максимальный порог АЭМ характеризует максимально возможное количество положительных информационных весов по всем факторам. Естественно, что в практике редко наблюдается такое полное совпадение всех благоприятных факторов, поэтому часто приходится снижать величину максимального порога АЭМ, добываясь, например, выделения площадей с порогом 0,9 или 0,75 максимального.

Изменяя сочетания признаков и варьируя их относительными информативными весами, специалист геолог получает возможность выполнять многовариантный анализ перспективности района.

Таким образом, предлагаемый подход открывает широкие возможности оперативной оценки конечных результатов прогнозирования на основе различных генетических концепций, задаваемых отдельными геологами.

Практическая проверка излагаемого подхода была выполнена в исследуемом районе. Модели комплексного поисково-оценочного критерия медного оруденения были составлены тремя геологами (рис. 21). Первый из них (геолог Т.) составил модель (рис. 21, а), используя лишь теоретические представления. Второй специалист (геолог Н.) построил модель с учетом десятилетнего опыта работы в исследуемом районе (рис. 21, б). Третий специалист (геолог О.), проработавший в исследуемом районе более 20 лет, составил третью модель (рис. 21, в).

Пример решающего правила по алгоритму эвристического моделирования
В. В. Марченко.

№ факторов	Границы интервалов значений факторов* и их относительные информационные веса	Наименование факторов
1	$-км \leftarrow \begin{array}{c} \text{«0»} \\ X_1^1 \end{array} \xrightarrow{\text{контакт}} \begin{array}{c} \text{«1»} \\ X_1^2 \end{array} \rightarrow +км$	Граниты комплекса «Иг.»
2	$-км \leftarrow \begin{array}{c} \text{«0»} \\ X_2^1 \end{array} \xrightarrow{\text{контакт}} \begin{array}{c} \text{«-1»} \\ X_2^2 \end{array} \rightarrow +км$	Отложения силура (S ₁)
3	$км \leftarrow \begin{array}{c} \text{«0»} \\ X_3^1 \end{array} \xrightarrow{\text{контакт}} \begin{array}{c} \text{«3»} \\ X_3^2 \end{array} \xrightarrow{\text{контакт}} \begin{array}{c} \text{«0»} \\ X_3^3 \end{array} \rightarrow км$	Зона разломов «В»
...
n-1	$\begin{array}{cccccccc} & 1 & 10 & 100 & 1000 & & & \\ & \text{«0»} & \text{«0»} & \text{«0»} & \text{«1»} & \text{«2»} & & \\ \text{«1»} & X_{n-1}^1 & X_{n-1}^2 & X_{n-1}^3 & X_{n-1}^4 & X_{n-1}^5 & & \end{array} \rightarrow \text{г/т}$	Геохимические ореолы свинца
n	$\begin{array}{cccccccc} & 2 & 5 & 8 & & & & \\ & \text{«1»} & \text{«2»} & \text{«3»} & \text{«5»} & & & \\ \text{«1»} & X_n^1 & X_n^2 & X_n^3 & X_n^4 & & & \end{array} \rightarrow \frac{\text{мг/л}}{\text{км}}$	Горизонтальные градиенты Δg

Здесь 1, 2, ..., n — номера отдельных факторов; X₁¹, X₁², ... — интервалы значений отдельных факторов, выделенные специалистом геологом из содержательных соображений, i_j = «-1», «0», «1», «2», ... — относительные информационные веса интервалов значений признаков, установленные специалистом геологом на основе теоретических предположений и опыта.

* Для картографических факторов могут быть выбраны интервалы в виде расстояний от контакта, для числовых факторов — любые заданные значения.

Вполне естественно, что выбор факторов в рассматриваемом примере был ограничен информацией, имеющейся в территориальном банке данных АИПС/ЕС.

Сравнение этих критериев показывает, что специалисты по-разному оценивают значения отдельных факторов. Так, например, геолог Н. в два раза выше оценивает влияние на рудоконтроль (локализацию оруденения) даек малых интрузий, нежели специалист Т. В то же время специалист Т. отрицательно оценивает фактор наличия отложений силура (фактор № 5). Геолог О. негативно оценивает фактор девонских интрузивов и т. п. (табл. 24).

Прогнозные карты, построенные по рассмотренным эвристическим критериям, приведены на рис. 22.

Анализируя результаты, можно, во-первых, отметить, что все известные в районе медные месторождения попадают в выделенные перспективные площади. Этот факт косвенно подтверждает достаточную представительность исходной информации и правильность определения гео-

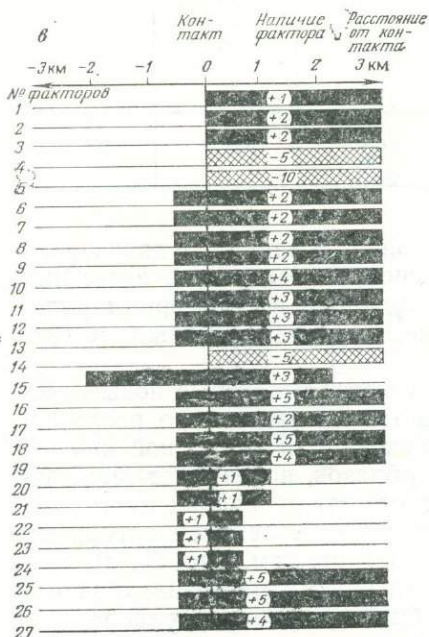
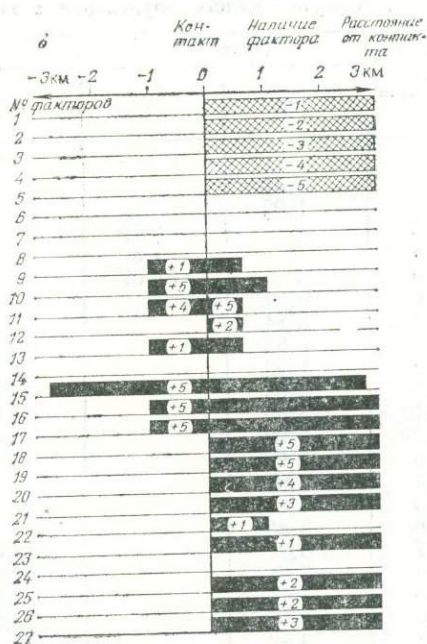
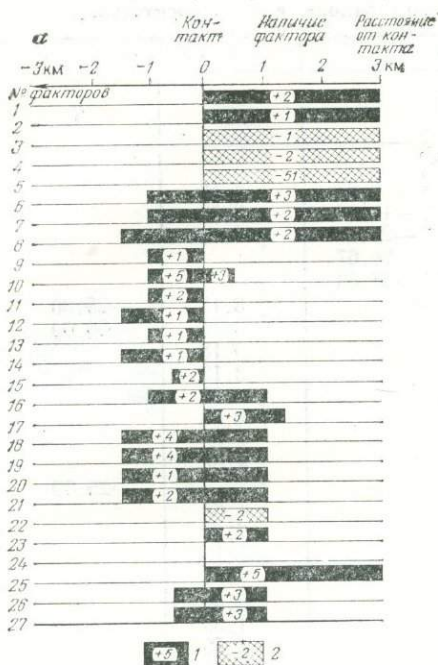


Рис. 21. Модели поисково-оценочного критерия медного оруденения (по алгоритму эвристического моделирования В. В. Марченко).

a — модель поисково-оценочного критерия медного оруденения, составленная геологом Н.; *в* — то же, составленная геологом О.

1 — области распространения (для картографического) или величины значений (для числового) факторов, благоприятных для локализации медного оруденения и их относительные информационные веса; *2* — то же, для факторов, неблагоприятных для распространения медного оруденения.

Сравнительная оценка относительной информативности факторов (i_j) локализации медного оруденения в исследуемом районе, в % относительных

№ факторов	Геолог Т.		Геолог Н.		Геолог О.	
	+ i_j	- i_j	+ i_j	- i_j	+ i_j	- i_j
1	3,84			-6,67	1,56	
2	1,92			-13,33	3,13	
3		-1,79		-20,00	3,13	
4		-3,57		-26,67		
5		-91,07		-33,33		
6	5,77		—	—	3,13	-25,00
7	3,84		—	—	3,13	-50,00
8	3,85		—	—	3,13	
9	1,92		1,81		3,13	
10	9,62		9,09		6,25	
11	3,85		9,09		4,68	
12	1,92		3,63		4,68	
13	1,92		1,81		4,68	
14	1,92		—	—		-25,00
15	3,85		9,09		4,68	
16	3,85		9,09		7,82	
17	5,77		9,09		3,13	
18	7,70		9,09		7,81	
19	7,70		9,09		6,25	
20	1,92		7,27		1,56	
21	3,84		5,45		1,56	
22		-3,57	1,81		1,56	
23	3,84		1,81		1,56	
24	—	—	—	—	1,56	
25	9,62		3,63		7,82	
26	5,77		3,63		7,81	
27	5,77		5,45		6,25	
	100	100	100	100	100	100

логами информационной значимости отдельных факторов рудоконтроля в принятом масштабе исследований. Успешное выделение известных месторождений на прогнозной карте, построенной по критериям рудоконтроля АЭМ специалиста геолога, является принципиальным методологическим положением.

Во-вторых, на прогнозных картах выделяются некоторые новые площади, поисковое значение которых должно быть определено после содержательного анализа. Сопоставление этих карт с прогнозной картой, построенной по алгоритму распознавания образов, показывает, что размеры перспективных площадей в районах известных мелких месторождений на прогнозных картах, выполненных с использованием алгоритма эвристического моделирования, несколько меньше (см. рис. 20, *г*). Повидимому, при использовании принятого алгоритма распознавания образов существенное влияние оказывает общая ситуация в районе эталонных объектов обучения.

В-третьих, прогнозные карты, построенные с использованием АЭМ, имеют удовлетворительную сходимость с прогнозной картой, построенной традиционным способом коллективом специалистов (см. рис. 17).

