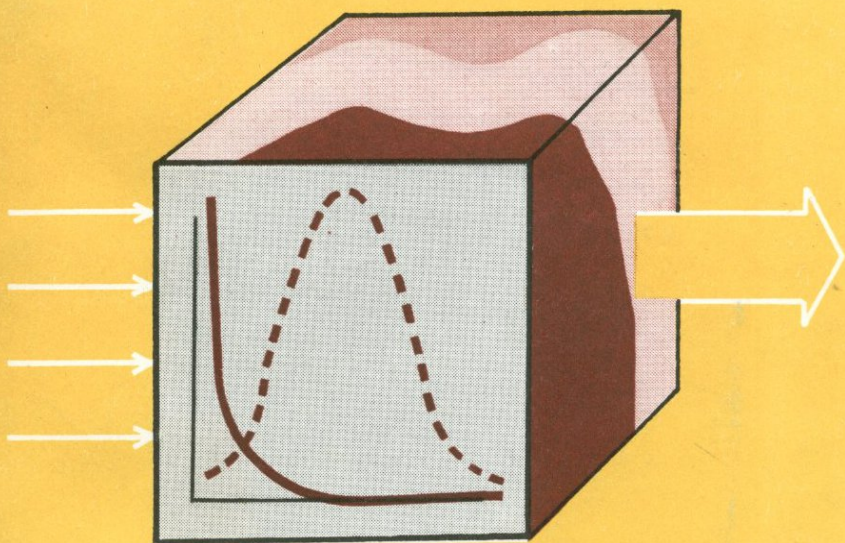


В. Ю. Раевский

МАТЕМАТИЧЕСКИЕ
РЕШЕНИЯ
ТРУДНО
ФОРМАЛИЗУЕМЫХ
ГЕОЛОГИЧЕСКИХ
ЗАДАЧ



«Наука»

4798

Раевский В. Ю.

Математические решения трудно формализуемых геологических задач.

М., 1987. 1 р. 30 к.

АКАДЕМИЯ НАУК СССР
МИНИСТЕРСТВО ГЕОЛОГИИ СССР
Всесоюзный научно-исследовательский институт
экономики минерального сырья
и геологоразведочных работ

В. Ю. Раевский

МАТЕМАТИЧЕСКИЕ РЕШЕНИЯ

трудно
формализуемых
геологических
задач

Ответственный редактор
доктор геолого-минералогических наук,
профессор Д. А. Родионов



МОСКВА
"НАУКА"
1987



4798
8674

Раевский В.Ю. Математические решения трудно формализуемых геологических задач. М.: Наука, 1987. 118 с.

Рассмотрены теоретические основы и методические приемы сбора, анализа и обработки смысловой информации применительно к геологическим задачам. Предложен способ преобразования смысловой (понятийной) информации в числовую, расширяющий возможности применения аппарата математической статистики при обработке качественной информации. Показаны примеры решения геологических задач на базе предлагаемой модели принятия решений

Для специалистов по системам обработки геологических данных.

Табл. 12. Ил. 18. Библиогр. 84 назв.

Рецензенты:

В.С. Бирюков, Н.И. Буялов

ВВЕДЕНИЕ

Современная наука располагает достаточным арсеналом разнородных средств, предназначенных для решения геологических задач самого различного характера. Совокупность геологических задач можно разделить на простые и сложные.

Простые геологические задачи, в которых используется небольшой комплекс признаков, решаются сейчас весьма успешно. Характерная особенность их заключается в том, что большинство признаков выражается числовыми данными и к ним могут быть применены количественные методы обработки информации, т.е. объективные методы. Такие задачи присущи поискам и разведке полезных ископаемых, изучению структурных особенностей геологических объектов — объектов, для которых имеются эталоны. Можно сказать, что к этой группе относятся задачи, для которых достигнут соответственный уровень знаний и опыта, или, что тоже самое, имеется необходимый информационный потенциал.

К сложным принадлежат задачи, для которых еще не накоплен достаточный информационный потенциал. Эта группа объединяет такие актуальные для практической и теоретической геологии вопросы, как генезис геологических явлений и их прогнозирование. При решении таких задач у геолога нет возможности экспериментального воспроизведения наблюдаемого явления. Историю процессов он восстанавливает путем сравнения и аналогий, опираясь на такие понятия, как больше, меньше, сильнее, слабее, раньше, позже, одновременно, возможно, невозможно и тому подобные расплывчатые, описательные показатели. Заключение геолога, следовательно, в подавляющем большинстве случаев являются качественными. В значительной степени качественный подход определяется и тем, что геолог имеет дело с процессами сложными, представляющими собой результат взаимного наложения обширного ряда более элементарных процессов. Выделить эти элементарные части и установить, какие из них являются основными и какие в той или иной мере случайными, на интуитивной основе чрезвычайно трудно или вообще невозможно.

Таким образом, одна из характерных черт сложных задач заключается в том, что некоторая часть информации (порой значительная), необходимая для принятия решения, отсутствует или не может быть полностью формализована, поскольку содержит множество качественных признаков. Поэтому при решении таких задач возникает ситуация, когда сторонники той или иной точки зрения, в равной мере оперируя одними и теми же фактами наблюдения в природе и не превосходя друг друга в системе аргумен-

тов, не в состоянии прийти не только к единому мнению, но и доказать преимущество одного предположения перед другим.

Сложившееся положение в первую очередь отражается на эффективности проведения прикладных геологических исследований, когда, например, не имея установившейся гипотезы происхождения какого-либо полезного ископаемого, при определении его перспектив приходится ориентироваться на мнение большинства. Однако в истории известны случаи, когда суждение небольшой группы исследователей, не согласующееся с мнением большинства, оказывалось верным.

Возникает проблема нахождения путей, позволяющих при выборе решения устранить подавляющее влияние "большинства", чтобы была возможность сопоставить достоверность "наиболее типичных" и "нетипичных" случаев. В геологии решению таких задач уделяется недостаточно внимания.

Кроме того, непрерывно увеличивающийся поток научно-технической информации заставляет изыскивать новые способы ее обобщения. Необходимо создание такой модели, которая объединяла бы не только теоретические и практические результаты собственно геологических исследований, но и результаты исследований из смежных областей науки, которые могли бы быть использованы при решении той или иной геологической задачи.

Поэтому представляется актуальным рассмотреть методологические принципы поиска объективных высказываний, определяющие по возможности неволевой (обоснованный) подход к решению трудно формализуемых геологических задач.

Предлагаемый в книге аппарат представляет собой модель, позволяющую объективно оценивать суждения исследователей о генетических явлениях и прогнозах в геологии. Разработки отличаются от традиционных исследований тем, что они, количественно сопоставляя "господствующие", "абсурдные" и "забытые" мнения ученых, позволяют не упустить знания, накопленные в мировой литературе, о природе конкретного геологического явления и выявить наиболее обоснованные, достоверные, объективные высказывания об этом явлении.

Приводимые в работе теоретические основы обобщения знаний о предмете исследования, некоторые методические приемы сбора, формализации, обработки информации, составляющие методологию решения трудно формализуемых геологических задач, а также практические результаты, полученные при реализации предлагаемой методологии, являются новыми в области применения математических методов и ЭВМ в геологии и повышают надежность геологических выводов при обработке качественной информации.

Книга состоит из 6 глав. Теоретические представления, развиваемые в первой главе, служили основой при разработке методов, описанных в следующих главах. Эти методы применялись при решении конкретной геологической задачи — выяснении генезиса нефти и газа и природе их миграции. Во всех разделах особое внимание уделяется вопросам обоснования необходимости применения математических методов и ЭВМ при решении задач, разработке специальных методов, а также выборе математического аппарата, наиболее соответствующего для применения в конкретной операции.

Глава 1

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ОБЪЕКТИВНОГО ПРИНЯТИЯ РЕШЕНИЙ

Целью этой части работы не является систематическое изложение всех методов, с помощью которых возможно решение сложных, трудно формализуемых задач. В ней затронут лишь тот короткий перечень вопросов, на котором базируется технологическая цепочка разработанного подхода к исследованию таких задач в геологии. К этим вопросам относятся основные понятия теории принятия решений, приемы, процедуры и методы экспертных оценок, а также принципы системного анализа — все они признаны рациональным инструментом исследования в ситуации неопределенности.

На рис. 1 изображена модель предлагаемого аппарата решения. Рассмотрим последовательность решения, которая хотя и не отличается от традиционных, но во многом определяется новизной операций, производимых в модульных блоках.

1. О ХАРАКТЕРЕ ИСХОДНЫХ ГЕОЛОГИЧЕСКИХ ДАННЫХ И УРОВНЕ ИХ НЕОПРЕДЕЛЕННОСТИ

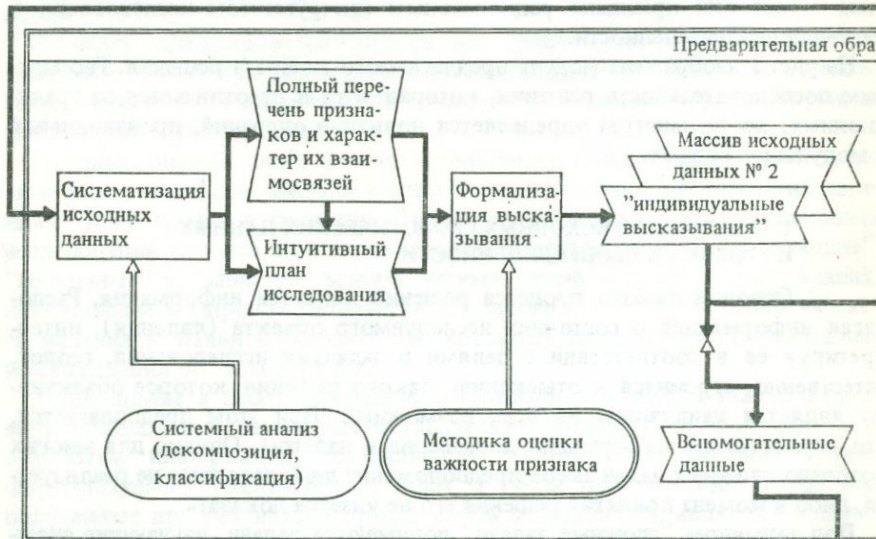
Основой любого процесса решения является информация. Располагая информацией о состоянии исследуемого объекта (явления), интерпретируя ее в соответствии с целями и задачами исследования, геолог, естественно, стремится к отысканию такого решения, которое объективно является наилучшим из всех возможных. При этом предполагается, что используемая информация достоверна и надежна. Однако для многих особенно сложных задач такое предположение либо заведомо не реализуется, либо в момент принятия решения его не удастся доказать.

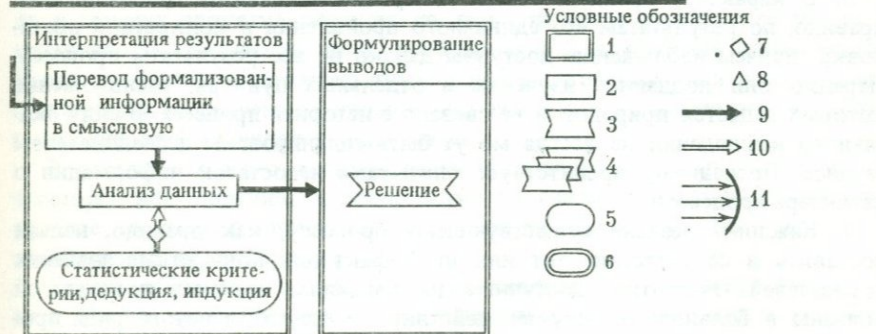
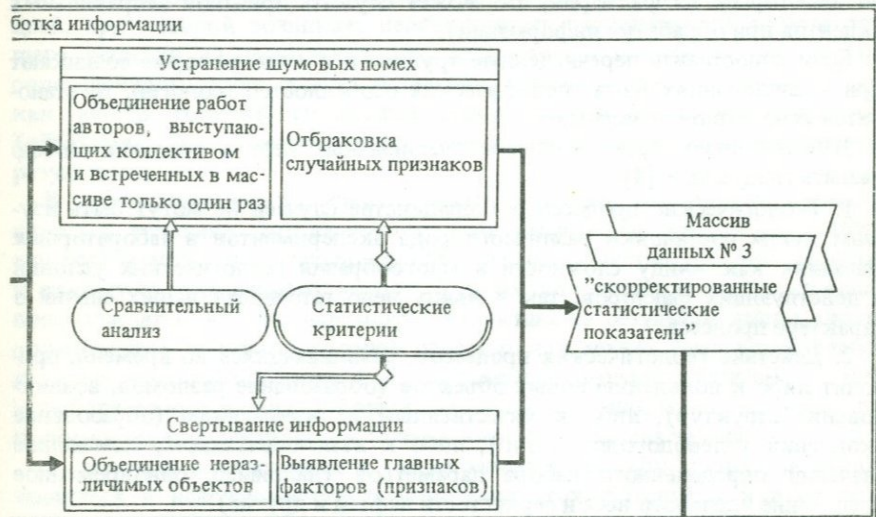
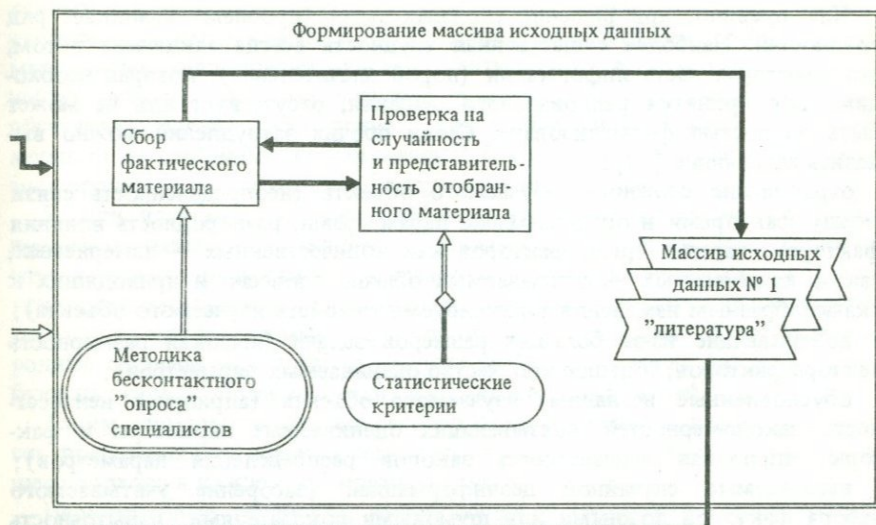
Под термином "сложные задачи" понимаются задачи, изучающие сложные природные явления (объекты), требующие многоаспектного подхода, в отличие от сравнительно простых задач, при решении которых достаточно собрать и соответствующим образом обработать информацию, относящуюся к небольшому количеству свойств изучаемых объектов.



Рис. 1. Модель процесса формирования решения при поиске объективных высказываний о предмете исследования

1 — граница модуля; 2 — операция или группа мелких операций; 3 — документ; 4 — компактное множество носителей данных одинакового вида; 5 — традиционные методы обработки информации, необходимые в конкретной операции; 6 — специально разработанные для модели методы и приемы обработки информации; 7 — альтернатива; 8 — выделение заданной выборки; 9 — направление потока информации; 10 — управляющее воздействие; 11 — внешняя среда





- Условные обозначения
- | | |
|-----|--------|
| □ 1 | ◇ 7 |
| □ 2 | △ 8 |
| ▭ 3 | → 9 |
| ▭ 4 | ⇨ 10 |
| ○ 5 | ⇨⇨⇨ 11 |
| ○ 6 | |

Как правило, при решении сложных задач (проблем) возникает ряд трудностей. Наиболее существенная трудность всегда заключена в том, что некоторая часть информации (порой значительная), которая необходима для принятия рационального решения, отсутствует или не может быть полностью формализована. Среди прочих затруднений можно выделить следующие [63]:

отражающие сложность изучаемого объекта (неопределенность связи между факторами и оцениваемыми параметрами, разнородность влияния факторов, наличие среди факторов как количественных — измеряемых, так и качественных — описываемых обычно словесно и приводящих к скачкообразным изменениям фиксируемых свойств изучаемого объекта);

возникающие из-за больших размеров задачи (высокая размерность вектора факторов, большое количество оцениваемых параметров);

обусловленные незнанием изучаемого объекта (априорная неизвестность закономерностей, связывающих оцениваемые параметры и факторы, априорная неизвестность законов распределения параметров);

вызываемые случайной дезинформацией (засорение учитываемого набора факторов ложными или шумовыми показателями; избыточность по некоторым из факторов, что может служить причиной неправильных акцентов при обработке информации).

Если сопоставить перечисленные трудности с теми, которые возникают при исследованиях из-за специфических особенностей геологии, то становится ясно, что они совпадают.

Действительно, среди наиболее специфических черт в геологии можно назвать следующие [4]:

1. Геологические процессы в большинстве случаев не могут быть изучены путем постановки различного рода экспериментов в лабораторных условиях как ввиду сложности и многообразия геологических условий и действующих факторов, так и ввиду недостаточности наших знаний о характере процесса.

2. Действие геологических процессов, развивающихся во времени, приводит либо к появлению новых объектов (образование разломов, возникновение структур), либо к качественным их изменениям (образование скоплений углеводородов и т.п.), либо к взаимосвязанному изменению значений определенного набора параметров (например, одновременное увеличение удельного веса и сернистости нефтей и другие).

3. О характере процесса геолог вынужден составлять суждение, как правило, по результатам его единичного проявления в конкретной обстановке, причем наблюдению доступны далеко не все результаты процесса. Нередко они поддаются изучению в отдельных пунктах, расположение которых задается природой и не связано с историей процесса. В силу сказанного наблюдения не всегда могут быть спланированы исследователем заранее. Последнему препятствует опять-таки недостаток информации о характере процессов.

4. Каждому реально существующему процессу, как правило, нельзя поставить в соответствие тот или иной факт или конкретные значения показателей. Напротив, доступные наблюдениям значения показателей обязаны в большинстве случаев действию не одного, а целого ряда процессов, находящихся друг с другом в различных временных и простран-

венных взаимоотношениях. При этом течение каждого из геологических процессов может искажаться и под действием случайных, неконтролируемых причин (или причин, пока не учитываемых в силу недостаточности наших знаний). Поэтому значения показателей, отражающих реальную обстановку, нужно рассматривать как сумму систематических (закономерных) и случайных составляющих, обязанных действию предполагаемых процессов и некоторых неконтролируемых случайных причин.

5. При изучении геологических явлений четкое выделение отдельных процессов зачастую практически невозможно (исключением являются случаи, когда присутствуют показатели-индикаторы процесса), а результаты действия подобной динамической системы не могут быть описаны хорошо интерпретируемыми функциональными связями. Изучаемые природные системы, как правило, зависят от большого числа факторов и требуют для своего описания многих параметров (показателей).

Таким образом, задача по расшифровке природы и сущности процессов представляется специфичной, весьма сложной и требует соответствующего подхода в каждом конкретном случае.

Перечисленные трудности, возникающие в процессе решения, создают условия, которые принято называть ситуацией неопределенности, ситуацией, при которой возникает необходимость рационального выбора лучшего решения из нескольких при недостаточности информации. Эта неопределенность чаще всего является следствием вероятностного, или, как иногда говорят, стохастического характера исследуемых явлений (объектов), т.е. невозможности точного предсказания окончательных результатов.

В противоположность этому принято говорить, что выбор осуществляется в условиях определенности, если имеется множество возможных решений и если относительно каждого из них известно, что оно приводит к какому-то конкретному исходу. В таких детерминированных ситуациях предполагается, что все параметры, влияющие на результат, имеют вполне определенное значение (которое известно или может быть установлено) и задача заключается лишь в перечислении возможных решений, а затем в выборе одного из них, который дает максимум или минимум некоторого показателя эффективности (в зависимости от смысла задачи).

В ситуации определенности также могут быть свои трудности, заключающиеся в невозможности перечисления всех решений или в подборе необходимого показателя эффективности. Могут быть также случаи, когда в момент принятия решения по тем или иным причинам отсутствует возможность получения всей необходимой для решения информации. Тогда прибегают к вероятностным оценкам, которые хотя и не связаны с природой исследуемого явления, но вытекают из возможностей сбора и особенностей имеющейся информации.

В отличие от этого в ситуации неопределенности всегда (!) имеются факторы или явления, вероятностные по своей природе и неконтролируемые со стороны принимающего решение, т.е. существуют задачи, не поддающиеся полной формализации.

Таким образом, проведенное сопоставление показывает, что сложные, трудно формализуемые геологические задачи находятся в условиях неопределенности. Вероятно, поэтому многие геологи вообще несколько

скептически относятся к поискам общих закономерностей и не находят каких-либо кардинальных возможностей по преодолению названных трудностей, считая, что геология неизбежно есть и будет наукой описательной, с нечеткой терминологией, с расплывчатыми заключениями, с интуитивно устранимой неопределенностью в познании основного в геологических процессах [79].

Однако невозможность полной формализации не исключает возможности применения точных наук — математико-статистического и логического анализов в процессах принятия рационального решения. Наоборот, вероятностный характер ситуации неопределенности определяет методы, которые используются при решении задач в этих условиях.

2. КРАТКИЙ ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ РЕШЕНИЯ СЛОЖНЫХ, ТРУДНО ФОРМАЛИЗУЕМЫХ ПРОБЛЕМ

На современном этапе развития науки и техники особое значение приобрели методы теории вероятностей и математической статистики, которые являются прогрессивным инструментом изучения сложных явлений. Эти дисциплины в сочетании с логическими методами и приемами обеспечивают количественный анализ и формализацию процессов подготовки рациональных решений. Исследование сложных задач сейчас немислимо также и без использования методов системного анализа. Важная роль в этом комплексе придается суждениям и предположениям, полученным от ученых и специалистов. Широкое использование интуиции и опыта геолога дает возможность сдвинуться с места при формализации изучаемых явлений даже в самых неопределенных ситуациях. Упомянутые ведущие методы решения сложных проблем в той или иной мере применяются в геологии. Однако их использование носит разрозненный характер. Уделяя большое внимание одному какому-либо методу, авторы порой не используют в комплексе исследований наиболее эффективные подходы, апробированные в других областях науки и техники при решении особенно сложных проблем. Поэтому кажется целесообразным сконцентрировать в одном месте самые прогрессивные методы изучения сложных, трудно формализуемых проблем и тем самым осветить новый подход к их решению.

Математико-статистические решения. Еще в недалеком прошлом основная роль статистики заключалась в сборе, систематизации и обобщении массовых эмпирических данных с целью выявления (и редко объяснения) тех или иных явлений. В настоящее же время основные задачи статистики все более сосредоточиваются на формулировании положений относительно характера и структуры изучаемых явлений, оценке их возможных свойств и других вопросах.

Поскольку решение таких задач всегда осуществляется в условиях неполноты необходимой информации и невозможности точного предсказания результатов, математическую статистику иногда рассматривают как науку, изучающую правила принятия решений в условиях неопределенности [77]. Теория принятия решений, как правило, позволяет оценить количественно возможные стратегии в случае, когда построение математической модели чрезвычайно затруднено или невозможно.

В современной теории статистических решений (теории принятия решений) принято рассматривать три типичные ситуации, различающиеся, по существу, степенью информированности о факторах или явлениях, неконтролируемых со стороны принимающего решение [1, 44, 73].

1. Детерминированный случай — определенная ситуация, когда относительно каждого действия известно, что оно неизменно приводит к некоторому конкретному, порой единственному, исходу. При этом предполагается, что факторы фиксированы, т.е. значения их известны.

2. Рискованная ситуация, когда каждое действие может привести к конечному или бесконечному множеству возможных частных исходов, причем каждый исход имеет определенную вероятность (или определенный закон распределения вероятностей). Предполагается, что принимающему решению эти вероятности (законы их распределения) известны.

3. Неопределенная ситуация, когда то или иное действие (или все действия) имеет множество возможных исходов, но вероятности этих исходов принимающему решению неизвестны.

Кроме перечисленных, конечно, возможны и некоторые промежуточные ситуации, когда, например, одновременно присутствуют случайные и неопределенные контролируемые факторы или известен вид закона распределения случайных факторов, но неизвестны его параметры (математическое ожидание, дисперсия). Однако и в таких, казалось бы, безнадежных ситуациях возможны количественная оценка и сравнение альтернативных решений.

Задача принятия решения всегда является многокритериальной. В ней по нескольким параметрам (показателям) эффективности отыскивается наилучший в каком-либо смысле вариант. Параметры эффективности в каждом конкретном случае выбираются исследователем в соответствии с желаемым результатом. Например:

необходимо минимальным набором признаков (факторов) описать какое-то явление; тогда показатель эффективности (критерий) — необходимый и достаточный минимум количества признаков;

предпринимается ряд мер по повышению надежности какого-либо высказывания; критерий эффективности может выражаться, например, максимальной суммой баллов по этому высказыванию (при условии, что количество признаков одинаково).

Другие примеры, которые характеризуют процесс назначения параметров эффективности, рассмотрены при решении конкретных геологических задач.

Отыскивается лучший вариант с помощью специальных методов и критериев, описание которых приводится в пятой главе при рассмотрении этапа "поиск объективных высказываний".

Полезность и вероятность в принятии решений. Часто вместо значений показателя эффективности в основной схеме принятия решений исследователи вынуждены оперировать с другим показателем, называемым полезностью (или выгодностью). Целесообразность такой замены возникает в ситуациях, когда необходимо оценить вероятность не массового, а индивидуального, единичного события или когда на первый план выдвигаются сугубо личные мотивы лица, принимающего решение, или при наличии риска проигрыша при выборе решения.

Рассмотрим сущность понятия "полезность" и способы ее вычисления [52]. Представим себе, что для любых двух объектов или любых двух мыслимых событий принимающий решение имеет отчетливое ощущение предпочтения. Более того, допустим, что он может сравнивать не только события, но и комбинации событий с заданными вероятностями. Под комбинацией двух событий будем понимать следующее.

Пусть два события обозначены через A_1 и A_2 ; принимаем для простоты, что заданное соотношение вероятностей составит 50% к 50%. Тогда их комбинацией является возможность наблюдать реализацию A_1 с вероятностью 50% и (если A_1 не происходит) реализацию A_2 с оставшейся вероятностью 50%. Подчеркнем, что обе альтернативы являются взаимоисключающими, так что здесь не остается возможности для чего-либо третьего. Кроме того, имеется абсолютная уверенность в том, что либо A_1 , либо A_2 произойдет. Таким образом, для событий, более или менее вероятных, можно установить упорядоченность.

Допустим далее, что принимающий решение обладает четким представлением о том, предпочитает ли он событие A_1 равновероятной комбинации событий A_2 и A_3 или, наоборот, события A_2 и A_3 предпочтительнее события A_1 . Ясно, что если для него A_1 является более предпочтительным, чем A_2 и чем A_3 , то A_1 будет также более предпочтительным, чем указанная комбинация. Аналогично: если A_2 и A_3 предпочтительнее, чем A_1 , то он предпочитает эту комбинацию.

Однако если бы для него A_1 было предпочтительнее, чем A_2 , а A_3 предпочтительнее, чем A_1 , то любое утверждение о его предпочтении A_1 по отношению к комбинации содержит существенно новую информацию. Более конкретно: если он предпочитает теперь A_1 равновероятной комбинации A_2 и A_3 , то это дает правдоподобные основания для численной оценки того, что его предпочтение A_1 по сравнению с A_2 превышает его предпочтение A_3 по сравнению с A_1 . Таким образом, получается критерий, по которому можно сравнивать предпочтение A_3 по отношению к A_1 с предпочтением A_1 по отношению к A_2 .

Введем численные измерители, т.е. выразим количественно степень предпочтения.

Предположим, что имеются три события A_1, A_2, A_3 , для которых выбранный порядок предпочтения их совпадает с тем порядком, в котором они записаны. Рассмотрим функцию $u(A_i)$ и примем $u_1 = u(A_1) = 1$ (как наилучший исход), $u_3 = u(A_3) = 0$ (как наихудший исход). Тогда значение функции $u(A_2)$ можно вычислить (точнее назначить), если использовать все возможные вероятностные распределения. Пусть p — вещественное число, заключенное между 0 и 1 и обладающее тем свойством, что событие $u(A_2)$ в точности столь же желательно, как и комбинарованное событие, составленное из $u(A_3)$ с вероятностью $1-p$ и $u(A_1)$ с остающейся вероятностью p . Тогда значение p предлагается принять за значение $u_2 = u(A_2)$, функция же $u(A_i)$, определенная таким образом, и называется функцией полезности, а ее значения u_i — полезностями соответствующих альтернатив.

Пусть теперь принимающий решение имеет выбор из m альтернатив, которые он упорядочил по степени их предпочтения следующим образом: $A_1 > A_2 > \dots > A_m$ (знак $>$ означает предпочтение). Описанным выше спо-

собом вычисляем полезности:

$$u_1 = 1 > u_2 > \dots > u_{m-1} > u_m = 0.$$

Здесь важно отметить, что во всех случаях полезности исходов определяются (назначаются) субъективно лицом, принимающим решение. Так, оценки вероятностей могут быть получены на основе личного опыта исследователя, или (и) на основе опыта принятия решений в аналогичных ситуациях в прошлом, или (и) из мнений экспертов и т.д. В общем можно сказать, что "... программа приписывания весов (полезностей) просто пытается уловить то, что разумные люди делают так или иначе" [49].

Определенные таким образом полезности играют роль показателя эффективности и для принятия решения можно воспользоваться одним из критериев, рассматриваемых в пятой главе.

Исчисление вероятности при описанном подходе настолько, по существу, основано на численном понятии вероятности, что будет весьма уместно сказать об этом понятии.

Понятие "вероятность" обычно связывается с классическим определением вероятности, существо которого заключается в том, что если некоторое событие может приводить к N единственно возможным, равновероятным и несовместимым элементарным исходам и если в m случаях появляется признак A , то вероятность события A определяется формулой

$$P(A) = m/N.$$

При рассмотрении большинства практических вероятностных задач часто встречаются события, которые также обладают свойством устойчивости частоты, но к ним неприменимо классическое определение частоты, так как каждое из них не распадается на конечное число единственно возможных, равновероятных и несовместимых случаев.

Принципиальные трудности в таких задачах преодолеваются при статистическом определении вероятности, где за вероятность события принимается относительная частота или число, близкое к ней. "Если о событии A известно, что оно может наступить в определенных условиях, которые могут повторяться неограниченное число раз, а в результате достаточно большого числа наблюдений установлено, что частота его почти всегда колеблется около некоторой (вообще говоря, неизвестной) постоянной величины p , то считают, что событие A имеет вероятность, которая равна p " [33].

Исчисление вероятностей при частотном подходе основано на законе больших чисел, который обеспечивает "объективную" обоснованность приписывания определенных численных значений вероятности возможным исходам при условии массового характера исследуемых явлений и при наличии достаточно большого числа подтверждений в прошлом. При таком понимании вероятности она, с одной стороны, становится такой же абстракцией, как и все понятия математики, а с другой — служит численной характеристикой некоторых сторон объективной реальности.

Однако при решении многих научных, технических, экономических и других задач исследователь часто не располагает достаточным числом наблюдений за изучаемыми явлениями в прошлом; существует (всегда остается!) неопределенность в отношении однородности и независимости

возможных исходов и отсутствует обоснованная преемственность между "прошлым и будущим". Использовать классический и частотный подход к оценке вероятности исходов в этой ситуации бывает трудно, а порой и просто невозможно. Вероятностные явления не имеют здесь массового характера (в том смысле, что они встречались в прошлом уже много раз), а следовательно, условия закона больших чисел не выполняются.

Вместе с тем очевидно, что проблема принятия решения в таких ситуациях не только не снимается, а наоборот, становится более острой. Для повышения обоснованности принимаемых решений в подобных ситуациях используют не только имеющиеся статистические данные, но и коллективный (индивидуальный) опыт. При этом информация должна быть выражена количественно в виде чисел, позволяющих, хотя бы относительно, оценить вероятность исходов с тем, чтобы выбрать один наиболее предпочтительный.

Правда, оценки вероятности свершения неповторяющихся событий субъективны. Действительно, два исследователя могут приписать различные числа одному и тому же возможному исходу. Но поскольку эти оценки базируются на постоянном анализе объективной действительности и накопленном опыте исследователя (что является следствием анализа объективной действительности), то предполагается, что при прочих равных условиях различие между полученными оценками не будет столь существенным, чтобы их нельзя было использовать для подготовки решения. Поэтому, сознавая различие между объективной и субъективной информацией, не следует считать, что последняя не представляет научной ценности.

Оценки вероятностей в таких случаях могут рассматриваться, например, как веса, позволяющие упорядочить возможные результаты по их относительной значимости. Утверждают [51], что в ситуации неопределенности полученная таким образом оценка будет наиболее рациональной. Существует подтверждаемое практикой предположение [7], сближающее субъективные оценки с объективными, в частности с частотными. Предположение состоит в том, что принимающий решение будет корректировать первоначальные оценки, приближая их к частотным, при получении дополнительной информации об исследуемой проблеме. Это логично, ибо при увеличении количества объективной информации в процессе познания нового всегда повышается правдоподобие вероятностных предположений и гипотез, происходит их трансформация в научное знание.

При всех обоснованиях возможности использования таких вероятностных оценок предполагается, что можно найти мысленный (или реальный) эксперимент, позволяющий моделировать ситуацию со случайными исходами, по отношению к которой лицо, принимающее решение, обнаружит определенное поведение и сумеет дать оценку вероятности возможных исходов. При этом если оно действует рационально, то при получении дополнительной информации его оценки будут приближаться к относительным частотам.

Психологи-экспериментаторы показали [51], что такого рода вероятностные оценки имеют определенный физический смысл. Во всяком случае, ясно, что человек как в научной, так и в повседневной деятельности, принимая решение при неполной информации, постоянно дает оценку вероятности различных событий. И хотя механизм этого явления до конца не понят,

устанавливая вероятностную оценку вполне объективных явлений, разумно поступать с ней, как и с вероятностью, найденной с помощью классического и частотного методов, полагая, что она обладает теми же свойствами и подчиняется тем же аксиомам. "Наше незнание объективных исходов событий, которые произойдут в будущем, и даже незнание объективных вероятностей их осуществления не имеет в этом случае особого значения, поскольку основной целью здесь является не оценка субъективных мнений отдельных людей, а объективные правила, позволяющие использовать эти мнения для выбора наиболее предпочтительного решения" [7, 51].

Системный анализ. Собирая фактический материал, геолог становится обладателем самой разнообразной по качеству и достоверности информации об изучаемом геологическом объекте. Нередко объемы ее превышают необходимые сведения и, оказываясь избыточными, затушевывают действительную картину. Как распорядиться обширным объемом информации, как извлечь из него максимум полезной информации, необходимой для решения конкретной задачи? Попытки ответить на этот вопрос неизбежно приводят к пониманию необходимости системного исследования геологических объектов или процессов. Рассмотрение предмета изучения в качестве системы дает возможность познать его как целое.

"В мире все без исключения находится во всеобщей взаимной связи, причем эта связь может быть либо непосредственной, либо более или менее опосредованной через ряд промежуточных звеньев, число которых может возрастать, практически до бесконечности" [34]. Поэтому для полного исследования какого-либо объекта (процесса) недостаточно описания только различных его сторон, свойств, а еще необходимо также изучение взаимосвязей, взаимоотношений между ними. Таким целям служит методология системного анализа.

В широком смысле термин "системный анализ" употребляют как синоним системного подхода. В свою очередь, системный подход — это изучение объекта (системы) как единого целого.

В узком смысле системный анализ трактуют как совокупность методологических средств, используемых для подготовки и обоснования решений по сложным, трудно формализуемым проблемам.

Методологический статус системного анализа как науки весьма необычен. С одной стороны, он располагает детализированными методами и процедурами, созданными специально для него, а также почерпнутыми из других областей науки. Это ставит его в ряд с другими прикладными направлениями современной методологии и даже, можно сказать, ставит его во главу этого ряда, поскольку трудно найти какое-либо современное прикладное исследование, в котором в той либо иной форме не были бы использованы принципы системного анализа.

С другой стороны, системный анализ еще не оформился в строгую и законченную теорию. Поэтому далеко не всегда обоснование решений с помощью системного анализа связано с использованием строгих формализованных методов и процедур; допускаются и суждения, основанные на личном опыте и интуиции.

Важнейшие принципы системного анализа сводятся к следующему: процесс принятия решений должен начинаться с выявления и четкого формулирования конечных целей;

необходимо выявлять и анализировать возможные альтернативы путей достижения целей;

необходимо рассматривать всю проблему как целое и выявлять все последствия и взаимосвязи каждого частного решения.

Центральным моментом в системном анализе является построение обобщенной модели, отражающей все факторы и взаимосвязи в реальной ситуации, которые могут проявиться в процессе осуществления решения [8]. Однако не все факторы и их взаимосвязи существенны, некоторые даже невозможно обнаружить. Поэтому сущность системного анализа заключается в том, что всегда предпринимается попытка учесть все возможные факторы, установить между ними связь, оценить их в целом и затем найти среди них существенные.

Построению модели предшествуют предварительные, приближенные представления о явлении (ситуации) в виде различных схем (например, различного рода классификаций или просто всего необходимого перечня факторов), отражающих мнение о наиболее важных параметрах и взаимосвязях их между собой. Такие представления называют концептуальными моделями. Иногда оказывается, что построение концептуальной модели достаточно для отыскания оптимального или по крайней мере рационального решения проблемы.

Однако чаще всего необходимо построение математической модели. Правда, трудно рассчитывать на появление моделей, которые полностью отражали бы природу явления, а также качественные и количественные взаимосвязи его параметров. Ясно, что реальная действительность всегда будет сложнее самых точных математических моделей, ясно также, что нельзя целиком опираться только на числовые данные — при принятии окончательного решения необходимо опираться на обычный здравый смысл.