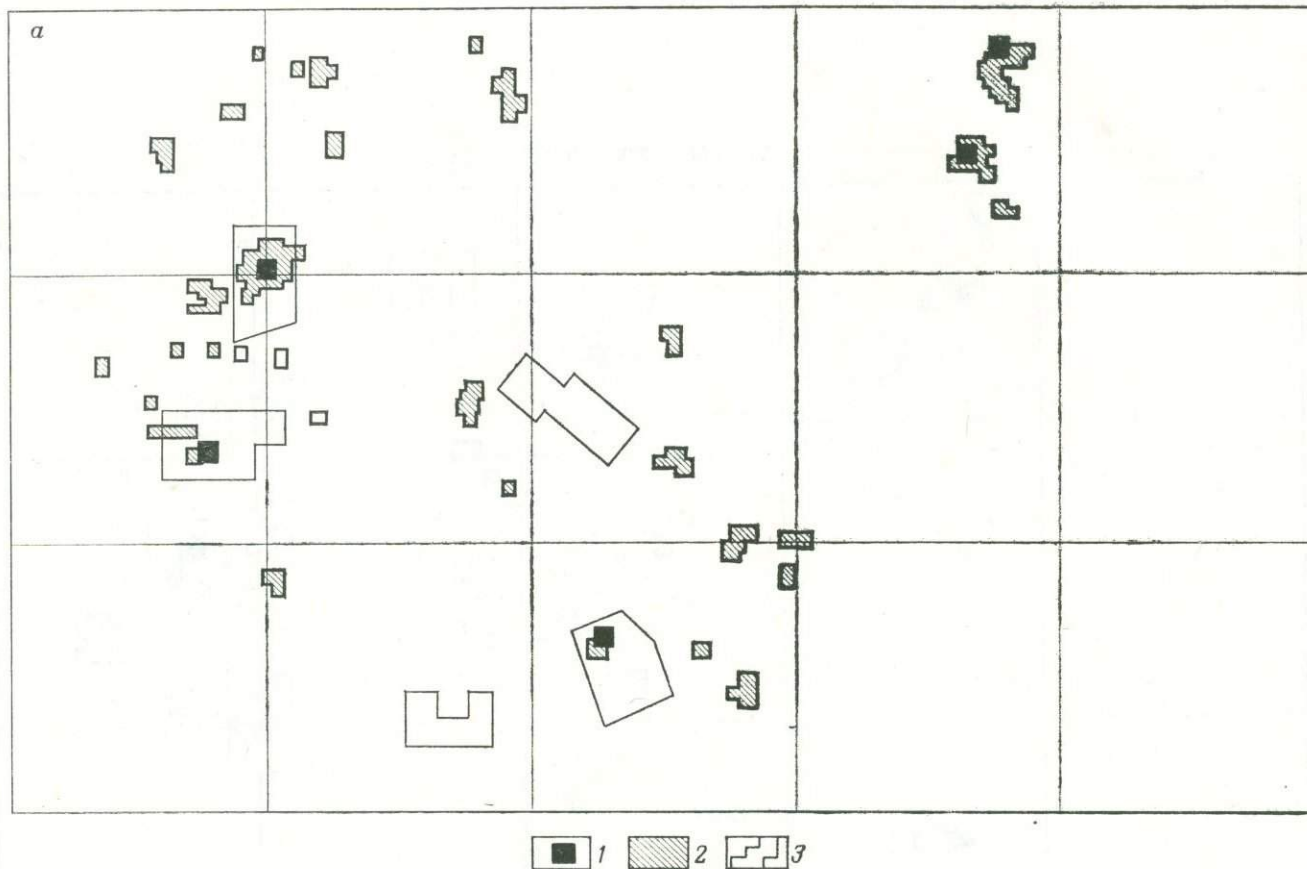
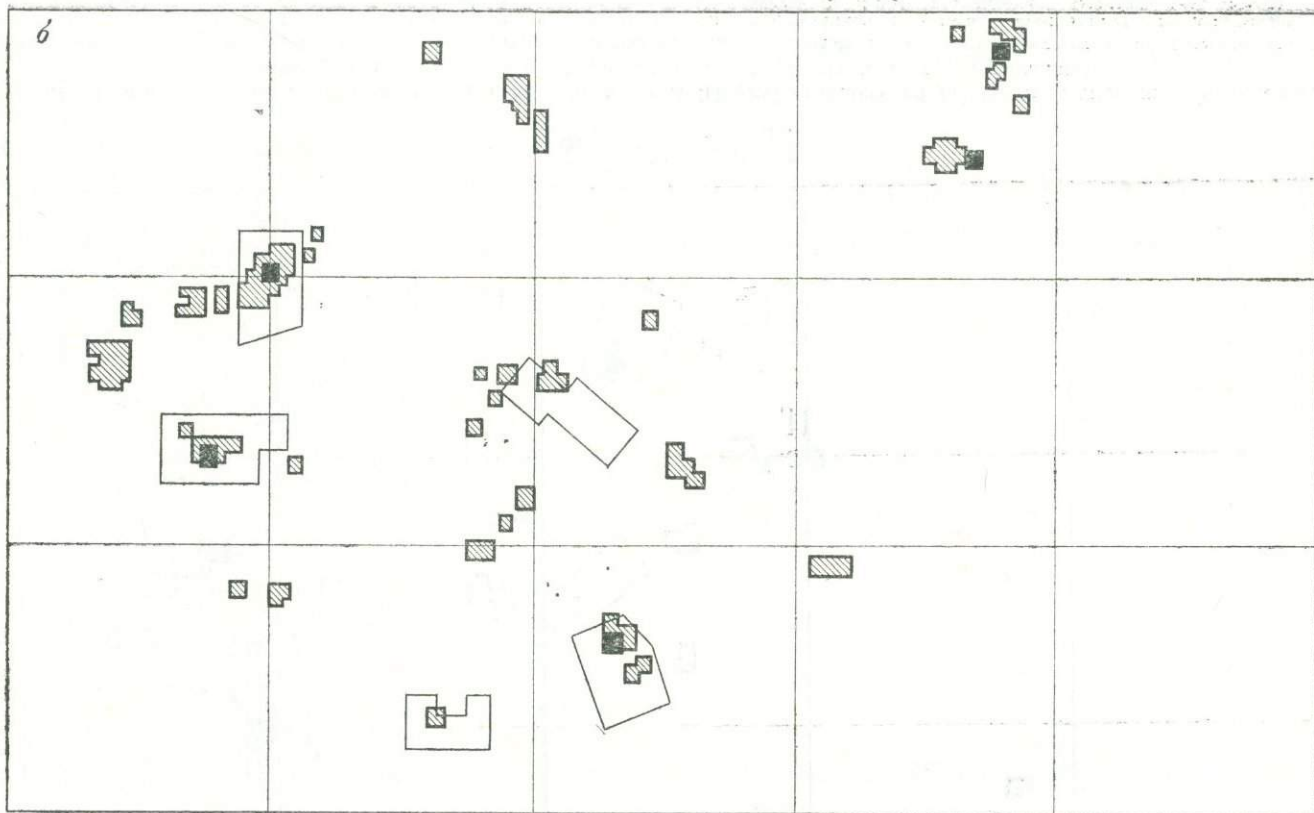
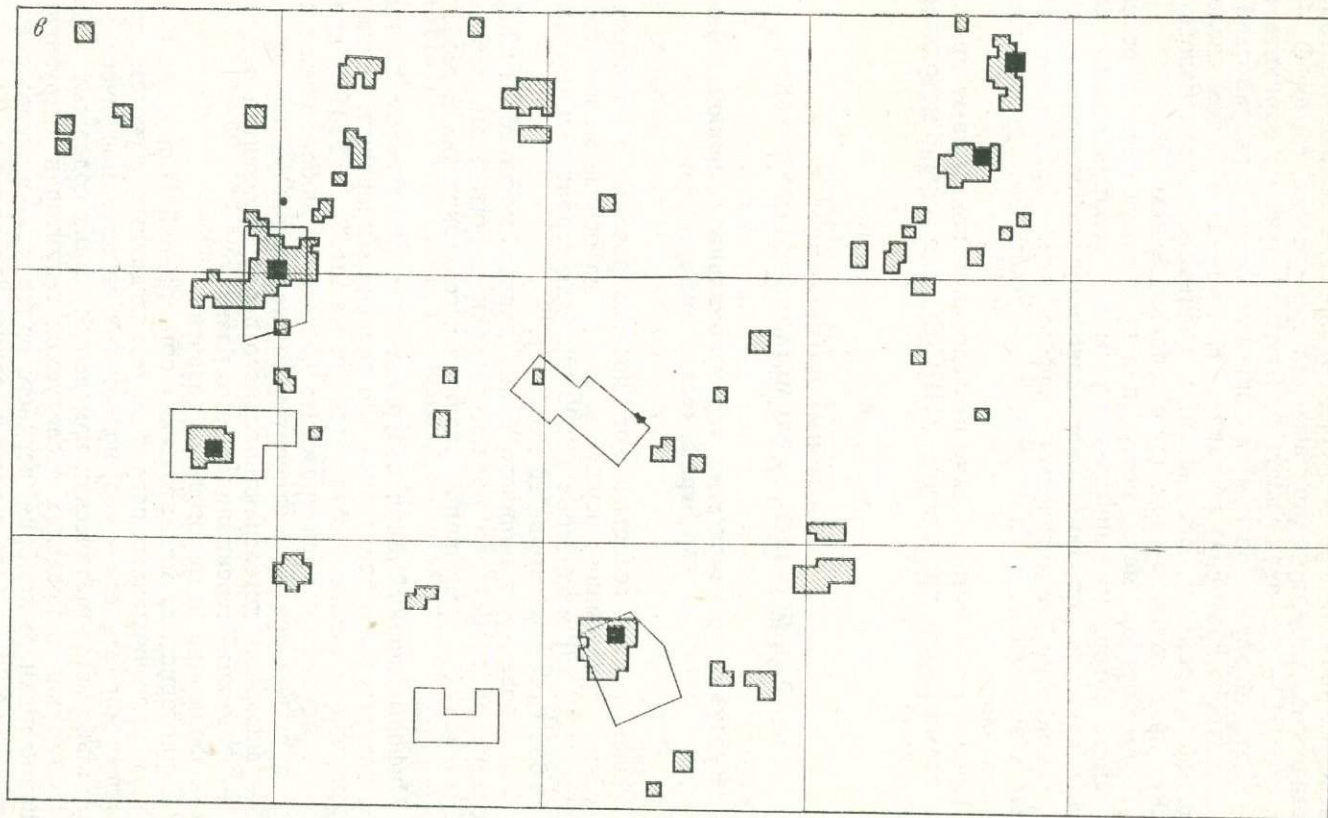


Рис. 22. Прогнозные карты, построенные с использованием АИПС/ЕС «Регион» на основе теоретических представлений и опыта специалистов геологов (по алгоритму эвристического моделирования).

a — карта, построенная по критерию геолога Т.; *б* — карта, построенная по критерию геолога Н.; *в* — карта, построенная по критерию геолога О.
 1 — известные медные месторождения, 2 — перспективные площади, характеризующиеся сочетанием критериев, благоприятных для локализации медного оруденения, 3 — перспективные площади первой очереди, выделенные коллективом специалистов на основе традиционной методики металлогенического прогнозирования (см. рис. 17)



Продолжение рис. 22.



Продолжение рис. 22.

На рис. 22 можно видеть, что из пяти перспективных площадей, выделенных коллективом специалистов на основе традиционной методики геологического прогнозирования и рекомендованных для проведения поисковых работ в первую очередь, геолог Т. с помощью алгоритма эвристического моделирования выделил три участка, геолог О. — четыре, а геолог Н. — все пять участков. Время получения варианта прогнозной карты по эвристическому алгоритму (вместе с разработкой поисково-оценочного критерия) составляет всего 1—2 ч (в том числе 7 мин машинного времени). Все указанные специалисты дали положительную оценку предложенному подходу с использованием АЭМ.

Таким образом, метод построения прогнозных карт на основе предлагаемого алгоритма эвристического моделирования можно рассматривать в качестве дополнительного современного инструмента оперативной обработки геологической информации. Применение новейших средств визуализации позволяет осуществить его дальнейшее совершенствование.

Проведенные методические исследования показывают принципиальные возможности применения АИПС/ЕС для научно-методических исследований.

2. ПРИМЕНЕНИЕ АИПС ДЛЯ ПРОГНОЗА РЕДКОМЕТАЛЬНОГО ОРУДЕНЕНИЯ

Краткая характеристика территориальных машинных банков картографической информации

Решение задач геологического прогнозирования и выполнение методических исследований осуществлялось на основе машинных банков данных [39, 95] масштабов 1 : 200 000 (для территории 105 тыс. км²) и 1 : 50 000 (для территории 20 тыс. км²).

Формированию машинного банка данных масштаба 1 : 200 000 по исследуемой территории предшествовали значительные работы по систематизации и унификации различных геологических и геофизических материалов.

Информационный банк содержал представительные данные, качество которых было предварительно оценено экспертами. Таким образом, с помощью машинного банка данных масштаба 1 : 200 000 представлялось возможным охарактеризовать положение любой точки в пределах исследуемой территории относительно следующих факторов:

- формаций догеосинклинального периода развития региона;
- формаций геосинклинального развития;
- формаций посторогенного развития.

Характеристика магматических образований выполнена с использованием многочисленных петрохимических данных, с учетом структурной позиции магматических тел, возможных автометаморфических и гидротермально-метасоматических изменений. Банк содержал сведения о расположении в пределах исследуемой территории интрузивных образований нижнего кембрия, верхнего ордовика — нижнего силура, верхнего силура — нижнего девона и верхнего девона.

Структурно-тектонические факторы были представлены главным образом геофизическими данными (зоны разломов, значения аномалий горизонтального градиента силы тяжести, наличие локальных гравита-

ционных и магнитных аномалий и др.). В целом банк данных содержал около 1 млн. чисел.

На основе этой исходной информации автоматически формировались различные варианты целевых массивов, используемых при решении конкретных задач геологического прогнозирования.

Первичный массив машинного банка масштаба 1 : 50 000 содержал сведения о расположении в пределах исследуемой территории 75 различных геологических и геофизических факторов, оказывающих (по мнению экспертов) наибольшее влияние на локализацию редкометалльного оруденения (фациальный состав толщ, гидротермально-метасоматические изменения, непосредственно предшествующие оруденению, первичные и вторичные ореолы рассеяния и т. д.). Таким образом, в банке сохранялись факторы, характеризующие поверхность доюрского фундамента, и факторы, отражающие состав и строение докембрийского фундамента, а также схемы интерпретации геофизических данных (модели строения верхней части земной коры).

Решение задач геологического прогнозирования

Выделение рудных узлов редкометалльных месторождений

Задача мелкомасштабного геологического прогнозирования в данном случае формулировалась следующим образом: «По имеющейся представительной информации банка данных выявить площади, аналогичные по комплексу признаков известным рудным узлам гидротермальных редкометалльных месторождений». Первичный массив банка данных содержал геологическую карту региона масштаба 1 : 200 000, а также геофизические материалы.

Для решения задачи был сформирован целевой массив, содержащий характеристики структурно-тектонических, литолого-стратиграфических и магматических факторов (всего 42 фактора).

В качестве эталонных объектов I класса при решении задачи были приняты рудные узлы, включающие известные месторождения. II класс был представлен площадями, в пределах которых, несмотря на достаточную геологическую изученность, были выявлены лишь мелкие месторождения и рудопроявления.

Задача решалась с использованием алгоритма распознавания образов последовательно в два этапа. На первом были выделены потенциально перспективные рудные площади (рудная выборка включала объекты I и II классов). На втором этапе в пределах перспективных площадей происходило выделение площадей, перспективных для обнаружения объектов I класса.

В результате выполненных исследований была составлена прогнозная карта для территории 100 тыс. км² в масштабе 1 : 200 000, при этом было выявлено 19 участков общей площадью 11,5 тыс. км². Надежность выделения перспективных площадей, определенная по результатам испытаний контрольной выборки, составила 85%. Экспертной оценкой специалистов подтверждена геологическая перспективность выделенных площадей. Все участки располагаются в пределах территорий, подлежащих проведению поисково-картировочных работ. Это обстоятельство позволяет сконцентрировать поисково-картировочные работы в первую очередь на наиболее перспективных районах.

Выделение рудных узлов месторождений благородных металлов

В качестве примера многоцелевого использования территориального машинного банка данных было выполнено геологическое прогнозирование рудных узлов месторождений благородных металлов. Задача прогнозирования состояла в выделении площадей, аналогичных по комплексу признаков известным в пределах исследуемого региона рудным узлам эндогенных месторождений благородных металлов. Общая территория, в пределах которой осуществлялось прогнозирование, — 60 тыс. км²; масштаб исследований — 1 : 200 000. Эталонными объектами I класса являлись известные промышленные эндогенные месторождения благородных металлов, расположенные в пределах изучаемой территории. Мелкие непромышленные месторождения и рудопроявления вошли во II класс эталонных объектов.

При формировании целевого массива для решения задачи была предварительно определена информационная значимость отдельных факторов банка данных. Для этого по всем данным, сохраняемым в банке, были построены гистограммы распределений значений отдельных факторов в пределах эталонных объектов. Эта процедура позволила выделить наиболее важные для решения этой задачи факторы (всего 37 факторов) и сформировать из них целевой массив. На прогнозном карте, составленной на основе алгоритма распознавания с обучением выделено 11 участков общей площадью 4,5 тыс. км². Эти участки включают известные месторождения, а также отдельные площади, в пределах которых детальных поисковых работ не проводилось.

В результате решения оказалось также возможным выделить факторы, наиболее благоприятные для локализации оруденения в пределах изучаемого региона. Так, установлено, что из 37 использованных для решения факторов 18 несут 70% полезной информации. К ним в первую очередь относятся некоторые особенности геофизических полей, характеризующие особенности глубинного геологического строения, наличие интрузивных и метаморфических пород определенных комплексов.

Выделение рудных полей редкометалльных месторождений

Среднемасштабное геологическое прогнозирование было выполнено на площади 20 тыс. км² в масштабе 1 : 50 000 в пределах участков, выделенных в качестве перспективных при мелкомасштабном (1 : 200 000) прогнозировании. Задача формулировалась следующим образом: «Выделить площади, аналогичные по комплексу признаков известным рудным полям редкометалльных месторождений для постановки поисково-съемочных работ».

Исходными данными для решения прогнозной задачи явились материалы, полученные в результате анализа имеющихся геологических и геофизических данных, выполненного по традиционной схеме: определение природы геофизических аномалий, проведение количественных расчетов, классификация аномалий, построение схем интерпретации и разрезов. Конечные результаты интерпретации вместе с геологической картой района были использованы для формирования целевого массива банка данных. Схема геологической интерпретации глубинного строения исследуемого района содержала характеристику магматиче-

ских образований, некоторые особенности строения метаморфических пород фундамента и основные элементы тектонического строения.

Интерпретация данных разрывной тектоники была выполнена на основе интерпретации гравимагнитных данных и дешифрования высотных аэрофотометрических снимков. Использование этих материалов обусловлено также и тем, что они сопоставляются с тектоническими схемами смежных районов. В качестве эталонных объектов были приняты известные месторождения и рудопроявления, расположенные в пределах исследуемой территории и относящиеся к одной рудной формации.

В результате прогнозирования были выделены участки, перспективные для проведения поисковых работ (около 5% от всей территории). Все участки располагаются в обстановке, перспективной, по мнению экспертов, для локализации оруденения. Вместе с тем общая площадь таких участков существенно меньше площадей, выделенных с использованием традиционной методики, что позволяет осуществить концентрацию поисково-оценочных работ.

Методические исследования

Многовариантный прогноз рудных зон редкометальных месторождений на основе различных генетических концепций

Наличие в системе АИПС машинного банка исходной информации и необходимого программного обеспечения позволяет оперативно формировать различные целевые массивы и изменять комплекс факторов в соответствии с выбранной геологической моделью.

Для решения задачи многовариантного прогнозирования в масштабе 1 : 200 000 были выбраны три конкурирующие гипотезы, в соответствии с которыми осуществлено исследование, позволившее получить конечные результаты геологического прогнозирования по каждой из гипотез. Эти гипотезы (модели) акцентируют внимание на различных сторонах процесса рудообразования. Первая из них предполагает, что оруденение обусловлено всем ходом геологического развития исследуемого участка земной коры и в общем виде соответствует идеям Ю. А. Билибина [8]. Вторая гипотеза предполагает доминирующую роль факторов структурного контроля оруденения и подчиненное значение литологических факторов; при этом не отрицается роль вмещающих пород как возможного источника оруденения. Третья гипотеза предполагает связь оруденения с формированием гранито-гнейсовых куполов. Не останавливаясь подробно на рассмотрении каждой из гипотез, отметим, что все они построены на соответствующем фактическом материале. Таким образом, исследование представлялось как работа, направленная на решение прогнозной задачи в рамках конкурирующих моделей. Решение было осуществлено на основе составления многовариантных прогнозных карт и выборе лучшего варианта в соответствии с заданными критериями (количество и соотношение ошибок I и II рода, размеры выделенных площадей, соответствие их существующим геологическим представлениям). Общая методическая схема исследования приведена на рис. 23.

Первая модель предполагает использование большого числа геологических характеристик, часть из которых, очевидно, не имеет прямого отношения к проблеме рудообразования. Поэтому для сокращения исходного пространства признаков строились одномерные распределения,

и дальнейшей обработке на ЭВМ подвергались лишь те из них, которые являются «делящими» для эталонных объектов I и II классов (промышленные месторождения и непромышленные оруденения). Такие одномерные распределения использовались для выявления статистических закономерностей пространственного положения месторождений относительно тех или иных образований, являющихся в отдельных случаях поисковыми критериями. По этой модели построен целевой массив 1 и на основании ее — рассчитаны прогнозные варианты 1.1, 1.2, 1.3 (см. рис. 23, а).

Вторая модель предполагала связь оруденения со становлением и развитием зон долгоживущих глубинных разломов. Принципиальными отличиями этой модели от предыдущей являются четкое выделение структурных факторов контроля, связь с глубинными источниками рудного вещества и, как следствие, пространственная приуроченность месторождений к вулканогенно-плутоническим комплексам, сформированным в завершающую стадию геосинклинали, т. е. с малыми интрузиями и дайками. В соответствии с этим при выборе и группировании факторов основное внимание уделялось расчленению вулканогенно-плутонических формаций орогенного периода развития, и, наоборот, сведения по формациям доорогенного периода представлялись в обобщенном виде. По данной модели был сформирован целевой массив 2 банка данных и рассчитаны варианты 2.1, 2.2 и 2.3.



Рис. 23. Методические схемы многовариантного прогнозирования редкометалльного оруденения.

а — среднемасштабное прогнозирование в рамках различных геологических моделей (варианты 1, 2 и 3 для каждого целевого массива формировались на основе различных эталонных выборок: 1 — все объекты, 2 — объекты, связанные с формацией березитов, 3 — объекты, связанные с формацией пропицитов); б — крупномасштабное прогнозирование на основе различных моделей строения верхней части земной коры

Третья модель устанавливает прямую зависимость между формированием и развитием гранитоидных массивов и оруденением. В такой постановке определяющими являются факторы метаморфизма и ультраметаморфизма и связанные с ними процессы миграции и локализации рудного вещества под воздействием тепловых потоков, идущих как из мантии, так и из гранитного слоя коры в ходе гранитообразования. В соответствии с данной моделью был сформирован целевой массив 3 банка данных и реализованы варианты 3.1 и 3.2.

Оценка вариантов прогнозных карт, построенных на основе различных моделей, приведена в табл. 25.

Содержательный анализ ошибок первого рода позволяет сделать вывод о том, что определенные идеи, развиваемые в работах различных исследователей, дают хорошие результаты при решении прогнозных задач только в конкретных ситуациях. Так, например, наилучшие результаты для центральной части региона дает «купольная» модель. В то же время она неприемлема для юго-западной и южной частей рассматриваемого региона. Аналогичный вывод можно сделать и в отношении модели 2.2 и др.

Таким образом, в результате анализа различных вариантов прогнозирования установлено, что в целом для рассматриваемого региона мо-

Таблица 25

Оценка вариантов прогнозных карт масштаба 1 : 200 000

Вариант	Число ошибок на контрольной выборке		Суммарные размеры перспективных площадей, % от общей площади	Особенности пространственного размещения перспективных участков	Соответствие принятой генетической гипотезе
	I рода	II рода			
1.1	1	7	24,0		На всей территории региона в целом соответствует принятой модели
1.2	1	8	16,5	Отмечается преобладание участков в северной части территории	На всей территории региона в целом соответствует принятой модели
1.3	2	11	60,0		Полностью не соответствует принятой модели
2.1	1	5	13,4	Отмечается преобладание в восточной части территории	На всей территории региона в целом соответствует принятой модели
2.2	3	10	12,6		Частично соответствует принятой модели
2.3	—	11	75,0		Полностью не соответствует принятой модели
3.1, 3.2	3	6	9,6		Соответствует частично на отдельных площадях региона

гут быть приняты только модели 1.1 и 2.1, предполагающие связь формирования месторождений со всеми процессами геологического развития земной коры (модель 1). Все другие гипотезы, в которых резко превалирует роль какого-либо одного из многих процессов геологического развития, приемлемы для решения прогнозных задач только в пределах отдельных частей региона.

В заключение представляется возможным отметить, что применение методической схемы АИПС позволило без дополнительной подготовки исходной информации оперативно получить окончательные результаты прогнозирования на основе различных геологических моделей, произвести экспертную оценку этих результатов, установить область соответствия отдельных гипотез реальной геологической обстановке и выбрать наилучшую модель, которая может быть принята за основу при проведении поисково-съёмочных работ в целом по рассматриваемому региону.

Вместе с тем это не исключает применение других моделей для отдельных частей региона.

*Многовариантное прогнозирование
рудных полей редкометальных месторождений
на основе различных моделей строения
верхней части земной коры*

В основу исследования, выполненного в масштабе 1 : 50 000, были положены четыре варианта признакового описания модели строения верхней части земной коры в пределах рассматриваемой территории.

Первые три варианта характеризовались постепенным увеличением полноты описания геологического строения верхней части земной коры по геологическим и геофизическим данным, в четвертой описании производилось лишь по геофизическим данным.

В первом варианте использовались геологические данные по поверхности палеозойского фундамента; во втором варианте эти данные были дополнены характеристикой магматических образований, залегающих ниже поверхности фундамента. Третий вариант дополнительно включал описание состава и строения докембрийского фундамента, а также соотношение структурных элементов в различных структурных этажах. Кроме того, здесь были использованы сведения о субвулканических и экструзивных фациях (по гравимагнитным данным). Наконец, четвертый вариант был построен лишь на основе геофизических данных (гравirazведки и магниторазведки), при интерпретации которых в первую очередь учитывался состав горных пород на различных глубинах [39, 93].

По факторам, отобранным при формировании описанных четырех моделей, были созданы четыре целевых массива. Методическая схема решения задачи приведена на рис. 23, б.

В качестве эталонных объектов были приняты известные рудные поля, вмещающие гидротермальные месторождения (I класс) и непромышленные месторождения и рудопроявления (II класс). Все объекты располагаются в пределах одной структурно-металлогенической зоны. Исследования по всем моделям выполнены при одной гипотезе рудообразования и постоянной выборке обучения. В итоге исследования были получены четыре варианта прогнозных карт. Результаты многовариантного прогнозирования схематически изображены на рис. 24.

Оценка вариантов осуществлялась с учетом ошибок на контрольной выборке и соответствия принятой модели, как это показано в табл. 26.

Оценка вариантов решения задачи

Вариант	Число ошибок на контрольной выборке		Суммарные размеры перспективных площадей, в % от общей территории	Соответствие принятой модели
	I рода	II рода		
1	1	4	16,0	Частично не соответствует (часть перспективных площадей располагается среди непродуктивных формаций)
2	—	3	10,0	Соответствует
3	—	1	9,0	„
4	—	3	12,0	„

Как следует из табл. 26, вследствие отсутствия сведений по глубинному строению результаты прогноза по первому варианту (см. рис. 24, а) частично не соответствуют реальной геологической обстанов-

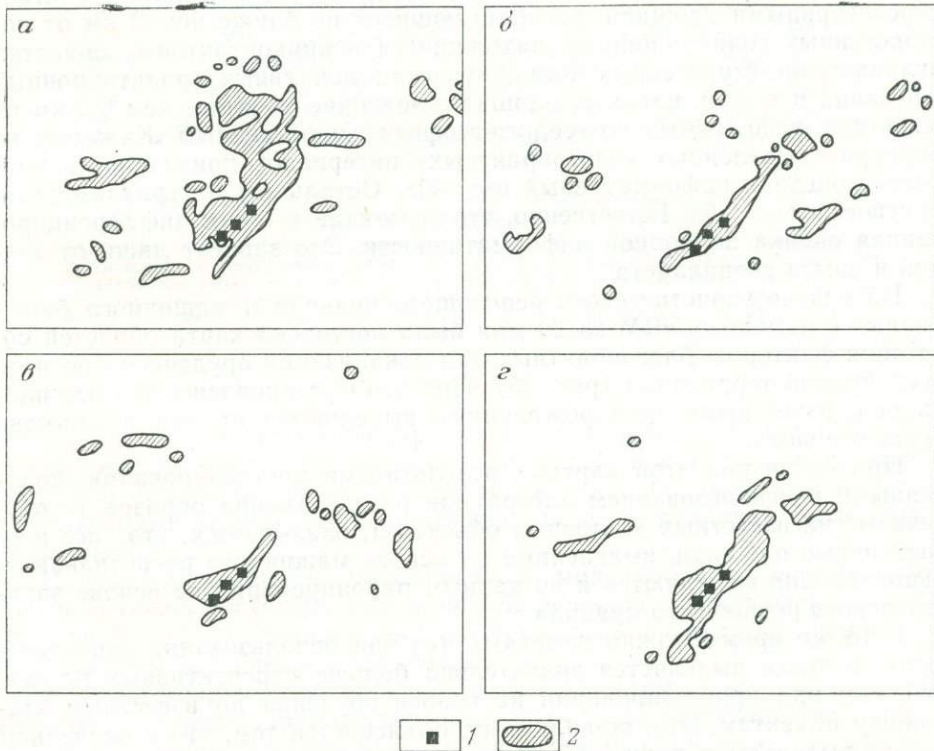


Рис. 24. Результаты многовариантного прогнозирования на основе 4 моделей строения верхней части земной коры (имитационное моделирование).

а — первый вариант, б — второй вариант, в — третий вариант, г — четвертый вариант.
1 — эталонные объекты, 2 — площади, аналогичные по комплексу признаков эталонным объектам

ке: некоторые перспективные участки располагаются среди непродуктивных формаций. Лучшие результаты получены по третьей модели (варианту), использующей наиболее полные данные о геологическом строении территории.

В то же время на модели, построенной по материалам интерпретации гравимагнитных данных, отражающим строение поверхности и глубинных горизонтов, получены результаты, более приемлемые, чем по первой модели, отражающей лишь данные о геологическом строении поверхности. Установлено, что геофизические данные, характеризующие глубинное геологическое строение рассматриваемой территории, вносят существенный вклад в решение задачи прогнозирования. Анализ выделенных по варианту 3 перспективных площадей, выполненный специалистами, устанавливает соответствие их существующим геологическим представлениям об особенностях локализации оруденения в районе.

*Выделение перспективных
и бесперспективных площадей
на основе эвристического моделирования*

Например, для определения площадей, перспективных на обнаружение редкометалльных гидротермальных месторождений конкретного типа, специалистом М. с помощью алгоритма эвристического моделирования было построено имитационное решающее правило (поисково-оценочный критерий). Из этого правила следует, что специалист считает перспективными площади, располагающиеся не ближе чем 2 км от посторогенных толщ; площади, находящиеся в приконтактных областях вулканогенно-терригенных толщ; участки локальных гравитационных аномалий и т. д. и, наконец, площади, лежащие не далее чем 0,5 км от разломов, выделяемых по геолого-геофизическим данным. Каждому из априорно выделенных «благоприятных» интервалов признаков специалист присвоил информативный вес «1». Остальным интервалам были присвоены веса «0». Естественно, что возможна и более дифференцированная оценка априорной информативности. Это зависит лишь от знаний и опыта специалиста.

На основе эвристического решающего правила и машинного банка данных с помощью ЭВМ за 20 мин была построена карта областей сочетания факторов, благоприятных для локализации оруденения по всей исследуемой территории (рис. 25). При этом установлено, что все площади с известными месторождениями выделяются на ней в качестве перспективных.

При сравнении этой карты с результатами прогнозирования, полученными с использованием алгоритмов распознавания образов (с обучением на известных эталонных объектах), оказывается, что все перспективные площади, выделенные на основе машинного распознавания, одновременно выделяются и по данным районирования на основе эвристического решающего правила.

В то же время можно заметить, что при использовании эвристического подхода выделяется значительно больше перспективных площадей, чем при прогнозировании на основе обучения по известным эталонным объектам. Это, по-видимому, объясняется тем, что в последнем случае выделяются лишь площади, аналогичные по сочетанию и информативному весу признаков площадям, в пределах которых располагаются конкретные эталонные объекты (месторождения), в то время

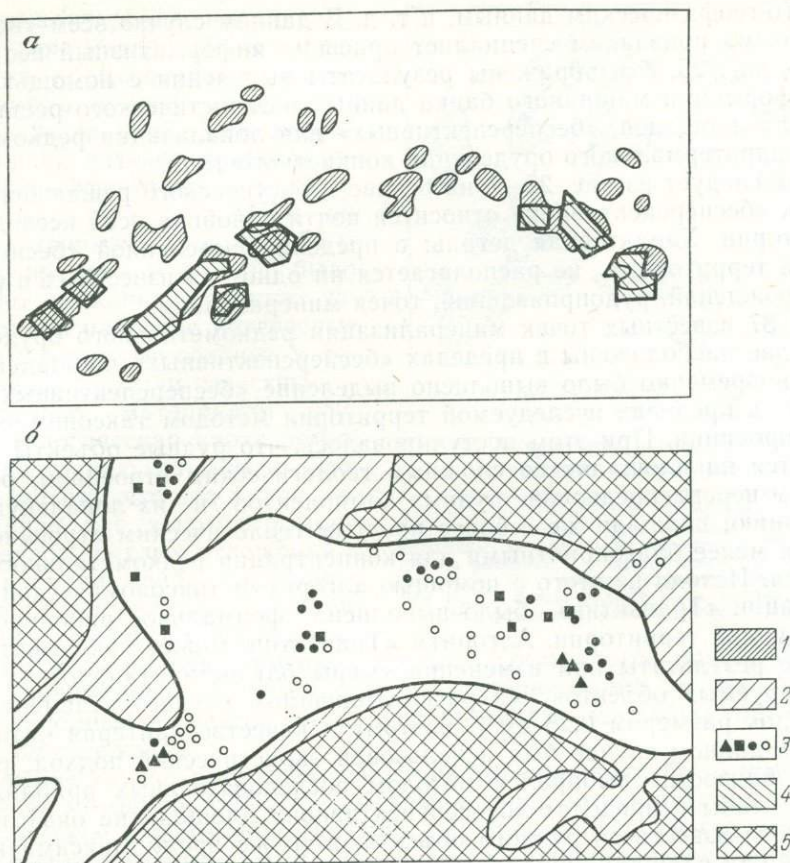


Рис. 25. Выделение перспективных (а) и бесперспективных (б) площадей на основе эвристического моделирования.