Центральным обстоятельством при построении математической модели является простота математического аппарата, что часто оказывается более важным, чем предполагаемая точность результата. Поскольку структура и процесс решения проблемы часто на первых порах не могут быть достоверно определены, точность результатов решения не может превышать того, что заключено в самой проблеме, а следовательно, применение более сложного математического аппарата отнюдь не гарантирует более точного результата. "Математический аппарат, — пишет один из крупнейших специалистов по применению математических методов Пьер Массе, — гарантирует, если отвлечься от ошибок в расчете, только то, что полученные выводы являются прямым следствием принятых предположений. Математический аппарат ни на йоту не добавляет истинности самих предположений" [45].

Вывод о том, что "подлинная математика заключается не в нагромождении искусственных вычислительных приемов, а в умении получать нетривиальные результаты путем размышлений при минимуме применяемого аппарата" [57], оправдывает современную тенденцию, когда используются простые математические методы в сочетании с априорной информацией, выдаваемой специалистами и учеными.

В числе наиболее употребительных и отработанных методов и форм системного анализа и отображения его результатов можно назвать методы, основанные на принципах декомпозиции и классификации, которые дополняют друг друга.

Декомпозиционный подход. Основной принцип декомпозиционного подхода заключается в расчленении (декомпозиции) процесса решения проблемы на несколько взаимосвязанных подпроцессов, решающих отдельные задачи проблемы как целого. Допустим, что можно построить "широкую" модель, описывающую одновременно весь набор параметров (признаков) изучаемого явления. Однако данное явление, в свою очередь, может состоять из самостоятельных явлений. Тогда получить решение на таких моделях редко удается и возникает необходимость разбиения или декомпозиции модели на ее части, для которых уже легче найти решение. Затем результат решения одной части служит исходными данными для следующей задачи и т.д. Эти задачи решают, как правило, независимо друг от друга, а результаты локальных решений в последующем координируют.

Декомпозиционные процедуры имеют итеративный характер. Это означает, что при таком подходе иногда приходится производить переоценку в тех полученных ранее решений, используя результаты решения последней части задачи. На практике решение отыскивается часто путем последовательного перехода от одной части модели к другой и повторения этого цикла, пока не будет найдено рациональное или по крайней мере удовлетворительное решение общей задачи.

При декомпозиционном подходе возникают три основные проблемы: формулирование отдельных подзадач как частей решаемой в целом задачи;

оформление этих задач в виде, наиболее удобном для координации полученных решений и эффективном с вычислительной точки зрения; построение самой процедуры координации.

Формирование локальных подзадач прежде всего влечет за собой формулирование локальных целей, при этом определяется сфера информации, необходимой для решения этой частной подзадачи, что дает возможность исключить информацию, не относящуюся к сформулированной частной цели, тем самым уменьшить неопределенность ситуации еще до начала решения.

Проблема формулирования локальных задач относительно обособлена. Как правило, при ее решении исходят из реально сложившейся структуры объекта изучения. Вторая и третья проблемы взаимосвязаны и содержат много общего. Так, при формировании локальных целей, формализации подзадач, назначении критериев и их сопоставлении необходимо учитывать не только критерии рационального решения локальных задач, но и критерии оптимальности изучаемой системы в целом, а также соответствующие им ограничения.

К л а с с и ф и к а ц и я. Под этим термином понимают систему понятий, описывающую совокупность каких-либо объектов, их классов, имеющих один или несколько характеристических признаков. Классификация служит средством для установления связей между классами объектов, а также для точной ориентировки в многообразии понятий.

Правильно построенная классификация, отобразив закономерности изучаемых объектов, вскрывает связи между этими объектами и помогает еще на начальных этапах исследования ориентироваться в сложных ситуациях, служит основой для прогнозов, а порой и обобщающих выводов.

Обычная форма представления результатов классификации — это различные схемы и таблицы, в которых фиксируются закономерные или возможные связи между классами объектов с целью определения места объекта в общей системе. Указав место объекта в системе, можно выявить его характерные свойства. В этом смысле классификация служит средством хранения и поиска информации, содержащейся в ней самой.

В том случае, когда классификация представляет собой систему соподчиненных понятий, ее структура может быть изображена в виде графа — перевернутого "дерева". Узлу этого графа, являющемуся "корнем", соответствуют наиболее общие понятия, "листьям" — самые частные, а узлам разветвлений — остальные названия классов (подклассов, групп и т.д.). Отрезки, соединяющие узлы графа, выражают отношение соподчинения, в котором находятся более общие и менее общие понятия. Маршруты, идущие от "корня" к "листьям", называют вертикальными рядами классификации, а узлы, одинаково отстоящие от общего, объединяющего их понятия, образуют горизонтальный ряд [83].

Имеются два пути разработки таблиц и схем классификации: дедуктивный и индуктивный. Дедукции отдают предпочтение при формулировании исходных, общих понятий, а также в случаях, когда определяют принципы, по которым может проводиться разбиение, дробление изучаемого объекта на части.

Индукции отдают предпочтение при обработке фактического материала и при оформлении его в виде таблиц и схем. При этом основываются на частных понятиях об отдельных предметах или их совокупностях. При таком подходе, как правило, труднее обеспечить логическое единство и устойчивость классификации, чем при дедуктивном.

Обычно классификации строятся на основе обоих подходов. Высшие классы, сосредоточенные в верхней части графа, как правило, образуются дедуктивно, низшие — индуктивно.

Иногда термином "классификация" обозначают процесс разнесения объектов по классам. В этом случае было бы правильнее употреблять слово "классифицирование". Основным принципом этого процесса является сравнение рассматриваемого множества объектов с заданными, эталонными представителями классов, групп и т.п.

Традиционный способ решения задачи классификации состоит в следующем. Отбирают из множества признаков, описывающих объект или явление, тот признак, который по каким-то соображениям исследователя должен быть положен в основу классификации. Затем производят разбиение множества объектов на подклассы (группы) в соответствии со значением данного признака.

В тех случаях, когда из теоретических соображений следует, что классификация должна быть проведена не по одному базовому признаку, а по нескольким (упорядоченным между собой по важности или же никак не упорядоченным), тогда построение классификации осуществляется в несколько этапов. Сначала образуются классы объектов в соответствии с градациями первого признака, затем каждый из полученных классов разбивается на подклассы по второму признаку и т.д. Важной методической особенностью данного приема является то, что на каждом этапе, на каждом шаге разбиения во внимание принимается только один признак.

Таким образом, происходит последовательное формирование, пошаговое уточнение, удлинение описания классов. Такой вид классификации называется иерархической классификацией, иногда — монотетической [26].

Классификация называется политетической, если для ее построения используется несколько признаков одновременно. Наиболее распространенным приемом такой классификации является конструирование некоторого обобщающего показателя, функционально зависящего от исходных признаков. Одним из математических методов, воплощающих основную идею этого варианта, является дискриминантный анализ, в котором в качестве интегрирующего показателя используется так называемая дискриминантная функция, обычно линейно зависящая от нескольких переменных.

Для получения обобщенных показателей используют также методы факторного анализа и главных компонент, а также методы экстремальной группировки параметров [65, 66, 69].

После процедуры конструирования обобщающего показателя дальнейшая задача построения классификации сводится практически к построению последовательной монотетической классификации. Применяют также процедуру выделения, так называемых естественных групп. Смысл этой процедуры заключается в следующем: исходное множество объектов, подлежащих классификации, проецируется в пространство первых двух-трех выделенных факторов. Компактные сгущения, обнаруживаемые либо визуально, либо с помощью специальных методов, отождествляются с "естественными" группами (классами).

Наиболее эффективными современными методами для решения задач классификации являются методы автоматической классификации. Главное преимущество этих методов в том, что при их использовании не требуется специально конструировать или априори назначать эталонные объекты. Это осуществляется в автоматическом режиме по заранее разработанным критериям. Выбор критерия, по которому выделяются однородные области, классы или группы, зависит от используемого метода. Еще одним достоинством методов автоматической классификации можно назвать то, что результаты группировки получаются на более объективной основе.

Методическая особенность политетической классификации заключается в том, что итоговое описание классов, как правило, не содержит в явном виде исходных признаков, в связи с чем интерпретация выделяемых классов становится затруднительной.

Любая схема классификации основывается на двух фундаментальных положениях [26]:

1) в один класс объединяются объекты, сходные между собой по некоторым показателям;

2) степень сходства между собой у объектов, принадлежащих к одному классу, должна быть больше, чем степень сходства между собой у объектов, относящихся к разным классам.

Для оценки результатов классификации следует попытаться определить (назначить) критерий качества — целевую функцию, значения которой дадут возможность сопоставить между собой различные классификационные схемы. Наилучшей группировкой в соответствии со сформулированными выше общими положениями будет считаться та, в которой имеет место наибольшая однородность внутри групп или возможно большее различие

между группами. При этом целевой функцией может быть, например, функция от потерь информации, что обычно происходит при замене индивидуальных значений признаков их средними значениями по образовавшейся группе. В данном случае целевая функция будет соответствовать минимуму потерь информации.

В тех же случаях, когда описать целевую функцию в формализованном виде не удастся, критерием качества классификации может служить "понятийный" критерий, с помощью которого становится возможным интерпретировать смысловое содержание найденных групп, согласовывать результаты классификации с теоретическими представлениями.

Критический анализ метода экспертных оценок. При изложении теоретических возможностей математико-статистических методов при решении сложных задач в ситуации неопределенности неоднократно в явной или неявной форме обращалось внимание на то, что значительную роль в достижении цели играет информация, полученная от специалистов. Практика последних лет показала, что даже простые статистические методы в сочетании с этой информацией при выборе перспективных решений часто приводят к более успешным результатам, чем "точные" расчеты с ориентацией на средние показатели и экстраполяцию существующих тенденций [6, 7, 22, 23].

Сейчас все более широкий круг исследователей приходит к пониманию того, что даже при отсутствии строгих теоретических обоснований уровень неопределенности можно снизить за счет умелого использования суждений специалистов и способности человека принимать рациональные решения в условиях невозможности их полной формализации. Опыт, понимание существа проблемы, чувство перспективы и интуиция помогают специалисту в ситуации неопределенности оценить значимость альтернативных исходов, выбрать лучший критерий их оценки, а следовательно, и более рациональное решение.

Таким образом, использование информации, основанной на предположениях и суждениях специалистов, позволяет частично компенсировать недостаток статистических данных.

Практика последних лет показывает, что использование информации, полученной от специалистов, может быть плодотворно только тогда, когда для ее сбора, обобщения и анализа применяются специальные логические и математико-статистические методы, получившие название методов экспертных оценок [7, 18, 25, 48]. Для широкого круга трудно формализуемых проблем в социально-политической и научно-технической сферах деятельности экспертные процедуры являются эффективным, а в ряде случаев и единственным средством их решения.

Сущность метода экспертных оценок заключается в проведении экспертами интуитивно-логического анализа проблемы с количественной оценкой суждений и формальной обработкой результатов. Полученное в результате обработки обобщенное мнение экспертов принимается как возможное решение проблемы. При этом эксперт рассматривается как качественный источник информации (своего рода "информационный измеритель") на основании предположения, что он является "хранилищем" большого объема рационально обработанной информации.

Факторы, на которых основана способность индивидуума давать полез-

ную информацию в условиях неопределенности, можно разделить на внутренние (индивидуальные) и внешние (социальные).

Внутренние факторы складываются из индивидуальности эксперта и зависят от его знаний, опыта, интеллекта, способности предвидеть будущее и ряда других факторов, измерение которых сложно или вообще невозможно. Индивидуальные особенности могут оказать влияние на информацию, получаемую от эксперта, и привести к отклонениям как непараметрическим, т.е. к отклонениям, связанным с излишне оптимистическим или пессимистическим отношением к проблеме, так и намеренным, зависящим от индивидуальной установки специалиста.

Внешние факторы, отражающиеся на информации, полученной от специалиста, зависят от его взаимодействия с обществом. Эти влияния могут быть вызваны, например, взглядами коллектива на изучаемую проблему, положением эксперта в структуре этого коллектива и т.п.

Устранить смещения, влияющие на точность и надежность информации, получаемой от экспертов, можно, если использовать формализованные процедуры сбора информации и ее обработки методами теории вероятностей и математической статистики, комплекс которых получил название метода экспертных оценок. Этот комплекс методов позволяет в определенной степени компенсировать смещение индивидуальных оценок, уточнить структуру решаемой проблемы и снизить уровень неопределенности, существовавшей до экспертизы.

При выполнении своей роли в процессе решения эксперты осуществляют две основные функции: формируют объекты (альтернативные ситуации, решения и т.п.) и производят измерение их характеристик (вероятности свершения событий, предпочтения решений и т.п.).

Характерными особенностями метода экспертных оценок как научного инструмента сложных, трудно формализуемых проблем являются, во-первых, научно обоснованная организация проведения всех этапов экспертизы и, во-вторых, применение количественных методов как при организации экспертизы и оценке суждений экспертов, так и при формальной групповой обработке результатов. Эти две особенности отличают метод от обычной, давно известной экспертизы, широко применяемой в различных сферах человеческой деятельности.

Можно ориентировочно наметить следующие основные этапы проведения экспертизы, последовательность и содержание которых могут изменяться в зависимости от реальных условий и ограничений:

- организация экспертного оценивания;
- отбор и формирование группы экспертов;
- проведение опроса;
- анализ и обработка информации, заключающиеся в синтезе объективной (статистической) и смысловой информации, полученной в результате экспертизы, с целью приведения их в форму, удобную для принятия решения.

Организация экспертного оценивания. На этапе организации работ по применению экспертного оценивания формулируется цель исследования и разрабатываются правила проведения опроса. Практика показывает, что наличие четко сформулированных целей и ясно понятых потребностей — это обязательное условие обеспечения надежного результата

экспертизы. Полезным инструментом формулирования целей исследования является обследование — описание предыстории и текущего состояния проблемы, что может служить для выбора целей экспертизы. При формулировании целей важно иметь представление о специфических особенностях и интересах специалистов, которые будут участвовать в экспертизе. После сбора таких данных можно приступать к определению целей и задач экспертизы, а также к разработке основных правил ее проведения.

Выбор целей и характер процедуры экспертизы во многом определяется существом проблемы, ее сложностью, полнотой имеющихся данных и вида требуемой информации, предполагаемыми конечными результатами и возможными способами их представления. Поэтому при формулировании целей экспертизы прежде всего необходимо установить (или по крайней мере наметить) критерий, по которому надлежит производить оценку.

Отбор экспертов. Работу по отбору экспертов обычно начинают с определения областей науки, которые затрагивают решение данной проблемы. Затем составляется список лиц, компетентных в этих областях. Этот список служит основой для отбора кандидатов в эксперты. Составляя список кандидатов в эксперты, учитывают следующие качественные характеристики: компетентность, креативность, конформизм, аналитичность и широту мышления, конструктивность мышления, самокритичность [25].

Важным требованием к характеристикам эксперта является измеримость этих характеристик. В настоящее время делаются попытки ввести количественные оценки некоторых из перечисленных характеристик. Рассмотрим сущность названных характеристик специалиста как эксперта и возможные пути количественного описания этих характеристик.

Компетентность — это степень квалификации эксперта в определенной области знаний. Компетентность может быть определена из анализа плодотворности деятельности специалиста, уровня и широты знакомства с достижениями мировой науки и техники, понимания перспектив развития. Некоторые отождествляют такие понятия, как "компетентность" и "научный авторитет".

Имеются предложения оценивать компетентность экспертов по ученой степени, занимаемой должности, количеству публикаций или ссылок на научные труды эксперта. Несомненно, эти показатели в определенной степени отражают компетентность эксперта, однако экспериментальные данные не показали однозначной зависимости между этими показателями и уровнем компетентности [6, 25].

В практике экспертного оценивания получила распространение оценка компетентности с помощью самооценки эксперта и оценки его другими экспертами, сущность и подробную методику вычисления которой можно найти в [47].

Креативность — способность человека решать творческие задачи. В настоящее время, кроме суждений, основанных на изучении деятельности эксперта, неизвестно каких-либо предложений по количественной оценке этой характеристики.

Конформизм — подверженность влиянию авторитетов. Это свойство проявляется в виде неустойчивости собственного мнения эксперта. Возможность частичного устранения такой черты эксперта, отрицательно влия-

ющей на точность оценки, рассматривается при описании процедуры формирования группы экспертов для решения проблемы генезиса нефти и газа.

Аналитичность и широта мышления — важные характеристики эксперта. Специалист, имеющий глубокие знания, но обладающий "профессиональной слепотой", не может качественно решать проблемы, требующие взгляда, выходящего за рамки сложившихся представлений.

Конструктивность мышления — прагматический аспект мышления. Существуют квалифицированные специалисты со слабо выраженной прагматической направленностью. Однако эксперт должен давать решения, обладающие свойством практичности. Учет реальных возможностей решения проблемы очень важен при проведении экспертного оценивания.

Самокритичность — эта характеристика эксперта проявляется при самооценке степени своей компетентности, что неизбежно сказывается на итоге решения рассматриваемой проблемы.

Перечисленные характеристики эксперта достаточно полно описывают необходимые качества, которые влияют на успешность решения задачи экспертного оценивания. Возникает проблема согласования характеристик между собой. Естественно, чем больше характеристик принимается во внимание, тем труднее принять решение о том, что важнее и что допустимо для эксперта. Кроме того, анализ каждой из характеристик требует кропотливой и достаточно трудоемкой работы по сбору информации и ее обработке.

Некоторые исследователи предлагают так называемую обобщенную характеристику эксперта, допускающую непосредственное измерение [25]. В качестве такой характеристики принимается достоверность суждения эксперта, которая количественно оценивается по формуле

$$D_i = N_{пi} / N \quad (i = 1, 2, \dots, m),$$

где $N_{пi}$ — число случаев, когда i -й эксперт дал решение, подтвержденное практикой; N — общее число случаев участия i -го эксперта в решении.

При таком подходе к оценке "качества" эксперта названные характеристики не анализируются в отдельности, а считается само собой разумеющимся, что все они содержатся в этой оценке. Применение "обобщенной характеристики" требует информации о прошлом опыте участия специалиста в экспертизе. Однако почти всегда, особенно при решении новых проблем, подобная информация отсутствует.

Таким образом, исчисление количественной оценки качественных сторон эксперта пока представляется ненадежным и, следовательно, при формировании экспертной группы большая роль все еще принадлежит субъективным оценкам.

Пр о в е д е н и е о п р о с а. Опрос является одним из главных этапов проведения экспертизы. В процессе опроса эксперты качественно и количественно оценивают заданные модели объектов, выбирают наиболее предпочтительную из них, обогащают информационное обеспечение изучаемого объекта, создают предпосылки для построения новых моделей объектов.

При проведении опроса могут использоваться самые разнообразные виды опроса, каждый из которых имеет свои достоинства и недостатки. Построение процедуры сбора информации, обмена ее между экспертами и в создании условий независимого творчества экспертов.

Среди наиболее эффективных, а поэтому и распространенных видов опроса являются анкетирование и интервьюирование, заключающиеся в предъявлении экспертам опросных листов-анкет. Оба вида опроса позволяют наилучшим образом сочетать информационное обеспечение экспертов с их самостоятельным творчеством.

Во время интервью происходит непосредственный контакт корреспондента и опрашиваемого, что позволяет оперативно получить большое количество информации и всесторонне, хотя и поверхностно, осветить объект экспертизы. Однако при этом появляется возможность сильного влияния опрашивающего на ответы эксперта, возникает дефицит времени для глубокого продумывания ответов и одновременно происходят большие затраты времени, необходимые на опрос каждого из экспертов — все это является недостатком интервьюирования.

Различия между интервью и анкетированием — только в форме получения ответа. Если в первом случае опрос осуществляется устно, то при анкетировании требуется ответ в письменной форме. Анкетирование может быть очным и заочным. Наиболее благоприятным, вероятно, является заочное анкетирование, при котором появляется возможность привлечь более широкий круг специалистов с меньшими затратами труда и времени. Кроме того, достоинством заочного анкетирования является отсутствие непосредственного контакта эксперта с организатором экспертизы, при котором может возникнуть серьезное влияние анкетировавшего на ответы эксперта.

Анкеты имеют обычную форму таблиц. Вопросы анкеты можно классифицировать по содержанию информации. Содержание вопросов анкеты по анализируемой проблеме, естественно, определяется сущностью проблемы. По форме вопросы делятся на открытые (или свободные), закрытые и с "веером" ответов.

Открытым называется вопрос, ответ на который может быть дан в произвольной форме. Вопросы такой категории целесообразны вначале опроса (например, на этапе обследования), так как позволяют широко охватить анализируемую проблему, выявить необходимый спектр вопросов и определить основные точки зрения экспертов.

Закрытый вопрос задается в форме, предлагающей лишь три возможных ответа: "да", "нет", "не знаю". Разновидностью закрытых вопросов является вопрос с "веером" ответов, который предоставляет эксперту возможность выбрать один из предлагаемых вариантов ответа. Например, эксперт может указать на любой из ранжированных заданных объектов, на оценку их весов, на оценку вероятности некоторого события, на оценку значимости фактора (признака) в баллах и т.п.

Кроме анкет-вопросников эксперт снабжается пояснительной запиской, в которой раскрывается цель экспертизы, задачи опроса; в ней содержатся сведения об объектах экспертизы, указываются необходимые организационные мероприятия и инструкции по заполнению анкет. После проведения опроса группы экспертов осуществляется анализ и обработка информации.

Среди прочих видов опроса можно назвать дискуссию и мозговой штурм, целесообразность использования которых зависит от конкретной ситуации экспертизы.

Так, дискуссию целесообразно проводить для решения задач, не требующих точной количественной оценки объектов, параметров, альтернатив. В дискуссии эксперты могут использовать различные приемы и методы анализа проблемы при высказывании своих соображений, доводов, мотивов и доказательств. Чаще всего применяются логические и математические методы, в последнее время в ходе дискуссии стали использоваться ЭВМ, в частности для проведения совместного моделирования проблемы, построения дерева решений, вычисления матриц "цели—средства".

Результатом дискуссии является проект (часто в нескольких вариантах), в котором подводятся итоги обсуждения и согласования решения. Как правило, единогласного принятия всех пунктов решения достичь не удается, поэтому в конце решения или в специальном приложении к нему должны быть изложены особые мнения с их аргументацией. Опыт проведения дискуссий [28] свидетельствует о возможности получения ценных предложений от участников примерно через сутки после дискуссии. В связи с этим целесообразнее на следующий день обратиться к участвующим в дискуссии с просьбой высказать мнение, если прежнее в чем-то изменилось или может быть дополнено. После этого результаты дискуссии подвергаются анализу и обработке (в том числе количественной).

Мозговой штурм представляет собой метод получения новых путей решения проблемы в результате коллективного творчества группы людей в ходе заседания-сеанса, проводимого по определенным правилам. Метод мозгового штурма иногда называют мозговой атакой, методом коллективной генерации идей, методом группового рассмотрения с оценочной оценкой.

Принципиальной особенностью метода является абсолютное исключение в ходе самого сеанса критики и какой-либо оценки высказываемых идей. Никто никого не должен стеснять, поэтому группу желательно составлять из людей, занимающих примерно одинаковое общественное и служебное положение.

В процессе генерирования идей происходит критика и оценка высказанных предложений в скрытой форме. Они выражаются в большей или меньшей поддержке высказанной идеи. Предполагается, что неперспективная идея, скорее всего, не получит поддержки и будет заменена новой. Важно, что субъективное мнение автора проблемы пройдет этот "фильтр" и будет скорректировано. Именно поэтому важны свобода и смелость в высказывании идей, их большое количество и разнообразие.

Метод мозговой атаки целесообразно использовать в ситуациях дефицита творческих решений, новых идей, свежих концепций.

Из высказанных во время заседания-сеанса предложений отбирают наиболее ценные, а затем подвергают их тщательному анализу и критике. Процесс анализа и отбора (отбраковки) высказанных идей и точек зрения состоит, как правило, из классификации этих высказываний по некоторым категориям и последующей оценки (качественной либо количественной) по системе критериев [25].

А н а л и з и о б р а б о т к а и н ф о р м а ц и и. Целью этого этапа является получение обобщенных данных, а также выявление новой информации, содержащейся в скрытой форме в экспертных оценках.

Рациональное использование информации, полученной от экспертов,

возможно только при условии преобразования ее в форму, удобную для дальнейшего анализа, направленного на подготовку и принятие решений.

Возможности формализации информации зависят от специфических особенностей изучаемой проблемы, надежности и полноты имеющихся данных, необходимого уровня принятия решения и других факторов. Все множество проблем можно разделить на два класса. К первому относятся проблемы, в отношении которых имеется достаточный уровень знаний и опыта, т.е. имеется необходимый информационный потенциал. При этом методы опроса и обработки основываются на использовании принципа хорошего в среднем измерителя. Под термином "хороший в среднем" понимается возможность получения результатов измерения, близких к истинным. Суждения экспертов формируются вблизи среднего значения. Отсюда следует, что для обработки результатов группового экспертного оценивания проблем первого класса можно успешно применять методы математической статистики.

Ко второму классу относятся проблемы, для решения которых еще не накоплен достаточный информационный потенциал. При решении задач из этого класса проблем эксперт уже не может рассматриваться как хороший измеритель. В этом случае наряду с количественными методами сбора и обработки информации должны применяться и качественные, поскольку не исключена возможность того, что мнение одного исследователя, сильно отличающееся от мнений остальных, может оказаться верным. Поэтому применение только методов "осреднения" может привести к большим ошибкам.

Для применения математических методов необходимо требование, чтобы исходная информация выражалась через числовые данные. Существует достаточное разнообразие приемов и способов присвоения чисел в практике формализации информации, полученной от экспертов [7, 25 и др.].

В настоящей работе такие вопросы не рассматриваются, поскольку специфические особенности процесса сбора информации потребовали совершенно иного подхода к преобразованию смысловой информации в числовую. Новый подход подробно рассматривается в четвертой главе.

Итак, формализованная информация готова к обработке ее математическими методами. Прежде всего оценивается согласованность высказываний экспертов. Если все оценки одинаковые, то проблема измерения решена. Экспертиза задачу выполнила. Обычно не все оценки совпадают, более того, мнения экспертов сильно разнятся. Тогда в задачу анализа входит выявление точек зрения экспертов, особенно тех, кто дал крайние оценки. Это можно делать, обращаясь к экспертам с просьбой объяснить свои оценки, но можно и формальными методами, на основе анализа их оценок.

Если полученные результаты не дают возможности принять решение, то экспертиза проводится повторно (с разъяснением имеющихся разногласий) до тех пор, пока организаторы не будут удовлетворены. Так работают по известному методу проведения экспертизы — методу Дельфи [43, 84]. Воспроизвести все многообразие использующихся процедур и методов анализа экспертных оценок не представляется возможным, да это и не входит в задачу настоящей работы. Теоретические вопросы некоторых из них будут рассмотрены при описании решения конкретных геологических примеров. Желаящим подробно ознакомиться с различными методами

можно рекомендовать монографию [48], в которой достаточно детально рассмотрены ключевые идеи и методы, составляющие математическое обеспечение экспертизы.

В результате обработки, собранной от экспертов информации, формируется решение проблемы (либо его альтернативы), рациональное с точки зрения теории принятия решений. При этом следует помнить, что назначение методов теории принятия решений — объективно разобраться в каждом явлении, численно оценить их и, возможно, предложить варианты решений, т.е. задача состоит в подготовке решения, а не в его принятии. Напомним, что принимающий решение должен учитывать, помимо рекомендаций, полученных в результате применения методов экспертизы и теории принятия решения, основанных на количественных оценках, и другие факторы, не поддающиеся формализации.

Г р у п п о в ы е р е ш е н и я. Предполагается, что проблема выбора решения стоит перед неким индивидуумом. В то же время ясно, что при решении сложных проблем один специалист не в состоянии учесть все факторы и взаимосвязи между этими факторами и оценить вероятности альтернатив. В таком случае в качестве основной предпосылки при решении проблемы включается простое правило: "ум хорошо, а два лучше".

При этом предполагается, что мнение группы специалистов надежнее, чем мнение отдельного индивидуума, т.е. что две группы одинаково компетентных специалистов с большей вероятностью дадут аналогичные ответы на ряд вопросов, чем два индивидуума. Предполагается также, что коллективная ответственность позволяет специалистам принимать более рискованные решения и что интервал оценок, полученных от группы специалистов, включает в себя истинную оценку.

При использовании мнений группы специалистов предполагается, что взаимодействие между ними позволит компенсировать смещение оценок отдельных членов группы за счет участия эрудированных специалистов, хорошо осведомленных во многих областях знаний. Это дает возможность разностороннего анализа количественных и качественных сторон решаемой проблемы. Преимущество групповой оценки заключается в уменьшении различий во мнениях, в получении в какой-то степени более представительного мнения, являющегося результатом так называемого стиля мышления (например, преобладания у индивидуума образных или словесно-логических компонентов мышления). В силу этого человек, обладая определенным стилем мышления, решает задачи одного типа лучше, а другого — хуже. Сочетание индивидуумов, обладающих разными стилями, позволяет повысить надежность решения.

Ясно, что сумма информации, имеющейся в распоряжении группы будет больше или по крайней мере не меньше, чем информация, полученная от любого члена этой группы. Так, сумма факторов, которые имеют отношение к данной проблеме и рассматриваются группой специалистов, как правило, больше суммы факторов, которые способен учесть отдельный исследователь.

Однако при всех достоинствах методы групповых оценок (решений) не свободны от недостатков. Очевидно, например, что мнения специалистов даже в одной узкой области могут расходиться, а следовательно, не исключено, что на один и тот же вопрос они могут дать различный ответ. Более

того, и однозначность ответов не является гарантией их обоснованности. Так, предположим, что несколько специалистов дали одинаковые оценки какому-то фактору. Тогда группа становится неразличимой с точки зрения данной оценки, поскольку пока не существует достаточно объективного способа для сравнения компетентности членов группы. В этом случае логично предположить, что взятое наудачу мнение любого из членов группы будет также достоверно, как и групповая оценка, поскольку никто не может сказать заранее, каков истинный результат. Следовательно, даже единодушие членов группы не является критерием достоверности: во всяком случае, нет способа проверить это, не проведя опыта.

К недостаткам методов групповых решений можно отнести тот факт, что количество неверной (шумовой) информации у группы исследователей будет большим или по крайней мере таким же, как и у одного из них. Кроме того, существует ряд причин психологического характера. Например, проблема авторитетов, когда имеет место ссыла на человека, — прием убеждения, рассчитанный на чувства публики и не опирающийся на объективные данные. Забегая несколько вперед, отметим здесь, что в процессе сбора фактического материала по проблеме происхождения нефти и газа такой аргумент, к сожалению, весьма распространен среди исследователей данной проблемы. Однако "там, где начинается догма и культ, кончается наука, которая не может двигаться вперед без критического пересмотра упомянутых высказываний, а культу вообще враждебна" [42]. Логика запрещает такой аргумент и требует, чтобы то или иное высказывание рассматривалось само по себе, независимо от того, кем оно сделано — "авторитетным" или "неавторитетным" ученым. Учет этого требования весьма важен.

Группа может оказывать влияние на отдельных ее членов, заставляя их соглашаться с мнением большинства, хотя полной уверенности в правильности этого мнения нет; то же самое может происходить в группе, где нет руководителя, тогда имеется возможность доминирования наиболее настойчивого индивидуума, игнорирующего аргументы остальных членов группы. Такая ситуация, искажающая действительное положение вещей, также должна быть учтена при подготовке экспертизы.

Несмотря на перечисленные недостатки, групповая оценка более правильна, чем индивидуальная, если соблюдать определенные требования. Это доказано экспериментально [51]. Наиболее важные требования, которые необходимо соблюдать при использовании методов групповой оценки, базируются на двух основных идеях [19].

Во-первых, подбор специалистов-экспертов осуществляется таким образом, чтобы максимально был выявлен "разброс мнений". Иными словами, необходимо "гладкое распределение оценок", полученных от экспертов, указывающее на независимость их мнений. В случае многомодального распределения должна быть установлена причина, по которой разные авторы по-разному интерпретируют одни и те же факты.

Вторая идея состоит в соблюдении принципа групповой надежности. Это означает, что две групповые оценки по конкретной проблеме, данные двумя одинаковыми подгруппами, выбранными случайным образом, должны быть близкими. Корреляция по ряду таких оценок должна быть достаточно высокой.

Подводя итог сказанному о методе экспертного оценивания, подчеркнем основные его недостатки.

Самый существенный из них заключается в том, что при сборе информации не предусмотрен элемент случайности, необходимый для дальнейшей статистической обработки полученных данных: состав специалистов разных направлений обычно формируется в каких-то заранее заданных пропорциях, не всегда достаточно обоснованных. Не менее существенным также то, что собираемая от экспертов информация носит одномоментный характер, при котором не учитывается история вопроса, что исключает возможность рассмотрения так называемых забытых гипотез. Кроме того, комплектуя группу специалистов, можно допустить ошибку, завысив значение сложившейся репутации какого-либо лица; чтобы выяснить степень эрудиции, компетентности и другие характеристики специалиста, нужно провести предварительное обследование экспертов с последующими специальными расчетами, что неизбежно влечет за собой затраты дополнительных средств, да и исчисление количественной оценки качественных сторон эксперта пока представляется ненадежным и, следовательно, при формировании экспертной группы большая роль принадлежит субъективным оценкам; и наконец, приемы сбора информации требуют непосредственного контакта со специалистами, что создает затруднения из-за ограниченных технических и административных возможностей.

Методы, частично устраняющие эти недостатки, изложены в третьей и четвертой главах настоящей работы.

Практическое использование перечисленных прогрессивных методов и их место в модели решения трудно формализуемых геологических задач рассматриваются на конкретных геологических примерах при поиске объективных высказываний о геологических явлениях.

Глава 2

ПОСТАНОВКА ЗАДАЧИ

Постановка задачи предполагает формулирование цели, определение объекта исследования и выбор средств достижения цели.

Цель исследования уже сформулирована во введении. Повторим здесь что она обусловлена необходимостью создания аппарата, способного объективно оценивать высказывания исследователей о конкретном предмете изучения. Возникла эта цель из-за существования множества "равноправных" гипотез о природе одного и того же геологического явления, оценит достоверность которого на интуитивной основе не представляется возможным.

Как выяснилось при обследовании, генетические и прогнозные задачи находятся в условиях большой неопределенности. Поэтому логично использовать методы и приемы, с помощью которых решаются задачи в таких условиях. Итак, средства изучения (или средства достижения цели) — это использование опыта (в широком смысле этого слова) предыдущих исследователей конкретной проблемы с применением комплекса математиче

ских и специальных логических методов, т.е. методов, наиболее рациональных для решения задач в ситуации неопределенности.

Решая поставленную задачу, определяем в качестве объекта исследования всю совокупность признаков, которой оперируют ученые при описании проблемы.

1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОБЪЕКТА ИССЛЕДОВАНИЯ

Процесс решения всегда складывается из локальных задач. Естественно, что объект наблюдения в каждой локальной задаче зависит от ее специфики и четкой формулировки некоторого комплекса условий, при которых производится решение. В одних задачах (стадиях решения) объектом изучения может служить высказывание, состоящее из группы признаков. В других – набор таких объектов, в третьих – только какой-то определенный признак, имеющийся у каждого объекта (высказывания). Тем не менее математическая модель объекта в любом случае остается одинаковой.

Изучаемую совокупность признаков (объектов) можно представить как некоторое множество T точек t , т.е. $t \in T$, причем T может быть как непрерывным, так и дискретным [65, 66]. В нашей модели можно ограничиться дискретным множеством, поскольку геолог, как правило, при решении задач располагает конечным числом наблюдений, которое хотя и может быть весьма большим, но не является бесконечным. Примером дискретности может служить конечный набор объектов, моделью которого будет просто натуральный ряд чисел $1, 2, \dots, t, \dots, N$, соответствующий номерам, образующим совокупность объектов (признаков).

Мы рассмотрели модель объекта исследования без учета каких-либо его свойств. Обозначим через J множество свойств j , которое можно измерить на каждом геологическом объекте. Естественно, что множество дискретно и $j = 1, 2, \dots, m$. Тогда любому геологическому объекту с номером $t \in T$ можно поставить в соответствие некоторое множество $J_k \subset J$, состоящее из $k \leq m$ элементов и являющееся подмножеством J .

Обозначим через x_{tj} результат наблюдения признака с номером j на объекте с номером t , а через $x_t(J_k)$ – набор наблюдений k признаков. Тогда

$$x_t(J_k) = \{x_{t1}, x_{t2}, \dots, x_{tj}, \dots, x_{tk} | J\} \quad (1)$$

будет представлять собой k -мерный вектор-строку или вектор-столбец, являющийся моделью геологического наблюдения при условии, что наблюдения проводятся над заданным множеством признаков J_k . Если набор J_k задан, то необязательно использовать его в записи, и формула (1) примет вид

$$x_t = \{x_{t1}, x_{t2}, \dots, x_{tj}, \dots, x_{tk}\}.$$

Необходимо отметить, что модель не зависит от того, с количественными, полуколичественными или качественными данными мы имеем дело [65, 66].

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПРИНЯТИЯ РЕШЕНИЯ

Термин "принятие решения" возникает в задаче при необходимости действовать в ситуации, известной не полностью (или иначе, в условиях неопределенности), когда по нескольким параметрам эффективности необходимо отыскать лучший в каком-то смысле вариант. Модель принятия решений имеет форму уравнений или неравенств, которые, хотя могут быть и весьма сложными, но с математической точки зрения отличаются очень простой структурой

$$W = f(a_1, a_2, \dots; x_1, x_2, \dots), \quad (2)$$

где W — "целевая функция", или значения критерия качества принимаемого решения; a_i — заданные (заранее известные) факторы, на которые мы влиять не можем; x_i — зависящие от нас факторы (управляемые переменные), значения которых можно в известных пределах выбирать по своему усмотрению.