1 — перспективные площади, выделенные специалистом геологом по алгоритму эвристического моделирования (АЭМ); 2 — бесперспективные площади, выделенные специалистом геологом по АЭМ; 3 — рудные объекты (месторождения, рудопроявления, точки минерализации); 4 — перспективные площади, выделенные при использовании алгоритма распознавания образов; 5 — бесперспективные площади, выделенные с использованием алгоритма таксономии

как при эвристическом моделировании выделяются все площади, благоприятные, с точки зрения специалиста, для концентрации гидротермального оруденения в более широком аспекте (в том числе и такого типа, который не известен в настоящее время в пределах исследуемой территории).

Используя алгоритм эвристического моделирования, специалистом М. было также построено имитационное решающее правило другого рода. Оно включало признаки «неблагоприятные», с его точки зрения, для локализации оруденения конкретного типа.

К «неблагоприятным» признакам в рассматриваемом аспекте специалист М. относит наличие посторогенных толщ, отсутствие или значительное удаление исследуемых участков от вулканогенно-терригенных толщ, размещение их в пределах «нормальных» геофизических полей, расположение участков далее чем 2,5 км от разломов, выделяемых по

геолого-геофизическим данным, и т. д. В данном случае всем «неблагоприятным» признакам специалист присвоил информативный вес «—1».

На рис. 25, б изображены результаты выделения с помощью ЭВМ по информации машинного банка данных и эвристического решающего правила площадей, «бесперспективных» для локализации редкометального гидротермального оруденения конкретного типа.

Как следует из рис. 25, б, на основе эвристического решающего правила к «бесперспективной» относится почти половина всей исследуемой территории. Характерная деталь: в пределах выделенной «бесперспективной территории» не располагается ни одно из известных в районе месторождений, рудопроявлений, точек минерализации.

Из 37 известных точек минерализации редкометального оруденения лишь две расположены в пределах «бесперспективных» площадей.

Одновременно было выполнено выделение «бесперспективных» площадей в пределах исследуемой территории методом таксономического районирования. При этом постулировалось, что рудные объекты располагаются на площадях со сложным геологическим строением, обусловившим перераспределение рудных компонентов и их локальную концентрацию; площади же с более простым геологическим строением являются менее благоприятными для концентрации редкометального оруденения. Исходя из этого с помощью алгоритма таксономической классификации «Транзитив» было выполнено формальное районирование исследуемой территории. Алгоритм «Транзитив» может показывать различные результаты при изменении «меры близости» (аналогии) классифицируемых объектов (в рассматриваемом случае — элементарных площадок размером 0,25 км²), поэтому в качестве критерия оптимальности в данном случае был использован эвристический подход: радиус «меры близости» выбирался методом последовательных приближений таким, чтобы в пределах «бесперспективной» площади не оказывалось ни одного известного рудного объекта; в то же время максимизировалась общая площадь, попадающая в «бесперспективный» таксон. Оптимальный вариант такого таксономического районирования показан на рис. 25, б. При сопоставлении этих данных с результатами выделения бесперспективных площадей на основе теоретических представлений и опыта специалиста (эвристическое моделирование) наблюдается достаточно хорошая сходимость. Однако специалист относит к «бесперспективным» несколько большие площади, как это видно на рис. 25, б.

Ретроспективный имитационный анализ эффективности использования геолого-геофизической информации

Суть настоящего исследования заключалась в том, чтобы на основе ретроспективного анализа совокупности геолого-геофизической и другой информации на различные временные периоды построить по методике АИПС прогнозные карты масштаба 1 : 200 000, соответствующие принятому периоду времени, и оценить качество этих карт со следующих позиций:

- размеры выделенных перспективных площадей;
- соответствие прогноза результатам проведенных в последующие годы геологоразведочных работ (в первую очередь положение вновь выявленных рудных узлов в пределах перспективных площадей, выделенных на основе ретроспективного анализа геолого-геофизической информации).

Проанализировав имеющиеся фактические материалы, историю открытия промышленных месторождений в районе, авторы исследования пришли к выводу, что наибольший интерес представляет оценка эффективности методики прогноза на основе АИПС по информационному уровню 1964 г. К тому времени в районе было известно несколько рудных узлов, причем наиболее интересные объекты, открытые позднее, к тому времени не были известны.

Исходным материалом служила геологическая карта региона по состоянию на 1964 г. Следует отметить, что эта карта, составленная для закрытого мощным чехлом рыхлых отложений к тому времени слабо изученного района, базировалась в значительной степени на материалах интерпретации гравимагнитных съемок. Кроме того, использовалась карта гравимагнитных съемок и схема интерпретации по ним. В качестве эталонных объектов были использованы известные на начало 1964 г. рудные узлы.

Решение задачи ретроспективного прогнозирования было выполнено на основе методики АИПС путем комплексной интерпретации геолого-геофизической информации (на 1964 г.), представленной геологическими факторами и некоторыми элементами глубинного строения, определенными по геофизическим данным.

Использованная для решения ретроспективного прогнозирования геолого-геофизическая информация с точки зрения современных знаний о районе была весьма схематичной. Это в первую очередь относилось к геологическим факторам, в частности к возрастной индексации отдельных формаций, а также к имеющимся эталонным объектам, принадлежащим к различным рудным формациям и отличающимся характером гидротермально-метасоматических изменений.

Решение задачи классификации, получение решающего правила и определение информационной значимости отдельных признаков было выполнено с помощью алгоритма «Гиперпласт» в соответствии с принятой методической схемой.

На основе полученного по исходной информации решающего правила была построена прогнозная карта для площади около 40 тыс. км², часть которой изображена на рис. 26.

Анализ полученной, таким образом, ретроспективной имитационной прогнозной карты показал следующее.

1. Наиболее важные в промышленном отношении объекты, открытые после 1964 г., располагаются в пределах контуров перспективных площадей, выделенных по методике АИПС на геолого-геофизических материалах 1964 г.

2. Общее количество перспективных площадей составило 20% от всей территории исследования.

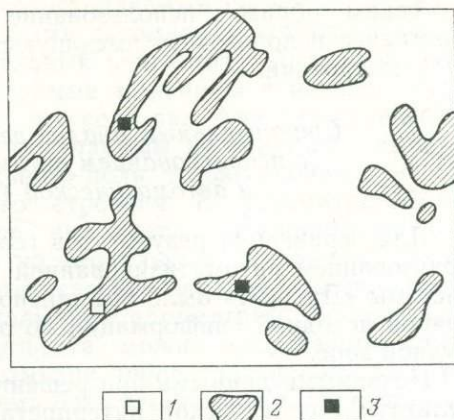


Рис. 26. Результаты ретроспективного геологического прогнозирования.

1 — объекты обучения — рудные поля гидротермальных месторождений, известные до 1964 г. (остальные объекты обучения распознаванию расположены за пределами территории, изображенной на рисунке), 2 — перспективные площади, выделенные по геолого-геофизическим данным 1964 г., 3 — новые рудные поля, открытые геолого-поисковыми работами после 1964 г.

Таким образом, использование АИПС позволяет существенно эффективнее использовать имеющуюся информацию для геолого-прогнозных построений.

*Сравнительный анализ результатов прогнозирования с использованием автоматизированной (АИПС) и автоматической («Днепр-4») систем**

Для сравнения результатов геологического прогнозирования с использованием автоматизированной системы АИПС и автоматической системы «Днепр-4» было выполнено специальное исследование по идентичной исходной информации по одной наиболее детально изученной рудной зоне.

Исходными данными при решении задачи в системе АИПС были результаты геологической интерпретации гравимагнитных данных, показывающие распространение геологических факторов как на уровне эрозионного среза, так и на некоторой глубине от него. Исходными данными для программы «Днепр-4» были значения гравитационных и магнитных полей в узлах равномерной сети точек. Варианты решения задачи, приведенные на рис. 27, отличаются только эталонными объектами. В вариантах 3 и 6 использованы как внешние, так и внутренние эталоны, в вариантах 2 и 5 — только внутренние, а в вариантах 1 и 4 — только внешние эталоны. Под внутренними эталонами понимаются в данном случае рудные объекты рассматриваемого рудного поля. В качестве внешних эталонов выбраны объекты с ближайшего рудного поля, расположенные в пределах той же рудной зоны. Оба рудных поля по основным элементам геологического строения близки между собой. При использовании внутренних и внешних эталонов рассматриваемое рудное

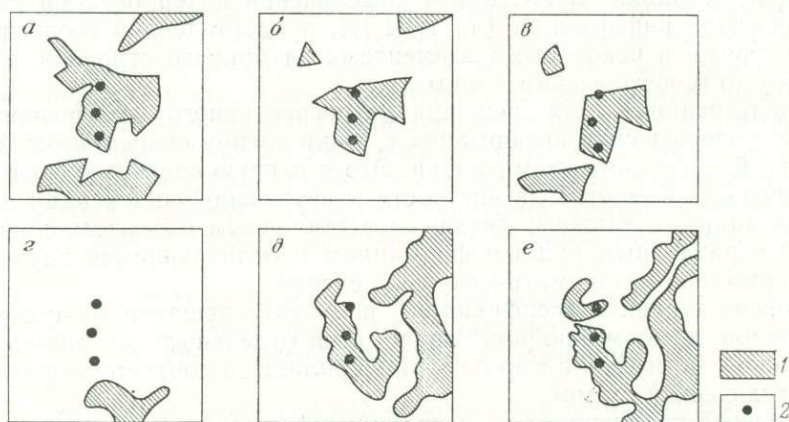


Рис. 27. Результаты сопоставления геологического прогнозирования с использованием автоматизированной (АИПС) и автоматической («Днепр-4») систем (по Б. Н. Павлову и др. [97]).

a—*в* — результаты прогнозирования на основе АИПС: *a* — использование внешних эталонных объектов обучения (вариант 1); *б* — использование внутренних эталонных объектов (вариант 2); *в* — использование смешанных эталонных объектов (вариант 3); *г*—*е* — результаты прогнозирования на основе «Днепр-4»: *г* — использование внешних эталонных объектов обучения (вариант 4), *д* — использование внутренних эталонных объектов (вариант 5); *е* — использование смешанных эталонных объектов (вариант 6).

1 — перспективные площади, 2 — известные рудные объекты

* Результаты исследования приведены по материалам Б. Н. Павлова и др. [97].

поле уверенно выделяется по обеим сопоставляемым системам. В варианте только внешних эталонов программа автоматической классификации «Днепр-4» показывает отрицательный результат. Анализ причин этого результата показал, что в программе «Днепр-4» в неявном виде заложено априорное предположение, что если два участка сходны по характеристикам физических полей, получаемым из заданного выбора трансформаций исходного потенциального поля, то они сходны и по основным особенностям геологического строения. В рассматриваемом случае это положение полностью не выполнено — эталонные рудные объекты сходны между собой лишь по «локальным» характеристикам геофизических полей, отображающим строение верхней части земной коры и резко различных по «региональным» параметрам. Естественно, поняв причины отрицательного результата, можно существенно улучшить результаты прогнозирования, изменив набор трансформаций и их параметров, но автоматическая система без участия специалиста оказывается не в состоянии выполнить эту задачу.

Таким образом, при решении задачи выделения перспективных участков автоматизированный режим обработки исходной информации, позволяющий осуществлять управление процессом обработки данных, в настоящее время существенно эффективнее, чем полностью автоматический.

VII. ПЕРСПЕКТИВЫ РАЗВИТИЯ ЧЕЛОВЕКО-МАШИННЫХ СИСТЕМ В ГЕОЛОГИИ

1. РАЗВИТИЕ ЭВРИСТИЧЕСКОГО МЕТОДА

Как отмечалось выше, современный этап внедрения математических методов и ЭВМ в прикладную геологию характеризуется прежде всего стремлением к интеграции систем хранения данных и систем обработки информации с использованием современных математических методов. В таких интегрированных системах в качестве активного начала используется непосредственно сам геолог. В интерактивных человеко-машинных системах удается в максимальной степени сочетать большие возможности современных ЭВМ по накоплению, хранению и обработке информации с опытом, знаниями и интуицией специалистов. Создание таких систем отвечает современному этапу развития науки и техники и является объективным результатом воздействия научно-технической революции на традиционную геологическую проблему — геологическое прогнозирование. Именно такие системы обеспечивают качественно новый скачок в развитии прикладной геологии. Они позволяют поднять производительность активного творческого труда специалистов при разработке ими вариантов прогнозирования, повысить надежность составления заключений путем сопоставительного анализа различных альтернативных вариантов и выбора оптимального варианта.

В настоящее время в геологии, как известно, большинство определенных носит трудно формализуемый характер либо вообще не поддается формализации. В этих условиях огромное значение приобретает личный опыт специалиста, его способность по некоторым характеристикам мысленно представить и проследить ход отдельных процессов, явлений и т. п. В связи с этим особое значение приобретают эвристические методы моделирования, позволяющие наиболее полно учитывать знания специалиста и эффективно использовать их при разработке вариантов прогноза.

До настоящего времени применение математических методов в прикладной геологии развивалось главным образом по пути использования классических подходов, употребление которых ограничено определенными типами распределения классов объектов в пространстве признаков. Такой подход применяется для решения широкого круга задач, несмотря на несоответствие характера исходных данных условиям успешного использования таких алгоритмов. Видимо, более широкое применение в геологических исследованиях эвристических способов, находящихся на стыке формальных и неформальных математических методов, является в настоящее время более перспективным. Эвристические методы прогнозирования в сочетании с классическими методами теории оптимальных решений позволяют эффективно использовать вычислительную технику при анализе и оценке геологических объектов. Они, в частности, позволяют учитывать специфику задач геологического прогноза и дополнить традиционные способы современными методами обработки информации.

В человеко-машинных системах при использовании эвристических методов ЭВМ превращается в помощника специалиста, помогающего

ему быстро рассчитать и увидеть конечные результаты своего творческого труда, т. е. при этом повышается производительность при сохранении качества конечного продукта. Вместе с тем при таком подходе открываются перспективы повышения качества геологических прогнозов. С одной стороны, это происходит за счет того, что специалист, накапливая опыт взаимодействия с системой, получает возможность все более точно (в количественной форме) оценивать значение отдельных признаков и явлений на процессы рудообразования и за счет этого он достигает все более точных конечных результатов. Таким образом, ЭВМ (вернее, машинная система диалогового типа) выступает в роли своеобразного «тренажера» специалиста, помогающего ему ускорить приобретение опыта.