Модель изучаемого процесса (ситуации) включает не только уравнение (2), но также одно или несколько ограничений на управляемые переменные (параметры). Ограничения записываются в виде уравнений или неравенств

$$a_i \leq \varphi_i(x_1, x_2, \dots) \leq b_i, \quad i = 1, 2, \dots \quad (3)$$

и означают, по существу, что для достижения поставленной цели мы располагаем ограниченными ресурсами (средствами).

Теперь задачу принятия решения можно математически сформулировать так: при заданных функциях f и φ_i и известных значениях параметров a_1, a_2, \dots найти такие значения управляемых переменных x_1, x_2, \dots , которые удовлетворяют ограничениям (3) и обращают целевую функцию (2) в максимум или минимум в зависимости от смысла задачи [73].

3. ПРОБЛЕМА ГЕНЕЗИСА НЕФТИ И ГАЗА

Следующие этапы предлагаемой модели поиска (выявления) объективных высказываний (см. рис. 1) рассмотрим на примере генезиса нефти и газа. Изложим технологическую цепочку, по которой обрабатывались многочисленные литературные источники, посвященные данной проблеме. Все неудачи и тупиковые ситуации, возникавшие в процессе решения здесь не приводятся. Зато приемы, процедуры и математические методы использованные при исследовании данной проблемы, излагаются подробно как может показаться, даже излишне подробно. Это сделано намеренно. Во-первых, для того, чтобы читатель мог получить возможность проверить каких-либо "сомнительных" мест. Во-вторых, поскольку проблема генезиса нефти и газа не является, как нам кажется, оригинальной среди многих нерешенных проблем в геологии, то подробное описание процесса может служить методическим руководством при поиске объективного решения подобных проблем.

В настоящее время проблема происхождения нефтяных углеводородов представляется крайне запутанной. До сих пор не найдено ее уверенного решения. Об этом может свидетельствовать такой факт, что в литературе только относительно характера исходного материала для нефти и газа в

сказывается множество предположений. На рис. 2 изображена схема взаимодействия элементов, из которых, как предполагают, образуются те или иные нефтяные углеводороды. О количестве всевозможных сочетаний элементов можно судить по числу пунктирных линий на схеме.

Можно показать также, что если проанализировать многие другие признаки, которыми оперируют исследователи для доказательства своей точки зрения, то ответы получатся такие же многочисленные. Например, анализ высказываний о признаке "оптическая активность" [59], который используется в качестве аргумента в пользу той или иной гипотезы, показал, что оптическая активность может быть свойством, приобретенным нефтью извне (рис. 3). При этом, если допустить, что оптическая активность нефти может быть вызвана каждым из перечисленных факторов либо любым их сочетанием, то, подсчитав число возможных сочетаний из 5, получим 31 возможный случай. Если же учесть, что почти каждый фактор допускает деление на подфакторы, то число сочетаний многократно увеличивается. Отсюда ясно, что сегодня не представляется возможным однозначно решить вопрос о влиянии того или иного фактора на природу оптической активности нефти, используя лишь эрудицию и интуицию исследователей.

Число примеров, в которых сторонники различных точек зрения по-разному понимают одни и те же факты наблюдения в природе, можно продолжить. В каждом из них на первый взгляд высказывания достаточно хорошо аргументированы. Такое положение породило ситуацию, при которой существует множество "равноправных" гипотез происхождения нефтяных углеводородов.

Между тем убедительное и правильное решение этого вопроса имеет большое не только теоретическое, но и практическое значение. Представления об условиях образования углеводородов во многих случаях являются основным принципом оценки перспектив нефтегазоносности территории, диктуют направление поисково-разведочных работ и тем самым влияют на распределение бюджета страны.

Цель, средства и объект исследования. Рассмотренные ранее положения позволяют сформулировать цель исследования как необходимость выявить на объективной основе наиболее обоснованную для сегодняшнего уровня знаний гипотезу происхождения нефтяных углеводородов.

Анализ прошлых исследований данной проблемы показал, что ситуация, характеризующаяся наличием нескольких равноправно существующих гипотез (более "смелые" авторы называют их теориями), возникла лишь потому, что изучение и обобщение фактического материала, а также аргументация выводов зиждется в основном на качественной основе. Здесь имеется в виду, что еще мало используются либо не используются вовсе объективные, количественные методы при исследованиях геологических явлений, происходящих при образовании нефти и газа и при формировании их скоплений, правда, если не считать задач поисков и разведки месторождений.

Сказанное о ситуации неопределенности в полной мере отражает условия, в которых находится проблема происхождения нефтяных углеводородов. Действительно, информация, используемая для описания гипотезы, не может быть полностью формализована, поскольку наряду с количественными показателями она содержит массу качественных показателей. Коли-

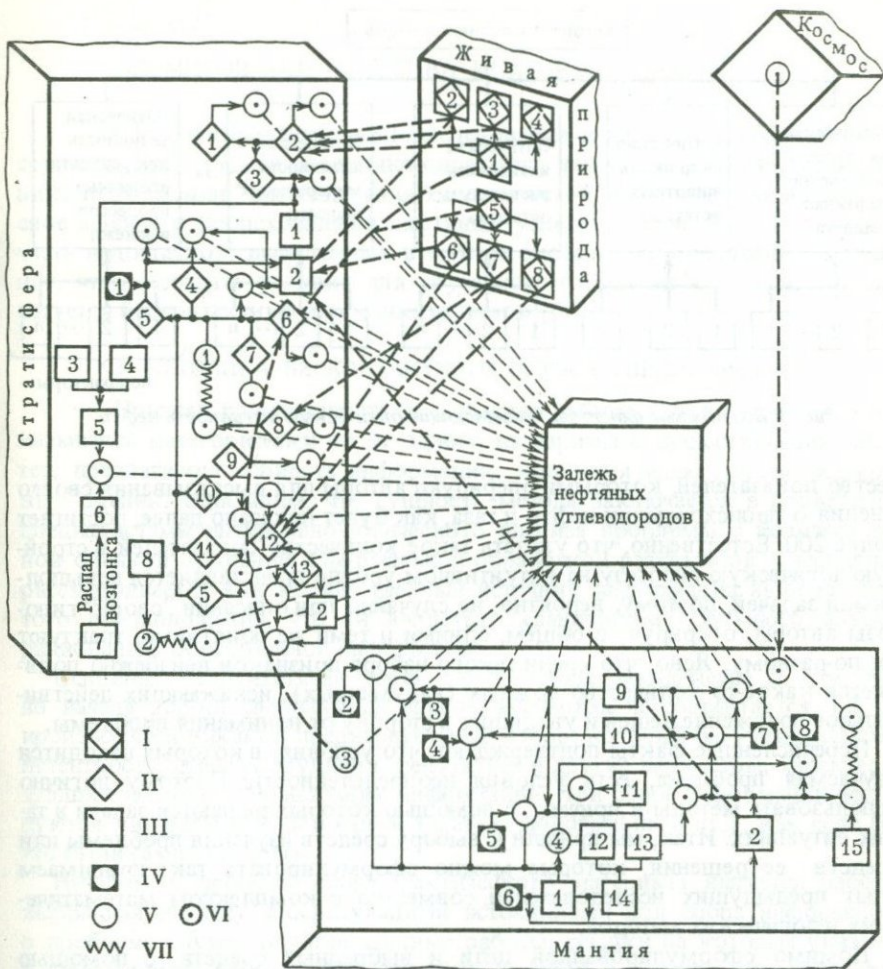


Рис. 2. Возможные источники образования нефтяных углеводородов

I – животные и растительные организмы: 1 – флора; 2 – флора (низшие); 3 – фитопланктон; 4 – флора (высшие); 5 – фауна; 6 – фауна (низшие); 7 – зоопланктон; 8 – фауна (высшие); **II** – органическое вещество: 1 – гумусово-сапротелевый тип; 2 – гумусовый тип; 3 – гуминовая кислота; 4 – уголь; 5 – олеиновая кислота; 6 – органическое вещество вообще; 7 – сланцы; 8 – раковины зоопланктона вообще; 9 – раковины фораминифер; 10 – липидный материал; 11 – стеариновая кислота; 12 – сапротелево-гумусовый тип; 13 – сапротелевый тип; **III** – неорганическое вещество: 1 – вода; 2 – водород; 3 – железо металлическое; 4 – силикаты FeO; 5 – ионы водорода; 6 – углерод; 7 – карбонаты; 8 – элементы, составляющие песчано-глинистые породы; 9 – двуокись углерода; 10 – окись углерода; 11 – соляная кислота; 12 – карбиды щелочных и щелочноземельных металлов; 13 – карбиды металлов (в основном железа); 14 – карбиды вообще; 15 – графит; **IV** – катализатор: 1 – силикаты алюминия; 2 – закись железа; 3 – магний; 4 – щелочные металлы; 5 – хлористый магний; 6 – радиоактивные элементы; 7 – песчаник; 8 – никель, железо, кобальт; **V** – углеводороды нефтяного ряда: 1 – радикалы (CH , CH_2 , CH_3); 2 – углеводородный газ вообще; 3 – метан; 4 – этилен; **VI** – углеводороды различного состава; **VII** – полимеризация

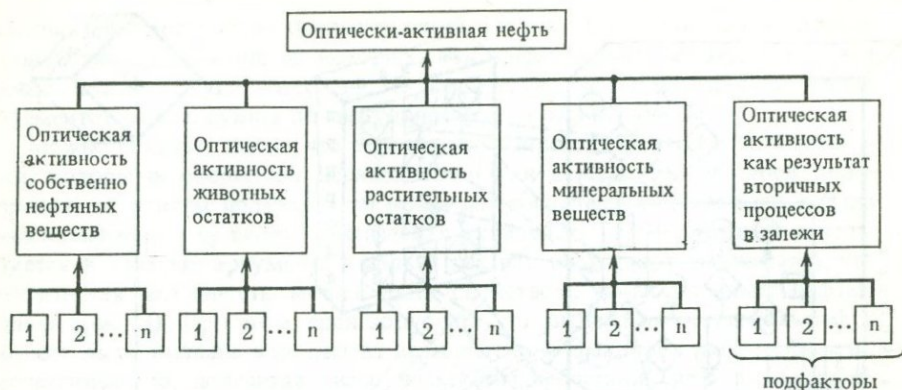


Рис. 3. Возможные факторы, вызывающие оптическую активность нефти

чество показателей, которыми оперируют авторы для высказывания своего мнения о происхождении нефти и газа, как будет показано далее, достигает более 200. Естественно, что увязать такое количество показателей в стройную логическую гипотезу на интуитивном уровне представляется невыполнимой задачей, поэтому, вероятно, не случайно при описании "своей" гипотезы авторы, оперируя, в общем, одними и теми же понятиями, трактуют их по-разному. Ясно, что среди такого набора признаков неизбежно появляется какое-то количество ложных (надуманных), искажающих действительное положение вещей и уводящих в сторону от понимания проблемы.

Перечисленные факты подтверждают, что условия, в которых находится изучаемая проблема, есть условия неопределенности. Поэтому логично использовать методы и приемы, с помощью которых решаются задачи в таких ситуациях. Итак, мы пришли к выбору средств изучения проблемы или средств ее решения, которые можно сформулировать так: принимаем опыт предыдущих исследователей, совместно с комплексом математических и логических методов.

Помимо сформулированной цели и выбранных средств, с помощью которых станет возможным достижение этой цели, еще необходимо определить (очертить) объект исследования. В нашей задаче объектом исследования будет полная совокупность признаков, которой оперируют исследователи для описания проблемы генезиса нефти и газа. Для характеристики объекта исследования необходимо собрать фактический материал, содержащий информацию, так или иначе относящуюся к достижению поставленной цели. Этому вопросу посвящен следующий раздел.

ФОРМИРОВАНИЕ МАССИВА ИСХОДНЫХ ДАННЫХ

Сформулированная цель исследования и характер выбранных для ее достижения средств позволяют определить (назначить) в качестве источника информации "высказывания" специалистов, в которых они выражают свое мнение о происхождении нефтяных углеводородов. Чтобы устранить часть недостатков, выявленных в методе экспертного оценивания, а главное, чтобы создать условия для случайного отбора информации, предлагается бесконтактный опрос специалистов [62].

1. МЕТОДИКА БЕСКОНТАКТНОГО ОПРОСА СПЕЦИАЛИСТОВ

Признав равноценность всех фактов наблюдений в природе, не отказываясь категорически ни от одного из вариантов существующих гипотез, предлагается собирать информацию, используя идею метода экспертных оценок. Напомним, что сущность метода заключается в проведении специалистами интуитивно-логического анализа проблемы с количественной оценкой суждений и формальной обработкой результатов. Специалист рассматривается как количественный источник информации на основании того предположения, что он является "хранилищем" большого объема рационально обработанной информации.

Основная отличительная черта этого метода заключается в том, что до начала опроса специальный список экспертов не составляется, т.е. не производится никакой предварительный отбор, который бы учитывал компетентность, степень эрудиции, репутацию и другие параметры, характеризующие эксперта как источник информации. При сборе информации исследуются литературные источники (статьи, монографии и др.), как если бы это были сами эксперты.

Началом в процессе сбора фактического материала является работа, выбранная наудачу. Последующими источниками для сбора информации о проблеме служат опубликованные работы, ссылки на которые указаны в предыдущей работе. Так, подобно "цепной реакции", накапливается информация об изучаемом объекте.

Получаемый спектр знаний дает возможность учитывать "нетипичные" случаи, которые в не меньшей степени важны, чем "наиболее типичные", и позволяет более разнообразно представить результаты наблюдений об объекте, создавая их большую информационную ценность.

Рассмотрим конкретную процедуру, с помощью которой осуществлялось обследование и сбор фактического материала по проблеме происхождения нефти и газа.

Фактический материал фиксировался в виде ответов на вопросы, поставленные самой проблемой, заранее анкеты-вопросники опрашиваемым не сообщались. Собранные высказывания, с помощью которых аргументируется мнение о гипотезе, легли в основу вопросника, используемого на этапе формализации информации (см. четвертую главу).

Сбор фактического материала осуществлялся с использованием метода "выборочного опроса". Этот метод по сравнению со сбором всей массы информации хорошо зарекомендовал себя как менее трудоемкий и обеспе-

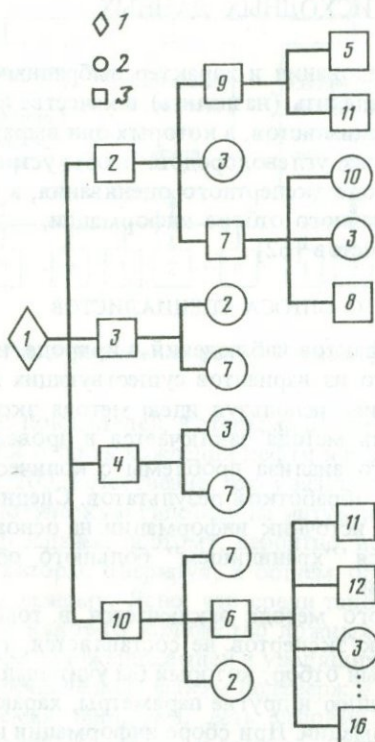


Рис. 4. Модель процесса сбора фактического материала

1 — работа, выбранная наудачу; 2 — работа, изученная на i -м шаге; 3 — работа, которую необходимо изучить на i -м шаге

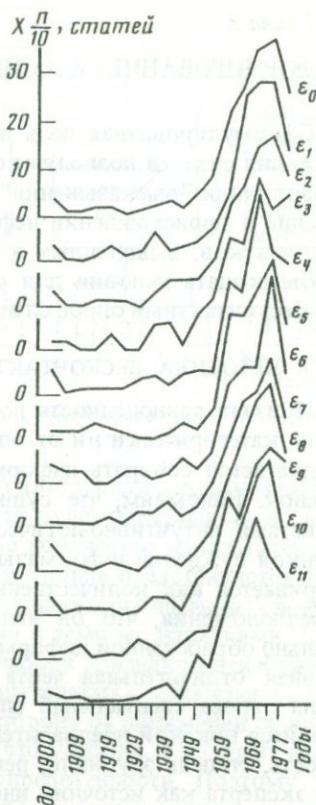


Рис. 5. Совмещение случайных выборок ϵ_i с изучаемой выборкой ϵ_0

чивающий более высокую ценность получаемых сведений. Доказано [50], что "на основе изучения выборки можно сделать более подробные и даже более качественные выводы, чем это возможно сделать на основе изучения всей большой совокупности".

При сборе информации от исследователей были использованы не только работы, непосредственно посвященные проблеме происхождения и формирования нефтяных и газовых скоплений, но также и те, в которых есть хотя бы упоминание о любом узком вопросе этой проблемы.

На рис. 4 представлена модель, наглядно поясняющая процесс сбора и накопления фактического материала. Изображение ромба с заключенной в него цифрой 1 означает работу, выбранную наудачу. С этой работы начался процесс по сбору фактического материала. От нее отходят "ветви", "листья" которых обозначают номера работ, подлежащих изучению на следующем шаге процесса. Эти работы являются наименованием ссылок, указанных в первой, исходной работе. Поскольку каждая научная работа практически всегда сопровождается списком использованной литературы, то

на каждом последующем шаге список работ, подлежащих изучению, удлиняется. "Листья" одних "веток" изображены квадратом, "листья" других — окружностью: первый случай означает, что работа встретилась в процессе сбора материалов впервые и, следовательно, ее необходимо изучить; второй случай — работа, на которую ссылаются, встречалась ранее на одном из предыдущих шагов, она зафиксирована и дальше рассматриваться не должна.

При таком подходе может сложиться ситуация, при которой на каком-то шаге все работы, которые могли бы служить дальнейшими источниками информации, уже были изучены ранее. В этой ситуации можно действовать следующим образом. Выбрать наудачу новую работу, как если бы это было начало описанного процесса, и продолжить сбор информации в соответствии с рис. 4.

Проверив полученную выборку на репрезентативность, в случае удовлетворительного результата можно завершить процесс сбора фактического материала.

Проверка гипотезы о случайном отборе фактического материала. Проверим, действительно ли удалось получить случайный набор работ, так необходимый для статистической обработки исходных данных. Для этого надо найти ситуацию, при которой известны результаты или по всей совокупности, или по выборке, полученной случайным образом, и сопоставить результаты отбора [37].

Случайно отобранные выборки, обозначим их как ϵ_i , можно получить с помощью таблицы равномерно распределенных случайных чисел, отобрав их из изучаемой совокупности (фактически собранного массива работ), которую обозначим как ϵ_0 .

Далее формулируем задачу следующим образом: если выборка ϵ_0 сходна по крайней мере с большинством из случайно отобранных выборок ϵ_i , то выборка ϵ_0 также отобрана случайным образом. Тогда предположение о представительности изучаемой совокупности будет справедливо.

Всего отобрано $m = 11$ выборок ϵ_i , каждая из которых составила 10% от исходного массива работ. На рис. 5 представлены графики, построенные по данным выборок ϵ_i , которые сопоставлены с графиком исходной выборки ϵ_0 — распределения работ по годам. Для облегчения последующих вычислительных процедур значения по оси абсцисс сгруппированы по пятилетиям, а для того чтобы соблюсти пропорцию с остальными графиками, в последнем значения по оси ординат уменьшены в 10 раз.

Сходство графиков обнаруживается даже при визуальном сравнении. Однако предположение о сходстве распределений можно проверить с помощью статистических критериев, дающих количественную оценку степени согласия.

Нулевая гипотеза формулируется так: если ряды ранжированы, то ряд значений ϵ_0 сходен с рядом значений ϵ_i , иначе говоря, расхождения между рядами несущественны. Можно было бы, используя методы парного сравнения, найти ранговую корреляцию между оценками каждой пары. Однако в нашем случае, когда имеется большой набор выборок, которые необходимо сравнить не только с исходной выборкой, но и каждую друг с другом, расчеты становятся чрезвычайно громоздкими.

В этой ситуации удобнее использовать коэффициент конкордации, позво-

ляющий дать оценку корреляции для всей группы. Вычисляется коэффициент конкордации по формуле [14, 30]

$$W = \frac{12S}{m^2(n^3 - n) - m \sum_{i=1}^m T_i}, \quad (4)$$

где

$$T_i = \sum_{k=1}^{H_i} (h_k^3 - h_k). \quad (5)$$

В формуле (4) S — сумма квадратов разностей, рассчитываемая по формуле

$$S = \sum_{i=1}^n \left(\sum_{j=1}^m x_{ij} - \bar{x} \right)^2,$$

где m — число сравниваемых выборок; n — число признаков (рангов); x_{ij} — ранг; \bar{x} — оценка математического ожидания, вычисляемая по формуле

$$\bar{x} = m(n+1)/2. \quad (6)$$

В формуле (5) T_i — показатель связанных рангов (одинаковых значений) в i -й ранжировке; H_i — число групп равных рангов в i -й ранжировке; h_k — число равных рангов в k -й группе связанных рангов при ранжировке в i -м ряду. При отсутствии связанных рангов $T_i = 0$.

Коэффициент конкордации равен единице, если все ранжировки i -го ряда одинаковы, и равен нулю, если совершенно нет совпадений.

Для оценки значимости коэффициента конкордации при $n > 7$ используют распределение χ^2 при числе степеней свободы $\nu = n - 1$. Этот критерий проверяет гипотезу о неравномерном распределении против альтернативы о равномерном.

Оценка рассчитывается по формуле

$$\hat{\chi}^2 = \frac{12S}{mn(n+1) - \frac{1}{n-1} \sum_{i=1}^m T_i}.$$

Если вычисленное значение $\hat{\chi}^2$ превышает табличное χ^2 , гипотеза о согласии принимается.

Для расчета значения коэффициента конкордации ранжируем частоты встречаемости количества работ, сгруппированных по пятилетиям. Результаты такого ранжирования представлены в виде матрицы, показанной в табл. 1. Затем найдем сумму оценок (рангов) по столбцам и квадрат отклонений от среднего значения для суммарных рангов ряда. Среднее значение рассчитаем по формуле (6). Для того чтобы учесть наличие связанных рангов, вычислим для каждого ранжированного ряда значения T_i . Теперь рассчитаем коэффициент конкордации между выборками ϵ_0 и ϵ_i , где $i = 1, 2, \dots, 11$.

Расчеты согласия между выборками показали, что $W = 0,86$, а $\hat{\chi}^2 =$

Таблица 1. Результаты ранжирования частот встречаемости количества работ по годам

Номер выборки	Ранг							
	1	2	3	4	5	6	7	8
0	4	1	2	3	5	6	7	11
1	4	2	1	3	5	6	8,5	10,5
2	3	1	2	4	5	6	7	8,5
3	2	3	1	5	4	6	7	13
4	3	2	1	5	4	7	7	9,5
5	5	1	2	3,5	3,5	6	11	7,5
6	4	1	2,5	2,5	5	6	10	7
7	4	1	2	3	5	6	8	11,5
8	4	1	2	3	5	6	8	15
9	2	1	3	5	4	6	7	13
10	4	1	2	3	5	6,5	6,5	10
11	4	2	1	3	5	7	6	12,5

Номер выборки	Ранг								
	9	10	11	12	13	14	15	16	17
0	13	10	14	15	12	13	16,5	16,5	8
1	12,5	7	12,5	14,5	10,5	14,5	16,5	16,5	8,5
2	8,5	10,5	16	13	13	13	16	16	10,5
3	8	9	15,5	15,5	10	15,5	15,5	11,5	11,5
4	12	9,5	11	14,5	14,5	14,5	14,5	17	7
5	14,5	16,5	16,5	14,5	11	7,5	11	11	11
6	13	15	13	13	16,5	16,5	10	10	8
7	7	9,5	9,5	15	11,5	15	15	15	15
8	9,5	9,5	16,5	15	11,5	16,5	15	11,5	7
9	9	9	13	16,5	13	13	13	16,5	9
10	16,5	10	8	14	14	16,5	10	12	14
11	8	9	12,5	16	16	16	12,5	12,5	10

= 218,13. Это значительно превышает допустимое значение $\chi^2 = 26,3$ для $\nu = 16$ и 5%-ного уровня значимости, поэтому делаем вывод, что гипотеза о согласии распределения работ в выборках принимается.

Таким образом, предположение о том, что работы, собранные для изучения проблемы происхождения нефти и газа, представляют собой выборку, отобранную случайным образом, не противоречит статистическим данным. Следовательно, отобранный фактический материал можно считать репрезентативным.

Для большей уверенности в полученном выводе рассмотрим этот же вопрос в другом аспекте. На рис. 6 изображено количественное распределение работ по годам. Судя по характеру распределения, может возникнуть предположение, что при отборе информации отдавалось предпочтение более современным работам и, следовательно, материал отобран тенденциозно.

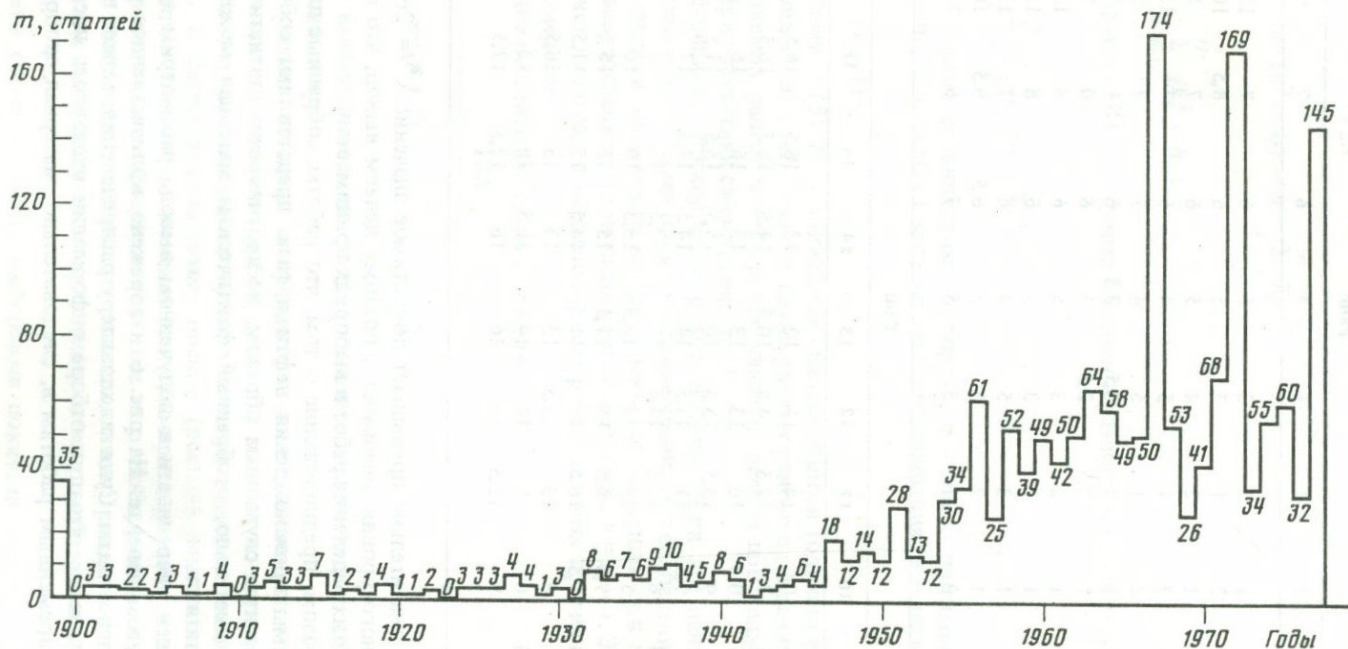


Рис. 6. График распределения количества работ по годам

Покажем, что это предположение не соответствует действительности и что благодаря выборочному методу, примененному при сборе фактического материала, нам удалось обеспечить равномерный отбор работ, написанных в разные годы и разными авторами на протяжении всей истории изучения данной проблемы.

Рассмотрим кривую полигона распределения случайной величины $\epsilon = (x_i)$, где x_i — количество собранных работ, принадлежащих i -му году (рис. 7). Сформулируем нуль-гипотезу H_0 , заключающуюся в том, что в основе выборки лежит равномерное распределение величины $\epsilon = (x_i)$, против альтернативной гипотезы H_A , заключающейся в том, что в основе выборки лежит любое другое распределение.

Пригодной для статистической проверки выдвинутой гипотезы является величина χ^2 -критерия, рассчитываемая по формуле

$$\chi^2 = \sum_{i=1}^k \frac{(x_i - \bar{x})^2}{\bar{x}}, \quad (7)$$

где x_i — наблюдаемая частота; \bar{x} — ожидаемая частота.

Суждение о рассчитанном значении χ^2 осуществляется на основании таблиц. Нуль-гипотеза отклоняется, если рассчитанное χ^2 равно или больше табличного значения χ^2 для заданного числа степеней свободы и принятого уровня значимости.

Проверка предположения о равномерном распределении количества собранных работ по годам с помощью описанного критерия показала, что такое предположение не подтверждается, т.е. нуль-гипотеза опровергнута.

В то же время из рис. 6 видно, что увеличение количества работ с годами носит ступенчатый характер. Логично интерпретировать такой факт, как согласованность с общей закономерностью, заключающейся в высокой скорости приращения информации со временем. Хотя тренд и выявляется на глаз, во избежание ошибочного вывода проверим его существование с помощью статистических критериев. Поскольку есть предположение, что изучаемое распределение является смешанным, то проверку тренда будем осуществлять с помощью критерия, не зависящего от вида распределения.

Так, фазочастотный критерий Виллиса и Мура [30] позволяет отличить отклонения ряда измерений $x_1, x_2, \dots, x_i, \dots, x_n$ ($n > 12$) от чисто случайной последовательности. Индексы 1, 2, ..., i , ..., n указывают на временную последовательность. Если рассматриваемая последовательность имеет случайный характер, то знаки разностей $(x_{i+1} - x_i)$ образуют случайную картину (нуль-гипотеза). Альтернативная гипотеза: последовательность плюсов и минусов значимо отличается от случайной. Последовательность одинаковых знаков назвали фазой. Критерий основан на частоте плюсовых и минусовых фаз. Если обозначить общее число фаз через h , причем начальная и конечная фазы опускаются, то в предположении случайности ряда статистика

$$\hat{z} = \frac{|h - (2n - 7)/3| - 0,5}{\sqrt{(16n - 29)/90}}$$

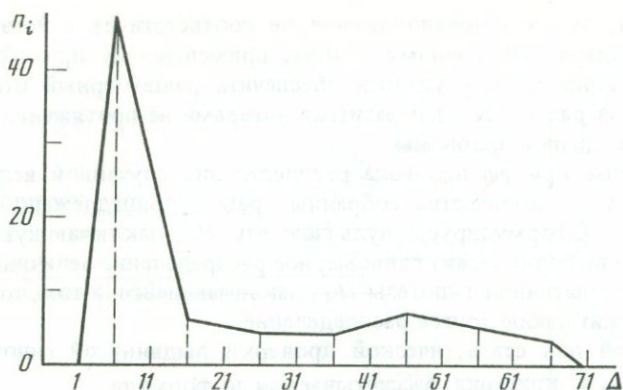


Рис. 7. Полигон распределения количества работ в i -м году

распределена нормально. При этом если $n > 30$, то поправка на непрерывность (0,5 в числителе) может быть опущена.

Нулевая гипотеза об отсутствии тренда принимается как подтвердившаяся, если вычисленная величина \hat{z} не превысит допустимой z , критические значения которой в зависимости от критерия (односторонний или двусторонний) соответственно равны: $z = 1,64$ и $z = 1,96$ для 5%-ного уровня значимости; $z = 2,33$ и $z = 2,58$ для 1%-ного уровня значимости.

Расчеты по приведенному критерию дали следующие результаты (см. исходные данные на рис. 6).

Случай 1 для двух вариантов:

1.А. Расчет произведен в условиях, при которых наблюдения ряда сгруппированы таким образом, что исследуемая величина представляет собой сумму работ за 5 лет. Нуль-гипотеза отклоняется, т.е. тренд установлен, поскольку $\hat{z} = 2,44 > z_{0,05} = 1,64$ при одностороннем критерии.

1.Б. В расчете принимает участие несгруппированный ряд наблюдений. Нуль-гипотеза сохраняется при $\hat{z} = 0,09 < z_{0,05} = 1,64$.

Противоречивость результатов обязывает использовать другие аналогичные критерии, служащие для выявления тренда. Критерий Кокса и Стюарта, также не зависящий от вида распределения, использует статистику [30]

$$z = \frac{|s - n/6| - 0,5}{\sqrt{n/12}},$$

которая распределена нормально, если S — сумма знаков плюс или минус, $n/6$ — ожидаемое значение, $\sqrt{n/12}$ — стандартное отклонение. Сумма знаков S определяется следующим образом. Ряд, состоящий из n значений, делят на три группы так, что первая и последняя содержат одинаковое число значений $n^* = n/3$. Средняя часть при объеме выборки, равном n и неделимом на три, уменьшается на одно или два значения. Далее, каждое наблюдение первой трети ряда сравнивается с соответствующим наблюдением последней трети ряда, и фиксируется знак плюс или минус соответственно при возрастающем и убывающем тренде. При $n > 30$ поправка (0,5 в числителе) не учитывается. Если обозначить число разно-

стей n^+ , то статистика критерия в точности равна приведенной выше статистике с n^+ наблюдениями, не равными нулю.

Результаты расчетов с использованием данного критерия для случая 2 такие:

2.А (условия те же, что и для случая 1.А). $\hat{z} = 2,24 > z_{0,05} = 1,64$, нуль-гипотеза отвергается, тренд установлен.

2.Б (несгруппированный ряд наблюдений) $\hat{z} = 5,00 > z_{0,05} = 1,64$, нуль-гипотеза отвергается, тот же результат.

Для большей убедительности выводов о существовании тренда в изучаемом ряду наблюдений воспользуемся еще одним критерием — итерационным, который также служит для проверки независимости случайного распределения выборочных значений. Итерация здесь трактуется как последовательность идентичных символов, перед или за которыми следуют другие символы. Например, последовательность Р, Р, Р, Р; О, О, О; Р, Р; О, О, О, О образует $\hat{r} = 4$ итерации при $n = 13$ наблюдениям. Или последовательность + + + +, — — —, + +, — — — —, отвечающая разности $x - Me$ (отклонение от медианы), образует также $\hat{r} = 4$ итерации при $n = 13$ наблюдениям. При одностороннем критерии нуль-гипотезе H_0 (последовательность случайная) противостоят альтернативные гипотезы H_{A1} — имеется гнездовой эффект (малое значение \hat{r}) или H_{A2} — имеется регулярное изменение (большое значение \hat{r}). Для проверки гипотезы используют специальные таблицы (критические границы $r_n = r_u$ и $r_b = r_o$, например, Закс Л., табл. 99) или для больших n (т.е. n_1 или $n_2 > 20$) аппроксимацию

$$z = \frac{|n(r-1) - 2n_1n_2|}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n-1}}}, \quad n = n_1 + n_2.$$

Односторонний критерий: H_0 должна быть отвергнута против H_{A1} и H_{A2} , когда соответственно $\hat{r} \leq r_u$ и $\hat{r} \geq r_o$ или $z > z_{одн}$. Расчеты по этому критерию показывают, что в случае 3.А (условие, как и в случае 1.А) нуль-гипотеза отвергается, поскольку при $n_1 = 6$ и $n_2 = 11$ проявляется гнездовой эффект (так как $\hat{r} = 3 < r_u = 4$ при 5%-ном уровне значимости), что подтверждает ранее выявленное скачкообразное увеличение информации со временем.

В случае 3.Б (несгруппированный ряд) нуль-гипотеза отвергается при $\hat{z} = 5,92 \gg z_{одн} = 1,64$, т.е. имеется явно выраженный тренд.

Результаты проверки исследуемого ряда после удаления из него экстремальных значений показали те же результаты, т.е. во всех случаях, кроме случая 1.Б, тренд установлен. Тот факт, что на одном и том же фактическом материале в 10 случаях из 12, или в 83%, тренд проявился, а в условиях случая 1.Б, характеризующегося более частой сменой фаз, критерий Виллиса и Мура "не сработал", говорит о том, что этот критерий мало чувствителен, а следовательно, не эффективен при исследованиях на подобном материале.

Вывод совпадает с исследованиями других авторов, изучавших этот критерий [35]. Интуитивно это понятно, поскольку наличие экстремальных точек есть локальное свойство, и на нем могли не отразиться

последствия легкого тренда, который, вероятно, проявляется в нашем ряду. Если же в качестве альтернативы взять цикличность (периодичность), критерий становится много лучше [35].

Проверка на цикличность в нашей работе не принципиальна и требует вычисления автокорреляционной функции, что без ЭВМ даже при малом числе наблюдений почти невыполнимо [24]. Поэтому ограничимся результатом, который получен: в исследуемом ряду тренд существует, поскольку на одном и том же фактическом материале, изученном с помощью различных статистических критериев, он проявился в 10 случаях из 12.

Таким образом, можно считать неопровергнутой закономерность увеличения количества информации по мере приближения к современности, что логично согласуется с общей тенденцией в век научно-технического прогресса.

Такая же закономерная связь подтверждается расчетами коэффициента ранговой корреляции Спирмена:

$$r_S = 1 - \frac{6\sum D^2}{n^3 - n}. \quad (8)$$

Для вычисления коэффициента ранговой корреляции сравниваемые ряды преобразуются с помощью рангов. Вычисляются разности D для n пар рангов, которые возводятся в квадрат и суммируются. Значение $\sum D^2$ используется в приведенной выше формуле.