С другой стороны, для составления поисково-оценочных критериев конкретного типа оруденения в виде эвристических решающих правил (моделей) можно привлекать самых высококвалифицированных специалистов, хорошо знающих регион и имеющих практику открытия месторождений. Тем самым открывается возможность более широко поставить на службу прикладной геологии знания таких специалистов, сделать их опыт общим достоянием различных геологических организаций. Иными словами, открывается возможность с большей эффективностью использовать самые высококвалифицированные кадры. Такие эвристические решающие правила, разработанные отдельными специалистами или группами экспертов для различных условий, составят «банк решающих правил», являющийся, по сути, «банком знаний специалистов».

Первый опыт использования эвристического подхода в человеко-машинной системе свидетельствует, что результаты прогнозирования при этом не хуже, чем при применении алгоритмов распознавания образов; в то же время качество прогноза во многом определяется квалификацией специалиста. Так, например, очевидно, что 3—4 специалиста высокой квалификации при использовании эвристического метода с помощью диалоговой машинной системы с территориальным банком данных смогут выделить перспективные площади (составить геологический прогноз) значительно быстрее, нежели это делается традиционным путем. При этом качество такого прогноза будет лучше, чем при использовании алгоритмов распознавания образов, и не хуже (а может быть и лучше), нежели у прогноза, составленного на традиционной основе. Первые результаты такого рода, изложенные в главе VI, подтверждают такую оценку возможностей человеко-машинных систем.

Таким образом, применение человеко-машинных систем позволяет аккумулировать коллективный опыт специалистов. При распознавании с помощью ЭВМ известных объектов по информации территориального банка данных специалисты получают оценки информативности геологических признаков, а применение эвристических алгоритмов позволяет непосредственно учитывать опыт специалистов, отображенный в процедуре решения. При этом в процессе решения практических задач последовательно осуществляются самообучение и самосовершенствование системы. Каждая новая информация о результатах прогнозирования вводится в ЭВМ, и последующее прогнозирование будет осуществляться с учетом всего накопленного опыта. Таким образом, чем больше накапливается результатов, тем более уточняются конкретные эвристические решающие правила. В этом процессе происходит не только одностороннее совершенствование системы, но и сами специалисты постепенно совершенствуют свои представления о закономерностях локализации оруде-

нения. В конечном итоге все это повышает точность и качество геологических прогнозов.

Учитывая современное состояние разработок формального языка описания геологических явлений, в настоящее время оказывается, по-видимому, предпочтительнее более широкое использование разработок вариантов геологических прогнозов на основе эвристических подходов. По всей вероятности, такое положение остается определяющим и на ближайшее будущее.

Видимо уже в 80-х годах использование диалоговых человеко-машинных систем, совмещенных с территориальными машинными банками данных, позволит более полно загрузить специалистов творческой работой, генерированием идей, обобщением материалов. Такой подход открывает возможность более строгой систематизации геологических исследований, помогает устанавливать наиболее благоприятные признаки конкретного типа оруденения и определять комплекс наиболее эффективных поисковых методов.

Человеко-машинные системы открывают практически неограниченные возможности оперативной оценки геологических гипотез и концепций. Становится вполне реальным фактом теоретическая проверка возможности нахождения в конкретном районе нового, неизвестного типа оруденения. На основе таких систем более качественно могут выполняться количественные прогнозы. Особо благоприятные перспективы открывают человеко-машинные системы в сочетании с территориальными банками данных для геолого-экономических исследований, составления прогнозных планов перспективного развития сырьевой базы отдельных отраслей промышленности и целых регионов.

2. ТЕХНИЧЕСКОЕ РАЗВИТИЕ

Интенсивное развитие всевозможного оборудования, облегчающего процесс человеко-машинного диалога, является весьма благоприятным фактором, обеспечивающим техническую основу для дальнейшего внедрения ЭВМ в область прикладной геологии.

Использование современного оборудования резко облегчает автоматизированный ввод разнообразной картографической информации (в том числе числовых полей) в память ЭВМ.

Видимо, в недалеком будущем окажется возможным осуществить полностью автоматический ввод в машинную память геологической, геофизической, геохимической, гидрогеологической информации и тем самым создать реальные предпосылки для автоматического формирования территориальных машинных банков картографической информации.

Имеющиеся уже в настоящее время автоматические графопостроители обеспечивают воспроизведение в исходном виде отдельных геологических контуров, геофизических карт в виде изолиний и др. В недалеком будущем, видимо, будут разработаны устройства, позволяющие автоматически воспроизводить из памяти ЭВМ сложные, многоцветные композиции геологических и прогнозных карт.

Особенно благоприятные перспективы для интенсивного использования человеко-машинных систем открываются при использовании таких технических средств диалогового взаимодействия, как многоцветные дисплеи автономного пользования. Система, содержащая территориальный банк данных, различные технологические схемы обработки информации и имеющая периферийные видеотермальные устройства, соеди-

ненными линиями связи с основным информационным центром, позволит специалистам непосредственно в полевых условиях пользоваться картографическими и другими фондовыми материалами, а также рассчитывать многочисленные варианты прогнозных решений. При этом каждый пользователь может иметь свой «личный архив» в виде целевых массивов банка данных. Кроме того, каждый специалист сможет использовать в своей практической работе поисковые критерии, разработанные высококвалифицированными экспертами. Использование такого подхода позволит существенно повысить качество и эффективность геологических исследований, интенсифицировать творческий труд специалистов и повысить его практическую отдачу. Отметим, что современный арсенал математических средств обработки данных уже в настоящее время является вполне достаточной методологической основой для практической реализации человеко-машинных систем указанного типа.

3. ОРГАНИЗАЦИОННОЕ РАЗВИТИЕ

Перспектива формирования территориальных машинных банков данных уже в настоящее время ставит ряд проблем организационного плана. С самого начала этот процесс, по-видимому, следует стандартизировать и направить в определенном направлении. В противном случае возникает опасность спонтанного развития многочисленных банков сугубо «местного» пользования. Видимо, целесообразно в качестве основы будущей организации таких банков данных использовать многоуровневую иерархическую систему. В такой системе целесообразно иметь, по крайней мере, четыре взаимосвязанных уровня:

- интернациональный информационный центр;
- национальный информационный центр;
- территориальный информационный центр;
- локальный информационный центр.

Банки данных в составе локального информационного центра могут содержать материалы по отдельному рудному району, месторождению, действующему горнорудному предприятию. Здесь целесообразно хранение материалов по отдельным скважинам, сведений о детальном геологическом строении, объемной геологической модели отдельного месторождения, данных о движении запасов, геолого-экономических характеристиках и т. п.

В территориальном банке данных должна, по-видимому, сохраняться геологическая, геофизическая, геохимическая, гидрогеологическая и другая информация, характеризующая отдельный регион, в масштабах 1 : 25 000—1 : 200 000. В таком банке должны также храниться материалы по опорным скважинам, наиболее типичным разрезам и т. п.

Национальный банк данных, по-видимому, должен содержать материалы региональных геологических и геофизических исследований, в том числе сведения о глубинном геологическом строении земной коры. Кроме того, в национальном центре наряду с материалами общегеологического характера целесообразно также хранить сведения об основных геологических характеристиках известных рудных месторождений (файл месторождений).

В национальном центре следует организовать накопление и хранение эвристических поисково-оценочных критериев различных типов оруденения, построенных высококвалифицированными специалистами-экспертами.

Наконец, в интернациональном центре следует, по-видимому, хранить геологические материалы глобального характера, в том числе дан-

ные космических съемок. Интернациональный банк геологических данных может содержать, например, сведения по территориям стран — членов СЭВ. Такая организация хранения геологической информации обеспечивает ее эффективное практическое использование и согласуется с трехзвенной структурой управления народным хозяйством.

4. ЭКОНОМИЧЕСКАЯ ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ ЧЕЛОВЕКО-МАШИННЫХ СИСТЕМ

Одной из наиболее актуальных задач прикладной геологии, которой в последнее время уделяется все возрастающее внимание, является повышение экономической эффективности геологоразведочных работ. Решение этой задачи в значительной мере связано с повышением качества геологического прогнозирования на каждой стадии геологоразведочного процесса.

Применение человеко-машинных систем оказывает влияние на эффективность геологоразведочного процесса путем ускорения и повышения качества геологического прогнозирования. Прежде чем перейти к оценке этого влияния, необходимо отметить, что специфика геологии, как отрасли производства, обуславливает существенные трудности при определении экономической эффективности геологоразведочного производства. Так, например, А. М. Марголин [66] отмечает: «...геологоразведочная деятельность не относится к числу типичных производственных отраслей материального производства. Это связано с тем, что, во-первых, целью функционирования геологоразведочной отрасли является не непосредственное производство тех или иных вещественных предметов, а выявление и разведка минерального сырья как потенциального предмета труда горнодобывающей промышленности ...Во-вторых, геологоразведочные работы во всех их звеньях имеют не чисто производственный, а скорее научно-производственный характер, когда производственные процессы бурения скважин ..., геофизических наблюдений и т. п. выполняются для сбора сведений о запасах минерального сырья, повышения точности представлений о свойствах, количестве и качестве разведанных запасов...».

Таким образом, продукция геологоразведочной отрасли имеет двойственный характер: с одной стороны, разведанные запасы, как товар, с другой — информация о запасах полезных ископаемых в недрах, как результат научной деятельности вне материального производства. По-видимому, этой двойственностью объясняется отсутствие в настоящее время общепринятой методики определения экономической эффективности геологоразведочного процесса.

С учетом изложенного перейдем к рассмотрению преимуществ применения человеко-машинных систем в геологии.

Отличительными особенностями таких систем, определяющими их преимущества по сравнению с традиционными методами прогнозирования, являются наличие машинного банка геологической, геофизической, геохимической и другой информации по исследуемому району и использование математического обеспечения в виде комплекса программ, реализующих все необходимые алгоритмы и процедуры обработки информации, хранящейся в банке данных. Эти особенности, в конечном счете, позволяют:

- улучшить информационное обеспечение процесса прогнозирования и геологических служб на данной территории;
- ускорить процесс составления геологического прогноза за счет автоматизации большинства рутинных операций по обработке данных.

а также составления различных вариантов прогнозных карт на базе различных геологических моделей; удешевить процесс составления геологического прогноза за счет уменьшения объема ручной работы по обработке данных;

— повысить точность геологического прогноза за счет наиболее полного информационного обеспечения и применения широкого комплекса математических методов.

Рассмотрим подробнее перечисленные составляющие положительного экономического эффекта от внедрения АИПС.

Улучшение информационного обеспечения

Человеко-машинные системы позволяют организовать централизованное хранение больших объемов разнородной геологической информации по исследуемому району, постоянное ее обновление и пополнение. Эта информация может использоваться не только непосредственно в целях прогнозирования, но и по запросу специалистов может быть оперативно обработана с помощью того или иного алгоритма и выдана в требуемом виде. Если учесть, что в настоящее время вся эта информация содержится в большом количестве разрозненных материалов, то эффект от улучшения информационного обеспечения геологических служб при использовании человеко-машинных систем представляется весьма значительным.

Ускорение и удешевление процесса геологического прогнозирования

Расчеты показывают, что при использовании человеко-машинных систем на этапе среднемасштабного прогнозирования время составления прогнозных карт уменьшается до 1—2 месяцев по сравнению с 1—1,5 годами при традиционных методах.

Очевидно, что составление прогнозных карт при минимальном отрыве во времени от проведения полевых работ и даже непосредственно в полевых условиях позволяет более оперативно учитывать новые данные и обеспечивать большую концентрацию и направленность поисково-разведочных работ.

Применение устройств автоматического ввода позволяет значительно уменьшить объем работы по подготовке данных. Все операции по математической обработке исходной информации, вплоть до построения вариантов прогнозных карт, выполняются с использованием ЭВМ. Это существенно сокращает количество вспомогательного труда технического и геологического персонала, занятого в настоящее время на второстепенных работах при составлении геологического прогноза.

Предварительные расчеты показывают, что при прогнозировании на основе человеко-машинной системы затраты на составление прогнозных карт среднего масштаба на один вид полезного ископаемого по одному региону снижаются по сравнению с традиционной методикой в 2—3 раза.

Повышение точности прогноза

Основной эффект при использовании человеко-машинных систем возникает от повышения точности прогноза. При этом существенно уменьшается количество площадей, которые рекомендуются для поста-

новки поисковых работ в качестве перспективных, что приводит, с одной стороны, к уменьшению непроизводительных затрат на опоскование неперспективных площадей и, следовательно, к снижению общей стоимости единицы разведанных запасов полезного ископаемого и, с другой стороны, к ускорению обнаружения новых объектов вследствие концентрации поисковых работ на наиболее перспективных участках.

При прогнозировании площадей, перспективных для постановки поисковых (оценочных, разведочных) работ, специалистами могут быть допущены ошибки двух видов: 1) перспективный участок может быть ошибочно признан неперспективным (ошибка I рода) и 2) неперспективный участок может быть оценен как перспективный (ошибка II рода). Очевидно, ошибки I рода при прогнозировании приводят к пропускам промышленных месторождений, а ошибки II рода влекут за собой увеличение непроизводительных затрат на изучение непромышленных объектов. Так, например, вероятность правильной оценки перспективности отдельного рудопроявления при традиционных методах прогнозирования практически не удается поднять выше 20—50%. Это означает, что в среднем из 10 выявленных рудопроявлений, рекомендуемых к оценке, 5—8 оказываются непромышленными, и вложенные в их детальную оценку средства следует относить к экономическим потерям. Из теории решений известно, что средние потери от n актов принятия решений R_n определяются следующим образом:

$$R_n = (C_1 P_1 + C_2 P_2),$$

где C_1, C_2 — цены ошибок I и II рода;

P_1, P_2 — вероятности ошибок I и II рода.

Если при традиционных методах интерпретации вероятности ошибок I и II рода равны P_1^1 и P_2^1 , а с привлечением человеко-машинных систем они будут уменьшены соответственно до P_1'' и P_2'' , то разность соответствующих величин определит экономическую эффективность, получаемую от внедрения указанных систем

$$\Delta = R_n' - R_n'' = C_1 (P_1' - P_1'') + C_2 (P_2' - P_2'').$$

Повышение точности прогнозных заключений приводит в конечном итоге к увеличению прироста запасов при фиксированных ассигнованиях на поиски и разведку, т. е. к повышению эффективности геологоразведочных работ. Кроме того, концентрация поисковых работ на наиболее перспективных участках обуславливает более раннее обнаружение новых месторождений.

Таким образом, эффект от использования человеко-машинных систем в геологии определяется в первую очередь тем ключевым положением, которое занимает процедура прогнозирования в геологоразведочном процессе. Практическое применение таких систем в конечном итоге позволит:

- централизованно хранить информацию, необходимую для составления геологических прогнозных заключений по определенной территории (участку, месторождению) и автоматизировать процесс обработки этой информации;

- автоматизировать процесс выдачи справок по запросам специалистов по данной территории;

- расширить комплекс математических и эвристических методов, применяемых при решении прогнозных задач;

- существенно ускорить процесс геологического прогноза;

— направить усилия специалистов на содержательный анализ информации, интенсифицировать их творческий труд.

При региональном прогнозе экономический эффект достигается за счет определения оптимального варианта развития и размещения производства на длительную перспективу. При прогнозе перспектив развития сырьевой базы в пределах отдельных территорий экономический эффект возникает за счет более целенаправленного проведения поисково-разведочных работ (концентрация усилий на наиболее перспективных участках).

Разработка многовариантных прогнозов с использованием человеко-машинных систем позволит создать основу для решения задач оптимального управления геологоразведочным процессом, задач планирования перспективного развития и размещения производства.

1. *Авгитов А. К., Беломар О. Д.* О результатах регионального прогнозирования ртутного оруденения в Донбассе с применением математических методов. — В кн.: Геологическая эффективность математических методов в геологии и геофизике. Минск, 1975, с. 66—68.
2. *Автоматизированная* информационно-прогнозирующая система и опыт ее эксплуатации на этапе мелкомасштабных исследований/В. Ф. Семенов, А. Н. Еремеев, Е. П. Власов и др. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с. 117—118.
3. *Агаджанян С. И.* Систематизация и типизация вопросов организации банков данных. Материалы семинара. М., МДНТП, 1975, с. 117—182.
4. *Аронов В. И., Невельская Э. Я., Самарин В. С.* Система поисковой и статистической обработки геологической информации на ЭЦВМ. — Тр. ВНИГНИ. М., 1971, вып. 103, с. 15—27.
5. *Аронов В. И., Невельская Э. Я.* Опыт эксплуатации системы поисковой и статистической обработки геологических данных на ЭВМ. — В кн.: Тезисы семинара «Применение математических методов и ЭВМ в геологии». Алма-Ата, 1974, с. 113—116.
6. *Бельшев Л. В., Смирнов Н. В.* Таблицы математической статистики. М., Наука, 1965, с. 265.
7. *Билибин Ю. А.* Общие принципы металлогенических исследований. — Изв. АН СССР, серия геол., 1947, № 5, с. 16—28.
8. *Билибин Ю. А.* Металлогенические провинции и эпохи. М., Госгеолтехиздат, 1955, с. 64.
9. *Бирюков В. И., Косов Б. М., Хрущов Н. А.* Рациональная последовательность геологоразведочных работ. — Советская геология, 1972, № 4, с. 18—26.
10. *Боганик В. Н., Шабельникова Т. Г.* Организация службы хранения и поиска геолого-геофизической информации. Организация и управление нефтяной промышленностью. — Реф. научно-экон. сб. М., 1974, № 5, с. 22—26.
11. *Бугаец А. Н., Мацак А. П., Садовский Ю. А.* Применение методов дискретного анализа при оценке месторождений полезных ископаемых на территории Казахстана. — Геология рудных месторождений, 1970, № 6, с. 62—72.
12. *Бугаец А. Н., Горбунов Л. Л., Подворчан В. И.* Математическое обеспечение систем обработки данных, — комплекс программ по методам автоматической классификации. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с. 67—68.
13. *Бугаец А. Н., Дуденко Л. Н.* Математические методы при прогнозировании месторождений полезных ископаемых. Л., Недра, 1976, с. 270.
14. *Бугаец А. Н.* Прогнозирование месторождений полезных ископаемых, как задача принятия решений в условиях неопределенности и математические методы ее решения. — В кн.: Математические методы при прогнозе рудоносности. М., Недра, 1977, с. 24—41.
15. *Бухникашвили А. В.* Некоторые критические замечания по поводу математизации геологии. — Советская геология, 1972, № 11, с. 34—35.
16. *Вапник В. Н., Лернер А. Я.* Узнавание образов при помощи «обобщенных» портретов. — Автоматика и телемеханика, т. XXIV, 1963, № 6, с. 36—42.
17. *Вапник В. Н., Лернер А. Я., Червоненкис А. Я.* Системы обучения распознаванию образов при помощи обобщенных портретов. — Техническая кибернетика, 1965, № 1, с. 19—24.
18. *Веселов В. В., Коростышевский М. Б., Курбанев М. С.* Принципы построения информационно-логической системы «АСУ-Поиск». — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с. 7—9.
19. *Веселов В. В., Спивак Л. Ф.* К алгоритму управления массивами геологических данных в информационно-логической системе. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с. 245—247.
20. *Вистелиус А. Б.* Красноцветные отложения полуострова Челекен. М., Недра, 1966, с. 142.
21. *Вольфсон Ф. И., Невский В. А.* О первичной зональности в гидротермальных месторождениях. — Изв. АН СССР, серия геол., 1949, № 1, с. 17—24.
22. *Воронин Ю. А., Гольдина Н. А.* Упрощенная схема математико-логического разбора геологических классификаций. — Геология и геофизика, 1965, № 2, с. 83—91.

23. *Воронин Ю. А., Еганов Э. А.* Методологические вопросы применения математических методов в геологии. Новосибирск—М., Наука, 1974, с. 85.
24. *Воронин Ю. А., Черемисина Е. Н.* О математическом обеспечении автоматизированных систем в геологии. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с. 51—52.
25. *Выбор первоочередных структур для бурения в Лено-Вилуйской нефтегазональной провинции с помощью алгоритмов направленного опробования/Е. Н. Черемисина, А. Д. Луговец, В. И. Берилко и др.* — В кн.: Применение математических методов и ЭВМ при поиске полезных ископаемых. Новосибирск, 1973, с. 101—103.
26. *Герасименко В. Я.* Об эмпирическом переходе текст-тезаурус-семантический код. — В кн.: Вопросы общей и теоретической тектоники. Хабаровск, 1974, с. 91—99.
27. *Глобальные закономерности размещения крупных рудных месторождений /М. А. Фаворская, И. Н. Томсон, В. А. Баскина и др. М., Недра, 1974, с. 193.*
28. *Глушakov В. М.* Вычислительная техника и проблема автоматизации управления. — В кн.: Будущее науки. М., 1971, вып. 4, с. 5—19.
29. *Голыздра Г. Я., Ахметшина А. К., Резенко А. Ф.* Опыт применения карт тектонических нарушений Украинского щита. Экспресс-информация. Серия: Математические методы исследований в геологии. М., ВИЭМС, 1972, № 8—9, с. 6—12.
30. *Головин И. В., Супруненко Э. И.* Некоторые приемы статистического анализа геофизических данных при мелкомасштабном прогнозно-металлогеническом районировании рудных регионов СССР. Л., Недра, 1971, с. 146.
31. *Горлицкий Б. А., Финкель Л. С.* Анализ массовой геохимической информации в автоматической системе обработки «Комплекс». — В кн.: Информационные системы, вып. 2. Киев, 1973, с. 20—22.
32. *Дмитриев А. Н., Журавлев Ю. И., Кренделев Ф. П.* О математических принципах классификации предметов и явлений. Дискретный анализ. Тр. Ин-та математики СО АН СССР, вып. 7. Новосибирск, 1967, с. 24—32.
33. *Дмитриев А. Н., Журавлев Ю. И., Кренделев Ф. П.* Об одном принципе классификации и прогноза геологических объектов и явлений. — Геология и геофизика, 1968, № 5, с. 50—64.
34. *Дмитриев М. К., Черкасов В. Е.* Методология комплексного анализа на ЭВЦМ материалов геолого-геофизических исследований. М., Недра, 1970, с. 70.
35. *Дородницын А. А.* Использование математических методов и ЭВМ в геологических исследованиях. — Изв. АН СССР, серия геол., 1966, № 1, с. 21—26.
36. *Дуденко Л. И.* Позиционные игры и металлогенический прогноз. — В кн.: Вопросы петрохимии. Л., 1969, с. 73—76.
37. *Дуденко Л. И.* Методы автоматизированного прогнозирования минерального сырья. — Тр. ВСЕГЕИ, вып. 223, 1975, с. 93—99.
38. *Еремеев А. Н.* Принципы методики поисков глубокопогребенных рудных месторождений. — В кн.: Глубинные поиски рудных месторождений, вып. 2. М. 1968, с. 37—55.
39. *Еремеев А. Н.* Системный метод при выделении и оценке геологических объектов. — Советская геология, 1976, № 2, с. 53—60.
40. *Заварицкий А. Н.* О картах прогноза. — Вестник АН СССР, 1939, № 8—9, с. 3—16.
41. *Загоруйко Н. Г.* Структура проблемы распознавания образов и методы ее решения. — В кн.: Автоматическое распознавание слуховых образов. Новосибирск, 1966, с. 8—14.
42. *Загоруйко Н. Г.* Методы распознавания и их применение. М., Наука, 1972, с. 146.
43. *Зеликовский Л. П.* Приемы определения понятий в геологии. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 2. М., 1977, с. 102—115.
44. *Зеликовский Л. П.* Синтез сложной системы оценки рудоносности площадей. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 2. М., 1977, с. 53—59.
45. *Использование математических методов при поисках месторождений урана/Г. И. Блассанов, Д. Я. Суражский, Б. А. Чумаченко и др.* — Атомная энергия, 1968, № 4, с. 274—280.
46. *Каждан А. Б.* Пути совершенствования методики геологоразведочных работ и повышения их экономической эффективности. — Советская геология, 1972, № 2, с. 26—34.
47. *Канцель А. В., Лернер А. Я.* Проблема использования вероятностно-статистических характеристик руд для классификации и прогнозирования рудных месторож-

дений. — В кн.: Математические методы в геологии. Доклады советских геологов на международном геологич. конгрессе. М., 1968, с. 76—79.

48. *Каратаев Г. И., Терлецкий В. В.* О построении теории эвристического прогноза геолого-геофизических объектов. — В кн.: Проблемы развития АСУ-геология и применение математических методов в геологии. Тезисы докладов, ч. 1. М., 1977, с. 46—49.

49. *Константинов Р. М.* Рудные формации, математические и кибернетические методы их изучения. — В кн.: Проблемы геологии минеральных месторождений. М., 1969, с. 134—139.

50. *Константинов Р. М.* Статистическое обоснование связей металлогенических факторов. — Советская геология, 1969, № 2, с. 119—124.

51. *Константинов Р. М.* Математические методы прогнозной оценки месторождений. — В кн.: Итоги науки и техники. Общая геология, т. 8. Математические методы в геологии. М., 1977, с. 5—58.

52. *Константинов Р. М., Щеглов А. Д.* Математические методы в металлогении. — Разведка и охрана недр, 1976, № 5, с. 6—10.

53. *Котеликов В. И., Кричевец Г. Н., Зеликовский Л. П.* О вероятности обнаружения эллиптических объектов неравномерными сетями наблюдений. — В кн.: Методика комплексной обработки информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 1. М., 1975, с. 140—145.

54. *Крейгер В. М.* Поиски и разведка месторождений полезных ископаемых. М. Госгеолтехиздат, 1961. 446 с.

55. *Куклин А. П.* Моделирование геологических условий, благоприятных для эндогенного оруденения на территории Чукотки. — Кольма, 1972, № 10, с. 42—44.

56. *Куклин А. П.* Электронно-вычислительная машина «читает» геологические карты. — Природа, 1972, № 2, с. 94—96.

57. *Кульбак С.* Теория информации и статистика. М., Наука, 1967. 352 с.

58. *Курош А. Г.* Лекции по общей алгебре. М., Наука, 1962. 315 с.

59. *Кусевич В. И., Зеликовский Л. П.* Опыт построения моделей формирования месторождений (на примере ртутных месторождений Южной Ферганы). — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 1. М., 1975, с. 49—74.

60. *Кухаренко Н. А.* Объекты информационных систем по металлогении. — Советская геология, 1974, № 7, с. 119—124.

61. *Лакно М. С., Шувев А. Н.* Стандартизация ввода, хранения, обработки и выдачи геологической информации на базе автоматизированной системы обработки информационных массивов «ОРИОН». — В кн.: Применение экономико-математических методов и средств вычислительной техники в управлении стандартизацией и качеством продукции. М., 1974, с. 40—41.

62. *Лбов Г. С.* Выбор эффективной системы зависимых признаков. — В кн.: Вычислительные машины, вып. 19, М., 1966, с. 16—22.

63. *Леонов Ю. Г., Шалло В. Н.* О некоторых методологических вопросах применения математических методов в геологии. — Геология и геофизика, 1973, № 4, с. 44—48.

64. *Лисичкин А. В.* Теория и практика прогнозтики. М., Наука, 1972, 224 с.

65. *Ломтадзе В. В.* О принципах построения автоматизированных систем обработки геолого-геофизических данных. — В кн.: Применение математических методов и ЭВМ при поиске полезных ископаемых. Новосибирск, 1973, с. 186—193.

66. *Марголин А. М.* Оценка запасов минерального сырья. Математические методы. М., Недра, 1974, 261 с.

67. *Математические методы в газонефтяной геологии и геофизике*/М. М. Элланский, А. И. Холин, Г. Н. Зверев и др. М., Недра, 1972. 208 с.

68. *Месарович М., Мако Д., Токахара И.* Теория иерархических многоуровневых систем. М., Мир, 1973, 346 с.

69. *Методика кодирования геологической информации для распознавания рудности с помощью ЭВМ*/М. М. Камилов, В. Я. Ковальчук, Т. М. Маринов и др. — Изв. АН УзбССР, серия техн. наук, 1973, № 4, с. 17—20.

70. *Методические указания по проведению отдельных этапов геологоразведочных работ.* М., Госгеолтехиздат, 1961. 86 с.

71. *Методы экспертных оценок при выделении существенных геологических признаков*/Е. П. Власов, А. Н. Еремеев, Л. П. Зеликовский и др. — Советская геология, 1973, № 1, с. 112—118.

72. *Митрофанов И. В.* Особенности геологоразведочного процесса как объекта автоматизации. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 2. М., 1977, с. 22—40.

73. *Нагорский В. А.* Пути и способы построения автоматизированной системы металлогенического прогнозирования. — В кн.: Повышение эффективности использования

вычислительной техники при геологоразведочных работах. Киев, 1975, с. 66—69.

74. *Некоторые* результаты разработки автоматизированной информационно-поисковой системы «Инженерная геология»/И. С. Комаров, А. Р. Зайдель, Я. И. Файн и др. — В кн.: Информационно-поисковые системы в инженерной геологии. М., 1975, с. 40—45.

75. *Немец В. В.* Автоматизация геологических и эксплуатационных моделей месторождений полезных ископаемых. — В кн.: Математизация и автоматизация в геологических исследованиях. Л., 1972, с. 81—86.

76. *Николаенко В. Н.* Автоматизированная система классификации и прогноза геологических объектов (АСКИП). — В кн.: Математизация и автоматизация в геологических исследованиях. Л., 1972, с. 86—90.

77. *Николаенко В. Н.* Опыт и использование автоматизированной системы классификации и прогнозирования геологических объектов и пути ее совершенствования. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с. 121—122.

78. *Нильсон Н.* Обучающиеся машины. М., Мир, 1967. 286 с.

79. *Об использовании* таксономии для анализа геологической информации/Е. П. Власов, Л. С. Лукьянов, В. В. Марченко и др. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 1. М., 1975, с. 99—101.

80. *Обработка* информационных массивов в автоматизированных системах управления/В. М. Глушков, В. П. Гладун, Л. С. Лозинский, С. Б. Погребинский. Киев, 1970. 284 с.