Равным значениям (связным рангам) соответствует среднее ранговое число. Если два ряда рангов равны, разности равны нулю и $r_S = 1$. Если ряды рангов обратны, то $r_S = -1$. Этот критерий позволяет тем самым ответить на вопрос о знаке имеющейся корреляции. Если в ряду значений часто встречаются связи (равные значения), то целесообразно применять статистику

$$r_{S, B} = 1 - \frac{6\sum D^2}{(n^3 - n) - (T_{x'} + T_{y'})},$$

$$T_{x'} = \frac{1}{2} \sum (t_{x'}^3 - t_{x'}), \quad T_{y'} = \frac{1}{2} \sum (t_{y'}^3 - t_{y'}),$$

где $t_{x'}$ — число членов в последовательных группах (с равными рангами) x' -ряда; $t_{y'}$ — число членов в последовательных группах (с равными рангами) y' -ряда (штрихи при x, y означают, что это ранги).

Значимость коэффициента r_S для пар значений при $n \geq 10$ можно установить по статистике

$$\hat{t} = |r_S| \sqrt{\frac{n-2}{1-r_S^2}},$$

распределенной по Стьюденту с $n - 2$ степенями свободы.

В нашей задаче коэффициент ранговой корреляции Спирмена использован для установления связи между количеством работ по годам и рядом натуральных чисел (годом наблюдений). Значение коэффициента составило $r_{S, B} = 0,99$ при допустимом значении $\hat{t} = 0,41$ на 5%-ном уровне значимости, т.е. положительная связь подтвердилась.

Таким образом, выявленный ступенчатый характер увеличения количества работ с годами согласуется с общей закономерностью — высокой скоростью приращения информации со временем. Аномальные значения количества работ на графике справа (см. рис. 6) соответствуют годам публикации сборников, посвященных республиканским и всесоюзным совещаниям по генезису нефти и газа.

Вернемся к рассмотрению полигона распределения случайной величины $\epsilon = (x_i)$ на рис. 7. Конфигурация его позволяет предположить, что функция распределения $F(x_i)$ складывается из нескольких видов распределения. Проверка этого предположения показала, что если весь ряд наблюдений разбить на три части и для каждого из вновь полученных эмпирических распределений подобрать теоретическую модель распределения, то окажется, что все они согласуются с равномерным распределением.

Проверка нуль-гипотезы о равномерном распределении $F(x_i)$ против любого неизвестного распределения $F(x_i)$ осуществлялась по χ^2 -критерию согласия (7). Для расчета были использованы данные рис. 6, но предварительно были исключены экстремальные значения величины ϵ , соответствующие 1967, 1972 и 1977 годам — годам публикаций выше названных сборников.

Проведенные расчеты дали следующие результаты:

1) $\epsilon^* = (x_i)$, где i — соответствует годам от 1900 до 1931, $\hat{\chi}^2 = 5,09 < \chi^2_{0,05; 3-1} = 5,99$;

2) $\epsilon^{**} = (x_i)$, где i соответствует годам от 1932 до 1946, $\hat{\chi}^2 = 4,80 < \chi^2_{0,05; 3-1} = 5,99$;

3) $\epsilon^{***} = (x_i)$, где i соответствует годам от 1947 до 1976, $\hat{\chi}^2 = 3,26 < \chi^2_{0,05; 6-1} = 11,07$.

Таким образом, нуль-гипотеза принимается, т.е. в основе всех трех выборок лежит равномерное распределение, поскольку рассчитанные χ^2 меньше табличных χ^2 значений для $n - 1$ степеней свободы и 5%-ного уровня значимости. Следовательно, предположение о тенденциозном отборе информации можно считать несостоятельным.

Остановимся еще на одном моменте, который, по нашему мнению, может косвенно свидетельствовать, что полученный набор работ отобран не предвзято.

Будем рассматривать не количество работ $\epsilon = (x_i)$, собранных по годам, а новую случайную величину r_{ij} , выражающую количественное соотношение сторонников различных точек зрения на проблему генезиса нефтяных углеводородов. Опишем алгоритм получения этой величины [25].

Пусть $r_{ij} = 1$, если в работе утверждается органическая гипотеза; $r_{ij} = 0$, если в работе утверждается неорганическая гипотеза; и наконец, $r_{ij} = 0,5$, если в работе обе гипотезы считаются равноценными. Тогда оценка математического ожидания случайной величины r_{ij} равна

$$x_{ij} = M[r_{ij}] = 1 \cdot \frac{n_i}{n} + 0,5 \cdot \frac{n_h}{n} + 0 \cdot \frac{n_j}{n}. \quad (9)$$

Общее количество работ в k -м году равно сумме

$$n_k = n_{ik} + n_{hk} + n_{jk}. \quad (10)$$

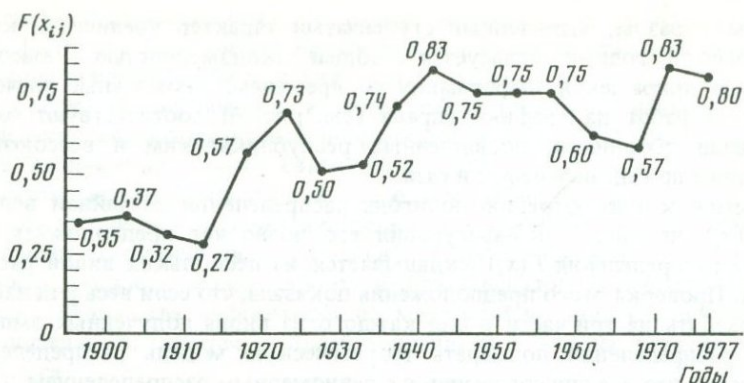


Рис. 8. Полигон распределения величины $\tau_{ij} = (x_{ij})$ — соотношение сторонников различных точек зрения

Определяя из (10) n_{hk} и подставляя его в (9), получаем

$$x_{ij} = \frac{1}{2} + \frac{n_{ik} - n_{jk}}{2n} \quad (i, j = 1, 2, \dots, m).$$

Совокупность величин x_{ij} образует ряд (рис. 8), который можно исследовать. Оказывается, что величина $r_{ij} = (x_{ij})$ в изучаемом ряду распределена равномерно, поскольку $\chi^2 = 5,20$ (рассчитан по (7)), что меньше, чем допустимое $\chi^2_{0,05; 4-1} = 7,81$.

Это значит, что при отборе информации не отдавалось предпочтения той или иной гипотезе происхождения углеводородов нефти и газа.

Итак, тот факт, что в пределах каждого периода, характеризующегося скачкообразным возрастанием информации, работы отбирались равномерно и, более того, пропорционально росту информации, свидетельствует о случайном отборе информации и опровергает предположение о какой-либо предвзятости.

Таким образом, применение предлагаемого способа бесконтактного опроса специалистов открывает возможность широкого использования методов математической статистики при обработке информации.

2. ХАРАКТЕРИСТИКА МАССИВА ИСХОДНЫХ ДАННЫХ "ЛИТЕРАТУРА"

Сведения накапливались в информационно-справочной картотеке "ручного управления", или, как ее еще называют, "малой механизации". Картотека представляет собой массив, состоящий более чем из 1800 карт с двухрядной краевой перфорацией и вмещающий 1768 источников информации. Несоответствие между количеством работ и количеством карт в картотеке объясняется тем, что некоторые конспекты не поместились на одной карте.

За основным признаком картотеки принят источник информации — литература. На внутреннем поле перфокарты записаны данные (признаки), собранные для этого источника информации, а именно:

название изучаемой работы (статьи, монографии и др.);
год и место издания;
автор (авторы);
конспект работы;

основные аргументы, с помощью которых высказывается мнение о гипотезе (теории), или ключевые слова, содержащиеся в тексте работы.

Для быстрого нахождения признаков каждый из них кодируется на перфорированных полях карты, а поиск осуществляется по ключевой карте. При разработке ключевой карты применены ключи, обычно употребляемые в перфокартотеках ручного управления и описанные в соответствующей литературе [2, 21]. Тип ключа выбирался в зависимости от типа поискового признака, его разнообразия и частоты обращения. Накапливать сведения можно и непосредственно, вводя их в ЭВМ.

Предлагаемая система сбора информации позволила в довольно короткие сроки сконцентрировать большой объем информации, содержащей сведения, не только прямо относящиеся к изучаемой проблеме, но и косвенно — данные физики, химии, биологии, астрономии, океанологии, вулканологии, геофизики, петрологии, металлогении, гидрогеологии и других наук, не касающиеся непосредственно вопросов генезиса углеводородов, но привлекаемые нефтяниками для подтверждения своей точки зрения либо опровержения противоположной.

Глава 4

ОБРАБОТКА ИНФОРМАЦИИ

1. СИСТЕМАТИЗАЦИЯ ИСХОДНЫХ ДАННЫХ

Предварительный анализ собранного фактического материала по проблеме производился с использованием принципов системного подхода. Это значит, что была осуществлена попытка учесть все без исключения факторы и соответствующие им признаки, а также их взаимосвязи, влияющие на проблему в целом, выявить наиболее существенные факторы и признаки, на которых в первую очередь следует остановиться при решении проблемы.

Использование системного анализа целесообразно, во-первых, чтобы показать действительную сложность проблемы, о которой было заявлено при постановке задачи, более подробно разобраться во всем многообразии вопросов, которые необходимо решить, и на этой основе сформулировать локальные цели. Во-вторых, применяя этот метод позволяющий выявить и наглядно представить отдельные части проблемы, можно добиться значительного упрощения формулировок как при постановке частных задач, так и при решении проблемы в целом, увязав решение этих частных задач в стройную последовательную цепочку.

Совокупность признаков, которыми оперируют исследователи, выражая свое мнение о проблеме происхождения нефти и газа, можно, пользуясь терминологией системного анализа, разбить на части, каждая из которых характеризуется набором факторов. Под термином "фактор" здесь

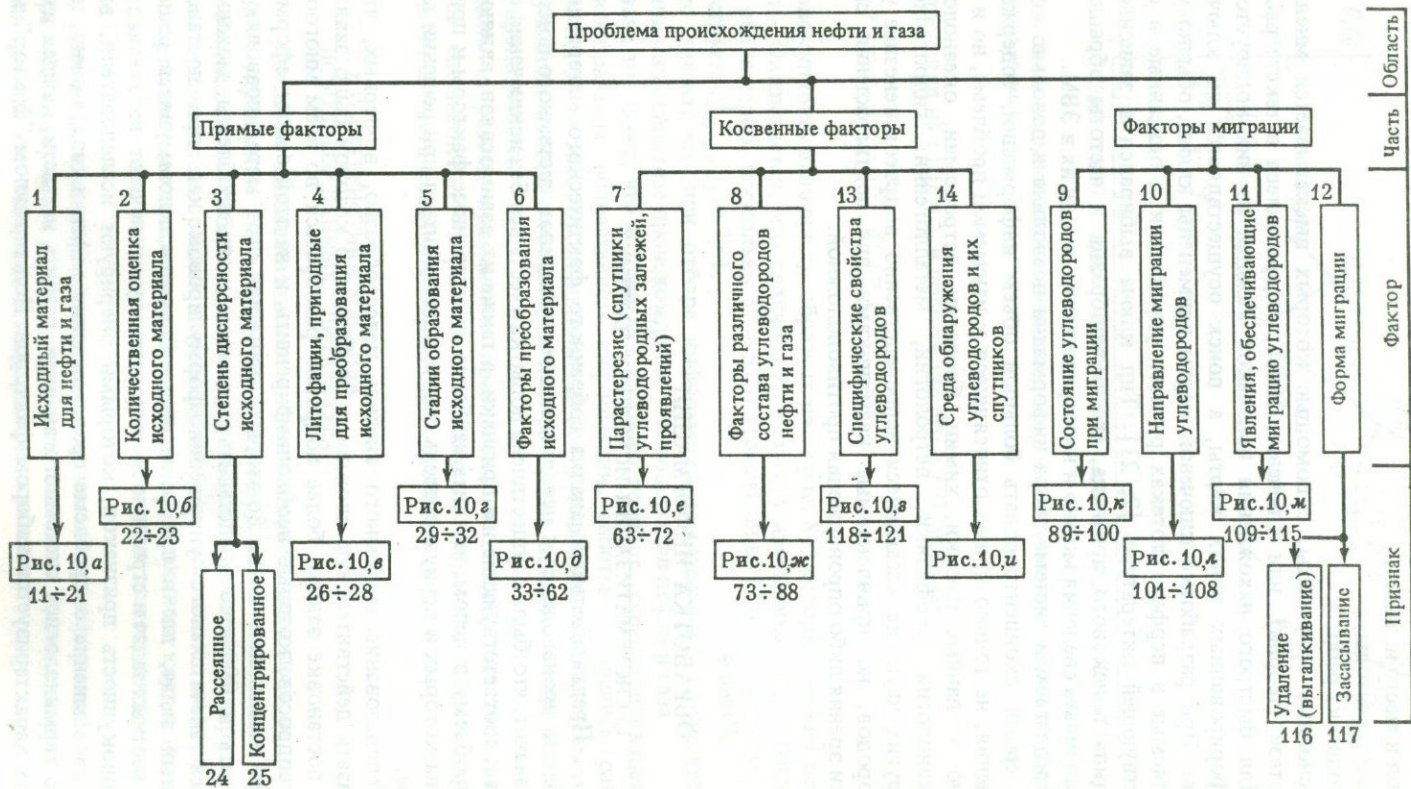


Рис. 9. Схема декомпозиции проблемы происхождения нефти и газа

используется понятие, высшее по иерархии, объединяющее совокупность признаков, сходных по существу вопроса. В тех случаях, когда необходимо деление признака на низшие иерархические разряды, употребляется термин "значение признака".

На рис. 9 представлена декомпозиционная схема проблемы происхождения нефти и газа, позволяющая увидеть, что широкий круг вопросов (факторов), относящихся к этой проблеме, разбит на три крупные части.

Первая часть объединяет группу факторов, непосредственно ответственных за образование нефти и газа.

Вторая часть объединяет вспомогательные факторы, способствующие объяснению различных точек зрения на происхождение нефти и газа.

К третьей части принадлежат факторы миграции углеводородов, ответственные за формирование их скоплений и месторождений.

Как конкретно сформулированы факторы, сколько их и к какой их части принадлежит каждый показано на рис. 9.

Для удобства описания всем факторам присвоен порядковый номер, указанный в верхней части соответствующего фактору блока. Дальнейшая декомпозиция проблемы заключалась в разбиении каждого выделенного фактора на признаки. Если декомпозиция осуществлялась с учетом принципов классификации, т.е. выдерживалась система соподчиненных понятий, то фактор сначала разделялся на группы, которые затем дробились на признаки. В некоторых случаях осуществлялось и обратное. Так если в процессе решения оказывалось, что слишком детальное дробление излишне, то признаки объединялись под одним названием фактора и тогда этому фактору присваивался ранг признака.

В нижней части блока соответствующего фактора указаны номера рисунков, на которых раскрывается содержание каждого фактора.

Так, на рис. 10, а изображено детальное деление фактора 1 — "исходный материал для образования нефти и газа". Из схемы видно, что в качестве возможных источников нефти и газа исследователи назвали более 21 элемента. Одни из выявленных элементов могут выступать самостоятельно в качестве источника углеводородов, другие — только в сочетании. Пронумеровав каждый элемент, отождествим его с признаком. Теперь при наличии подробного перечня признаков задача выявления взаимосвязи между ними существенно упрощается.

Такие же детальные схемы построены для каждого фактора (рис. 10, б-м). При их построении выдерживался один и тот же принцип во всех случаях дробление фактора доведено до конечного элемента, который пронумерован и фигурирует в дальнейшем в качестве признака. Нумерация признаков, как и некоторых факторов, не всегда соответствует порядковой нумерации. Это связано с тем, что некоторые признаки ранее были отнесены в другие группы факторов. Итерационный характер декомпозиционных построений позволил более рационально упорядочить набор признаков, не изменяя при этом их нумерацию, поскольку они с этим номером уже участвовали в предварительных вычислительных процедурах.

Декомпозиционные схемы дали возможность получить весь перечень признаков, которыми оперируют авторы, изучая, описывая и решая проблему. Кроме того, анализ этих схем уже на этапе предварительной обрабо-

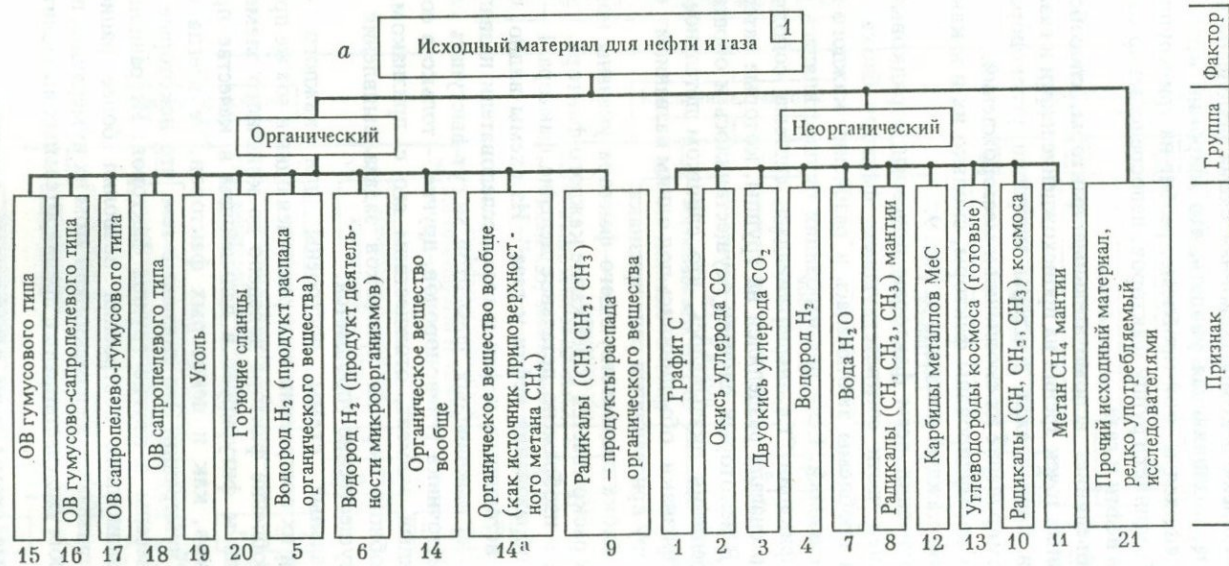
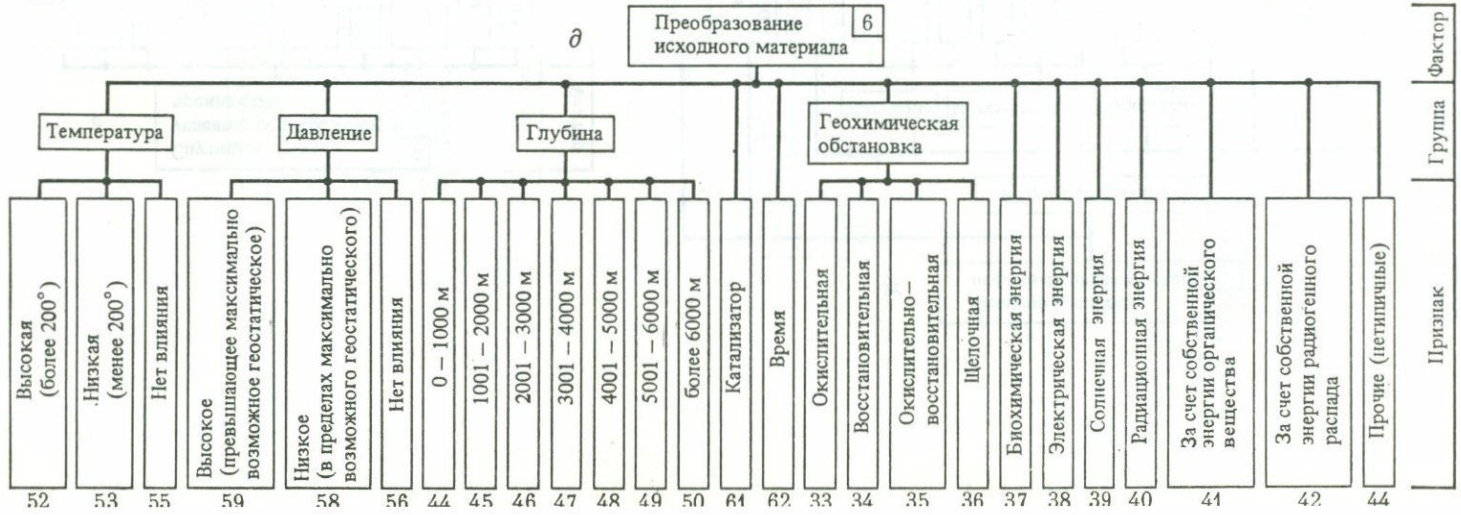
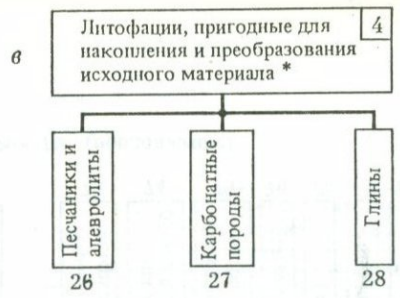


Рис. 10. Схемы декомпозиции факторов

а — исходный материал для нефти и газа; б — количественная оценка; в — литофации; г — стадии преобразования исходного материала; д — преобразование исходного материала; е — парастерические ассоциации углеводородов; ж — различный состав углеводородов; з — специфические свойства нефтяных углеводородов; и — среда обнаружения углеводородов и их спутников; к — состояние углеводородов при миграции; л — направление миграции углеводородов; м — силы и явления, обеспечивающие миграцию углеводородов; *фактор употребляется только в связи с исходным материалом для органических гипотез



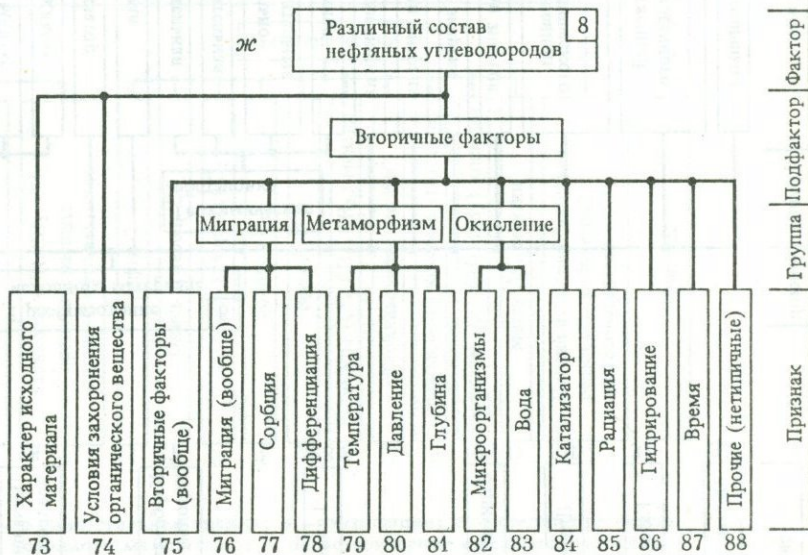
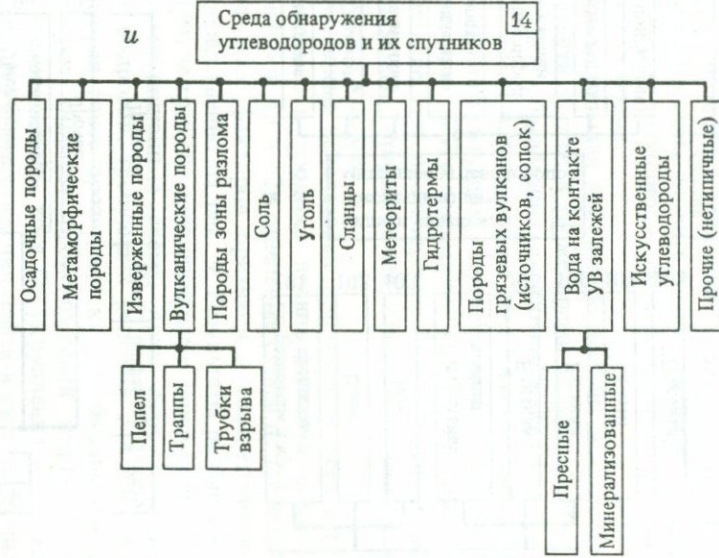


Рис. 10. (продолжение)

з



и



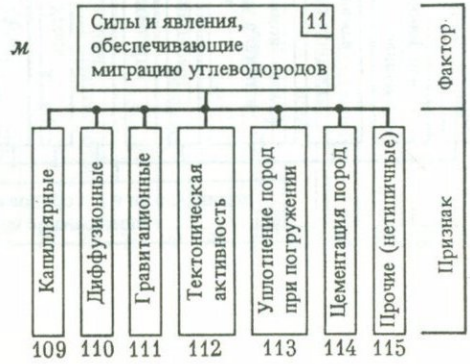
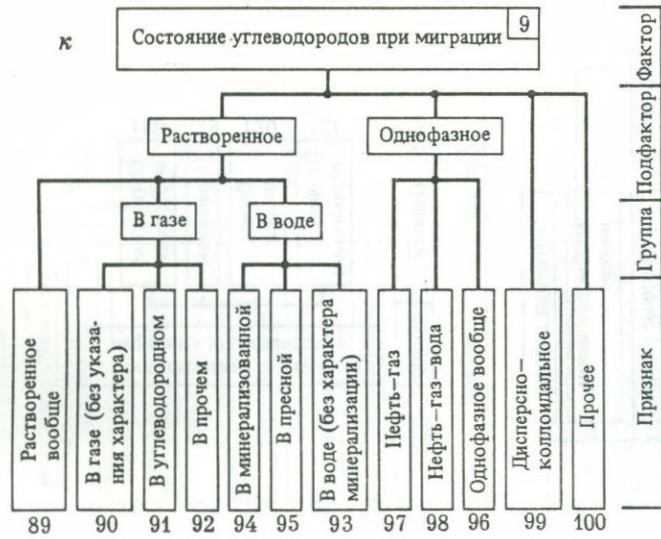


Рис. 10. (окончание)

ки дает возможность выявить (наметить) признаки, которые подлежат первоочередному изучению. Сказанное поясним на примере анализа схем на рисунке 10, б, изображающем декомпозицию фактора "количественная оценка".

Данная схема позволяет проследить взаимосвязь, по существу, двух факторов и предназначена уловить оценку авторов с точки зрения достаточности количества исходного материала для образования скопления нефтяных углеводородов.

Иногда авторы, называя какой-либо элемент в качестве источника для образования нефти и газа, в то же время оценивают его количество как недостаточное для образования скоплений каких-то определенных категорий запасов. Тогда использование такой взаимосвязанной схемы позволяет без труда выявить ситуацию, запрещающую совместное существование опровергающих друг друга признаков, и направить усилия в первую очередь на разрешение этих противоречий. Решение задачи даст возможность исключить "шумовые" признаки, что, естественно, скажется на упрощении решения проблемы в целом.

Описанный подход использован при анализе всех декомпозиционных схем, составленных в процессе предварительной обработки материала. Для двух факторов (3 и 12) отдельные схемы не приводятся. Из-за простоты деления этих факторов признаки и соответствующие им номера указаны непосредственно на рис. 9. Результаты разбиения остальных факторов, вероятно, не требуют дополнительных пояснений.

Итак, декомпозиционные схемы ярко демонстрируют сложность многообразия вопросов, требующих решения для достижения поставленной цели, которая ранее прозвучала как необходимость выявить наиболее обоснованную гипотезу происхождения нефтяных углеводородов. Если попытаться подсчитать, хотя бы укрупненно, количество существующих сегодня гипотез происхождения нефти и газа, то их число составит 24. На рис. 11 представлена классификационная схема таких гипотез, которая построена на базе высказываний, собранных от предыдущих исследователей данной проблемы.

При разработке схемы применен монотетический (иерархический) вид классификации. За основной классификационный признак при выделении класса гипотезы выбран признак, определяющий характер исходного материала. Согласно этому признаку выделены:

Класс А. Органические гипотезы — используется совокупность элементов органического вещества.

Класс Б. Неорганические гипотезы — источниками для нефтяных углеводородов служат элементы неорганического вещества.

Класс Г. Смешанные гипотезы — источниками являются органические и неорганические вещества совместно.

Класс "Органические гипотезы" можно разбить на четыре подкласса, если учитывать такой признак, как степень дисперсности органического вещества:

Подкласс А.1 — гипотезы, в которых предполагают органическое вещество в виде дисперсного материала.

Подкласс А.2 — гипотезы, в которых предполагают концентрированное органическое вещество.

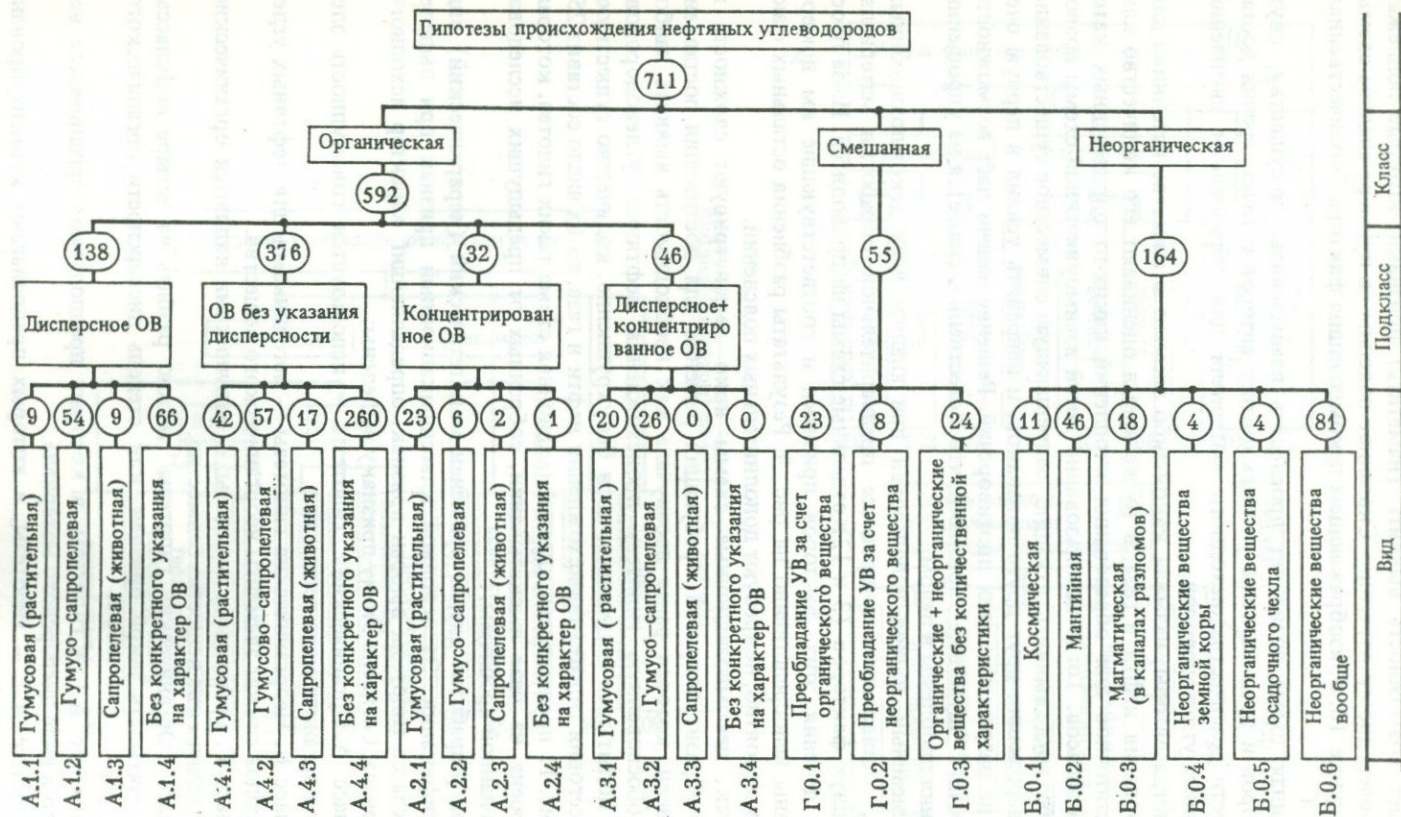


Рис. 11. Классификационная схема гипотез происхождения нефтяных углеводородов

① — количество мнений

Подкласс А.3 — гипотезы, в которых концентрированное и дисперсное органическое вещество участвуют совместно.

Подкласс А.4 — гипотезы, в которых упоминание о степени дисперсности (концентрации) исходного материала отсутствует вовсе.

При дальнейшей декомпозиции на виды каждый выделенный подкласс органических гипотез разбит по признаку, который имеет конкретное название органического вещества: будь то гумусовое, гумусово-сапроплевое, сапропелевое, либо без указания таких. Например, в подклассе А соответственно обособлены виды органических гипотез:

Вид А.1.1 — исходным материалом для образования нефти и газа служат растительные организмы.

Вид А.1.2 — растительные и животные совместно.

Вид А.1.3 — исключительно животные организмы.

Вид А.1.4 — не указан характер органического вещества, хотя и говорится, что оно должно быть рассеянным.

Класс "неорганические гипотезы" подразделяется на виды, минуя по классы, которые присутствуют в предыдущем классе гипотез, на основании признака, характеризующего место образования углеводородов. Согласно этому признаку выделены следующие виды:

Вид Б.0.1 — космическая гипотеза.

Вид Б.0.2 — мантийная гипотеза.

Вид Б.0.3 — магматическая (в каналах разломов) гипотеза.

Вид Б.0.4 — гипотеза, в которой процессы преобразования исходного материала предположительно протекают в земной коре.

Вид Б.0.5 — гипотеза, в которой процессы преобразования исходного материала предполагаются в осадочном чехле.

Вид Б.0.6 — гипотеза, в которой о месте образования углеводородов не говорится, хотя ясно, что исходный материал — неорганические вещества.

Класс "смешанные гипотезы" разбит на виды по признаку, обозначающему долю участия вещества (органического либо неорганического) в качестве исходного материала для образования нефти и газа:

Вид Г.0.1 — выделяется в случае количественного преобладания органических веществ.

Вид Г.0.2 — при преобладании неорганических веществ.

Вид Г.0.3 — когда нет прямых указаний на количественную оценку того или иного компонента, но ясно, что в наличии должны быть как органические, так и неорганические вещества.

Следует отметить, что некоторые гипотезы или их низшие разряды объединяют, по существу, несколько разнородных высказываний. Например, подкласс А.4 объединяет случаи, в которых из анализа высказываний не ясно, то ли: 1) автор не может уверенно назвать степень дисперсности исходного материала; 2) автор не придает решительно никакого значения степени дисперсности органического вещества; 3) автор считает само собой разумеющимся тот факт, что исходным материалом для нефти и газа однозначно может служить только определенной степени дисперсности органическое вещество. Но какой ...? Такая же степень неопределенности и у выделенного вида Б.0.6 — неорганических гипотез.

Необходимость выделения такого рода гипотез является ярким при-

мером недостаточной формализованности языка, используемого геологами и позволяющего неоднозначно истолковывать одни и те же факты, понятия, явления и т.п.

Классификацию гипотез можно было бы продолжить, используя для этого любые другие признаки. Однако это, вероятно, неблагодарный труд, поскольку ясно, что, используя только логические приемы и интуицию, вряд ли можно добиться рационального решения поставленной цели. Ничего не даст нового и такой анализ, как оценка той или иной гипотезы по количеству ее сторонников. Действительно, на рис. 11 возле каждой гипотезы указано число ее сторонников. Отметим, что цифры указывают не на количество исследователей, а на количество работ, в которых отмечается конкретная гипотеза. Если ориентироваться на эти цифры, то получается, что наибольшее число (наиболее часто) исследователей называют гипотезу вида А.4.4, т.е. органическую гипотезу, в которой не указывается ни степень дисперсности органического вещества, ни его характер, т.е. отдают предпочтение гипотезе, характеризующейся наибольшим элементом неопределенности.

Подводя итог, отметим, что цель, которая была отведена настоящему разделу, достигнута. Более детально показана сложность проблемы происхождения нефти и газа, выявлены основные ее разделы и намечены первоочередные этапы решения, в частности:

1) начинать обработку фактического материала следует с первой части (см. рис. 9), объединяющей только факторы, непосредственно ответственные за образование нефти и газа. Если этого окажется недостаточно для окончательного решения проблемы, то полученные результаты затем необходимо увязать с третьей частью, в которой объединены миграционные факторы, и подключить косвенные факторы;

2) перед началом этапа количественной обработки фактического материала необходимо осуществить формализацию, которая бы позволила подготовить информацию к виду, удобному для ее статистической обработки. Описанию такой работы посвящен следующий раздел.

2. ФОРМАЛИЗАЦИЯ СМЫСЛОВОЙ ИНФОРМАЦИИ

Правила оценки качества "истинности" высказывания. Прежде чем приступить к преобразованию смысловой информации в числовую, необходимо произвести ее качественный анализ, т.е. "очистить" информацию от "шумовых" помех. Дело в том, что, оценивая какой-либо признак, с помощью которого описывается объект исследования, эксперты могут высказываться с различной степенью обоснованности. Одни высказывания базируются на опыте, другие — на теоретических обоснованиях, а третьи не имеют под собой вообще каких-либо обоснований.

Для обоснования качественного критерия при оценке суждений используем работу [9], в которой предложен вариант обязательных правил, необходимых при истинной оценке естественнонаучного высказывания.