81. *О количественной* оценке исходной информативности геологических карт — основе пространственных прогнозных исследований/В. В. Богацкий, М. А. Чурилин, Б. И. Суганов и др. — В кн.: Новые методы крупномасштабного геологического картирования и поисков твердых полезных ископаемых на Сибирской платформе. Иркутск, 1976, с. 64—69.

82. *Олейников А. Н.* Кодирование диагностических признаков. — В кн.: Цифровое кодирование систематических признаков древних организмов. М., 1972, с. 12—45.

83. *Омелин В. М.* Решение геологических задач в системе «человек — ЭВМ». — Тр. ВНИГНИ, вып. 356, 1976, с. 32—38.

84. *Омелин В. М., Охотин С. И., Румянцева В. К.* Кодирование карт для ввода в ЭВМ. — В кн.: Геофизическая аппаратура, вып. 48. Л., 1972, с. 114—119.

85. *Определение* априорной информативности геологических признаков на основе метода экспертных оценок/Е. П. Власов, Л. В. Голикова, А. А. Дерягин и др. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 1. М., 1975, с. 75—98.

86. *Оптнер С.* Системный анализ для решения деловых и промышленных проблем (Перевод с английского и вступительная статья С. П. Никанорова). М., Советское радио, 1968. 216 с.

87. *Опыт* применения математических методов в геологии/М. С. Арабаджи, Ю. М. Васильев, В. С. Мильничук и др. М., Недра, 1969. 248 с.

88. *Опыт* разработки банков данных для автоматизированных систем прогнозирования/Б. А. Чумаченко, Е. П. Власов, В. В. Марченко и др. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, КазИМС, 1974, с. 248—250.

89. *Опыт* разработки и использования классификаторов и унифицированных форм-носителей в Уральском ТГУ/В. Я. Топорков, А. Б. Мурмухаметова и др. — В кн.: Вопросы разработки информационной и технической обеспеченности АСУ-геология. Свердловск, 1975, с. 20—21.

90. *Организация* автоматизированной системы для количественной оценки перспектив территории при поисках месторождений полезных ископаемых/В. В. Давыденко, В. С. Романов, Г. И. Деревянченко, Г. Г. Стихарная. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с. 123—125.

91. *О системном* подходе к планированию и управлению поисковоразведочными работами/Е. П. Власов, А. А. Дерягин, А. Н. Еремеев и др. — Тезисы докладов Всесоюзной конференции «Применение математических методов и ЭВМ при поиске полезных ископаемых». Новосибирск, 1973, с. 5—6.

92. *Основные* принципы составления и условные обозначения металлогенических прогнозных карт рудных районов/Е. Т. Шаталов, А. В. Орлова, К. В. Яблокова и др. М., Недра, 1963. 194 с.

93. *Особенности* обработки гравимагнитных данных при среднемасштабном прогнозировании эндогенного оруденения с использованием методов математического моделирования/Б. Н. Павлов, А. И. Корниенко, Л. П. Зеликовский и др. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 1. М., 1975, с. 102—104.

94. *О создании* информационной системы приема, хранения и обработки информации при подсчете запасов на примере месторождения Коунрад/А. А. Колаченко, Н. П. Скворцов, Б. Л. Чепрасов и др. В кн.: Геология медно-порфировых месторождений. Тезисы докладов 2-го Всесоюзного совещания по медно-порфировым месторождениям. Алма-Ата, 1971, с. 83—85.
95. *О структуре* и организации банков данных, используемых при решении задач геологического прогнозирования/Б. А. Чумаченко, Е. П. Власов, В. В. Марченко и др. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 1. М., 1975, с. 35—43.
96. *Оценка* оловорудных и ртутных месторождений с помощью алгоритмов распознавания образов/Н. Г. Демидова, Р. Г. Пинелис, С. Г. Рабинович и др. — В кн.: Математические методы в геологии, вып. 1. М., 1974, с. 47—56.
97. *Павлов Б. Н., Корниенко А. И., Зеликовский Л. П.* Сравнительный анализ методик обработки гравимагнитных данных при оценке рудоносности площадей. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 2. М., 1977, с. 73—86.
98. *Пинелис Р. Г.* Алгоритм распознавания «перебор». — В кн.: Математические методы в геологии, вып. 1. М., 1974, с. 26—29.
99. *Принципы* построения автоматизированной информационно-прогнозирующей системы (АИПС)/В. Ф. Семенов, А. Н. Еремеев, Е. П. Власов и др. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 1. М., 1975, с. 14—27.
100. *Применение* дискретного анализа при оценке горного хрусталя Памира/И. Л. Комов, А. Н. Бугаец, Л. Н. Хетчиков и др. — В кн.: Прогнозирование, поиски и разведка месторождений горного хрусталя. М., 1975, с. 18—26.
101. *Применение* методов распознавания образов при оценке месторождений и рудопроявлений/И. Д. Савинский, П. П. Астанин, Н. Г. Патык-Кара и др. — В кн.: Математические методы в геологии, вып. 2. М., 1975, с. 24—32.
102. *Проблемы* использования вероятностно-статистических характеристик руд для классификации и прогнозирования рудных месторождений/Э. М. Богушевский, А. В. Канцель, И. М. Коновалов и др. — В кн.: Применение математических методов в геологии. Алма-Ата, 1968, с. 140—141.
103. *Растрингин Л. А.* Случайный поиск в задачах оптимизации многопараметрических систем. Рига, Знание, 1965. 146 с.
104. *Растрингин Л. А.* Статистические методы поиска. М., Наука, 1968. 232 с.
105. *Результаты* сопоставления эффективности различных алгоритмов распознавания образов при решении некоторых геологических задач/А. Н. Еремеев, Р. Г. Пинелис, Е. П. Власов и др. — В кн.: Математические методы в геологии, вып. 2. М., 1975, с. 5—19.
106. *Решение* геологических задач с помощью программы распознавания/М. М. Бонгрд, Ш. А. Губерман, Н. А. Извеннова и др. — Советская геология, 1963, № 7, с. 42—44.
107. *Родионов Д. А.* Проблема выбора поисковых признаков и ее статистическое решение. — В кн.: Математические методы в геологии. М., 1968, с. 23—29.
108. *Розовский Л. Б.* Введение в теорию геологического подобия и моделирования (применение природных алгоритмов и количественных критериев подобия в геологии). М., Недра, 1969. 128 с.
109. *Савинский И. Д.* Некоторые вопросы разработки алгоритмов распознавания образов при решении геологических задач и описание алгоритма «Потенциал-2», БР. — В кн.: Математические методы в геологии, вып. 1. М., 1974, с. 8—24.
110. *Сазерленд А.* Система ввода-вывода информации для вычислительных машин. В кн.: Информация. М., 1968, с. 47—68.
111. *Серединский А. Н., Цуккерман И. И.* Способ представления контуров изображений в натуральных координатах. Авт. свидетельство № 170214 06К-9/00. — Бюллетень изобретений, 1970, № 8.
112. *Сидоров В. М.* О взаимоотношении информационных систем и автоматизированных систем обработки данных в геологии. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с. 231—233.
113. *Системный* подход к управлению поисково-разведочными работами/А. Н. Еремеев, В. Ф. Семенов, Е. П. Власов и др. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании, вып. 1. М., 1975, с. 7—13.
114. *Смирнова А. С.* Автоматизированные системы накопления, хранения, поиска и обработки информации по месторождениям полезных ископаемых. В кн.: Итоги

науки и техники. Общая геология, т. 8. Математические методы в геологии. М., 1977, с. 59—122.

115. *Смирнова А. С.* Запросно-ориентированная система программирования (метод «Запрос»). — В кн.: Разработка систем математического обеспечения автоматизированных систем управления. М., 1973, с. 67—75.

116. *Стрижевский В. Н.* Об одном подходе к определению информационных признаков. — Тр. МИФИ, 1970, вып. 36, с. 22—28.

117. *Ступак Н. К., Наугольников В. Б.* Использование ЭВМ при качественной интерпретации результатов комплексных геофизических наблюдений. — Геофизический сборник, № 61. Киев, 1974, с. 8—12.

118. *Суражский Д. Я.* Методы поисков и разведки месторождений урана. М., Атомиздат, 1960. 240 с.

119. *Тамм Б. Г., Тыгуэ Э. Х.* О создании проблемно-ориентированного программного обеспечения. — Кибернетика, 1975, № 4, с. 76—85.

120. *Тестовый* подход к количественной оценке геолого-структурных факторов и масштабов оруденения/С. В. Яблонский, Н. Г. Демидова, Р. М. Константинов и др. — Геология рудных месторождений, 1971, № 2, с. 64—72.

121. *Томсон И. Н.* Прогнозные построения на картах крупного масштаба на примере Кавалеревского рудного района. — В кн.: Основные принципы составления, содержания и условные обозначения металлогенических и прогнозных карт рудных районов. М., 1964, с. 173—184.

122. *Торопов С. М., Ломтадзе В. В.* Использование системы оперирования данными для ЕС ЭВМ в процессах ведения и наращивания информационной базы АСУ. — В кн.: Вопросы разработки информационного и технического обеспечения АСУ-геология. Свердловск, 1975, с. 36—40.

123. *Турбович И. Т.* Об оптимальном методе опознавания образов при взаимокоррелируемых признаках. — В кн.: Опознавание образов. Теория передачи информации. М., 1965, с. 18—24.

124. *Урсул А. Д.* Отражение и информация. М., Мысль, 1973. 231 с.

125. *Фокина И. А., Чумаченко Б. А., Власов Е. П.* Метод информативных интервалов и его применение для классификации магматогенных пород Северного Казахстана. — Минеральное сырье, 1970, № 22, с. 76—82.

126. *Формальная* постановка задачи прогноза нефтегазоносности. Краткие сведения об алгоритмах и программах/Ю. А. Воронин, А. Э. Канторович, Э. Э. Фотиади и др. — В кн.: Применение математических методов и ЭВМ для решения задач нефтяной геологии. Новосибирск, 1972, с. 12—20.

127. *Харкевич Л. А.* О выборе признаков при машинном опознавании. — Изв. АН СССР, технич. кибернетика, 1963, № 2, с. 36—39.

128. *Человек* и вычислительная техника (под ред. Глушкова В. М.). Киев, Наукова думка, 1971. 294 с.

129. *Чурилин М. А., Богацкий В. В., Витязь В. И.* Пространственно-статистический анализ геологического строения как базы прогноза полезных ископаемых (на примере юго-западной части Сибирской платформы). — В кн.: Новые методы крупномасштабного геологического картирования и поисков твердых полезных ископаемых на Сибирской платформе. Иркутск, 1976, с. 82—86.

130. *Шагалов Е. Т.* Основные задачи металлогении рудных районов, краткая характеристика методов их исследований. — В кн.: Основные принципы составления, содержания и условные обозначения металлогенических и прогнозных карт рудных районов. М., 1964, с. 32—61.

131. *Шибанов Г. П.* Распознавание в системах автоконтроля. М., Машиностроение, 1973. 424 с.

132. *Ширяев Е. Е.* Новые методы картографического отображения и анализа геoinформации с применением ЭВМ. М., Недра, 1977. 182 с.

133. *Школьников А. Д., Вуль В. А., Омелин В. М.* Автоматизация обработки графической информации при решении горных и геолого-геофизических задач. Л., Недра, 1976. 186 с.

134. *Штайн Л. М.* Диалоговая специализированная система обработки данных и программирования. — В кн.: Применение математических методов и ЭВМ в геологии, Алма-Ата, 1974, с. 65—66.

135. *Якубович К. И.* Выявление смысловых (ассоциативных) связей при создании ИПЯ. — В кн.: Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений, вып. 2. М., 1977, с. 87—102.

136. *Agterberg F. P.* Computer based statistical analysis of geological data: past and current activities. „Pap. Geol. Surv. Can.“, 1974, N 74—70, p. 15—18.

137. *Anderton M. R.* A drillhole data bank. — Appl. Comput. Meth. Miner. Ind. Proc. 10th Int. Symp. Johannesburg, 1972. — Johannesburg, 1973, p. 79—82, Discuss., 377, 378.