В естественных науках приняты по крайней мере два критерия истинности высказывания: 1) в высказывании содержится экспериментальное подтверждение определенной идеи; 2) в высказывании имеется убеждающее теоретическое объяснение этой идеи.

Поясним, что имеется в виду под "убеждающим теоретическим объяснением" на примерах. В начале развития нефтепромыслового дела был выдвинута теория, согласно которой утверждалось, что глинистые породы не могут являться коллектором. Теория звучала логично и убедительно и в силу этого стала общепризнанной. Многие десятилетия истинность теории не подвергалась сомнению, несмотря на то, что при проходке бурением глинистых толщ нередко отмечались нефтегазопроявления порой значительные. Только относительно недавно обнаруженные месторождения нефти и газа с глинистым коллектором показали ложность теории. В настоящее время новая теория о коллекторских свойствах глинистых толщ звучит логично и убедительно.

Еще пример. Существует теоретическое обоснование того, что жизнь способность прекращается при высоких температуре и давлении. Однако организмы, обнаруженные на глубине 9670 м (в глубокой скважине в Кольском полуострове), где температура и давление выше предельных теоретически допускаемых учеными, поколебали истинность теории*. Невого теоретического объяснения этому факту пока не найдено.

В обоих примерах приведены высказывания, истинные в силу одного критерия и ложные в силу другого. Положение, при котором высказывание одновременно является истинным и ложным, недопустимо. Это влечет за собой сохранение одного критерия суждения об истинности и отказ от другого. Естественно, сохранить экспериментальный критерий.

Сформулируем правила суждения об истинном значении высказывания. Потребуем, чтобы всякое высказывание содержало два элемента: 1) описание эксперимента (или, что то же самое, процедуры проверки высказывания); 2) предсказанный результат эксперимента.

Высказывание, содержащее оба элемента, назовем правильно построенным высказыванием. Так, утверждение о том, что мамонт — самое сильное животное, когда-либо существовавшее на Земле, не является правильно построенным. Такое высказывание не содержит и даже не подразумевает определенной процедуры проверки.

Итак, правило суждения об истинности высказывания гласит: если результаты предыдущих процедур проверки каждый раз совпадали с предсказанным результатом, то высказыванию присваивается истинное значение, а если хотя бы один раз не совпали, то ложное.

В первых двух примерах было показано, что убеждающее теоретическое утверждение как критерий истинности может вступать в противоречие с экспериментальным критерием истинности. Такое положение служит основанием для введения дополнительного (косвенного) правила суждения об истинности высказывания. Это правило диктует: никакие другие способы, кроме экспериментальной проверки, для установления истинного значения высказывания не допускаются, т.е. факты, полученные в результате эксперимента, могут приниматься без теоретического обоснования.

Кроме того, из тех же примеров вытекает еще одно косвенное правило

*Вероятно, нельзя допустить, что все находки микроорганизмов на глубинах, превьюшающих теоретические допуски, проведены в условиях, не исключающих загрязнения опыта.

ло: до экспериментальной проверки истинное значение высказывания считается неопределенным (не опровергнутым, не подтвержденным), поскольку любое предсказание, сбывшееся сегодня, может не сбыться в будущем.

Перечисленные критерии качественной оценки информации предлагается обязательно учитывать при подготовке ее к этапу формализации, в конкретном случае к переводу смысловой информации в числовую.

Метод количественной оценки качественных признаков. Рассмотрим сущность и обоснование предлагаемой числовой оценки высказывания о признаке, учитывающую специфический подход к сбору и анализу информации.

Логично допустить, что чем чаще автор высказывается о признаке при описании "своей" гипотезы, тем более важным он считает этот признак по сравнению со всеми, с помощью которых описывается гипотеза. Отсюда о важности признака в той или иной гипотезе можно судить по частоте упоминания о нем.

Пусть событие A заключается в том, что автор при описании гипотезы упомянул об изучаемом признаке один или несколько раз в одной и той же работе. При этом нас не интересует число упоминаний о признаке в конкретной работе, мы фиксируем только сам факт упоминания о нем.

Вместе с тем высказывания о признаке неравноценны по обоснованности. В одних работах утверждения о признаке высказываются со ссылкой на опыт, в других содержатся теоретические обоснования, в третьих признак упоминается без какого-либо обоснования. Важно отметить, что если автор упоминает о признаке, он обязательно присваивает ему ту или иную оценку обоснованности.

Пусть B_i (где $i = 1, 2, \dots, l$) — событие, заключающееся в том, что автор, упоминая о признаке, дает ему оценку обоснованности. Отметим, что в одной и той же работе события B_i несовместимы.

Кроме оценки обоснованности высказывания о признаке, автор одновременно присваивает признаку оценку качества. Например, в одной статье автор мог отрицать право на существование какого-либо признака при описании своей гипотезы, а в другой (ранней или поздней) он ставил этот признак во главу угла.

Обозначим через C_j (где $j = 1, 2, \dots, h$) событие, заключающееся в том, что автор, упоминая о признаке, присваивает ему оценку качества. События C_j также несовместимы в одной работе.

Для того чтобы выяснить точку зрения автора на важность признака с учетом степени обоснованности и качества, которое он придает признаку при высказывании, необходимо рассматривать совместное появление событий A , B_i и C_j . Обозначим совместное появление их через AB_iC_j . Тогда мерой объективной возможности наступления этого события будет величина, которую назовем вероятностью $p(AB_iC_j)$:

$$p(AB_iC_j) = p(A) \cdot p_A(B_i) \cdot p_{AB}(C_j). \quad (11)$$

При условии, что из общего числа n единственно возможных, равно-возможных и несовместимых работ событию A благоприятствуют m случаев, из которых t_i благоприятствуют событию B_i , и из t_i событию C_j

благоприятствуют k_j случаев, получим

$$p(AB_iC_j) = \frac{m}{n} \frac{t_i}{m} \frac{k_j}{t_i} \frac{k_i}{n}. \quad (12)$$

Очевидно, что $\sum t_i = \sum k_j = m \leq n$.

Для того чтобы количественно оценить важность признака, введем случайную величину ϵ , принимающую значения ϵ_h , когда происходит событие AB_iC_j с вероятностью, равной $p(AB_iC_j)$. Оценка математического ожидания этой величины

$$\overset{A}{M}(\epsilon) = \frac{1}{n} \sum_{h=1}^l \epsilon_h k_h,$$

где k_h — число упоминаний о ϵ_h ; $h = 1, 2, \dots, l$.

Эта оценка учитывает, помимо частоты упоминания о признаке, еще и степень обоснованности, и качество, которое автор придает признаку, упоминая о нем. Такое число необходимо определить для каждого признака, участвующего в описании гипотезы. Числовая характеристика признака служит основой для применения количественных методов при решении проблемы.

Продемонстрируем применение предлагаемого метода формализации смысловой информации на наглядном числовом примере.

Введем некоторые обозначения и понятия. Обозначим через A событие, когда автор упомянул о признаке. Пусть событие B_1 происходит в том случае, когда высказывание подтверждается экспериментом, проведенным непосредственно автором, B_2 — когда в высказывании содержится ссылка на эксперимент, проведенный предыдущими исследователями, B_3 — когда высказывание сопровождается теоретическими обоснованиями.

Таблица 2. Пример формализации высказывания для признаков 1-4

Номер работы	Признаки			
	1	2	3	4
1	1	4	1	—
2	—	0	5	—
3	2	1	4	—
4	—	—	—	1
5	2	1	—	—
6	—	3	0	—
7	5	0	5	—
8	5	5	1	—
9	—	—	—	1
10	3	—	1	—
11	1	—	—	—
12	3	-2	1	—
k_h	8	8	8	2
ϵ_h	1,75	1,00	1,50	0,17

ми, B_4 — когда в высказывании содержится ссылка на предыдущие теоретические обоснования, B_5 — когда высказывание о признаке интуитивное, т.е. производится без какого-либо теоретического или опытного обоснования.

Примем, что событие C_1 происходит в том случае, когда автор использует признак в качестве аргумента, подтверждающего защищаемую им гипотезу, C_2 наступает в том случае, когда не ясно из высказывания, в поддержку какой гипотезы упоминается признак, C_3 происходит в том случае, когда признак упоминается как несостоятельный в развиваемой автором гипотезе.

Используя формулы (11) и (12), определим меру объективной возможности всех сочетаний совместных событий. Для количественной оценки важности признака введем случайную величину ϵ , принимающую следующие значения:

$$\epsilon = \begin{cases} 5, p = k_1/n & \text{при } AB_1C_1 \\ 4, p = k_2/n & \text{при } AB_2C_1 \\ 3, p = k_3/n & \text{при } AB_3C_1 \\ 2, p = k_4/n & \text{при } AB_4C_1 \\ 1, p = k_5/n & \text{при } AB_5C_1 \\ 0, p = k_6/n & \text{при } AB_i C_2 \\ -1, p = k_7/n & \text{при } AB_5C_3 \\ -2, p = k_8/n & \text{при } AB_4C_3 \\ -3, p = k_9/n & \text{при } AB_3C_3 \\ -4, p = k_{10}/n & \text{при } AB_2C_3 \\ -5, p = k_{11}/n & \text{при } AB_1C_3 \end{cases}$$

Обозначив положительные значения величины ϵ через ϵ_x , а отрицательные через ϵ_y , вычислим математическое ожидание

$$\hat{M}(\epsilon) = \frac{1}{n} \left(\sum_{h=1}^5 \epsilon_{x_h} k_h - \sum_{h=1}^5 \epsilon_{y_h} k_h \right). \quad (13)$$

Высказывание автора можно представить в виде табл. 2, по строкам которой расположены наблюдения — имеются в виду отдельные работы (статьи, монографии и др.), а по столбцам — признаки, которыми оперирует автор при описании гипотезы.

Для простоты примера ограничимся рассмотрением только четырех признаков, хотя признаков, используемых для описания гипотезы, несравнимо больше. Условимся, что признаку 1 соответствует высказывание о гумусовом органическом веществе (ОВ), признаку 2 — о сапропелевом ОВ, признаку 3 — о гумусовом и сапропелевом ОВ совместно и, наконец, признаку 4 — об ОВ вообще, т.е. без конкретного указания на его характер.

Необходимо определить: какому именно типу органического вещества как исходному материалу для нефти и газа автор отдает предпочтение.

В клетках табл. 2 фиксируются упоминания о соответствующих признаках. При этом присвоение балльной и качественной оценок производится согласно схеме возможных значений величины ϵ , которые соответствуют определенным высказываниям (событиям). Например, в клетке на пересечении первого столбца и седьмой строки цифра 5 означает, что

произошло событие AB_1C_1 , т.е. автор, упоминая о гумусовом веществе в качестве источника образования углеводов, подтверждает свое высказывание экспериментом. Высказывание, помещенное во втором столбце и второй строке, характеризуется баллом 0. Это означает неуверенность автора в том, что сапропелевое ОВ может участвовать в процессе образования нефти и газа. Клетка на пересечении второго столбца и двенадцатой строки содержит оценку, равную -2 , означающую, что автор ссылается на теоретические расчеты предыдущих исследователей, которые отвергают возможность участия сапропелевого вещества в качестве исходного материала для нефти и газа. Заполнение остальных клеток таблицы не требует дополнительных пояснений. Прочерки в клетках указывают на отсутствие упоминания о признаке.

Визуальный анализ таблицы еще не дает возможности однозначно судить о преимущественном отношении автора к тому или иному признаку. Действительно, если проследить по столбцам упоминания автора о признаках, то получится, что каждый признак им назван как возможный источник образования углеводов.

Однако, суммируя число высказываний о каждом признаке, можно выявить предпочитательность одного признака перед другим. Так, в табл. 2 из строки k_n , означающей сумму упоминаний о признаке, явствует, что автор, описывая исходный материал для образования углеводов, в равной мере отдает предпочтение признакам 1, 2, 3 перед признаком 4. Отсюда можно предположить: 1) автор считает первые три признака одинаково важными; 2) проведенная формализация недостаточна для ранжирования изучаемых признаков.

Для более тонкого анализа наряду с частотой высказывания о признаке необходимо учитывать оценку обоснованности высказывания совместно с качественной характеристикой признака с точки зрения автора. Используя данные табл. 2, по формуле (13) вычислим оценки важности каждого признака. Так, оценка признака 2 (сапропелевое вещество) составит:

$$\hat{\epsilon}_2 = \frac{1}{12} (0 \cdot 2 + 1 \cdot 2 + 3 \cdot 1 + 4 \cdot 1 + 5 \cdot 1 - 2 \cdot 1) = 1,00.$$

Оценки других признаков, вычисленные таким же образом, записаны в нижней строке таблицы.

Теперь, ранжируя полученные оценки важности признаков, можно однозначно установить, что автор отдает предпочтение гумусовому ОВ в гипотезе происхождения нефтяных углеводов, поскольку оценка важности признака 1 имеет максимальное значение, равное 1,75.

Таким образом, используя степень обоснованности высказывания и учитывая качественную характеристику признака, мы добились более четкой формализации высказывания. Оценив таким методом каждый признак, участвующий в описании той или иной гипотезы, можно добиться формализации проблемы в целом, что даст возможность при исследовании применять обширный аппарат методов математической статистики.

3. ХАРАКТЕРИСТИКА МАССИВА ДАННЫХ "ИНДИВИДУАЛЬНЫЕ ВЫСКАЗЫВАНИЯ"

Собранный фактический материал представлен в таком виде, что содержащиеся в нем суждения принадлежат, как правило, не отдельному индивидууму, а соавторам. Дело в том, что в массиве "литература" за основной признак при сборе информации принята работа, в написании которой, как правило, участвует несколько авторов. Такой источник информации, аналогичный источнику, используемому при групповой экспертизе, как показано в первой главе, обладая определенными преимуществами перед исследованиями, применяемыми при опросе индивидуумов, имеет и свои недостатки. Информация, получаемая от группы авторов, может привести к искажениям и, следовательно, смещениям групповых оценок из-за каких-либо причин психологического или даже морального характера. Например, группа может оказывать влияние на отдельных ее членов, которые, не будучи уверены в правильности навязываемого мнения, вынуждены соглашаться с мнением большинства. Или наоборот, мнение большинства может подчиниться мнению наиболее настойчивого индивидуума, игнорирующего аргументы остальных членов группы.

Искажающий эффект при использовании информации, полученной от группы, может возникать еще и по такой причине. Допустим, что один и тот же автор на протяжении нескольких лет выступал в разных коллективах. При этом мнения коллективов контрастно отличаются. Тогда уловить личное мнение автора не представляется возможным, хотя оно может оказаться очень важным при наших исследованиях. Такой случай не выдуман для примера, а существует в действительности. Подтверждением этому служит тот факт, что среди собранного фактического материала нередко встречаются работы, в которых автор, выступая индивидуально, менял свою точку зрения на радикально противоположную порой несколько раз.

Однако какие бы противоречивые высказывания автор не приводил, он все равно употребляет одни и те же "стержневые" признаки при описании своей гипотезы. Выявить эти признаки, отсеяв случайные, и составляет задачу исследований.

Следовательно, "очистить" информацию от помех, содержащихся в коллективном высказывании, вероятно, можно, если в качестве элементарного источника информации использовать не отдельную работу, а отдельного автора — участника этой работы. Поэтому следующий шаг при обработке фактического материала заключался в формировании нового массива (картотеки), названного "индивидуальные высказывания", содержащего "статистические показатели геологии нефти и газа" (см. рис. 1).

Создание картотеки состояло в следующем. На базе предыдущих исследований, а именно на основе декомпозиционных схем, была разработана поисковая кодовая карта, включающая в себя все факторы и соответствующие им признаки, выявленные на этапе обследования и относящиеся к проблеме происхождения нефти и газа. Поисковая карта служит одновременно двум целям. Во-первых, как можно судить по названию карты, она предназначена для поиска необходимых сведений, содержащихся в картотеке. Во-вторых, она представляет собой модель анкеты либо опросного листа, подобие которых применяется обычно в методах экспертизы на этапе сбора фактического материала (табл. 3).

Таблица 3. Классификация показателей, участвующих в статистическом анализе гипотез происхождения нефтяных углеводородов

Фактор	Подфактор	Признак	Номер показателя
1	2	3	4
Исходный материал		Графит	1
		Оксид углерода	2
		Двуокись углерода	3
		Водород неорганический	4
		Продукт распада органического вещества (ОВ)	5
		Продукт деятельности микроорганизмов	6
		Вода	7
		Радикалы (СН, СН ₂ , СН ₃) мантии	8
		Радикалы (продукты распада ОВ)	9
		Радикалы космоса	10
		Метан мантии	11
		Карбиды металлов	12
		Углеводороды космоса (готовые)	13
		Органическое вещество вообще	14
		Органическое вещество вообще (как источник приповерхностного метана)	14а
		ОВ гумусового типа	15
		ОВ гумусово-сапропелевого типа	16
		ОВ сапропелево-гумусового типа	17
		ОВ сапропелевого типа	18
		Уголь	19
		Горючие сланцы	20
Прочий (нетипичный)	21		
Количественная оценка	Неорганический исходный материал	Достаточно	22а
		Недостаточно	22б
	Органический исходный материал	Достаточно	23а
		Недостаточно	23б
Степень дисперсности ОВ	—	Рассеянное	24
		Концентрированное	25
Литофации	—	Песчаники и алевролиты	26
		Карбонатные породы	27
		Глины	28
Стадии преобразования	—	Седиментогенез	29
		Диагенез	30
		Катагенез	31
		Метаморфизм	32
Геохимическая обстановка	—	Окислительная	33
		Восстановительная	34
		Окислительно-восстановительная	35
		Щелочная	36
Преобразование исходного материала	Энергия	Биохимическая (микроорганизмы)	37
		Электрическая	38
		Солнечная	39
		Радиация	40
	Спонтанно	Собственная энергия ОВ	41
		Радиогенный распад	42

Таблица 3 (продолжение)

1	2	3	4	
	—	Прочие нетипичные	43	
	Глубина, м (интервал)	0—1000	44	
		1001—2000	45	
		2001—3000	46	
		3001—4000	47	
		4001—5000	48	
		5001—6000	49	
		>6000	50	
	Температура	Конкретные значения	51	
		Высокая	52	
		Низкая	53	
		Влияние вообще	54	
		Нет влияния	55	
	Давление	Нет влияния	56	
		Влияние вообще	57	
		Низкое	58	
		Высокое	59	
		Конкретные значения	60	
		—	Катализатор	61
		Время (в геологическом масштабе)	62	
Парастерезис (спутники УВ- залежей, -проявле- ний)	—	Негорючий газ (вообще)	63	
		Окись углерода	64	
		Двуокись углерода	65	
		Водород	66	
		Сероводород, сера и ее соединения	67	
		Азот и его соединения	68	
		Инертные газы	69	
		УВ непередельные	70	
		Микроэлементы	71	
		Прочие (нетипичные)	72	
Различный сос- тав УВ	—	Исходный материал	73	
		Условия захоронения ОВ	74	
	Вторичные факто- ры	Вторичные факторы (вообще)		75
	Миграция	Миграция вообще		76
			(Ад) сорбция	77
			Дифференциация	78
	Метаморфизм	Температура		79
			Давление	80
			Глубина	81
Окисление	—	Микроорганизмы	82	
		Вода	83	
		Катализатор	84	
		Радиация	85	
		Гидрирование	86	
		Время (в геологическом масштабе)	87	
Прочие (нетипичные)	88			
Состояние УВ при миграции	Растворенное в газе	Растворенное состояние (вообще)	89	
		В газе (вообще)	90	
		В углеводородном	91	
		В неуглеводородном	92	
	Растворенное в воде	—	В воде (вообще)	93
			В минерализованной	94

Таблица 3 (окончание)

1	2	3	4
		В пресной	95
	Однофазное состояние	Однофазное состояние (вообще)	96
		Нефть-газ (струйная миграция)	97
		Нефть-газ-вода	98
	-	Дисперсно-коллоидальное	99
		Прочие (нетипичные)	100
Направление миграции УВ	Вертикальное	Вертикальное (вообще)	101
		Вверх	102
		Вниз	103
	Латеральная	Латеральная (вообще)	104
		Близкая	105
		Дальняя	106
		in situ	107
	"Боковая"	108	
Факторы, влияющие на миграцию УВ	-	Капиллярные явления	109
		Диффузионные явления	110
		Гравитационные явления	111
		Тектоническая активность	112
		Уплотнение пород при погружении	113
		Цементация пород	114
		Прочие (нетипичные)	115
Процесс миграции	-	Удаление	116
		Засасывание	117
Специфические свойства УВ	-	Оптическая активность	118
		Порфирины (их наличие)	119
		Споры, пыльца (их наличие)	120
		Прочие (нетипичные)	121
Среда обнаружения УВ	-	Рассматривается как признак	122

Анкета включает следующий перечень вопросов.

1. Возможный исходный материал для образования нефтяных углеводородов (УВ).
2. Достаточно ли количества названного исходного материала для образования крупных месторождений УВ?
3. Необходимая степень дисперсности исходного материала.
4. Имеют ли значение литофации осадочного чехла при образовании УВ. Если да, то какие именно?
5. На какой стадии формирования осадочного чехла образуются УВ.
6. Возможные факторы преобразования исходного материала.
7. Возможные парастерические ассоциации, встреченные при изучении УВ-залежей (проявлений).
8. Необходимые факторы, определяющие различный состав УВ.
9. Физико-химическое состояние УВ, способствующее их миграции.
10. Возможное направление миграции УВ.
11. Необходимые силы или явления, обеспечивающие миграцию УВ.

б

NN п/п	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96
1																1	1	1															
2										1																							
3	$1^{(5)}$	$1^{(5)}$	$1^{(5)}$		$1^{(5)}$					1			1			1															1		
4																																	
5		1			1						1																						
6	$1^{(5)}$	$2^{(1,5)}$	$1^{(5)}$		$2^{(1,5)}$					2			1			2	1	1													1		
7	1,0	1,2	1,0		1,2					0,4			0,2			0,4	0,2	0,2													0,2		
NN п/п	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	т					
1																																	8
2						1		1																									8
3																																	19
4																																	6
5																																	12
6						1		1																									29
7						0,2		0,2																									

Примечание:

Рис. 12. Пример заполнения рабочей перфокарты "Статистические показатели геологии нефти и газа"
 а — лицевая сторона; б — оборотная сторона

12. Какие процессы (выталкивание или засасывание) происходят при взаимодействии УВ с вмещающими породами?

13. Специфические свойства, отмеченные при изучении состава УВ.

14. Среда, в которой обнаружены скопления УВ со специфическими свойствами.

Ответы на вопросы фиксируются в рабочей перфокарте, пример заполнения которой показан на рисунках 12, а, б. Такая карта заведена для каждого автора, участника хотя бы одной работы, попавшей в список картотеки "Литература", собранной на этапе обследования проблемы.

На поставленный вопрос требуется ответ в "закрытой форме". В поисковой карте-вопроснике содержится весь возможный перечень ответов, которым присвоен номер, соответствующий номеру признака на декомпозиционных схемах. Учтены также случаи совсем уж "необычных" ответов, для чего отведено специальное поле на карте, соответствующее признаку "прочий (нетипичный)". Такой признак может быть возведен в ранг фактора, если не потребуются его градация на отдельные элементы в процессе решения проблемы.

На каждый вопрос может быть получено только четыре несовместимых ответа: "да", "нет", "не уверен", "игнорирую", которые отмечаются определенным образом в клетках рабочей карты. Первые два ответа отмечаются знаками "+", "-", при этом одновременно учитывается балльная система, описанная в предыдущем разделе, где показано вычисление количественной оценки признака.

Ответ "не уверен" употребляется в том случае, когда из текста изучаемой работы неясно, в пользу опровержения или подтверждения назван конкретный признак. При фиксировании этого ответа в карте клетка соответствующего признака заполняется нулем (0).

Ответ "игнорирую" соответствует случаю, при котором описание гипотезы обходится без употребления данного признака. При этом отведенная для этого признака клетка в карте остается пустой.

Правда, последние два ответа становятся неразличимыми, содержат одинаковое количество информации, равное нулю, в случае, когда для обработки информации, предварительно формализованной по предлагаемой шкале, используется описанная методика оценки важности признака. Однако, как стало ясно в процессе обработки информации, особенно на его предварительном этапе, учитывать различия между такими ответами необходимо.

Так, на этапе предварительного (огрубленного) анализа проблемы возникла необходимость рассчитать оценку важности признака только лишь на основании частоты упоминания, т.е. без учета степени обоснованности и качественной характеристики признака, необходимых в предлагаемой методике. В таком случае принятая градация ответов принесла определенный эффект. Действительно, ответ "не уверен" означает, что признак упомянут, а это можно трактовать как факт, что данный признак более достоин внимания по сравнению с каким-то неупомянутым.

Проследим на конкретном примере, как осуществляется процесс заполнения рабочей перфокарты из картотеки "Индивидуальные высказывания".

Пусть автор А является участником какой-либо работы, зафиксирован-

ной в картотеке "Литература". Анализируя реферат этой работы и ориентируясь на возможные ответы, перечисленные в карте-вопроснике (см. табл. 3), перенесем все признаки, которые использованы при высказывании о гипотезе, в карту автора *A* (см. рис. 12, *a, б*). (Точно такие же признаки фиксируются в картах, заведенных для других авторов, если эти авторы являются участниками данной работы.) Затем аналогичная операция проводится с другой работой, и так до тех пор, пока список всех работ, составляющих картотеку "Литература", не будет исчерпан.

Допустим в какой-то работе утверждается, что гумусовое органическое вещество, преобразуясь на стадии катагенеза, может служить материалом для образования нефтяных углеводородов. При этом объясняется: поскольку органическое вещество может быть представлено как рассеянной, так и концентрированной формой, то количества этого вещества вполне достаточно для формирования любых скоплений нефти и газа. Отмечая различный состав углеводородных скоплений, относят этот факт на счет различного влияния глубины, температуры и давления, при которых обнаружены эти скопления.

В формализованном виде такое высказывание соответствует набору признаков 15, 23, 24, 25, 31, 79, 80 и 81, отмеченных в первой строке рабочей карты автора *A*. Эта строка соответствует работе, написанной в 1977 г.

Рассмотрим другую работу, в написании которой также участвовал автор *A*. Признаки, выявленные при чтении этой работы, занимают третью строку в рабочей карте (рис. 12, *a, б*). По сравнению с предыдущей работой в этой используется значительно большее число признаков для описания гипотезы. Причем высказывание о некоторых признаках сопровождается опытными и теоретическими исследованиями. В частности, в этой работе описан эксперимент, в котором уголь, другие гумусовые и сапропелевые вещества при высоком (?) давлении подвергались нагреванию до $150 \div 250^\circ\text{C}$. В результате получено нефтеподобное вещество, в составе которого обнаружены концентрации CO , CO_2 , H_2 и N_2 . Приведены расчеты, обосновывающие достаточность количества названных исходных веществ для образования крупных месторождений нефти и газа. Клетки, отвечающие упомянутым признакам, заполнены соответствующим баллом. Признак 59, хотя и указан среди названных в опыте признаков, фиксируется в карте с баллом, равным единице, т.е. его оценка приравнивается к интуитивному высказыванию, поскольку не ясно, при каком конкретном давлении производился опыт.

Кроме того, в этой же работе указано, что преобразование исходного материала начинается на стадии диагенеза, а завершается при катагенезе осадочных пород. Образовавшиеся углеводороды затем мигрируют в растворенном в воде состоянии. Отмечают также, что колебания в составе углеводородов, наблюдаемые в разных залежах, зависят в основном от характера исходного материала, от миграции, в процессе которой происходит изменение состава нефти и газа, а также от температурных условий, влияющих на залежь.

В следующей работе, датированной 1968 годом, автор уделит внимание также опровержению противоположной гипотезы, где он отрицает возможность образования нефтяных углеводородов из неорганических веществ, таких, как окись и двуокись углерода в сочетании с водородом.

Сводка всех признаков, необходимых с точки зрения автора *A* для опи-

сания гипотезы происхождения нефти и газа, приведена в предпоследней строке карты. В ней указаны сумма упоминаний о каждом признаке, а в скобках число упоминаний о них с соответствующим баллом. В последней строке записана оценка признака, рассчитанная по формуле (13) и учитывающая как качественную характеристику признака, так и степень обоснованности высказывания о нем.

4. СВЕРТЫВАНИЕ ИНФОРМАЦИИ

Устранение шумовых помех. На рис. 12, а, б видно, что некоторые признаки, названные в отдельных работах, не учтены в сводке, характеризующей мнение автора о гипотезе в целом. При анализе информации с позиций "качественного критерия оценки истинности высказывания" эти признаки оказались отбракованными, поскольку в высказываниях об этих же признаках в других работах содержатся противоположные утверждения, сопровождаемые опытными обоснованиями. Отрицание же данных признаков сопровождается только лишь интуитивными высказываниями, и не встретилось ни одной работы, в которой этот факт подтверждался бы опытным или расчетным обоснованием.

Для отбраковки необоснованных значений признаков удобно использовать табл. 4, представляющую собой сводку всех значений признаков, к помощи которых обращаются исследователи при описании своей точки зрения на образование углеводородов, формирование их скоплений и другие вопросы, относящиеся к проблеме нефти и газа.

В таблице некоторые значения помечены звездочкой, которая указывает, что данное значение противоречит опытным результатам и, следовательно, при дальнейшей обработке информации не должно учитываться.

Всего для описания проблемы, как следует из табл. 4, используется более 203 значений признаков*. Однако после того как необоснованные значения признаков были устранены, осталось только 160 значений, которые будут участвовать при последующем исследовании проблемы в процессе решения поставленной задачи.

Таким образом, проведенный анализ с позиций качественного критерия оценки истинности высказывания позволил "очистить" информацию от некоторых "шумовых" помех, что дало возможность сократить объем информации.

Однако "шумовые" помехи все еще содержатся в изучаемых объектах в виде случайных признаков, не отвечающих истинному мнению автора, но упомянутых при высказывании о проблеме в работах, написанных в соавторстве с непостоянным коллективом.

У с т р а н е н и е с л у ч а й н ы х п р и з н а к о в. Исходя из предположения, что, чем выше оценка признака, тем большее значение автор придает ему при высказывании о гипотезе, будем считать шумовыми те из них, оценки которых значимо ниже средней оценки, рассчитанной для всех признаков, употребляемых автором. Следует оговориться, что такое предполо-

* Признаки под названием "прочий" (нетипичный) имеют, как правило, несколько значений, хотя в данной таблице, преследующей вполне определенную цель, их количество не указано.

Таблица 4. Сводка признаков, участвующих в описании проблемы, и их максимально возможный балл

Номер признака	1	2	3	4	5	6	7	8	9	10	11	12	
Значение признака	1	2	-2*	3	-3*	4	-4*	5	-5*	6	-6*	7	
Максимально возможный балл	5	5	-	5	-	5	-	5	-	3	-	5	
Номер признака	13	14	15	16	17	18	19	20	21	22	23	24	
Значение признака	-7*	8	-8*	9	10	11	-11*	12	-12*	13	-13	14	
Максимально возможный балл	5	3	-	3	3	5	-	5	-	3	3	3	
Номер признака	25	26	27	28	29	30	31	32	33	34	35	36	
Значение признака	-14	15	-15	16	17	18	-18	19	-19	20	-20	21	
Максимально возможный балл	3	5	5	3	3	5	5	5	5	5	5	5	
Номер признака	37	38	39	40	41	42	43	44	45	46	47	48	
Значение признака	22	-22*	23	-23	24	-24*	25	-25*	26	-26	27	-27	
Максимально возможный балл	3	-	3	3	5	-	5	-	3	3	3	3	
Номер признака	49	50	51	52	53	54	55	56	57	58	59	60	61
Значение признака	28	-28	29	-29	30	-30	31	-31	32	-32*	33	-33*	34
Максимально возможный балл	3	3	5	5	5	5	3	3	3	-	5	-	5
Номер признака	62	63	64	65	66	67	68	69	70	71	72	73	74
Значение признака	35	-35	36	37	-37	38	39	40	-40*	41	42	43	44
Максимально возможный балл	3	3	3	5	5	3	3	5	-	3	3	5	3

Таблица 4. (окончание)

Номер признака	75	76	77	78	79	80	81	82	83	84	85	86	87
Значение признака	-44	45	-45	46	-46	47	-47	48	-48	49	-49	50	-50*
Максимально возможный балл	3	3	3	3	3	3	3	3	3	3	3	3	—
Номер признака	88	89	90	91	92	93	94	95	96	97	98	99	100
Значение признака	52	-52	53	-53*	54	55*	56	57	58	-58*	59	-59	61
Максимально возможный балл	5	5	5	—	3	—	5	3	5	—	5	5	5
Номер признака	101	102	103	104	105	106	107	108	109	110	111	112	
Значение признака	-61*	62	-62*	62	-62	63	64	65	-65*	66	-66*	67	
Максимально возможный балл	—	5	—	3	3	3	5	5	—	5	—	5	
Номер признака	113	114	115	116	117	118	119	120	121	122	123	124	
Значение признака	-67*	68	69	70	-70*	71	72	73	-73	74*	-74	75	
Максимально возможный балл	—	5	5	5	—	5	5	5	5	—	5	5	
Номер признака	125	126	127	128	129	130	131	132	133	134	135	136	
Значение признака	-75*	76	-76*	77	-77*	78	-78*	79	-79*	80	-80*	81	
Максимально возможный балл	—	5	—	5	—	5	—	5	—	5	—	5	
Номер признака	137	138	139	140	141	142	143	144	145	146	147	148	
Значение признака	-81	82	-82	83	-83*	84	85*	-85	86	87	-87	88	
Максимально возможный балл	5	5	5	5	5	5	—	3	5	5	5	5	

Номер признака	149	150	151	152	153	154	155	156	157	158	159	160
Значение признака	89	-89	90	-90*	91	-91	92	-92	93	-93	94	-94
Максимально возможный балл	3	3	5	-	5	5	5	5	5	5	5	5
Номер признака	161	162	163	164	165	166	167	168	169	170	171	172
Значение признака	95	-95	96	-96*	97	98	99	-99	100	101	-101*	102
Максимально возможный балл	5	5	5	-	3	3	5	5	5	5	-	5
Номер признака	173	174	175	176	177	178	179	180	181	182	183	184
Значение признака	-102*	103	-103	104*	-104	105	106	-106	107	-107*	108	-108*
Максимально возможный балл	-	5	5	-	3	3	5	5	5	-	5	-
Номер признака	185	186	187	188	189	190	191	192	193	194		
Значение признака	109	110	111*	-111	112	-112*	113	-113	114	115		
Максимально возможный балл	3	5	-	3	3	-	5	5	5	5		
Номер признака	195	196	197	198	199	200	201	202	203			
Значение признака	116	-116	117	-117	118	119	120	121	122			
Максимально возможный балл	5	5	3	3	5	5	5	5	5			

Примечание. Признаки, отмеченные звездочкой, не участвуют в дальнейших исследованиях.

жение проверено статистическими методами, для чего сравнивались средние оценки признаков, полученные в различных их совокупностях. Изучались:

Выборка 1 — вариационный ряд, составленный из признаков, обоснованных баллом.

Выборка 2 — вариационный ряд тех же признаков, но без учета степени обоснованности.

Выборка 3 — вариационный ряд только тех признаков, которые употребляются без всякого обоснования.

Выборка 4 — вариационный ряд признаков всех без исключения (охарактеризованных и не охарактеризованных балльной оценкой).

Выборка 5 — вариационный ряд всех признаков без учета баллов.

Проверка осуществлена с помощью непараметрического критерия Вилкоксона W , поскольку в некоторых сравниваемых выборках нормальность эмпирических распределений и равенство дисперсий резко нарушены. Метод чрезвычайно прост, эффективен и заключается в следующем [72].

Обозначим объем первой выборки $n_1 = m$, а второй $n_2 = n$ при $m \leq n$, в противном случае выборки поменяем местами. Составим из двух выборок общий вариационный ряд объемом $N = m + n$ и припишем членам этого ряда номера от 1 до N . Членам вариационного ряда с равным значением оценки признака присвоим один и тот же средний номер (ранг). Определим сумму номеров (рангов), принадлежащих меньшей по числу наблюдений выборке. Обозначим эту сумму через $W = \sum \tilde{r}$, где \tilde{r} — номера (ранги), принадлежащие выборке с числом $m (m \leq n)$.

Нижние и верхние критические значения, соответствующие уровню значимости q , m и n наблюдениям (при этом m или $n > 25$) в совокупностях определяют по приближенным формулам:

$$W_1 = \frac{1}{2} [m(m+n-1)] - \psi_{(1-q/2)} \sqrt{\frac{mn(m+n-1)}{12}},$$

$$W_2 = m(m+n-1) - W_1,$$

где $\psi_{(1-q/2)}$ — значение обратной функции нормального распределения с параметрами 0 и 1. Для уровня значимости $q = 0,05$ значение $\psi_{(1-q/2)} = 1,96$.

Проверяемая гипотеза $H_0: \mu_1 = \mu_2$ (μ — математическое ожидание случайной величины) принимается как не противоречащая эмпирическим данным, если величина W не выйдет за пределы, образованные критическими значениями $W_1 \leq W \leq W_2$, и отклоняется как неподтвердившаяся, если величина W выйдет за допустимые пределы $W < W_1$ или $W > W_2$.

Результаты сравнения выборок 3 и 5, не охарактеризованных баллом, показали, что проверяемые гипотезы о равенстве средних не противоречат эмпирическим данным. Значения же средних оценок признаков, которые используются в высказывании с обоснованием (выборка 1), выше, чем такие же оценки в остальных выборках. При этом выборка 2, в которой изучались оценки "обоснованных" признаков без учета степени их обоснованности, также имеет повышенные значения по сравнению со средней оценкой всего ряда признаков, несмотря на то, что в нем учитывались баллы (выборка 4). Среднее значение признаков в выборке 2 значительно превышает среднее значение в выборке 5. Во всех случаях расчеты проведены с уровнем значимости $q = 0,05$.

Такие же выводы получены в результате проверки равенства средней оценки признаков, несущих различную информацию о степени обоснованности высказывания у отдельных авторов.

Таким образом, проведенный анализ позволил удостовериться в справедливости предположения о том, что признаки, имеющие низкие оценки, можно считать излишними, создающими "шумовые" помехи.

Следующий этап работы заключался в отбраковке случайных признаков, встреченных в каждом объекте. Для этой цели по объектам рассчитана средняя арифметическая признака, а затем с помощью интервальной оценки, использующей односторонний критерий, отсеяны признаки, значения которых отличаются от наилучшей их средней оценки.

Выражение доверительного интервала для μ с вероятностью $1 - q$ выглядит следующим образом:

$$p = \left[\bar{x} - \frac{\sigma}{\sqrt{n}} t_{q/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} t_{q/2} \right] = 1 - q.$$

В этой формуле μ — неизвестная средняя распределения; \bar{x} — выборочная средняя; σ — стандартное отклонение для всей выборки; n — число элементов в выборке. Дробь σ/\sqrt{n} равна стандартному отклонению средней.

Определив на основе предыдущих исследований, что среднее значение признаков, достойных дальнейшего изучения, больше, чем среднее значение всей их совокупности, формулируем нуль-гипотезу $H_0: \mu_1 \leq \mu_2$ при соответствующей альтернативной гипотезе $H_1: \mu_1 > \mu_2$, и на этой основе все признаки, которые не фиксирует альтернативная гипотеза, исключаем из дальнейшего рассмотрения.

Сокращение числа объектов. В результате разбраковки высказываний, содержащихся в одной работе, написанной коллективом авторов, на "самостоятельные" высказывания, которые на данном этапе стали объектами изучения, их число увеличилось с 1768 до 2409.

Ясно, что обработка такого огромного объема информации представляет собой очень трудоемкий процесс. Сокращение же числа исследуемых объектов по возможности без потери содержащейся в них информации позволит решать задачу более дешевыми средствами.

Процедура уменьшения числа объектов изображена на рис. 13.

Процесс операций А и Б уже подробно описан в начале этого параграфа и, вероятно, здесь не требует дополнительных пояснений.

О п е р а ц и я В. В объект объединены работы одного автора, написанные в разные годы. Такая форма учета, во-первых, полнее характеризует индивидуальное высказывание о гипотезе, во-вторых, дает основание для расчета количественной оценки важности признака (фактора).

О п е р а ц и я Г. В том случае, если объекты, полученные при формировании картотеки "Индивидуальные высказывания", не имели ни одной оригинальной работы, т.е. не имели никакой новой информации по сравнению с групповым высказыванием, они вновь объединялись. Такая процедура необходима по той причине, что при подсчете числа упоминаний о том или ином признаке оно может оказаться завышено по сравнению с фактическим, поскольку единичное высказывание, не несущее дополнительной информации, но размноженное искусственным путем, увеличит частоту

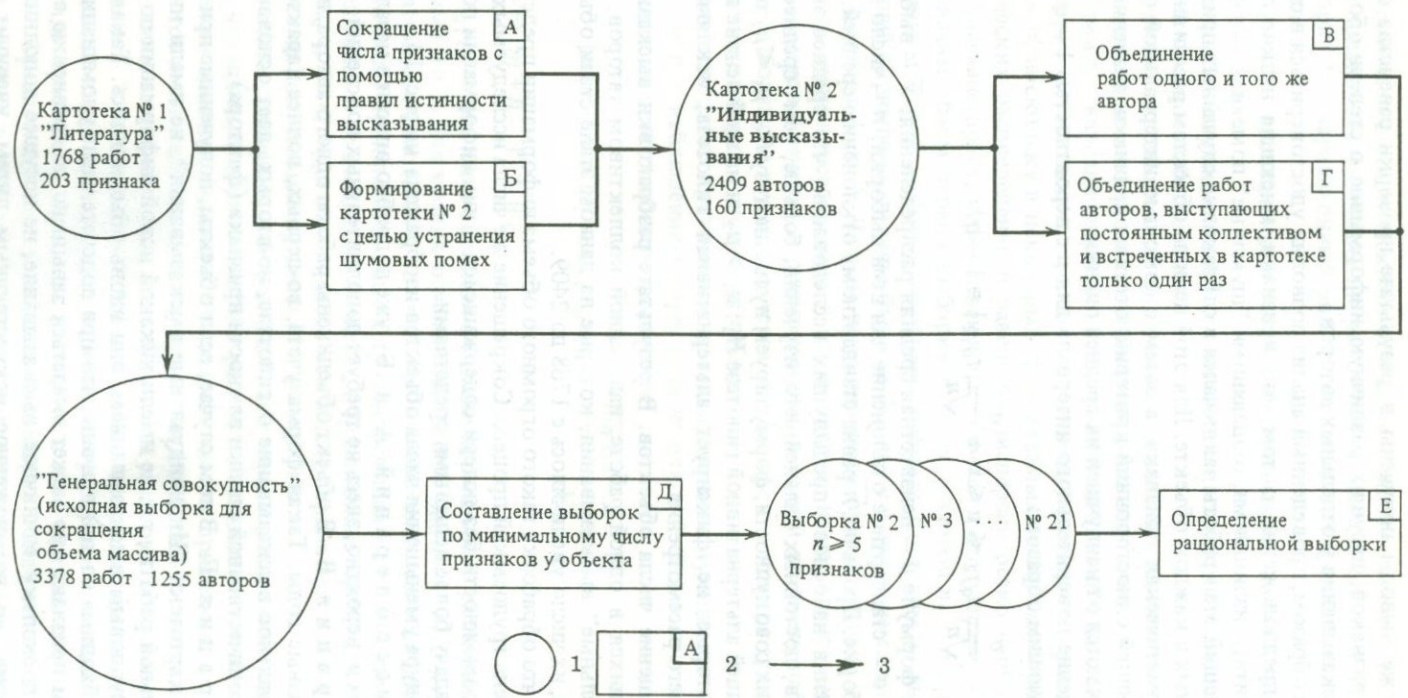


Рис. 13. Блок-схема формирования представительной выборки сокращенного объема
1 — объекты изучения; 2 — операция и ее индекс; 3 — функциональные связи

встречаемости признака. За счет операций объединения число объектов уменьшилось до 1255.

О п е р а ц и я Д. Каждый из оставшихся в массиве объектов содержит неравномерное количество информации. Об этом можно судить по различному числу признаков, употребляемых для выражения точки зрения автора относительно изучаемой проблемы. График на рис. 14 указывает на резко выраженную положительную асимметрию ряда распределения количества признаков, названных в индивидуальных высказываниях. Наблюдаемую картину можно интерпретировать двояко: автор либо считает, что небольшой набор признаков достаточен для описания всех сторон гипотезы, либо описывает лишь отдельные фрагменты проблемы генезиса нефти и газа.

Тщательный анализ фактического материала показал, что, как правило, малое число признаков используется теми авторами, которые описывают вспомогательные стороны проблемы — процессы миграции и преобразования углеводородов. Поэтому следующий шаг состоял в отбраковке объектов с малым числом признаков.

Было составлено 20 выборок, различающихся количеством признаков, содержащихся в изучаемых объектах (характеристику выборок см. в графах 2—4 табл. 5). Уменьшая объем выборки при увеличении числа признаков в объекте, мы несем определенные потери информации. Для того чтобы вести учет этим потерям, каждая выборка сравнивалась с эталонной с помощью коэффициента ранговой корреляции Спирмена по формуле (8).

Эталоном служила выборка, включающая все собранные данные о проблеме генезиса нефти и газа, предварительно очищенные от шумовых помех, с помощью правила оценки качества "истинности" высказывания. Эту выборку назовем генеральной совокупностью, поскольку все сравниваемые выборки составлены из нее. Так, в выборку 2 попали все объекты, имеющие 5 признаков и больше, в выборку 3—8 и более признаков и т.д. При этом каждая предыдущая выборка включает в себя последующую. Сравнимые выборки ранжированы по частоте упоминания о каждом признаке. Результаты вычисления коэффициентов корреляции Спирмена сведены в табл. 5 (графа 5).

При сокращении числа объектов, кроме того, неизбежно теряется информация и в виде балльной оценки, присвоенной высказыванию. Эти потери также оценены с помощью коэффициента корреляции Спирмена, только в данном случае ранги назначались по сумме баллов для каждого признака. Вычисленные коэффициенты корреляции указаны в графе 6 табл. 5.

Ранее, характеризуя сложность проблемы, мы указывали, что наиболее важными факторами, непосредственно ответственными за образование нефтяных углеводородов, являются так называемые прямые факторы (см. рис. 8). По ним (без учета остальных) также проведено сравнение отобранных массивов информации.

Аналогично предыдущему сравнивались искажения, возникающие в зависимости от сокращения объема выборки как по частотам упоминания о признаках, так и по их балльной оценке. Эти результаты приведены соответственно в графах 7 и 8 табл. 5.

Так как оценка важности признака базируется на частотной основе, важным моментом при формировании лучшей для наших целей выборки является учет представительности объекта изучения (высказывания о ги-

Таблица 5. Расчет наилучшей представительной выборки сокращенного объема

Номер выборки	Характеристика выборки			Матрица оценки				
	n	m	k	r_1	r_2	r_3	r_4	\bar{n}
1	2	3	4	5	6	7	8	9
1	Исходная выборка	1255	3378	1,00	1,00	1,00	1,00	2,69
2	5	545	1944	0,992	0,944	0,992	0,996	3,57
3	8	332	1382	0,989	0,937	0,990	0,920	4,16
4	9	285	1314	0,986	0,901	0,984	0,860	4,61
5	10	235	1214	0,982	0,885	0,986	0,847	5,16
6	11	192	1072	0,981	0,874	0,981	0,841	5,59
7	12	176	1026	0,979	0,860	0,983	0,835	5,83
8	13	141	910	0,975	0,849	0,981	0,832	6,45
9	14	127	864	0,972	0,849	0,980	0,827	6,80
10	15	126	863	0,971	0,825	0,980	0,810	6,85
11	16	114	802	0,969	0,824	0,977	0,809	7,04
12	17	100	750	0,968	0,828	0,971	0,804	7,50
13	18	87	674	0,963	0,808	0,972	0,786	7,75
14	19	79	640	0,960	0,812	0,967	0,810	8,10
15	20	75	602	0,960	0,820	0,968	0,791	8,02
16	21	66	559	0,958	0,811	0,964	0,806	8,47
17	22	58	515	0,954	0,799	0,963	0,803	8,88
18	23	52	477	0,949	0,801	0,961	0,814	9,18
19	24	45	444	0,946	0,797	0,961	0,807	9,87
20	25	43	432	0,943	0,719	0,961	0,789	10,05
21	29	31	337	0,934	0,646	0,939	0,730	10,87

Обозначения. n — минимальное количество признаков; m — количество авторов; k — количество статей; *наилучшая выборка.

потезе). В данном случае представительность определяется количеством различных работ, сосредоточенных в одном объекте.

Судя по рис. 15 количество работ у авторов, а следовательно, и у объектов изучения, которые они представляют, распределено неравномерно. Этот график по форме повторяет график распределения количества признаков у авторов (см. рис. 14). Может сложиться впечатление, что большое число признаков у автора накапливается за счет большого числа статей, в которых он использует различные признаки. Однако анализ фактического материала свидетельствует, что это не так. В картотеке есть достаточно большое число авторов, имеющих мало статей, но использующих много признаков для описания гипотезы.

Такой факт лег в основу при формировании показателя "представительности объекта". За лучший показатель принято не максимальное число работ, а среднее их содержание в массиве, объекты которого имеют 29 и более признаков. Таких объектов оказалось 31.

В свою очередь, полученный объем массива принят лучшим по показателю представительности исходя из следующих соображений. Допустим,

Стандартизированные значения					Мак- симин	Матрица потерь					Мини- макс
r_1^1	r_2^1	r_3^1	r_4^1	\bar{n}^1	W_B	r_1^1	r_2^1	r_3^1	r_4^1	\bar{n}^1	W_C
10	11	12	13	14	15	16	17	18	19	20	21
1,88	2,13	1,86	2,69	-1,97	-1,97	0	0	0	0	3,72	3,72
1,41	1,39	1,28	2,14	-1,57	-1,57	0,47	0,74	0,58	0,55	3,32	3,32
1,24	1,30	1,14	1,40	-1,30	-1,30	0,64	0,83	0,72	1,29	3,05	3,05
1,06	0,83	0,71	0,44	-1,10	-1,10	0,82	1,30	1,15	2,25	2,85	2,85
0,82	0,62	0,86	0,22	-0,84	-0,84	1,06	1,51	1,00	2,47	2,59	2,59
0,76	0,47	0,50	0,13	-0,65	-0,65	1,12	1,66	1,36	2,56	2,40	2,56*
0,65	0,29	0,64	0,05	-0,54	-0,54	1,23	1,84	1,22	2,64	2,29	2,64
0,41	0,14	0,50	-0,02	-0,26	-0,26	1,47	1,99	1,36	2,71	2,01	2,71
0,24	0,14	0,43	-0,10	-0,10	-0,10*	1,64	1,99	1,43	2,79	1,85	2,79
0,18	-0,17	0,43	-0,37	-0,08	-0,37	1,70	2,30	1,43	3,06	1,83	3,06
0,02	-0,18	0,21	-0,39	0,01	-0,39	1,86	2,31	1,65	3,08	1,74	3,08
0,00	-0,13	-0,21	-0,47	0,22	-0,47	1,88	2,26	2,07	3,16	1,53	3,16
-0,29	-0,39	-0,14	-0,76	0,33	-0,76	2,17	2,52	2,00	3,45	1,42	3,45
-0,47	-0,34	-0,50	-0,37	0,49	-0,50	2,35	2,47	2,36	3,06	1,26	3,06
-0,47	-0,24	-0,43	-0,68	0,46	-0,68	2,35	2,37	2,29	3,37	1,29	3,37
-0,59	-0,36	-0,71	-0,44	0,66	-0,71	2,47	2,49	2,57	3,13	1,09	3,13
-0,82	-0,51	-0,78	-0,48	0,84	-0,82	2,70	2,64	2,64	3,17	0,91	3,17
-1,12	-0,49	-0,93	-0,31	0,98	-1,12	3,00	2,62	2,79	3,00	0,77	3,00
-1,29	-0,54	-0,93	-0,42	1,30	-1,29	3,17	2,67	2,79	3,11	0,45	3,17
-1,47	-1,56	-0,93	-0,71	1,38	-1,56	3,35	3,69	2,79	3,40	0,37	3,69
-2,00	-2,53	-2,50	-1,66	1,75	-2,53	3,88	4,66	4,36	4,35	0	4,66

что каждый содержащийся в массиве объект имеет только одну работу и необходимо оценить важность признака, используя частотную характеристику. Тогда число 30 оказывается вполне надежным при статистических исследованиях, если оно служит в качестве числа наблюдений [12]. Это положение отвечает нашим условиям. Объекты же, содержащие более 30 признаков, составляют массив, в котором не набирается 30 единичных наблюдений. Кроме того, потери информации оказались сравнительно большими. Например, для выборки 30 и более признаков при сопоставлении частот всего набора признаков коэффициент корреляции Спирмена \hat{r}_1 составил 0,724. Несмотря на то что согласованность остается значимой, слишком большой скачок по сравнению с предыдущим значением не дает основания использовать данный массив в дальнейших исследованиях.

Таким образом, в добавление к параметрам, характеризующим потери информации, присоединяется параметр "представительность объекта исследования". Этот параметр, указанный в графе 9 табл. 5, выражен через среднее арифметическое числа работ в объектах, составляющих выборки.

О п е р а ц и я Е. Теперь, имея набор показателей, можно выбрать

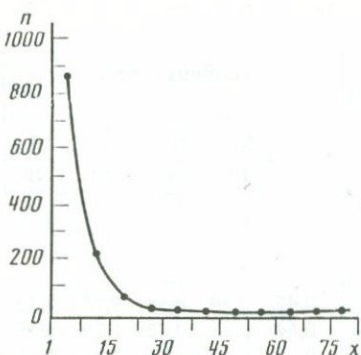


Рис. 14. Распределение количества признаков у авторов (у объекта)

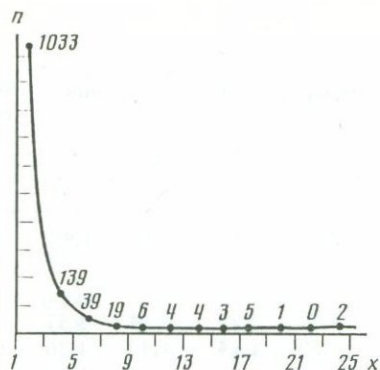


Рис. 15. Распределение количества работ у авторов (у объекта)

лучший массив: собранные в нем объекты являются наиболее представительными и содержат максимальное количество информации. Анализируя таблицу, легко увидеть, что при сокращении объема выборки растет среднее число работ, иными словами, увеличивается представительность объектов, но одновременно ухудшаются их информационные свойства.

Таким образом, перед нами возникла необходимость принять компромиссное решение. Для объективного выбора представительных объектов, несущих наиболее полную информацию, но сосредоточенных в массиве максимально сокращенного объема, воспользуемся методами из теории принятия решений, позволяющих делать оценку одновременно по нескольким параметрам (см. пятую главу).

Показатели r_1, r_2, r_3, r_4 и \bar{n} во взаимосвязи с исследуемыми выборками 1, 2, ..., 21 составляют матрицу оценки. Для того чтобы сравниваемые распределения одной переменной с другой были сопоставимы, преобразуем их в стандартную форму с помощью вычитания из каждого наблюдения соответствующего среднего арифметического и деления разности на стандартное отклонение [24]:

$$r_i = (x_i - \bar{x})/\sigma. \quad (14)$$

После этого преобразования новые переменные имеют нулевое среднее значение и единичную дисперсию. Соответствующие стандартизированные значения приведены в табл. 5.

Применяя максиминный критерий, определим лучшую выборку, ориентируясь на худшие условия. Сначала по каждой выборке найдем наихудший, характеризующий ее показатель (графа 15), затем среди них отыщем наилучшую выборку. Такой оказалась выборка 9. Эта выборка содержит объекты с 14 и более признаками.

Применяя минимаксный критерий, получим наилучшую выборку, включающую объекты с 11 и более признаками (графа 21 табл. 5).

Расхождения в оценках, полученных с помощью обоих критериев, незначительны, и мы вправе использовать любую из названных выборок. Однако для дальнейших исследований выбираем массив, оцененный в качестве наилучшего с помощью минимаксного критерия, поскольку

Таблица 6. Расчет оценок важности факторов, участвующих в проблеме происхождения нефти и газа

Номер фактора	1	2	3	4	5	6	7	8	9
Сумма частот	54,43	16,59	14,63	14,38	11,87	43,36	17,81	41,37	15,58
Ранг	1	6	8	9	12	2	5	3	7
Номер фактора	10	11	12	13	14				
Сумма частот	27,29	13,57	12,11	10,07	8,37				
Ранг	4	10	11	13	14				

сущность этого критерия в том, чтобы любыми путями избежать большого риска при принятии решения. Кроме того, оценка выборки по критерию Севиджа проведена строго по одному параметру, который оказался худшим, в то время как критерий Вальда указывает на два худших параметра. Таким образом, после применения операции E массив сократился до 196 объектов.

Выявление важнейших факторов. При декомпозиции проблемы происхождения нефти и газа были выявлены 14 факторов, так или иначе участвующих в описании образования и формирования нефтяных и газовых месторождений. При этом отмечалось, что при описании происхождения нефтяных углеводородов решающими являются факторы исходного материала и его преобразования.

Такое интуитивное предположение можно проверить с помощью статистических критериев, тем самым обосновав его объективность. Для этой цели использованы данные, содержащиеся в отобранной в качестве наилучшей для дальнейшего изучения выборке. В каждом объекте, в данном случае в каждом высказывании о гипотезе, принадлежащем автору, по формуле $p_i = n_i/N$, где p_i — мера важности i -го признака (фактора); n_i — число упоминаний об i -м признаке (факторе); N — количество работ у объекта, рассчитана частота встречаемости упоминания о факторе, характеризующая оценку его важности. Затем оценки просуммированы по всей выборке и получен ряд чисел, ранжировка которого позволила выявить наиболее важные факторы с позиций исследователей проблемы. Данные сведения указаны в табл. 6, из которой явствует, что выдвинутое предположение о лидирующей роли указанных факторов не опровергнуто.

В таблице номера факторов соответствуют номерам, указанным на схеме декомпозиции проблемы происхождения нефти и газа (рис. 9).

Полученные выводы были перепроверены более сложными статистическими методами, в частности корреляционным анализом, с помощью которого изучают характер связи и зависимости в многомерных совокупностях.

Научная основа теории корреляции и корреляционного анализа изложена в многочисленной математической литературе и здесь не рассматривается. Напомним только, что если над двумерной случайной величиной проведены n наблюдений $x_1, y_1; x_2, y_2; \dots; x_n, y_n$, то оценкой коэффициента кор-

Таблица 7. Оценки парных коэффициентов корреляции r между факторами, описывающими проблему происхождения нефти и газа (праводиагональная часть), и трансформированные оценки парных коэффициентов корреляции (леводиагональная часть $\alpha = \arccos r$)

	1	2	3	4	5	6	7
1		0,32	0,40	0,32	0,34	0,47	0,07
2	1,245		0,29	0,34	0,32	-0,01	0,09
3	1,154	1,276		0,33	0,44	0,13	-0,15
4	1,245	1,224	1,234		0,46	-0,06	-0,26
5	1,224	1,245	1,115	1,092		0,12	-0,17
6	1,081	1,580	1,440	1,510	1,450		0,10
7	1,500	1,480	1,720	1,832	1,740	1,470	
8	1,500	1,970	1,610	1,670	1,580	1,580	1,640
9	1,781	1,791	1,720	1,761	1,730	1,650	1,530
10	1,832	1,480	1,570	1,895	1,885	1,660	1,540
11	1,781	1,610	1,450	1,640	1,450	1,690	1,730
12	1,490	1,308	1,318	1,126	1,266	1,589	1,938
13	1,570	1,520	1,740	1,791	1,740	1,580	1,359
14	1,690	1,520	1,740	1,843	1,781	1,520	1,159

реляции r является величина \hat{r} , определяемая по формуле

$$\hat{r} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y},$$

где \bar{x} и \bar{y} — средние арифметические; n — число парных измерений x_i и y_i , S_x и S_y — оценки стандартных отклонений для x_i и y_i в выборке.

Оценку коэффициента корреляции можно вычислить другим, более удобным способом, который дает значительную экономию во времени при использовании настольных вычислительных машин:

$$\hat{r} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i)}{\sqrt{\left[\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2 \right] \left[\sum_{i=1}^n y_i^2 - \frac{1}{n} (\sum_{i=1}^n y_i)^2 \right]}}.$$

По выборочному значению коэффициента корреляции \hat{r} необходимо проверить гипотезу $H_0 : r = 0$ при альтернативе $H_1 : r \neq 0$. Если будет принята нулевая гипотеза, то делать вывод о наличии корреляционной зависимости нельзя, если же H_0 будет отвергнута и принята H_1 , то следует, что рассматриваемая пара случайных величин связана корреляционной зависимостью.

В качестве статистического критерия для проверки гипотезы $H_0 : r = 0$ обычно используется величина

$$t = \frac{|\hat{r}|}{\sqrt{1 - \hat{r}^2}} \sqrt{n-2},$$

где n — число наблюдений.

8	9	10	11	12	13	14
0,07	-0,21	-0,26	-0,21	0,08	0,00	-0,12
-0,39	-0,22	0,09	-0,04	0,26	0,05	0,05
-0,04	-0,15	0,00	0,12	0,25	-0,17	-0,27
-0,10	-0,19	-0,32	-0,03	-0,43	-0,22	-0,13
-0,01	-0,16	-0,31	0,12	0,30	-0,17	-0,21
-0,01	-0,08	-0,09	-0,12	-0,02	-0,01	0,05
-0,07	0,04	0,03	-0,16	-0,36	0,21	0,40
	-0,05	-0,25	-0,07	-0,08	-0,10	-0,26
1,620		0,08	0,29	0,21	-0,18	-0,16
1,822	1,490		0,09	-0,21	0,01	0,03
1,640	1,276	1,480		0,49	-0,14	0,00
1,650	1,359	1,781	1,058		-0,22	-0,18
1,670	1,750	1,560	1,710	1,791		0,29
1,832	1,730	1,540	1,540	1,750	1,276	

Эта величина в условиях нулевой гипотезы распределена по закону Стьюдента с $n-2$ степенями свободы. Таким образом, H_0 отвергается (т.е. зависимость считается установленной), если t превысит допустимо значение $t_{q, n-2}$ при уровне значимости q и $n-2$ степенях свободы.

Опираясь на значения оценок важности факторов, определенных для каждого объекта, построим корреляционную матрицу, при интерпретации которой следует руководствоваться принципом: сходство (различие) изучаемых факторов проявляется во взаимосвязи их значений.

Найденные коэффициенты корреляции, образующие матрицу R (табл. 7) несут основную информацию о структуре изучаемых факторов. Однако закономерности в такой матрице находятся все еще в скрытом состоянии. Для того чтобы скрытая в ней закономерность стала заметной, используем методы исследования структуры корреляционной матрицы, которые правды, в отличие от методов коэффициентов связи далеко не всегда: полностью формализованы; здесь нередки эвристические решения.

Один из эффективных методов прямого анализа корреляционной матрицы предложил Л.К. Выхану [15]. В основу метода положен принцип "максимального корреляционного пути", реализация которого определенным образом упорядочивает элементы системы. Результаты анализа изображают в виде так называемого взвешенного графа, т.е. графа с приспаянными длинами ребер. Метод опирается на весьма простой алгоритм, который описывается следующим образом [69].

В матрице R отыскивают максимальный коэффициент корреляции r_{kl} ($k \neq l$). Соответствующие ему элементы $a^{(k)}$ и $a^{(l)}$ образуют начало графа. Ребру, соединяющему вершины $a^{(k)}$ и $a^{(l)}$, может быть приспаянна длина, равная $1 - r_{kl}$. Зафиксировав положение элементов $a^{(k)}$ и $a^{(l)}$ вычеркивают в таблице связи коэффициент r_{kl} . В прореженных таким образом столбцах k и l вновь ищут максимальный коэффициент корреляции. Допустим, что таким коэффициентом является r_{jk} , при этом

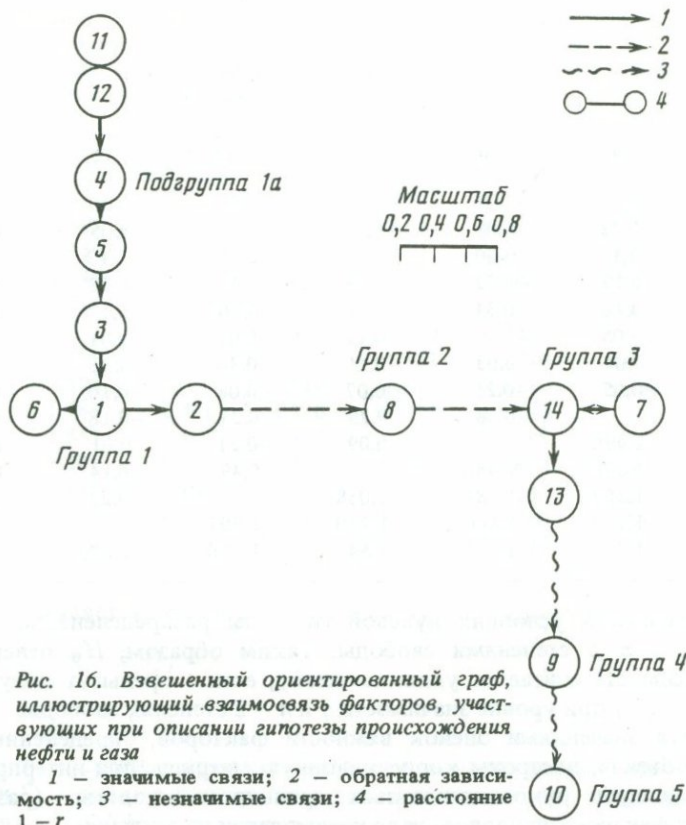


Рис. 16. Взвешенный ориентированный граф, иллюстрирующий взаимосвязь факторов, участвующих при описании гипотезы происхождения нефти и газа

1 — значимые связи; 2 — обратная зависимость; 3 — незначимые связи; 4 — расстояние $1 - r$

$f \neq k, r_{fk} \leq r_{lk}$. Элемент $a^{(f)}$ вводится в граф в виде новой вершины, соединяющейся с вершиной $a^{(k)}$ ребром длиной $1 - r_{fk}$. Удалив из матрицы коэффициент r_{fk} , а также столбец l (либо столбец k), переходят к поиску максимальных коэффициентов в столбцах f и k (либо l).

Процедура повторяется до полного исчерпания матрицы R . Затем, выбрав пороговое значение r_0 , делят полученный взвешенный граф на подграфы. Описанный алгоритм обеспечивает разбиение множества A на непересекающиеся группы.

На рис. 16 в виде ориентированного взвешенного графа, построенного с помощью метода "максимального корреляционного пути", показаны результаты исследования структуры матрицы (табл. 7). На графе отчетливо прослеживаются пять групп. Наблюдаемая картина свидетельствует об определенной их разграниченности и в то же время подчеркивает связь между элементами, относящимися к одной и той же группе. Выявленные ассоциации изучаемых факторов вполне согласуются с представлениями, полученными на интуитивной основе при классификации проблемы.

Действительно, в одну группу попали все факторы, ранее выделенные априори в качестве "прямых", характеризующих проблему генезиса нефти и газа. Причем среди них одна из наиболее тесных связей проявляется у ин-

тересующих нас факторов 1 и 6. Правда, статистические данные обнаруживают тесную связь их с двумя факторами, как бы посторонними — не принадлежащими к разряду "прямых". Это фактор 11, объединяющий природные явления и физико-химические условия, обеспечивающие миграцию углеводородов, и фактор 12, характеризующий форму их миграции.

Такой факт позволяет скорректировать прежние представления о классификации и выделить самостоятельную "вспомогательную часть" факторов. Дело в том, что связь этих двух факторов с "исходным материалом" проявляется не непосредственно, а через другие факторы 3—5. Но последние при описании проблемы используются не всеми авторами, а только теми, кто защищает органические гипотезы происхождения нефтяных углеводородов. Эти же исследователи прибегают и к помощи факторов 11 и 12. Целевая направленность названных факторов (3—5, 11 и 12) позволяет выделить их в самостоятельный комплекс.

Отметим, что описанная группа, кроме перечисленных факторов, включает также фактор 2 — "количественная оценка исходного материала для нефти и газа".

Названные три фактора связаны с остальными обратной зависимостью. Это фактически подтверждает, что для описания проблемы происхождения нефти и газа выделенные факторы используются неравномерно: либо усилия авторов сосредоточиваются на "прямых" факторах, либо они описывают только миграционные условия углеводородов, либо уделяют внимание только лишь закономерностям различного их состава.

Таким образом, основная цель проведенного анализа достигнута. Показано, что предыдущие представления о решающих факторах в проблеме происхождения нефти и газа не опровергнуты. Тесная связь выявлена также с фактором, определяющим достаточность либо нехватку исходных компонентов для образования крупных скоплений углеводородов.

С той же целью применен еще один не менее простой и эффективный метод, дающий возможность представить результаты анализа корреляционной матрицы в наглядном виде, метод называется "иерархический агрегативный" [40]. Характерной чертой его является иерархическая структура результатов классификации. Графические иерархические группировки удобно представлять в виде дендрограммы, т. е. связанного графа, не содержащего циклов. В общем, иерархическое группирование осуществляется следующим образом [69].

Предположим, что имеется P классов объемом не более одного элемента $A_j \subset A; j = 1, 2, \dots, m; m = p$.

Рассмотрим процедуру, которая будет повторяться неоднократно (а точнее $p-1$ раз), заключающуюся в объединении групп (классов), полученных на предыдущем шаге. Нарастивание групп производится на основе специальных мер сходства (близости), вычисляемых в общем случае заново на каждом шаге. Две наиболее сходные группы объединяются.

Этот процесс может быть продолжен до получения единой группы, включающей все элементы системы

$$A_j = A; j = m; m = 1.$$

Очевидно, с содержательных позиций бессмысленно доводить процедуру объединения до крайне малых значений числа групп m , так как в этом

случае выделяемые подмножества оказываются слишком "рыхлыми" из-за включения в их состав слабо связанных элементов. Поэтому разумнее остановиться на некотором промежуточном этапе. Правда, момент остановки с чисто формальных позиций выбрать весьма затруднительно [69].

Как правило, классифицирование прекращается в момент получения удовлетворительного (или противоположного) результата — это зависит в основном от опыта и интуиции исследователя.

Основным моментом описываемого метода является выбор меры междугруппового сходства. Этой мерой он собственно и отличается от многочисленной группы методов, объединенных под общим названием методы "кластер-анализа".

Если рассматривать иерархическую классификацию как определенное упорядочение элементов системы в некотором метрическом пространстве, то, вероятно, нецелесообразно использовать в качестве критерия объединения коэффициенты связи r_{kl} , поскольку такие не обладают свойствами дистанционных параметров. Поэтому задача иерархического группирования обычно решается с помощью таких критериев, которые можно выразить в метрической форме. Дистанционный коэффициент подобного рода получают путем несложного преобразования мер связи r_{kl} , составляющих матрицу R :

$$\alpha = \arccos r_{kl}. \quad (15)$$

Мера близости (сходства) между элементами или группами элементов имеет свойства метрики и удовлетворяет условиям

$$\begin{aligned} \alpha(a^{(k)}, a^{(k)}) &= 0, \quad \alpha(a^{(k)}, a^{(l)}) \neq 0; \\ \alpha(a^{(k)}, a^{(l)}) + \alpha(a^{(l)}, a^{(j)}) &\geq \alpha(a^{(k)}, a^{(j)}). \end{aligned}$$

В результате преобразования (15) корреляционной матрицы R она трансформируется в матрицу D . Анализ матрицы D удобнее начать с выбора минимального значения α , соответствующего паре сходных (т. е. наиболее тесно связанных, если выражаться в терминах корреляционного анализа) элементов системы. В результате получим первую группу A_1 , отвечающую первому иерархическому уровню:

$$\min \alpha^{(1)} = \arccos r_{kl}; \quad a^{(k)}, a^{(l)} \in A_1; \quad p_1 = 2.$$

Следующий этап — формирование матрицы D' размерностью $(p-1) \times (p-1)$, в которой по сравнению с матрицей, полученной на предыдущем шаге, изменяются лишь значения α , связанные с членами группы A_1 . Так, мера сходства между группой $A_1 = \{a^{(k)}, a^{(l)}\}$ и элементом $a^{(j)}$ ($a^{(j)} \in A_1$) вычисляется следующим образом:

$$d'(A_1, a^{(j)}) = [d(a^{(k)}, a^{(l)}) + d(a^{(k)}, a^{(j)}) + d(a^{(l)}, a^{(j)})] / (p_1 + 1).$$

В матрице D' вновь отыскиваем минимальное значение, в результате чего формируется группа A_2 (при этом $\min d < \min d'$). Процедуру продолжают до тех пор, пока полностью не исчерпываются элементы системы. Отметим, что каждый шаг этой процедуры приводит либо к наращиванию групп, полученных ранее, либо формирует новую группу, состоящую из элементов, не участвующих в построении классов.

В общем виде формулы вычисления меры близости между группами

A_i и A_j , слагающихся из p_i и p_j элементов, имеют следующий вид:

$$d(A_i, A_j) = \frac{\{p_i(p_i - 1)/2 \bar{d}(A_i) + [p_i(p_j - 1)/2 \bar{d}(A_j) + p_i p_j h(A_i A_j)]\}}{[p_i(p_i - 1)/2 + p_j(p_j - 1)/2 + p_i p_j]} \quad (16)$$

В этой формуле символами $\bar{d}(A_i)$ и $\bar{d}(A_j)$ обозначены внутригрупповые усредненные меры, т. е. усредненные внутригрупповые расстояния между парами элементов, входящих в группу A_i или A_j :

$$\bar{d}(A_i) = \frac{\sum_{a^{(k)}, a^{(l)} \in A_i} d(a^{(k)}, a^{(l)})}{[p_i(p_i - 1)/2]},$$

$$\bar{d}(A_j) = \frac{\sum_{a^{(f)}, a^{(h)} \in A_j} d(a^{(f)}, a^{(h)})}{[p_j(p_j - 1)/2]}, \quad k < l, f < h.$$

Величины $\bar{d}(A_i)$, $\bar{d}(A_j)$ отражают однородность внутри групп и являются, по сути дела, мерой их компактности. По мнению специалистов, подобная усредненная групповая мера хороша тем, что на ее величину не слишком сильно влияют элементы, вводимые на поздних стадиях классификации [69].

Третье слагаемое в числителе формулы (16), обозначенное символом h , представляет собой усредненное межгрупповое расстояние

$$h(A_i, A_j) = \frac{1}{p_i p_j} \sum_{a^{(k)} \in A_i} \sum_{a^{(f)} \in A_j} d(a^{(k)}, a^{(f)}).$$

Чем больше величина h , тем "дальше" друг от друга находятся сопоставляемые группы, тем выше степень различия между ними. Заметим, что для групп, состоящих из одного элемента

$$A_i = a^{(k)}, \quad A_j = a^{(l)},$$

справедливо равенство

$$d(A_i, A_j) = h(A_i, A_j).$$

Результаты группирования изображают графически в виде дендрограмм. На одну из осей наносят символические обозначения элементов системы, другая ось служит для масштабного представления иерархических уровней (d' , d'' , d''' и т. д.). Содержательная ценность каждого такого древовидного графа может быть существенно повышена, если на нем дополнительно отразить информацию о межгрупповых расстояниях (h' , h'' , h''' и т. д.).