138. *Beckett P. H. T., Webster R., Michell G. M.* Terrain evaluation by means of data bank. — *Geographical Journal*, 1972, vol. 138, N 4, p. 430—456.
139. *Bie Stein W., Heer Hans de, Stalduinen Cornelis Ivan, Tinck Joris R. E., Dam Willem M.* Automated data capture poorly structured geological borelogs. — *Geol. Jahrb.* 1975, N 28, p. 15—25.
140. *Berner H., Ekstrom T. e. t. c.* Data storage and processing in geological mapping. Field data sheet. — *Geol. foran, Stockholm forhandl.* 1971, vol. 93, N 1, p. 85—101.
141. *Bouille F.* Application de la theorie des graphes an traitement de la carte geologique. „*Rev. Inst. Franc. petrole*“, 1974, 29, N 2, p. 173—216.
142. *Chow C. K.* A classe of nonlinear recognition procedures. „*JEEE Trans. System Science and Cybernetics*“, December 1966, vol. ssc—2.
143. *Cover T. M., Hart P. E.* Nearest Neighbor Pattern Classification. „*JEEE Trans. on Inf. Theory*“, Jan. 1967, vol. IT—13.
144. *Elovan I. E.* Selection of target areas by factor analysis. *Wst. Miner*, 1968, vol. 41, N 2, p. 44—54.
145. *Fix E., Hodges J. L.* Discriminatory Analysis, Nonparametric Discrimination. Consistency Properties. Texas, February 1951.
146. *Fabbi A. J.* A geological data bank for statistical analysis. „*Pap. Jeol. Surv. Can.*“, 1973, N 1, Part B, p. 29—37.
147. *Fabbi A. G., Divi S. R., Wong A. S.* A data base for mineral potential estimation in the Appalachian region of Canada. „*Pap. Geol. Surv. Can.*“, 1975, N 1, Part c, p. 123—132.
148. *Forgotson James M., Stout John L.* Future of well data formation system. — *Comp. Appl. Earth Sci. Int. Sympt. Proc. Conf.*, Lawrence, 1969, New York—London, 1969, p. 61—72.
149. *Forgotson James M., Stark Philip H.* Well-data files and the computer, a case history from northern rocky mountains. *Bull. Amer. Assoc. Petrol. Geolog.*, 1972, 56, N 6, p. 1114—1127.
150. *De Geoffroy J., Wignall T. K.* Design of a statistical data-processing system to assist regional exploration planning. Part 1. — *Can. Mining J.*, 1973, 94, N 11, p. 30—35.
151. *De Geoffroy J., Wignall T. K.* A probalistic appraisal of mineral resources in a portion of the Grenville province of the Canadian Shield. — *Econ. Geol.*, 1971, vol. 66, N 3, 466—479.
152. *Gose E. E.* An Adaptive Network for Producing Real of Binary Inputs. — *Information and Control*, vol. 8, p. 111—123, April 1965.
153. *Gose E. E.* A Synthesis Technique for Networks Consisting of Logical Functions. Feeding a Lear, Summation Element. *JEEE Trans. on El. Computers*, vol. EL—14, p. 254—256, April 1965.
154. *Greder G. C., Rappoport L. A., Seegeres R. G.* Experiment in quantitative geology modeling. — *Bull. Amer. Assos. Petrol. geol.*, 1974, 58, N 3, p. 488—498.
155. *Hutchinson W. W., Roddick J. A.* Sub-Area retrieval system used on the Coast Mountain. Project of the Geological Survey of Canada. *Pap. Geol. Surv. Can.* 1974, N 74—60, p. 29—33.
156. *Jelinková J.* Vytvóření a využití ložiskové Databanky v Severočeskem hnedouhelném revíru. — *Symp. pracov. bank. pram. Příbram*, 1972, Sek. M. S. I. s. a. 161—178, 403—404.
157. *Kendall M. G.* Ranc Correlation Methods, New York, 1955.
158. *Knížková J., Buzkova H.* Automatizované zpracování geologických informací a tvorba databank v Odvetvovém informačním středisku Geofond Praha. — *Symp. pracov. bank. pram. Příbrav.* 1972, Sek. M. S I., s. a. 303—307, 423—425.
159. *Kuchnle Manfred R., Kok Jurgen, Jochinson Diter.* Data card and method of encoding same (Elliot Business Machines, Inc.—), Пат. США, кл. 235—61 (GOGK 19/06), № 3705294, заявл. 8.04.70. Оpubл. 5.12.72.
160. *Leech J. B.* Project Appalachia (reports 22—29—), „*Pap. Geol. Surv. Can.*“, 1975, N 1, Part C, p. 121—122.
161. *Montgomery J. H., Sinclair A. J., Wynne-Edwards H. J., Fox A. C., Giroux J. H.* Date capture in the construction of MINDEPs computer based mineral deposits files. „*Can. Earth. deposits files Sci.*“, 1975, 12, N 4, p. 798—803.
162. *Patric E. A., Fisher F. P.* Nonparametric Feature Selection. — *JEEE Transactions on Information Theory*, March, 1969, vol. IT—15.
163. *Wiatr Inez, Stenzel Przemyslaw.* Metodyka zastosowan matematycznego modelu liniowego do analizy map geologicznych. „*prz. geol.*“ 1973, 21, N 8, p. 449—452.
164. *Wong A. S., Chung C. F., Fabbi A. G.* Data managment and display for project Appalachia „*Pap. Geol. Surv. Can.*“ 1975, N 74, P. 62—67.
165. *Wynne-Edwards H. R., Sinclair A.* MINDEP; „*West. Miner.*“, 1976, 49, n. 2, p. 10—18.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Автоматизированная информационно-прогнозирующая система 25, 35, 141, 147, 168, 176, 226
- Автоматизированная система прогнозирования 23, 24
- Автоматический режим решения задач 143, 144, 226
- Алгоритмы распознавания образов 36, 44, 149, 226
 - Гиперпласт 75, 103, 183
 - детерминистские 49, 75
 - Обобщенный портрет 105
 - Перебор 105
 - Потенциал-2 104
 - туниковых тестов 103
 - статистические 49, 67
- Анализ информации: 37, 41, 171, 176, 185
 - геологических гипотез 177
 - информативности 109, 185, 201, 213
 - многовариантный 199
 - структуры выборки 180
- Анизотропия околорудного пространства 21
- Аппарат бинарных отношений 100
- Архив преобразованной информации 151
- Ассемблер 132, 136, 164, 173
- Банк данных 25, 26, 28, 37, 113, 169, 173
 - картографических 113, 116, 150
 - каталог 118, 173, 177
 - линейных 115, 118
 - объемных 115, 119
 - общего назначения 129
 - по скважинам 118
 - статистического анализа 28
 - территориальный 31, 148
- Безэталонная классификация 149
- Библиотека паспортов и правил 152, 155
 - программ 164,
- Библиотечный файл 152, 160
- Бинарная матрица 117
- Блок диалоговых процедур 145
 - разработки вариантов решений 148
 - обработки информации 31
- Видовой тип информации 134
- Выбор объектов обучения 61
- Выборка обучения 48
- Геологическая аномалия 124
- Геолог-интерпретатор 163, 169, 174, 180, 182, 183, 187, 188
- Геологические модели 117, 199, 217, 218, 222
- Геологический прогноз 18, 31, 169, 176, 228
- Геологический объект 21
- Генетический подход 34, 122, 217
- Гипотеза простоты 97
- Гистограммы распределений 182, 216
- Графа вершины 73
- Графа ветви 73
- Графоматематическая модель 123
- Двухуровневая структура планирования 14
- Диалог 139
- Диалоговые задачи 122, 139, 168
- Диалоговая система обработки 24, 145, 162, 229, 230
- Дисплей 230
- Задача оптимизации 16
- Идентификатор кода 137
- Иерархическое описание процесса 12
- Имитационное моделирование 141, 156
 - решающее правило 200
- Имитационный режим 158
 - ретроспективный анализ 224
- Интегрированная человеко-машинная система 25, 144, 228
- Интерактивный режим 158
- Интерфейс 148, 159, 161
- Информационная модель 178, 200
- Информативность факторов 51, 109, 185, 186, 201, 213
- Историко-геологический подход 32
- Картографическая информация 116, 121, 122, 128
- Картографическое преобразование 124, 125, 126, 127
- Картографическое моделирование 124
- Картографический признак 126
- Картографический фактор 126, 152
- Кодирование 125, 151
- Количественная интерпретация геологических данных 230

- Комплексные ореолы косвенных признаков 21
- Коэффициент компактности 63, 88, 90
- конкордации 53
 - ранговой корреляции 53
 - представительности 63, 88, 90
 - согласованности 56
 - удаленности 63, 88, 90
- Критерий Байеса 68
- Кусочно-линейная гиперповерхность 46
- Литологическая анизотропия 128**
- Личные библиотеки пользователей 252
- Личный целевой массив 179, 251
- Логическая структура данных 130
- Макроструктура системы 147**
- Математическое обеспечение 156
- Массив месторождений 28, 29
- признаков 127, 150, 152
 - прогнозных и классификационных карт 151, 155
 - решающих правил 155
 - структурный 120
 - трансформированной информации 150
- Матрица 36
- Машинная классификация 43
- Металлогеническое прогнозирование 18, 124
- Метод дифференциального ранжирования 54
- корреляционный 60
 - линейной классификации 75
 - информативных гиперсвер 89
 - нелинейной классификации 77
 - интервалов 61, 86
 - ранговой корреляции 51
 - распределений 58
- Минимальная единица информации 118
- Многосвязные совокупности классов 46
- Непараметрический метод обучения распознаванию 47**
- Области решений распознавания образов 47**
- Обобщенный язык функций 148, 162, 163, 165
- Объемная модель геологического строения 120
- Односвязные совокупности классов 46
- Описание листа 118, 152
- региона 118, 153
 - точки 118, 154
- Оптимальное количество информации 112
- решающее правило 67
- Оптимальной аппроксимации алгоритм 73
- Организация банка данных 40
- Открытая система 159
- Отношение эквивалентности 101
- Отождествляющие признаки 38
- Ошибка первого рода 19, 184, 219, 221, 234
- второго рода 19, 184, 219, 221, 234
- Параметрический метод обучения распознаванию 47**
- Первичный массив данных 117, 152, 173
- Перенос информации на машинные носители 40, 173
- Перфокарты 117,
- План 16
- ограничения 17
 - оптимальный 17
- Показатель эффективности геологоразведочного процесса 17
- Порог классификации 184, 207
- Постановка задачи геологического прогнозирования 40, 176, 215, 216
- Правило Байеса 68
- ближайших величин 50
- Признак 36, 38, 51, 127
- объединяющий 186
 - разъединяющий 187
 - шумовой 39, 187
- Проблемно-ориентированная система 156
- Прогноз 18, 20, 177
- многовариантный 199, 217, 218, 220
- Прогнозная карта 18, 20, 188, 195, 214, 217
- Прогнозные запасы 9
- Прогнозные заключения 20, 188
- Прогнозирование оруденения 34
- Процесс рудообразования 21
- Радиус близости 196**
- Радиус влияния факторов 127, 173
- Разделяющая поверхность 46
- Ранжирование признаков 51
- Распознавание образов 44
- алгоритмы 42, 44, 103
 - без учителя 99
 - геометрическая интерпретация 45
 - кусочно-линейные методы 46
 - линейные методы 76
 - методы 45, 67, 75
 - с обучением 45
 - с учителем 48

- Региональный геологический файл 28
 Родовой вид информации
- Сегмент данных** 133
- Системная модель геологоразведочного процесса 10
- Системный подход 11, 140
- Системы управления базами данных 136
- Специальное математическое обеспечение 164
- Стадия диагноза 19
 — ретроспективного анализа 20
 — прогноза 20
- Статистическая обработка данных 28
- Статистические характеристики геологических факторов 127, 197
- Страница данных 118
- Структура данных 133
 — математического обеспечения 159
 — массива
 — процесса управления поисково-оценочными работами 14
 — построения человеко-машинной системы 35
- Субрезидентный диск 164
- Таксономия** 99, 155, 181, 196, 198, 223
- Таксономический анализ 180, 181, 196
- Таксономическая классификация 180, 181, 196, 223
- Территориальный машинный фонд 169
- Технология диалоговых процедур 168
- Технологический модуль 141
- Технология человеко-машинного взаимодействия 168, 170, 174, 175
- Тип сегментов данных 133
- Транзитив 99
- Транслятор обобщенного языка функций 148, 162
 — языка пользователей 149
- Файл** 152
 — месторождений 29, 231
- Фактор** 125, 150, 172, 192
 — картографический 125, 150, 173
 — книга 118
 — интервалы 127, 182, 185, 200
 — лист 154, 157
 — числовой 150, 151
- Формализованный геологический язык 112
- Формирование банка данных 37, 122, 150, 169, 192
 — геологических прогнозов 19
 — матриц 36
 — рабочей гипотезы 177
 — целевого массива 155
- Формат данных 133
- Формула Муавра — Лапласа 96
 — Пуассона 95
 — Шеннона 112
- Фортран 132, 164
- Функции дикриминантные 46
 — полиномиальные 46
 — решающие 46
 — целевые 17
- Функционал качества 98
- Целевой массив** 117, 145, 148, 151, 155, 157, 200, 218
- Цена ошибок 234
- Человеко-машинные системы** 22, 35, 138, 145, 147, 168, 228
- Эвристический подход** 30, 31, 149, 188, 228
 — алгоритм Марченко 200, 207, 208, 209
- Эвристическое моделирование 141, 151, 169, 188, 200, 213, 222
- Экспертные оценки 51
- Экспертный фильтр 171
- Элементарная площадь 126
 — объем 27, 115, 120
 — ячейка 27, 115, 116, 125, 153, 154
- Этап геологопрогнозный 8
 — геологопоисковый 8
 — геологосъемочный 8
 — детальной разведки 9
 — оценочный 8
 — поисковый 8
 — предварительной разведки 9
 — региональных металлогенических исследований 8
 — эксплуатационной разведки 9
- Язык манипулирования данными** 130
 — описания структур данных 133
 — пользователей 163
- Ячейка сбора информации 116

ОГЛАВЛЕНИЕ

Предисловие	3
I. Системный анализ геологопоисковых исследований	7
1. Схемы этапности процесса геологопоисковых и разведочных работ.	7
2. Построение системной модели управления геологоразведочным процессом.	11
3. К постановке задачи оптимизации управления поисково-разведочными работами	15
4. Некоторые методологические аспекты прогнозирования полезных ископаемых при создании автоматизированных информационно-прогнозирующих систем	18
II. Человеко-машинные системы геологического прогнозирования как средство повышения эффективности прогнозов	22
1. Современное состояние разработок по созданию человеко-машинных систем геологического прогнозирования	22
2. Системный анализ процесса геологического прогнозирования	31
3. Структура и принципы построения человеко-машинной системы геологического прогнозирования	35
III. Математические методы решения задач геологического прогнозирования.	44
1. Общие вопросы применения методов распознавания образов в задачах геологического прогнозирования	44
2. Ранжирование и выделение информативных признаков по результатам опроса экспертов	51
3. Ранжирование признаков по данным выборки обучения в предположении их независимости	58
4. Ранжирование признаков по данным выборки обучения в предположении существования зависимости между ними	64
5. Статистические методы распознавания образов	67
6. Детерминистские методы распознавания образов	75
7. Оценка качества обучения распознаванию образов с учителем	94
8. Метод распознавания образов без учителя — «Транзитив»	99
9. Результаты сопоставления эффективности некоторых алгоритмов распознавания образов с учителем	103
IV. Методы организации и формирования геологических банков данных	111
1. Современное состояние информационного обеспечения задач геологического прогнозирования с использованием ЭВМ	111
2. Принципы организации банка картографических данных для решения задач геологического прогнозирования	113
3. Картографическая информация — основа формирования территориальных банков данных	122
4. Структура и организация универсального банка данных общего назначения (БОНД)	129
V. Принципы функционирования человеко-машинной системы геологического прогнозирования АИПС/ЕС «Регион»	138
1. Методика решения задач прогноза и оценки геологических ситуаций с применением ЭВМ	138
2. Структура и схема функционирования автоматизированной человеко-машинной системы АИПС/ЕС «Регион»	147
3. Технология человеко-машинного взаимодействия в системе АИПС/ЕС «Регион»	168

VI. Решение задач прогноза и оценки геологических ситуаций на основе человеко-машинной системы	190
1. Применение АИПС/ЕС для прогноза медного оруденения	190
2. Применение АИПС для прогноза редкометального оруденения	214
VII. Перспективы развития человеко-машинных систем в геологии	228
1. Развитие эвристического метода	228
2. Техническое развитие	230
3. Организационное развитие	231
4. Экономическая эффективность использования человеко-машинных систем	232
Список литературы	236
Предметный указатель	243

*БОРИС АЛЕКСЕЕВИЧ ЧУМАЧЕНКО,
ЕВГЕНИЙ ПЕТРОВИЧ ВЛАСОВ,
ВЯЧЕСЛАВ ВАСИЛЬЕВИЧ МАРЧЕНКО*

**СИСТЕМНЫЙ АНАЛИЗ ПРИ ГЕОЛОГИЧЕСКОЙ ОЦЕНКЕ
ПЕРСПЕКТИВ РУДОНОСНОСТИ ТЕРРИТОРИИ**

*Редактор издательства В. В. Кузовкин
Переплет художника Л. А. Шиханова
Художественный редактор Е. Л. Юрковская
Технические редакторы Т. П. Локтионова, Н. Ю. Якунинская
Корректор Э. Г. Агеева*

ИБ № 2897

Сдано в набор 17.01.80. Подписано в печать 04.06.80. Т-12116. Формат 70×100^{1/16}. Бумага тип. № 1.
Гарнитура «Литературная». Печать высокая. Усл.-печ. л. 19,99. Уч.-изд. л. 20,09. Тираж 1500 экз.
Заказ 38/7401—14. Цена 3 р. 50 к.

Издательство «Недра», 103633, Москва, К-12, Третьяковский проезд 1/19
Ленинградская картографическая фабрика объединения «Аэрогеология»

3050

3359

BEAFA