На рис. 17 изображена иерархическая структура матрицы сходства (см. леводиагональную часть табл. 7), полученной путем соответствующего преобразования корреляционной матрицы R (см. ту же таблицу — праводиагональная часть). Рассчитанные меры сходства $d = \arccos r$ между отдельными факторами (и их группами) и соответствующие этим мерам сходства межгрупповые расстояния h сведены в табл. 8, которая явилась основанием для построения дендрограммы на рис. 17.

При ее построении учитывались не только меры близости d , но и межгрупповые расстояния h , что нашло свое отражение в степени сближенности или удаленности групп, а также отдельных элементов в этих группах.

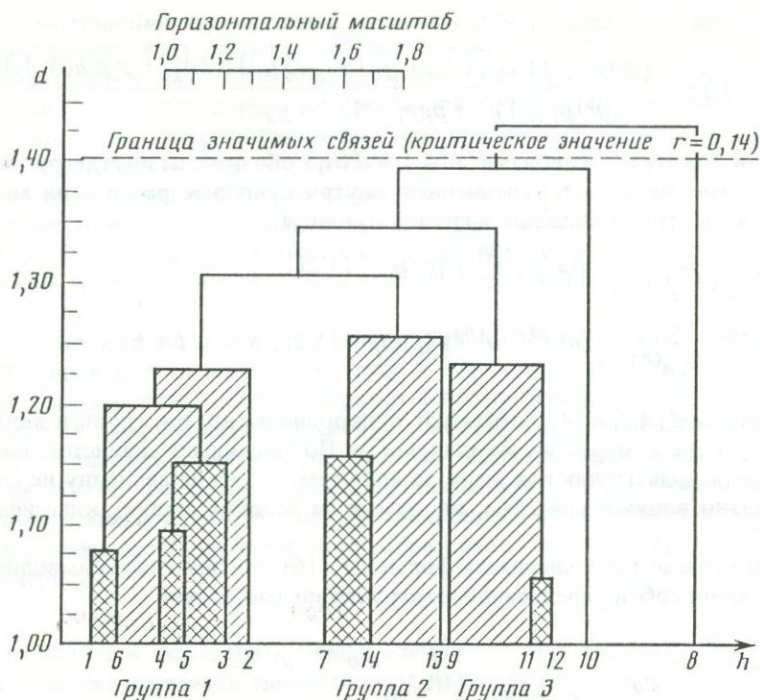


Рис. 17. Дендрограмма, иллюстрирующая взаимосвязь факторов, участвующих при описании гипотезы происхождения нефти и газа

На дендрограмме проявляется более дифференцированная картина, чем при анализе корреляционной матрицы методом "максимального корреляционного пути". На ней четче выделяются группы. Вновь обособленно стоят факторы 1 и 6. К этой же группе тяготеет и фактор 2. Факторы же 3–5, 11 и 12, выделенные в предыдущем анализе в самостоятельную подгруппу, здесь разделились. В целом деление, наблюдаемое на дендрограмме, ближе к интуитивным представлениям о характере и взаимосвязи изучаемых факторов.

Подводя итог, отметим, что, сопоставляя результаты, полученные с помощью обоих методов, убеждаемся в их вполне удовлетворительном совпадении. Следовательно, выявленные ассоциации действительно имеют место. Выделение же групп, подтвержденных обоими методами, можно считать надежным.

Объединение дублирующих объектов. При дальнейшей обработке материала сосредоточим внимание только на трех главных факторах, непосредственно ответственных за образование нефти и газа. Как выяснилось в предыдущем разделе, к ним относятся:

- исходный материал для нефти и газа (1);
- преобразование исходного материала (6);
- количественная оценка исходного материала (2).

Сформируем новую выборку. Для этого, используя всю совокупность,

Таблица 8. Меры компактности (d), межгрупповые расстояния (h) и усредненные коэффициенты корреляции ($\tilde{r} = \cos d$). Таблица составлена на основании обработки трансформированной матрицы D (леводиагональная часть табл. 7)

Номер шага	Состав группы	d	h	\tilde{r}
1	11 + 12	1,058	1,06	0,49
2	1 + 6	1,081	1,08	0,47
3	4 + 5	1,092	1,09	0,46
4	(4,5) + 3	1,147	1,17	0,41
5	7 + 14	1,159	1,16	0,40
6	(3, 4, 5) + (1, 6)	1,194	1,34	0,37
7	(1, 3, 4, 5, 6) + 2	1,230	1,31	0,33
8	(11, 12) + 9	1,231	1,32	0,33
9	(7, 14) + 13	1,265	1,32	0,30
10	(1, 2, 3, 4, 5, 6) + (7, 13, 14)	1,313	1,65	0,25
11	(1, 2, 3, 4, 5, 6, 7, 13, 14) + (9, 11, 12)	1,337	1,61	0,23
12	(1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14) + 10	1,387	1,64	0,18
13	Все факторы + 8	1,433	1,68	0,14

которая названа "генеральной", выберем только такие объекты, в которых обязательно присутствуют факторы 1 и 6. При этом достаточно, чтобы каждый из факторов был представлен хотя бы одним признаком.

Фактор 2 из-за недостаточности информации в каждом объекте рассчитывается искусственным путем. Его величина, как будет показано в пятой главе, зависит от характера исходного материала в конкретном объекте.

Среди отобранных объектов в выборке оказались высказывания, совершенно не отличающиеся по набору признаков. Сюда не относятся объекты, которые ранее уже отбракованы как некачественные или несущие дублирующую информацию, имеются в виду высказывания соавторов, выступающих постоянным коллективом и встреченных в картотеке только один раз (см. операцию Γ в разделе "Сокращение числа объектов").

Объединим одинаковые объекты, что позволит сократить размер выборки. В качестве исходных данных для процедуры группирования можно использовать информацию о близости объектов изучения, которая задается одной из мер сходства. Достаточно полный перечень мер сходства можно найти в работах [11, 18, 26], в которых рекомендуются три типа мер: 1) коэффициенты подобия (или квантифицированные коэффициенты связи); 2) коэффициенты связи (корреляции); 3) показатели расстояния в метрическом пространстве. Выбор того или иного типа мер в значительной степени зависит от характера исходной информации. В наших условиях, когда информация представлена булевыми переменными, наиболее подходящей мерой являются коэффициенты подобия.

Раскроем характер информации, обрабатываемой на данном этапе исследования. Пусть множество объектов — носителей информации, подлежащих группированию $u = \{u_j\}$, конечно с объемом n . Каждый объект $u_j \in u$ обладает числом степеней свободы m , равным числу признаков, с помощью которых его можно описать. Таким образом, можно считать, что любой

объект u_j описан с помощью многомерной случайной величины (или многомерного признака) $x_j = (x^{(1)}, x^{(2)}, \dots, x^{(m)})_j$. При этом под $x_j^{(k)}$ понимается результат измерения k -го признака на j -м объекте. Каждый признак (переменная) может принимать значения 1 в случае присутствия его в объекте или 0 в случае его отсутствия, следовательно, признак $x^{(k)}$ булевый*.

Известно много разнообразных измерителей подобия. Сравнение их показывает, что разные коэффициенты подобия, рассчитанные для одних и тех же объектов u_i и u_j будут неодинаковы по величине [26]. Выбор того или иного коэффициента является во многом субъективным.

В настоящей работе использован коэффициент Роджерса–Танимото, так как он наименее трудоемкий при вычислении.

Коэффициент рассчитывается по формуле

$$S_{ij} = n_{ij} / (n_i + n_j - n_{ij}),$$

где n_{ij} — число совпадающих единичных признаков у обоих объектов; n_i, n_j — число единичных признаков i -го и j -го объектов соответственно.

С помощью этого коэффициента проведено сравнение каждой пары объектов. Теперь любой объект $u_j \in U$ можно описать вектором $x = (x^{(1)}, x^{(2)}, \dots, x^{(m)})$, каждая из m компонент которого принимает значение коэффициента подобия (меры сходства) между сравниваемыми объектами.

В результате расчетов составлена матрица, по строкам которой расположены объекты изучения, а по столбцам — соответствующие им коэффициенты подобия. В таком виде материал готов для выявления однородных объектов.

Для группирования выбран метод "разграничения объектов по комплексу признаков", детально рассмотренный в монографиях [65, 66]. Апробация этого метода [60, 75 и др.] показала его высокую "работоспособность" в условиях, когда:

- заранее неизвестны ни число однородных групп, ни их границы;
- отсутствуют эталонные выборки;
- отсутствуют четкие правила классификации;
- среди однородных групп возможно присутствие групп, в которых число наблюдений меньше числа оцениваемых параметров.

Метод основан на использовании статистики V , распределенной (при условии нулевой гипотезы — гипотезы об однородности рассматриваемой совокупности) по закону Пирсона и представляющей собой взвешенное евклидово расстояние. Вычисление критерия производится по формуле

$$V = \sum_{j=1}^m V_{kj} = \sum_{j=1}^m \frac{[\bar{x}_j^{(1)} - \bar{x}_j^{(2)}]^2 n_1 n_2}{(S_j^{(1,2)})^2 (n_1 + n_2)}, \quad (17)$$

где $\bar{x}_j^{(1)}$ и $\bar{x}_j^{(2)}$ — оценки средних арифметических j -й составляющей в сравниваемых группах; n_1, n_2 — число наблюдений в этих группах, S_j^2 — мера изменчивости j -й составляющей в пределах всей выборки.

* На этом этапе работы оценка важности признака во внимание не принимается.

Статистика V при нулевой гипотезе имеет распределение χ^2 с f степенями свободы, где f — число изучаемых составляющих (признаков). В случае когда максимальное значение не превышает значения χ^2 , соответствующего данному числу степеней свободы f и принятому уровню значимости q , гипотеза об однородности как непротиворечащая эмпирическим данным принимается. В противном случае максимальное значение критерия и определяет границу.

В итоге последовательного разграничения исследуемая совокупность объектов расчленяется на однородные группы, но их число будет превышать число реально существующих. Это связано с тем, что мы имеем дело не с истинными границами, а с их приближенными статистическими характеристиками. Заключительным этапом работы по определению границ (выявлению однородных групп) является устранение ложных разграничений.

Для всевозможных пар групп рассчитывается статистика V по формуле (17), где n_1, n_2 — число объектов каждой группы, входящей в пару. Если

$$\min V \leq \chi_{q, f}^2,$$

где $\chi_{q, f}^2$ — табличное значение, то пара, соответствующая этой статистике, признается однородной. Процедура повторяется до тех пор, пока либо на каком-то шаге все границы признаются значимыми, т.е. имеем

$$\min V > \chi_{q, f}^2,$$

что соответствует значимым различиям в совокупности свойств объектов, объединенных в разрозненные группы, либо на $k - 1$ -м шаге все группы объединяются в одну. Последнее соответствует случаю полной однородности исследуемой совокупности.

С целью проверки достоверности результатов разграничения по описанному критерию была использована другая статистика [65], которая описывается формулой

$$T = (V - m) / \sqrt{2m},$$

где V — критерий, рассчитанный по формуле (17); m — число признаков, на основе которых проводится разграничение. Статистика T в условиях нулевой гипотезы также распределена по закону Пирсона. Поэтому все сказанное о принятии и соответственно отбрасывании нулевой гипотезы для V -критерия здесь также имеет место.

Сопоставление результатов группирования по обоим критериям показало их хорошую согласованность. Однако некоторые объекты, объединившиеся по одному критерию, по другому оказались в разных группах. Чтобы устранить этот недостаток, была проведена корректировка результатов разграничения. В основание корректировки было положено ограничение, заключающееся в том, что в качестве достоверных принимались только те группы, которые одинаково выделились по обоим критериям. Остальные объекты, оказавшиеся неразличимыми по данным критериям, оставлены необъединенными и далее рассматривались как самостоятельные наряду с групповыми.

Благодаря процедуре объединения значительно уменьшился объем выборки и улучшилось ее качество. Последнее обстоятельство выразилось

в том, что при группировании одинаковых объектов увеличилась их надежность (возросло число наблюдений над признаками), а это сказалось на оценке важности признаков, которыми оперируют исследователи для выражения своей точки зрения о происхождении нефтяных углеводородов. В то же время количество информации в групповом высказывании не изменилось, поскольку объединились только те объекты, которые не различаются ни одним признаком.

Предлагаемый подход к группированию высказываний выгодно отличается от традиционного, применяемого при групповой экспертизе, когда в группу собирают единомышленников (в широком смысле этого слова), которых объединяют какие-то общие понятия о проблеме, что зачастую засоряет информацию, подлежащую изучению, затушевывая главные стороны проблемы.

В результате применения описанного алгоритма группирования получено 289 (вместо 594, содержащих факторы 1 и 6) новых объектов, среди которых необходимо выбрать наилучшие.

Глава 5

ПОИСК ОБЪЕКТИВНЫХ ВЫСКАЗЫВАНИЙ

Казалось бы, целесообразно выбрать из числа оставшихся высказываний (объектов изучения) только такие, которые охарактеризованы высшим баллом, и считать их наиболее приближенными к истине, поскольку они подтверждены экспериментом. Однако такой подход может привести к существенным ошибкам.

Во-первых. Не все высказывания о гипотезе сегодня можно проверить экспериментальным путем. Среди них встречаются гипотезы, подтверждаемые теоретическими расчетами, а также чисто умозрительные заключения, никак не отмеченные другими исследователями, и до экспериментальной проверки они должны считаться истинными (см. четвертую главу); из-за введенной балльной оценки обоснованности такие высказывания окажутся в последних рядах, не будут учтены и, таким образом, потеряется информация, которая может привести к истине.

Во-вторых. Среди группы наиболее обоснованных высказываний лидирующее положение может занять мало представительное, приведенное лишь только в одной работе, в то время как высказывание, основанное на большом количестве работ, но незначительно меньше обоснованное (например, из-за неравномерного употребления балльной оценки по каким-то техническим причинам), выпадет из рассмотрения. Практически это означает, что количество наблюдений над каждым признаком в группе должно быть достаточно многочисленным, а эксперимент должен иметь необходимую повторяемость, чтобы в статистических данных могло проявиться действие закона больших чисел. Отдельные факты наукой, как правило, в расчет не принимаются. "... Доверия заслуживает по преимуществу то, что добыто во многих наблюдениях и опытах, — писал Д.И. Менделеев. — Единичные опыты немного говорят для науки, и всеисильны для предубеждения, пото-

му что ни в одном опыте нельзя устранить всех влияний, а предубежденному глазу и видно только одно желаемое влияние" [46].

Во избежание возможных ошибок при выборе наиболее обоснованных высказываний использованы критерии из теории принятия решений, позволяющие объективно (количественно) оценивать суждение одновременно по нескольким параметрам.

1. ТИПИЧНЫЕ МЕТОДЫ И КРИТЕРИИ ПРИНЯТИЯ РЕШЕНИЙ

При записи целевой функции W (см. формулу (2)) удобнее пользоваться векторной формой

$$W = f(a, x), \quad (18)$$

где $x = x(x_1, x_2, \dots, x_l)$ — управляемый вектор, т.е. совокупность управляемых переменных; $a = a(a_1, a_2, \dots, a_q)$ — неконтролируемый вектор. Если значение неконтролируемого вектора известно, то задача сводится к отысканию вектора x , удовлетворяющего ограничениям (3) и обращающего целевую функцию W в максимум. Для определенности в дальнейшем будем считать, что мы всегда стремимся к достижению максимума (18); случай минимума сводится к предыдущему введением новой целевой функции

$$W' = -f(a, x).$$

Возникающие при выборе решений ситуации удобнее представить в виде таблицы, называемой матрицей оценки (платежной матрицей) (табл. 9).

В этой таблице каждый столбец определяет возможный результат a_i — значения, которые может принимать неконтролируемый фактор (с вероятностями p_i); каждая строка определяет возможную стратегию $A(x_j)$, заключающуюся в выборе значения x_j управляемого вектора; $W(a_i x_j)$ — соответствующее значение показателя эффективности.

Напомним, что ситуации определенности, риска и неопределенности различаются между собой лишь степенью знания о распределении вероятности, которую может принимать фактор или событие, неконтролируемые со стороны принимающего решение. Разберем каждую ситуацию в отдельности с точки зрения существа вопроса и различных критериев и методов принятия решения (выбора определенной стратегии $x = x_j$).

В простейшем, детерминированном случае, когда вектор a имеет фиксированное значение a_{i_0} , получаем, очевидно, следующее распределение вероятностей: $p_{i_0} = 1, p_i = 0 (i \neq i_0)$, и задача выбора оптимального решения сводится к определению наибольшего значения из совокупности

$$W(x_j, a_{i_0}), \quad j = 1, 2, \dots, m.$$

На практике, как правило, нет необходимости перебора всех стратегий с целью выбора наилучшей, а решение дается алгоритмом вычисления значения целевой функции

$$W^0 = \max_x W(x, a_{i_0}).$$

Именно таково большинство моделей принятия решений [1, 13, 39, 44].

Таблица 9. Матрица оценки

Стратегия	P_1	P_2	...	P_n
	a_1	a_2	...	a_n
$A(x_1)$	$W(a_1 x_1)$	$W(a_2 x_1)$...	$W(a_n x_1)$
$A(x_2)$	$W(a_1 x_2)$	$W(a_2 x_2)$...	$W(a_n x_2)$
...
$A(x_m)$	$W(a_1 x_m)$	$W(a_2 x_m)$...	$W(a_n x_m)$

В условиях риска, когда a представляет собой случайный вектор с известным законом распределения (в дискретном случае все вероятности p_i известны и отличны от нуля), в принципе возможны следующие три подхода при выборе оптимального решения.

1. Осреднение по случайностям, т.е. замена неконтролируемого вектора a его средним значением (математическим ожиданием):

$$a_{cp} = \sum_{i=1}^n a_i p_i.$$

Тогда принятие оптимального решения становится эквивалентным определению величины

$$W^0 = \max_x W(x, a_{cp}),$$

и мы приходим к уже рассмотренному детерминированному случаю.

Описанный подход к выбору оптимального решения в условиях риска эффективен лишь тогда, когда имеет место приближенное равенство

$$W(x, a) \approx W(x, a_{cp}),$$

т.е. использование математического ожидания в качестве критерия для оценки эффективности стратегии правомерно при условии, что имеется достаточная информация о прошлом для того, чтобы можно было рассчитать вероятности возможных исходов решения в будущем, или, как их иногда называют, ожидаемые значения эффективности.

Таким образом, эффективность такого подхода к решению проблемы в большинстве случаев крайне мала, поскольку сумма взвешенных средних (математическое ожидание) вряд ли может служить критерием эффективности стратегий.

2. Применение максиминного критерия, который состоит в следующем: для каждой стратегии $A(x_j)$ находят величину

$$W'(x) = \min_a W(x, a),$$

т.е. наименьший (гарантированный) уровень "выгоды", а затем выбирают ту стратегию x_0 , для которой эта гарантированная выгода является наибольшей, т.е.

$$W^0 = W'(x_0) = \max_x \min_a W(x, a), \quad (19)$$

отсюда ясно название критерия "максимин".

3. Наиболее эффективным подходом при выборе решения в рискованных ситуациях является осреднение по случайностям самого показателя эффективности, когда $W(x, a)$ заменяется на критерий

$$W_{\text{ср}}(x) = \sum_{i=1}^n W(x, a_i) p_i,$$

и тем самым задача принятия решений опять сводится к детерминированному случаю.

Наиболее распространенными критериями выбора решения в ситуации неопределенности являются: максиминный, критерий Гурвица, минимаксного риска и недостаточного основания. Разберем каждый из них в отдельности.

Максиминный критерий (Вальда). Об этом критерии уже упоминалось при рассмотрении рискованных ситуаций. Отметим только, что выбор решения с помощью этого критерия в условиях неопределенности более оправдан, чем в ситуации риска, поскольку о законе распределения вероятностей неконтролируемых факторов действительно ничего неизвестно. Данный критерий ориентирует лицо, принимающее решение, на наихудшие условия и рекомендует выбирать то из них, для которого в худших условиях выигрыш максимален, т.е. "всегда надо рассчитывать на худшее". Максиминный критерий (Вальда) часто так и называют — критерий крайнего пессимизма [1, 13].

Критерий Гурвица (обобщенный максимин). Гурвиц [1, 13], желая сделать максиминный критерий более оптимистичным, видоизменил условие (19) следующим образом:

$$W^0 = \max_a [\gamma \max_a W(x, a) + (1 - \gamma) \min_a W(x, a)],$$

где $0 \leq \gamma \leq 1$. Величину γ часто называют показателем "оптимизма". При $\gamma = 0$ критерий Гурвица сводится к максимину, при $\gamma = 1$ критерий становится максимаксом, т.е. будет приводить к выбору решения, соответствующего наибольшему из возможных выигрышей. Большой недостаток критерия Гурвица — отсутствие критерия выбора собственно показателя оптимизма γ , точнее сказать, выбор его субъективен. Он зависит от темперамента исследователя и его оценки ситуации: чем опаснее ситуация, чем больше в ней хотелось бы подстраховаться, тем ближе к единице выбирается величина γ [13].

Критерий минимаксного риска (Севиджа). Выясним смысл этого критерия непосредственно на абстрактном примере [73]. Пусть схема принятия решения дается в табл. 10, где "выигрыш" (значения целевой функции) указаны в условных единицах.

Тогда, если бы принимающий решение знал, что неконтролируемый вектор примет значение $a = 0$, то он выбрал бы стратегию $A(x_2)$, а при выборе стратегии $A(x_1)$ он понес бы потери, равные $1(3-2=1)$. Аналогично, если бы заранее было известно, что неуправляемый вектор примет значение $a = 1$, то мы предпочли бы стратегию $A(x_1)$, а выбор стратегии $A(x_2)$ привел бы к потерям, равным $2(4-2=2)$. В итоге подобных вычислений можно вместо матрицы выигрышей (см. табл. 10) составить другую — матрицу потерь (табл. 11).

Таблица 10. Матрица выигрышей

Стратегия	p_1	p_2	p_3	p_4	p_5
	$a_1 = 0, p_1$	$a_2 = 1, p_2$	$a_3 = 2, p_3$	$a_4 = 3, p_4$	$a_5 = 4, p_5$
$A(x_1)$	2	4	3	3	2
$A(x_2)$	3	2	4	4	3

Таблица 11. Матрица потерь

Стратегия	$a_1 = 0, p_1$	$a_2 = 1, p_2$	$a_3 = 2, p_3$	$a_4 = 3, p_4$	$a_5 = 4, p_5$
	$A(x_1)$	1	0	1	1
$A(x_2)$	0	2	0	0	0

Применим к новой матрице минимаксный критерий для выбора решения, т.е. принимаем ту стратегию, которая приводит к наименьшим максимальным потерям. В рассмотренном примере наибольшие возможные потери для стратегии $A(x_1)$ составляют 1, для $A(x_2)$ — 2. Следовательно, стратегия $A(x_1)$ предпочтительнее.

Таким образом, сущность этого критерия в том, чтобы любыми путями избежать большого риска при принятии решения. Критерий Севиджа, как и критерий Вальда, — это критерий крайнего пессимизма, но только пессимизм здесь проявляется в другом: худшим считается не минимальный выигрыш, а максимальный риск — минимальная потеря выигрыша по сравнению с тем, чего можно было бы достичь в данных обстоятельствах [73].

Критерий недостаточного основания (Байеса—Лапласа). Это означает, что если нам совершенно неизвестно, какое из значений a_i принимает неконтролируемый вектор, то нужно поступать так, как будто все значения равновероятны, следовательно, задача принятия решения рассматривается, как в условиях риска с равномерным распределением вероятностей. Таким образом, оптимальным считается действие, которому соответствует

$$W = \max_x \frac{1}{n} \sum_{a=1}^n W(x, a).$$

Легко проверить, что для предыдущего примера по этому критерию стратегия $A(x_2)$ оказывается выгоднее, чем $A(x_1)$.

Отсюда возникает вопрос: какой из перечисленных критериев предпочтительнее? Тщательный анализ этих критериев приводит к выводу, что этот вопрос не обоснован, так как все они в той или иной мере субъективны и каким критерием воспользоваться в конкретной ситуации — это, так сказать, дело вкуса и степени оптимизма лица, принимающего решения [44, 73].

2. НАЗНАЧЕНИЕ (ВЫБОР) ПАРАМЕТРОВ ЭФФЕКТИВНОСТИ

Процедура определения параметров эффективности зависит от характера решаемой задачи. В примере с проблемой нефти и газа параметры эффективности назначены следующим образом.

1. Исходя из положения о том, что, чем важнее употребляемый признак (фактор), тем выше он оценивается автором, естественно принять его в качестве лучшего. Найдем для каждого объекта оценки важности отдел по факторам:

а) фактор 1 — исходный материал для нефти и газа, оценка определяется по формуле

$$\bar{x}_M = \frac{1}{n} \sum_{i=1}^m \bar{\epsilon}_i,$$

где $\bar{\epsilon}_i$ — оценка важности i -го признака у объекта, рассчитывается по формуле (13), n — число признаков у фактора;

б) фактор 6 — преобразование исходного материала

$$\bar{x}_\Pi = \frac{1}{n} \sum_{i=1}^m \bar{\epsilon}_i,$$

где ϵ_i и n означают то же, что и в п. а, но принадлежат фактору 6;

в) фактор 2 — количественная оценка исходного материала

$$\bar{x}_K = \frac{1}{n} \sum_{i=1}^m \bar{\epsilon}_i,$$

в формуле для оценки важности признака $\bar{\epsilon}_i$ использованы не собственные оценки изучаемого объекта, а коллективные, представляющие собой среднюю арифметическую, выявленную при анализе точек зрения всех изучаемых авторов, которые когда-либо упоминали о количестве (достаточности или недостаточности) того или иного исходного компонента для образования нефти и газа. "Коллективная" оценка употребляется из-за недостатка информации об этом факторе у единичных объектов (табл. 12).

Выпишем теперь параметры эффективности (критерии оценки) в том же порядке, как они рассмотрены:

$$W_M = f(\bar{x}_M) \rightarrow \max,$$

$$W_\Pi = f(\bar{x}_\Pi) \rightarrow \max,$$

$$W_K = f(\bar{x}_K) \rightarrow \max,$$

где \bar{x}_M , \bar{x}_Π , \bar{x}_K — оценки соответственно факторов "исходный материал. . .", "преобразование. . ." и "количественная оценка. . .".

2. При высказывании о факторе употребляются различные по обоснованности признаки. Одни из них могут быть оценены высшим баллом — в пять единиц, другие — только в три. Последнее происходит в тех случаях, когда невозможно к описанным явлениям применить экспериментальную проверку. Максимально возможные оценки для соответствующих признаков указаны в табл. 4 (каждая 3-я строка). Среднюю арифметическую максимально возможных оценок признаков назовем идеальной оценкой фактора обозначим \bar{x} . Естественно считать, что чем выше эта оценка, тем она лучше.

Таблица 12. Сводка "коллективных" оценок фактора 2 – количественная оценка исходного материала для нефти и газа

Номер признака	$\bar{\epsilon}_i$	Номер признака	$\bar{\epsilon}_i$	Номер признака	$\bar{\epsilon}_i$
1	0,35	8	0,57	15	0,60
2	0,59	9	0,06	16	0,53
3	0,51	10	0,59	17	0,48
4	0,49	11	0,29	18	0,56
5	0,38	12	0,38	19	0,29
6	0,31	13	0,87	20	-0,47
7	0,49	14	0,60	21	0,00

тем более обоснованные признаки участвуют в высказывании. Поэтому

$$W'_M = f(\bar{x}_M) \rightarrow \max,$$

$$W'_N = f(\bar{x}_N) \rightarrow \max,$$

где индексы при \bar{x} означают то же, что и в пункте 1.

3. О представительности высказывания, или его надежности, будем судить по количеству работ у объекта изучения (автора либо группы авторов). Тогда

$$W_n = f(n) \rightarrow \max,$$

где n – количество работ (статей, монографий и т.п.) у объекта.

4. Лучший объект поможет выявить и разность между идеальной и истинной оценками $\Delta x = \bar{x} - \bar{x}$. Понятно, что, чем меньше разница, тем ближе автор к истине. Следовательно:

$$W''_M = f(\Delta x_M) \rightarrow \min$$

$$W''_N = f(\Delta x_N) \rightarrow \min.$$

5. Равномерность обоснования высказывания характеризуется параметром $\Delta = \Delta x_M - \Delta x_N$. Если один фактор описывается со ссылкой на эксперимент, а другой – исходя только из теоретических соображений, то критерий уловит эти расхождения. Ясно, что лучшей оценкой следует считать минимальное значение Δ по абсолютной величине, т.е.

$$W_\Delta = f(|\Delta|) \rightarrow \min.$$

Таким образом, всего назначено 9 критериев эффективности. Рассчитанные их значения для каждого объекта составляют матрицу оценки. Затем значения параметров стандартизируются. Нормирование проведено с помощью z -оценки по формуле (14). При этом для удобства сравнения переменных целевые функции параметров, значения которых стремятся к минимуму, заменены новыми функциями с помощью преобразования:

$$W_c \rightarrow \min,$$

$$W_n = W_c \cdot (-1) \rightarrow \max,$$

где индексы c и n при критериях W обозначают соответственно новое и старое значения.

Для выбора наилучшей стратегии (в нашем случае наиболее обоснованного высказывания) использованы:

- максиминный критерий Вальда (W_B^*);
- критерий недостаточного основания Байеса—Лапласа ($W_{н.о.}^*$);
- минимаксный критерий Севиджа (W_C^*), применение которого потребовало построения специальной таблицы — матрицы потерь. Критерий Гурвица из-за недостаточной обоснованности показателя оптимизма γ не применялся.

Предварительный анализ результатов, полученных по различным критериям, показал, что если значения, рассчитанные по критериям Вальда (W_B^*) и недостаточного основания ($W_{н.о.}^*$), способствуют поставленной цели, т.е. дают возможность выявить лучшие объекты, то критерий Севиджа (W_C^*) не "сработал". Дело в том, что специфика последнего заключается в измерении возможных потерь при принятии решения. Среди же параметров эффективности, которые назначены для выбора лучших объектов один параметр W_n , характеризующий "представительность высказывания" оказывает слишком сильное влияние именно на этот критерий. Значительные вариации параметра W_n , несопоставимые с другими, как бы уравнивая оценки, определяемые по критерию Севиджа (W_C^*), что затрудняет поиск лучших объектов.

В связи с этим для выбора лучших объектов использованы не только оценки, учитывающие полный набор установленных параметров эффективности, но и оценки, рассчитанные по точно таким же критериям, одна учитывающие сокращенный набор параметров, из которого исключен параметр W_n . Такой прием, как будет показано ниже, приводит к удовлетворительному результату. А именно, используя весь комплекс оценок (конкретно 6 оценок вместо 3), удалось устранить доминирующее влияние параметра "представительность высказывания", однако отведенная ему роль в поиске лучших объектов сохранена. Теперь полученный результат не противоречит логике и здравому смыслу.

3. АЛГОРИТМ СОПОСТАВЛЕНИЯ ОЦЕНОК, ПОЛУЧЕННЫХ ПО РАЗНЫМ КРИТЕРИЯМ ОПТИМАЛЬНОСТИ

Результат расчета по названным критериям эффективности сопоставляется с общепринятым воззрением о несовместимости оценок, полученных по различным критериям оптимальности. Считают, что использованное каждым из них в какой-то мере субъективно, признавая при этом, что принятое решение на основе любого из них будет более обоснованным, чем волевой выбор. Однако логично допустить, что надежно определенными можно считать только те объекты, которые выделены в качестве лучших по нескольким критериям. Сопоставление результатов по этим критериям с большей вероятностью позволит объективно выбрать лучший объект, одновременно используя все разнообразие критериев, что в значительной мере оградит нас от принятия субъективных решений.

Дальнейшая работа с оценками, полученными по различным критериям заключалась в следующем. Каждый объект охарактеризован 6 независимыми оценками. Представим, что значение каждой оценки соответствует месту, которое занимает объект в ранжированном ряду. Примем, что

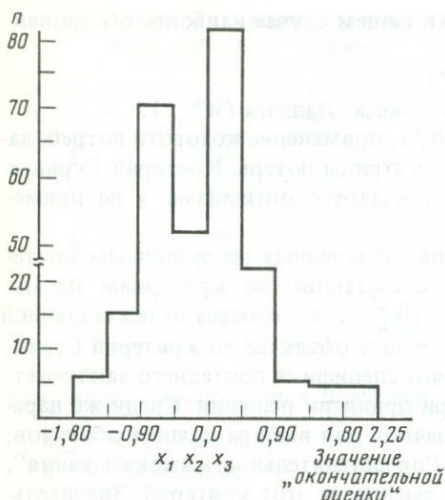


Рис. 18. Гистограмма, иллюстрирующая характер распределения "окончательной оценки" обоснованности высказывания о гипотезе происхождения нефти и газа

чем дальше (вправо или влево в зависимости от сущности рассматриваемого критерия) он удален от модального значения соответствующего ряда оценок, тем лучше этот объект с точки зрения нашего поиска, т.е. является наиболее обоснованным, представительным, достоверным. Тогда, чем большую сумму мест имеет объект, тем он больше удовлетворяет поставленным требованиям.

Прежде чем суммировать оценки (места), их необходимо привести к виду, удобному для сравнения, поскольку они имеют неодинаковую "цену" из-за различной сущности критериев, с помощью которых эти оценки получены. Как и раньше, воспользуемся преобразованием

$$z = (x_i - \bar{x})/\sigma$$

и найдем для каждого ряда новые оценки, пригодные для сравнения.

Составим таблицу оценок, представленную стандартизированными значениями прежних шести ($W_{В}^*$, $W_{н.о.}^*$, $W_{С}^{**}$, $W_{В}^{**}$, $W_{н.о.}^{**}$, $W_{С}^{**}$). По строкам этой таблицы находим алгебраическую сумму $\sum W_i$, которая является окончательной оценкой изучаемых объектов.

Теперь осталось только установить границу, за пределами которой располагаются лучшие объекты. Для этого рассмотрим характер распределения "окончательной оценки". Построим гистограмму (рис. 18), предварительно проведя проверку аномальных значений изучаемого ряда распределения по критерию

$$\epsilon(\eta, \sigma_n) = \frac{\max_i (\eta_i - \bar{\eta})}{\sigma_n},$$

где η_i — аномальное значение; $\bar{\eta}$ — среднее арифметическое изучаемого ряда оценок; σ_n — смещенная оценка стандартного отклонения.

Если значение ϵ больше, чем значение, рассчитанное по статистике [10]

$$u \sqrt{\frac{2(n-1)}{2n-5+u^2+(3+u^2+2u^4)/6(2n-5)}},$$

где $u = \Psi(1-Q/200n)$, то η_i считается аномальным, не принадлежащим данному распределению. В нашем случае аномальных значений оценок, из числа рассматриваемых, не обнаружено.

Визуальный анализ гистограммы, изображенной на рис. 18, свидетельствует о неоднородности распределения оценок: на ней выделяются две

вершины. Это можно трактовать как наличие двух самостоятельных статистических видов, каждый из которых соответствует своей вершине. Для проверки выдвинутого предположения (о статистической неоднородности изучаемых объектов) в работе использован критерий, проверяющий неоднородность выборки по многовершинности эмпирических функций распределения [20].

Предлагаемый критерий эффективен, когда ничего не известно о допустимых разбиениях совокупности на однородные части. Это как раз и соответствует разбираемому случаю.

Если распределение измеряемого признака x унимодально (одновершинно), то не существует таких последовательных (не обязательно смежных) и равных по величине интервалов x_1, x_2 и x_3 , что

$$p(x_1) > p(x_2) \text{ и } p(x_2) < p(x_3), \text{ где } p(x_k) = p\{x \in x_k\}$$

— вероятность попадания x в интервал x_k . Отсюда следует, что если найдутся такие интервалы x_1, x_2 и x_3 (равные по длине), что число попаданий признака в x_1 и в x_3 значимо больше числа попаданий в x_2 , то гипотеза унимодальности должна быть отвергнута. Критерий для сравнения всех трех частот оказывается трудоемким, поэтому полезно критерий свести к проверке двух частот.

Обозначим n_i — число попаданий признака в интервал x_i . Если n_1 и n_3 значимо не различаются и таким образом можно принять гипотезу $p(x_1) = p(x_3)$, то гипотеза унимодальности должна быть отвергнута, если $n_1 + n_3$ значимо больше $2n_2$. Следовательно, проверка унимодальности сводится к сравнению двух частот в интервалах x_2 и $x_1 \cup x_3$. Если же n_1 значимо меньше, чем n_3 , то для проверки гипотезы унимодальности частоту n_2 достаточно сравнить только с n_1 , так как если n_1 значимо больше n_2 , то тем более n_2 значимо меньше n_3 . Таким образом, для проверки гипотезы унимодальности можно ограничиться сравнением двух частот в двух непересекающихся интервалах. Критерий для такой проверки имеет вид

$$u = \frac{n_1 - n_2}{\frac{1}{1-p} \sqrt{N_p(1-2p)(1+p)}}. \quad (20)$$

Эта величина распределена асимптотически нормально с нулевым математическим ожиданием и единичной дисперсией. Поэтому для проверки нулевой гипотезы $H_0: p_1 = p_2$ против конкурирующей гипотезы $H_1: p_1 > p_2$ можно использовать правило: нулевая гипотеза отвергается с уровнем значимости q , если

$$n_1 - n_2 > u_q \frac{1}{1-p} \sqrt{N_p(1-2p)(1+p)},$$

где u_q — q -квантиль нормального распределения.

Для оценки неизвестного числа p можно воспользоваться формулой

$$\hat{p} = (n_1 + n_2) / 2N,$$

являющейся оценкой максимума правдоподобия для p при гипотезе

$p_1 \neq p_2$. Строго говоря, при неизвестном p разность $n_1 - n_2$ имеет отличное от нормального распределение, но при больших N это отличие невелико.

Используя критерий (20), при малых N полезно иметь в виду, что максимальное значение множителя при u_q равно $\sqrt{N/3}$ (при $p = 1/3$). Поэтому если p неизвестно, но $n_1 - n_2 > u_q \sqrt{N/3}$, то выражение (20) и подавно имеет место при любом p . Описанный критерий непосредственно может быть использован для проверки унимодальности, если обе вершины изучаемого распределения значимо различны ($n_3 \gg n_1$), что мы и имеем в конкретном случае.

Расположение интервалов x_1, x_2 и x_3 показано внизу гистограммы рис. 18. Уровень значимости принимаем равным 0,05, следовательно, $u_q = 1,96$. Воспользовавшись критерием (20), имеем: $n_1 = 70, n_2 = 52, N = 269, p = 122/538 = 0,227$ и тогда $u = 1,04 < 1,96 = u_q$. Таким образом, отклонение $n_1 - n_2$ является незначимым, т.е. распределение унимодальное.

Далее, учитывая смысловое содержание оценок, составляющих исследуемую совокупность (а именно, зная направление, в котором располагаются лучшие оценки), и используя приемы отсечения "хвостов", выделим лучшие объекты.

Существующие формулы для вычисления доверительных интервалов основаны на предположении о нормальном виде распределения изучаемой переменной. В случае, когда имеются отклонения от нормального вида, непосредственное использование этих формул может привести к ошибкам.

Для того чтобы устранить их, воспользуемся следующим алгоритмом обработки несимметричного (не согласующегося с нормальным) вида распределения.

1. Простой метод получения доверительных интервалов основывается на использовании предположения, что несимметричное распределение состоит из двух частей: его левая часть представляет собой одно нормальное распределение, а правая часть — другое. Причем каждое из этих распределений имеет одну и ту же моду (среднее значение), но различные стандартные отклонения [38].

Исходная асимметричная гистограмма распределения изучаемых объектов достраивается от модального значения симметричными фиктивными наблюдениями и вычисляются два стандартных отклонения: одно для объектов, равных или превышающих моду, второе для объектов, равных или меньших, чем мода:

$$\hat{\sigma}_i = \sqrt{\frac{\sum_{n_i} (x - M_0)^2}{n-1}}, \quad i = 1, 2;$$

$$n_i - \frac{1}{2}$$

тогда доверительные интервалы для средних значений определяются соотношением

$$p(M_0 - t \hat{\sigma}_1 \leq \mu \leq M_0 + t \hat{\sigma}_2) = S,$$

где $\hat{\sigma}_1$ и $\hat{\sigma}_2$ являются двумя значениями $\hat{\sigma}_i$; t — коэффициент распределения Стьюдента с $n-1$ степенями свободы и надежностью $1-\alpha$.

Однако из-за искажений, возникающих при работе с ассиметричным (а не с нормальным) распределением, вероятностная оценка доверительных интервалов будет смещена в сторону увеличения.

2. Для определения этой вероятности: $p(\xi > M)$, где случайная величина ξ — модель значения "окончательной оценки высказывания", воспользуемся системой Грама—Шарлье типа А [38]:

$$p(\xi > M) = 1 - p(\xi \leq M) = 1 - p\left(\frac{\xi - Mo}{\hat{\sigma}_i} \leq \frac{\mu - Mo}{\hat{\sigma}_i}\right) =$$

$$= 1 - \left[\Phi\left(\frac{\mu - Mo}{\hat{\sigma}_i}\right) - \frac{A}{3!} \Phi^{(3)}\left(\frac{\mu - Mo}{\hat{\sigma}_i}\right) + \frac{E-3}{4!} \Phi^{(4)}\left(\frac{\mu - Mo}{\hat{\sigma}_i}\right) \right], \quad i = 1, 2,$$

где $\Phi(z)$, $\Phi^{(3)}(z)$, $\Phi^{(4)}(z)$ — соответственно функция стандартного нормального распределения и ее третья и четвертая производные, которые табулированы и приведены, например, в таблицах [10];

$$z = (\mu - Mo) / \hat{\sigma}_i;$$

$$A = \frac{1}{n \hat{\sigma}_i^3} \sum_{j=1}^n (\mu_j - \bar{x})^3; \quad E = \frac{1}{n \hat{\sigma}_i^4} \sum_{j=1}^n (\mu_j - \bar{x})^4, \quad i = 1, 2.$$

Таким образом, подставив в разложение Грама—Шарлье $z = (\mu - Mo) / \hat{\sigma}_i$, можно оценить фактический уровень значимости отсеченного "хвоста" вместо грубоориентировочного, характеризующего симметричное распределение.

Глава 6

ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

В результате соответствующих расчетов в качестве лучших с точки зрения объективности, достоверности, обоснованности определены всего 13 гипотез из 269 высказываний о генезисе нефти и газа, которые были исследованы на последнем этапе обработки, причем 8 выделены на уровень значимости 0,029, а остальные 5 — на уровне 0,053. Они расположены в порядке достоверности с точки зрения теории принятия решений. Гипотезы (высказывания о них), указанные с одним и тем же порядковым номером занимают одинаковые места в ранжированном ряду (имеют одинаковые ранги), а содержащееся в них смысловое различие отмечено штрихом.

1. ОБЗОР НАИЛУЧШИХ ГИПОТЕЗ ПРОИСХОЖДЕНИЯ НЕФТЯНЫХ УГЛЕВОДОРОДОВ

Выявленные наилучшие высказывания о гипотезе генезиса нефти и газа представлены в формализованном виде. На данном этапе исследования вернемся к более привычному для геолога "описательному" языку и раскроем сущность каждого признака, содержащегося в том или ином высказывании.

Гипотеза 1. Описываются результаты по превращению фитола в присутствии алюмосиликатного катализатора при температуре 200°C в смесь изопреноидных углеводородов с открытой цепью состава $\text{C}_{14}-\text{C}_{20}$ (за исключением C_{17}). Кроме изопреноидных углеводородов, из фитола образуется также сложная смесь циклоалканов. Как считают сторонники этой гипотезы, циклизация происходит за счет процессов димеризации первоначально образующегося фитадиена. В том случае, когда исходным веществом является смесь фитола и олеиновой кислоты, реакция циклизации уменьшается и выход изопреноидных углеводородов увеличивается. В заключении авторы утверждают, что "проведенная работа экспериментально подтверждает образование изопреноидных углеводородов в нефтях из фитола" [56].

Гипотеза 2. Опыты получения нефти состояли в том, что некоторые животные жиры (особенно рыбий) или отдельные их составные части (триолин, тристеарин, олеиновая кислота) подвергались перегонке под давлением 10 атм и температуре 400°C . Полученная нефть при дробной перегонке давала керосин, смазочные масла, бензин и все те продукты, которые вырабатываются на заводах. Искусственная нефть при долгом хранении полимеризуется, и тогда ее сходство с природной усиливается. Так, продукт, имевший вскоре после изготовления удельный вес $0,815 \text{ Н/м}^3$, через 9 лет хранения показывал удельный вес $0,8221-0,8295 \text{ Н/м}^3$ [81].

Гипотеза 2'. Специально поставленными опытами показано, что при нагревании непредельных жирных кислот (олеиновой кислоты) до $150 \div 200^{\circ}\text{C}$ в присутствии алюмосиликатов происходит глубокое преобразование углеродистого скелета исходных кислот и образование разнообразных (более 50) углеводородов [54].

Гипотеза 3. Экспериментально исследовалась возможность получения углеводородов из окиси углерода и водорода в условиях, близких к естественным (на природных катализаторах — песчаниках нижнекембрийского возраста). При температуре 470°C и давлении 30 кг/см^2 , которое задавалось в автоклаве с водородом и окисью углерода в отношении 1 : 1, получены газообразные предельные и непредельные углеводороды, следы масел и органических кислот [3, 55].

Гипотеза 4. Синтез широкой гаммы углеводородов из CO и H_2 протекает при температуре более 200°C в присутствии катализаторов, в качестве которых могут служить природные руды и различные минералы [36, 53].

Гипотеза 5. Эксперименты по гидрированию CO . Вместо водорода применялась вода. При этом удалось получить (на базе CO и H_2O) при температуре $210-260^{\circ}\text{C}$, при нормальном давлении и азотсодержащих катализаторах Fe , Co или Ni значительные выходы бензиновых и даже твердых углеводородов [64].

Гипотеза 5'. При нагревании до 200°C с бентонитом бегеновая кислота образует углеводороды. В присутствии воды отношение изобутана или изопентана к нормальному бутану или пентану гораздо выше, чем при реакции с водой. Кроме C_{21} и более низкомолекулярных *n*-алканов, образуются также *n*-алканы с $22 \div 34$ атомами углерода. После нагревания при температуре 200°C без бентонитовой глины бегеновая кислота

еще не дает углеводов. Бентонит содержал всего лишь около 0,1% органического углерода, однако холостой опыт, проведенный с ним, позволил получить незначительные количества углеводов [82].

Гипотеза 6. Излагается серия опытов по исследованию широкой известной реакции Фишера-Тропша. Пропуская под различным давлением окись углерода и водород над слабо активными катализаторами (железо кобальт) при температуре 250–300°С, получили этан, бутан, пропан и другие углеводороды. Опытами установлено, что увеличение давления и понижение температуры способствуют синтезу высших углеводородов. Установлено также, что действие различных катализаторов влияет на конечный продукт синтеза. В результате этих опытов были получены газообразные предельные и непредельные углеводороды, масла, органические кислоты эфиры, порфирины и т.п. [76].

Гипотеза 7. "На измельченном дымчатом кварце Волыни исследовалась температурная зависимость выхода газовой фазы и сделан вывод что вначале нагрева первыми появляются H_2O , CO_2 , N_2 причем вода резко преобладает. Начиная с 200°С появляются углеводородные газы с максимумом выделения в области 500 ÷ 600°С. В этой же области наблюдается максимальное содержание водорода [68].

Гипотеза 7'. Экспериментальное исследование процесса образования углеводородов при синтезе смеси, состоящей из углекислого газа и воды, в присутствии металлического магния в качестве восстановителя подтвердило возможность получения таким путем смеси углеводородов состоящей из метана, этана, пропана, бутана [16]. К сожалению, в этом опыте авторы не называют конкретную температуру, при которой происходит образование углеводородов. Однако в высказывании употреблен термин "синтез", что свидетельствует о создаваемой при опыте температуре не менее 200°С.

Гипотеза 8. При нагреве горючего сланца до 190 ÷ 490°С установлен образование углеводородных и битуминозных веществ, количество которых составило несколько процентов от исходного органического материала [70].

Гипотеза 8'. Аналогично предыдущей гипотезе в этой также подогревались горючие сланцы, но при более низкой температуре (185°С). При этом использовались сланцы Грин ривер, шведские сланцы алум (квасцовые). Авторы поясняют: "... в условиях, подобных обычным в природе, наблюдается выделение ряда насыщенных углеводородов с прямой цепочкой, особенно распространенных в природных веществах от метана (CH_4) до додекана ($C_{12}H_{26}$)" [17].

Гипотеза 9. Проведены экспериментальные исследования по моделированию процессов газообразования в недрах с породами, содержащими органическое вещество гумусового, сапропелевого и смешанного типов. В опытах предварительно десорбированные образцы подвергались длительному (? — В.Р.) нагреванию от 100 до 300°С и сжатию до 2000 кг/см² в вакуумированной камере. Состав свободно выделяющейся газовой фазы обнаруживает определенные зависимости от типа органического вещества, его начальной степени метаморфизма и температуры воздействия. Во всех случаях активное газообразование фиксируется при температуре 100°С

и, как правило, возрастает с увеличением температуры. Увеличивается при этом и доля углеводородных газов, достигая 40–50%.

Неуглеводородная часть представлена окисью и двуокисью углерода (преобладает в гумусовом типе органического вещества), азотом (преобладает в сапропелевом типе органического вещества) и водородом (содержание в газах достигало 20–30%) [27].

К сожалению, из-за небрежности, допущенной при описании данного опыта, остается неясным, с какого именно температурного интервала фиксировалось выделение углеводородных газов. Правда, другая группа сторонников этой же гипотезы отмечает: "На буроугольной стадии интенсивное газообразование связано с образованием большого количества CO_2 (более 80%). Долевое участие углеводородов в новообразованных газах этой стадии при температуре 200°C ничтожно мало. При дальнейшем повышении температуры до 230°C концентрация углеводородов (преимущественно метана), оставаясь незначительной, несколько увеличивается ($0,95 \div 1,75\%$). Относительное увеличение углеводородов до 6,5% в опытах с бурными углями наблюдается лишь при температуре 300°C ".

Далее высказывается предположение, что "в температурных условиях, когда бурые угли, очевидно, переходят в последующие стадии катагенеза, количество углеводорода в этих условиях может быть значительным". Так, "в опытах с углями стадий длиннопламенной и газовой процентное содержание углеводородов в газе возрастает до 40–50% (при температуре $150 \div 250^\circ\text{C}$). В том числе тяжелые углеводороды составляют 20,4%".

Вновь остается неопределенность из-за нечеткого изложения опыта — указан слишком широкий диапазон температур. Не ясно, с какой же температуры начинает увеличиваться углеводородная фракция в газе.

Для газовой фазы угля стадии жирной характерно резкое увеличение концентрации водорода и снижение доли тяжелых углеводородов. Аномальность состава газовой фазы угля, вероятно, обусловлена существенным изменением химической структуры органического вещества углей. Газовая фаза в опытах с углями стадий коксовой и отошено-спекающейся характеризует новую по химическому содержанию фазу превращения органического вещества углей. При этом газовая составляющая отличается ничтожно малым содержанием метана и исключительно высокой концентрацией тяжелых углеводородов (до 88% при температуре 200°C), среди которых идентифицированы в значительных концентрациях (до 23%) бутан, пентан, гексан. В опытах с образцами углей последующих стадий (тощая, полуантрацитовая и антрацитовая) наблюдается новая вспышка метанообразования (до 85%) при температуре 150°C . Тяжелые углеводороды практически исчезают [71].

На основании описанных опытов, привлекая также обширные геологические наблюдения, сторонники данной гипотезы заключают. "Установлена относительная обособленность скоплений газа и скоплений нефти при сходстве или единстве материальных источников их генерации, т.е. различных форм и типов органического вещества осадочных толщ. Подавляющая часть разведанных в мире запасов природного газа (более 90%) заключена не в совместных с нефтью, а в чисто газовых или газоконденсатных залежах (месторождениях), и эти последние встречаются не изолированно, а группами, которые содержат газоносные или преимущественно газоносные

зоны и районы. Более того, обнаруживаются преимущественно газоносные (или нефтеносные) провинции и даже бассейны. Таким образом, в распространении залежей газа, так же как и нефти, — ископаемых единого генетического ряда — наблюдается пространственная упорядоченность, региональная обособленность или зональность” [27].

2. ВЫВОДЫ О СОВРЕМЕННОМ СОСТОЯНИИ ПРОБЛЕМЫ ГЕНЕЗИСА НЕФТИ И ГАЗА

С точки зрения теории принятия решений все перечисленные гипотезы находятся в равных условиях. Они в одинаковой степени объективны, в такой же степени достоверны, находятся на одном и том же уровне по обоснованности.

Однако следует помнить, что назначение теории принятия решений — объективно разобраться в каждом явлении, численно оценить их и предложить варианты решений, т.е. задача состоит в подготовке решения, а не в его принятии. Принять решение — задача специалиста-заказчика, в данном случае специалиста-нефтяника. Он, принимая решение, обязан учитывать, помимо рекомендаций, полученных в результате применения методов экспертной оценки и теории принятия решений, основанных на количественных оценках, и другие факторы, не поддающиеся формализации.

Из приведенного “обзора наилучших гипотез” явствует, что авторы, используя самые разнообразные элементы исходного материала для получения искусственной нефти и газа, а также различные катализаторы, и подвергая испытуемые смеси давлению или обходясь без него, единодушно считают фактор температуры главным при образовании углеводородов.

Обращает на себя внимание тот факт, что только в одном случае для начала образования углеводородов потребовалась температура 150°C . Во всех остальных указываются более высокие температурные режимы.

Отсюда формулируем вывод: вне зависимости от природы исходного материала, каталитических и динамических условий процесс нефтегазообразования (но не метана) начинается при температуре не менее 150°C .

Температура, равная 150°C , в природе встречается на глубине 6–7 км, да и то только в пределах молодых, геотермически активных платформ. (Сюда, конечно, не входят локально разогретые участки вулканов, термальных источников и прочее.) В пределах же древних платформ температура даже у кровли кристаллического фундамента редко достигает 100°C [41, 58, 74]. Следовательно, необходимо признать, что начало процесса нефтегазообразования на древних платформах происходит на еще больших глубинах.

Поэтому заключаем: в природных условиях процесс нефтегазообразования протекает по крайней мере не выше, чем на глубине 6–7 км.

Это вызывает вопрос, который не удалось формализовать в начале решения. В какой же среде происходит формирование углеводородов: в осадочной, как утверждают сторонники органических гипотез, или вне ее, как утверждают ее противники?

Допустим, что формирование нефти и газа происходит только в осадочном чехле. При этом если учесть, что “общий объем осадочной обо-

лочки, по данным последних измерений, составляет около 11% объема земной коры" [67], то суммарная площадь благоприятных (в нашем определении) для образования углеводородов участков осадочного чехла составляет незначительную долю от общей территории, на которой обнаружены скопления нефти и газа. Тогда факт нахождения даже самых незначительных скоплений углеводородов, не говоря о нефтяных и газовых месторождениях-гигантах, внутри сравнительно маломощных осадочных толщ иначе как явлением дальней латеральной миграции объяснить нельзя, поскольку в процессе решения все остальные аргументы, свидетельствующие в пользу такой возможности, отброшены как несостоятельные (не достаточно обоснованные).

Пример исследования по выяснению суждений специалистов о природе миграции углеводородов. Имеющийся фактический материал позволил провести специальные исследования по выяснению мнений экспертов о природе миграции углеводородов вообще и, в частности, о явлении дальней латеральной миграции.

На базе высказываний авторов изучалась статистическая связь между фактором "исходный материал для нефти и газа" и фактором, который при декомпозиции проблемы условно назван "направление миграции углеводородов" (см. рис. 10, а, л). В зависимости от особенностей и качества фактического материала использовались либо коэффициент парной корреляции, либо коэффициент сопряженности [72]

$$\hat{r} = \frac{AB \cdot \bar{A}\bar{B} - \bar{A}B \cdot A\bar{B}}{\sqrt{A \cdot B \cdot \bar{A} \cdot \bar{B}}},$$

где AB – наличие соответственно признаков A и B ; $\bar{A}\bar{B}$ – отсутствие этих признаков; $\bar{A}B$ – одновременное наличие признака A и отсутствие признака B и т.д. Для установления существенной зависимости между признаками оценку коэффициента необходимо сопоставить с допустимым значением для заданного уровня значимости из выражения [30, 72]

$$r_{\text{доп}} = 3 \frac{1 - \hat{r}^2}{\sqrt{N}},$$

где N – число наблюдений. Если абсолютное значение коэффициента выше допустимого, то есть основание считать, что связь признаков не случайна.

Отдельно обрабатывались мнения: сторонников органического происхождения нефти и газа; сторонников неорганического их происхождения; всей группы экспертов, в том числе и занимающихся исключительно вопросами миграции углеводородов. Ранее мнения этих авторов были устранены из рассмотрения как шумовые для определенного этапа исследований. На данном этапе их мнение о генезисе нефти и газа выяснялось из контекста высказываний.

Результаты проведенного статистического анализа позволяют сделать следующие заключения.

Всевозможные варианты расчетов коэффициента корреляции между отдельными признаками изучаемых факторов не дали возможности однозначно установить точку зрения экспертов относительно приоритета того

или иного миграционного явления. Во всех вариантах расчета значения коэффициента корреляции оказались незначимыми.

Коэффициент сопряженности \hat{r} , рассчитанный для фактора "исходный материал для нефти и газа", который в данном случае фигурировал в качестве признака, и признака, характеризующего вертикальную миграцию при нарушениях, оказался равным 0,987, что значительно превышает допустимое значение $r_{\text{доп}} = 0,13$ при $q = 0,05$.

Такая же связь статистически проверена и для случая вертикально молекулярной миграции, идущей поперек напластования и направленно к земной поверхности, возможность существования которой допускают при отсутствии явных тектонических нарушений. При этом коэффициент сопряженности между "исходным материалом..." и данным признаком составил $\hat{r} = 0,940$ при допустимом значении $r_{\text{доп}} = 0,19$ на 5%-ном уровне значимости.

Явно проявившаяся почти функциональная связь дает право заключить, что сейчас не осталось авторов, возражающих против свободно (по трещиноватым вертикальным либо наклонным каналам) или затрудненной (сквозь слабо проницаемые породы) разгрузки углеводородов к поверхности Земли.

Особое внимание авторы в своих высказываниях уделяют пластовой или, как еще ее называют, латеральной миграции, идущей по пористым песчаным и другим пластам. По масштабам движения ее разделяют на локальную (ближнюю) и региональную (дальнюю). С локальной миграции связывают процессы перераспределения углеводородов в пределах одного и того же структурного элемента, а чаще только внутри отдельного месторождения в основном пластового типа. При региональной миграции допускают перетоки флюидов из одного структурного элемента в другой в зависимости от поперечных антиклинальных, литологических либо тектонических препятствий, встречаемых на их пути.

Коэффициент сопряженности для случая локальной латеральной миграции $\hat{r} = 0,764$ при допустимом значении $r_{\text{доп}} = 0,23$ и уровне значимости $q = 0,05$. При этом если рассматривать мнения только сторонников органического происхождения нефти и газа, то коэффициент становится даже несколько ниже ($\hat{r} = 0,721$ при $r_{\text{доп}} = 0,340$ на 5%-ном уровне значимости).

Что же касается явления дальней латеральной миграции, то статистически оно не подтвердилось. И более того, среди собранного фактического материала не оказалось ни одной работы, в которой бы это явление обосновывалось экспериментом; имеются только теоретические расчеты, в основе которых лежат неестественно высокие перепады давления. Выявилась степень достоверности существования таких явлений, как "вертикальная миграция вниз", "in situ" и "боковая" (строго горизонтальная, межрезуарная), с помощью статистического анализа не представляется возможным ввиду малого числа высказываний о них. Однако, рассуждая логически, можно заключить, что на современном этапе знаний о проблеме исследователи полностью отказались от признания этих природных явлений, поскольку упоминания о них зафиксированы только в ранних работах, в период становления науки о нефти.

Резюмируя сказанное о миграции углеводородов, заключаем, что только вертикальную разгрузку к поверхности Земли и внутрирезервуарную ми

грацию в пределах ловушки можно считать неопровержимыми природными явлениями.

Имея это в виду, теперь можно ответить на ранее поставленный вопрос: "В какой же среде происходит формирование нефти и газа — в осадочной толще или вне ее?" следующим образом.

Существование в природе огромных скоплений битумов (Атабасского, Мелекесского и, по-видимому, Анабарского), приуроченных к древним плитам, характеризующимся маломощным чехлом осадочных пород (0,5—2 км), и далеко удаленных от глубоких (6—7 км) впадин, иначе как первичностью их образования вне осадочного чехла на современном научном уровне знаний объяснить нельзя.

Часть же месторождений, фиксируемых над районами глубоких геосинклиналей и краевых прогибов, таких, например, как Прикаспийская или Южно-Каспийская впадины, заполненных мощной толщей осадочного слоя, где возможны геотермические условия, необходимые для образования углеводородов, следует признать продуктами осадочного чехла.

Таким образом получается, что основная масса известных скоплений нефти и газа на земном шаре мигрировала из нижних слоев земной коры, не принадлежащих осадочному слою. Какой именно слой коры (гранитный или базальтовый), а может быть, и не кора, а мантия является источником основной массы углеводородов — на современном уровне знаний остается загадкой.

Конечно, интересно знать действительный механизм образования углеводородов. Поэтому факт, что нам частично удалось однозначно решить вопрос об условиях формирования нефти и газа по крайней мере в части температурных режимов, позволяет более рационально направить дальнейшие поиски этих полезных ископаемых.

ЗАКЛЮЧЕНИЕ

В геологических исследованиях при обобщении научно-технической информации, как правило, используются теоретические представления, полученные ранее, о конкретно решаемой задаче и мало или вовсе не учитываются результаты исследований из смежных областей науки, которые могут быть полезны при решении этой задачи. Такое положение особенно характерно для информации, получаемой при экспертном опросе, поскольку собираются мнения специалистов определенной отрасли знания. К экспертным оценкам обращаются еще и в тех случаях, когда информация об изучаемой задаче выражается в основном качественными (неколичественными) показателями. Однако метод экспертных оценок имеет ряд недостатков.

Указанные обстоятельства обусловили необходимость разработки новых научных приемов обработки смысловой информации. Работа проведена в трех направлениях, и получены следующие результаты.

1. Обоснован принцип обобщения опубликованной научно-технической литературы, дающий возможность накопить обширные сведения, не только прямо относящиеся к изучаемой проблеме, но и косвенные, не касающиеся ее непосредственно, а привлекаемые исследователями для подтверждения своей точки зрения или опровержения противоположной. Его достоинство состоит в том, что случайный сбор работ исключает тенденциозность отбора материала и поэтому дает возможность исследовать информацию, полученную от любого специалиста. Это расширяет возможности учета как "нетипичных", так и наиболее "типичных" фактов наблюдений и позволяет со стороны представить объект; кроме того, создает большую информационную ценность его характеристик.

2. Для обработки накопленной информации предложен подход, позволяющий полностью формализовать высказывание о предмете изучения количественно оценить качественные показатели. Такой подход должен способствовать более широкому внедрению аппарата методов математической статистики при исследовании геологических задач.

3. Разработана технология комплексирования методов для использования при оценке суждений о предмете изучения. Комплекс методов адаптирован к модели принятия решения, которая может быть использована в качестве методического руководства при изучении сложных генетических и прогнозных геологических задач.

Отметим основные особенности предлагаемой системы.

При принятии окончательного решения недостаточно использовать ка

кой-либо один критерий оптимальности. Необходимо учитывать одновременно весь возможный их комплекс, каждый из которых, указывая на различные лучшие стороны объекта, позволяет при выборе существенно ограничить волевые решения.

В случае, когда нет эталонов для оценки и мнения ученых рассматриваются как случайные, прогноз по средней оценке вообще может оказаться хуже любого индивидуального прогноза. В работе такому явлению уделено особое внимание. Во многом по-новому рассматривается процедура сбора и обработки фактического материала, полученного от специалистов, что помогло исключить традиционное влияние "большинства" и при выборе окончательного решения учитывать и "нетипичные" случаи.

Предложенная технология комплекса методов обработки информации была использована при исследовании вопроса о состоянии проблемы нефти и газа. В результате из множества высказываний (2409) удалось выявить 13 наиболее объективных и фактически обоснованных при современном уровне знаний гипотез.

Дальнейшее применение полученных выводов, как и свертка большого объема исходной и обработанной информации, могут значительно сэкономить средства и время последующего изучения данной проблемы и сосредоточить внимание только на выявленных, наиболее обоснованных высказываниях, которые были рассеяны в многочисленных литературных источниках.

В заключение хотелось бы отметить, что применение предложенной технологии обработки информации позволяет существенно сократить время и затраты на изучение сложных проблем, связанных с анализом большого объема информации. Это особенно важно в условиях быстрого развития науки и техники, когда информация накапливается с огромной скоростью. Предложенная технология может быть использована в различных областях науки и техники, где требуется анализ большого объема информации и принятие обоснованных решений.

ЛИТЕРАТУРА

1. Акоф Р., Сасиени М. Основы исследования операций. М.: Мир, 1971. 534 с.
2. Александрова Т.Д. Семинар по применению перфокарт в геологии и географии//Изв. АН СССР. Сер. геогр. 1965. № 4. С. 145-147.
3. Башкиров А.Н. О каталитической активности некоторых осадочных горных пород в синтезе углеводородов из окиси углерода и водорода// Докл. АН СССР. 1965. Сер. геол. Т. 161. С. 947-948.
4. Белонин М.Д., Иванова К.П., Наливкин В.Д. и др. Математические методы анализа при изучении закономерностей формирования и размещения углеводородных скоплений. М.: ВИЭМС, 1973. С. 1-46. (Мат. методы исслед. в геологии).
5. Белоусов В.В. Пути развития науки о Земле//Взаимодействие наук при изучении Земли. М.: Изд-во АН СССР, 1963. С. 11-24.
6. Бешелев С.Д., Гурвич Ф.Г. Экспертные оценки. М.: Наука, 1973. 159 с.
7. Бешелев С.Д., Гурвич Ф.Г. Математико-статистические методы экспертных оценок. М.: Статистика, 1980. 263 с.
8. Блауберг И.В., Юдин Э.Г. Становление и сущность системного подхода. М.: Наука, 1973. 270 с.
9. Божич С.П. О способах истинной оценки естественнонаучного высказывания//Логика и эмпирическое познание. М.: Наука, 1972. С. 243-255.
10. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1968. 474 с.
11. Боннер Р.Е. Некоторые методы классификации//Автоматический анализ сложных изображений. М., 1968. С. 205-234.
12. Вайнберг Дж., Шумекер Дж. Статистика. М.: Статистика, 1979. 389 с.
13. Венцель Е.С. Исследование операций. М.: Сов. радио, 1972. 552
14. Власов Е.П., Голикова Л.И. Дерягин А.А. и др. Определение априорной информативности геологических признаков на основе методов экспертных оценок//Методика комплексной обработки геолого-геофизической информации с помощью математических методов и ЭВМ при прогнозировании и оценке месторождений. М.: ВИЭМС, 1973. С. 75-98.
15. Выханду Л.К. Об исследовании многопризнаковых биологических систем//Применение математических методов в биологии. Л.: Изд-во Ленинг. ун-та, 1964. Сб. 3. С. 19-22.
16. Гавриленко Е.С. Гидрогеология тектоносферы. Киев: Наук. думка, 1973. 195 с.
17. Гедберг Х.Д. Геологические аспекты происхождения нефти. М.: Недра, 1966. 124 с.
18. Гладких Б.А. Некоторые проблемы классификации//Распознавание образов в экономико-статистическом исследовании. Новосибирск: Ин-т экономики и организации промышленного производства, 1974. С. 5-30.
19. Глушков В.М. Математизация научного знания и теория решений. Вопросы философии. 1978. № 1. С. 28-33.
20. Гольдин С.В. О проверке однородности совокупностей геологических объектов//Математические методы в геологических исследованиях в Западной Сибири. Тюмень, 1968. С. 108-12 (Тр. Зап.-Сиб. науч.-исслед. геол.-развед. нефтяного ин-та; Вып. 18).
21. Гусельников И.И., Турпигтько А. Перфокарты с краевой перфорацией. М.: Высш. шк., 1967. 291 с.
22. Гурвич Ф.Г. Методы и процедуры получения экспертной информации

Экономика и мат. методы. 1973. Т. 9. Вып. 5. С. 962-975.

23. *Гурвич Ф.Г.* Методы обработки экспертной информации//Там же. 1973. Т. 9. Вып. 6. С. 1157-1169.

24. *Девис Дж.* Статистика и анализ геологических данных. М.: Мир, 1977. 571 с.

25. *Евланов Л.Г., Кутузов В.А.* Экспертные оценки в управлении. М.: Экономика, 1978. 133 с.

26. *Елисеева И.И., Руковишников В.О.* Группировка, корреляция, распознавание образов//Статистические методы классификации и измерения связей. М.: Статистика, 1977. 143 с.

27. *Жабров И.П., Ермаков В.И., Орел В.Е.* и др. Генезис газа и прогноз газоносности//Геология нефти и газа. 1974. № 9. С. 1-8.

28. *Жариков Е.С., Арутюнов В.Х., Чуримов С.И.* О некоторых способах повышения эффективности научных дискуссий//Гез. докл. III симпозиум по науковедению и науч.-техн. прогнозированию. Киев: Наук. думка, 1970. Вып. 5. С. 18-19.

29. *Жуковская В.М., Мучник И.Б.* Факторный анализ в социально-экономических исследованиях. М.: Статистика, 1976. 152 с.

30. *Закс Л.* Статистические оценивания. М.: Статистика, 1976. 598 с.

31. *Ивахненко А.Г.* Самообучающиеся системы распознавания и автоматического управления. Киев: Техника, 1969. 392 с.

32. *Кадибур Т.С., Елисеева И.И.* Вопросы методики аналитической группировки//Науч. зап. Ленингр. фин.-экон. ин-та. Л., 1975. Вып. 42. С. 216-223.

33. *Карасев А.И.* Теория вероятностей и математическая статистика. М.: Статистика, 1977. 279 с.

34. *Кедров Б.М.* Научная концепция детерминизма//Современный детерминизм: Законы природы. М.: Мысль, 1973. С. 6-35.

35. *Кендалл М.Дж., Стюарт А.* Многомерный статистический анализ и временные ряды. М.: Наука, 1975. 736 с.

36. *Коган Ю.Б., Крюков Ю.Б., Башкиров А.Н.* О механизме синтеза углеводородов из окиси углерода и водорода//Докл. АН СССР. 1951. Т. 78. № 2. С. 275-276.

37. *Кокрен У.* Методы выборочного исследования. М.: Статистика, 1976. 440 с.

38. *Коуден Д.* Статистические методы контроля качества. М.: Физматгиз, 1961. 623 с.

39. *Кофман А.* Методы и модели исследования операций. М.: Мир, 1966. 523 с.

40. *Крамбейн У., Кауфмен М., Мак-Кеммон Р.* Модели геологических процессов. М.: Мир, 1973. 150 с.

41. *Красковский С.А.* О тепловом поле шитов//Изв. АН СССР. Сер. геофиз. 1961. № 3. С. 387-392.

42. *Кудрявцев Н.А.* О книге А.Я. Креста "Вопросы формирования залежей нефти и газа"//Геология и геохимия, 1 (VII). Л.: Гостехиздат, 1957. С. 248-254.

43. *Лисичкин В.А.* Теория и практика прогностики. Киев: Наук. думка, 1972. 224 с.

44. *Льюс Р.Д., Райфа Х.* Игры и решения. М.: Изд-во иностр. лит., 1961. 642 с.

45. *Массе П.* Критерии и методы оптимального определения капиталовложений. М.: Статистика, 1971. 503 с.

46. *Менделеев Д.И.* Об опытах императорского Вольного экономического общества над действием удобрений//Тр. имп. Вольного экон. об-ва. СПб., 1872. Т. 1, вып. 4. С. 413-414.

47. Методика программного прогнозирования развития науки и техники: Постановление ГКНТ СМ СССР № 462 от 26.II.1971. 113 с.

48. *Миркин Б.Г.* Проблема группового выбора. М.: Наука, 1974. 256 с.

49. *Моррис У.* Наука об управлении. М.: Мир, 1971. 304 с.

50. *Мостеллер Ф., Рурке Р., Томас Дж.* Вероятность. М.: Мир, 1969. 431 с.

51. *Налимов В.В.* Теория эксперимента. М.: Наука, 1971. 207 с.

52. *Нейман Дж., Моргенштейн О.* Теория игр и экономического поведения. М.: Наука, 1970. 707 с.

53. *Нефедов Б.К., Эйдус Я.Т.* Развитие каталитических синтезов органических соединений из окиси углерода и водорода//Успехи химии. 1965. Т. 34, вып. 4. С. 630-652.

54. *Петров А.А.* Строение нефтяных углеводородов и проблема происхождения нефти//Генезис нефти и газа. М.: Недра, 1967. С. 133-140.

55. *Пиковский Ю.И., Башкиров А.Н., Новак Ф.И.* О каталитической активности некоторых осадочных горных пород в синтезе углеводородов из окиси угле-

рода и водорода//Докл. АН СССР. Сер. геол. 1965. Т. 161, № 4. С. 947–948.

56. *Пустильникова С.Д., Цедилина А.Л., Красавченко М.И.* и др. Образование изопреноидных углеводородов нефтей из фитола//Геология нефти и газа. 1973. № 12. С. 56–63.

57. *Радемахер Г., Теплиц О.* Числа и фигуры. М.: Физматгиз, 1962. 263 с.

58. *Раевский В.Ю.* Использование тренд-анализа при геотермических исследованиях в нефтегазовых областях//Изв. АН СССР. Сер. геол. 1973. № 6. С. 119–126.

59. *Раевский В.Ю.* О несостоятельности использования оптической активности в проблеме генезиса нефти и газа. Киев: Наук. думка, 1979. № 2. С. 15–20.

60. *Раевский В.Ю., Спивак Т.Ю., Грабенко В.И.* и др. Применение комплекса математических методов при выявлении аномалий по результатам газометрической съемки//Экспресс-информ. "Мат. методы исслед. в геологии". М.: ВИЭМС, 1980. С. 1–10.

61. *Раевский В.Ю.* Метод экспертной оценки при решении плохо формализуемых геологических задач//Сов. геология. 1981. № 6. С. 30–39.

62. *Раевский В.Ю.* Метод оценки качества экспертной информации в геологических исследованиях//Там же. 1983. № 7. С. 41–47.

63. *Резников А.П.* Обработка накопленной информации в затрудненных условиях. М.: Наука, 1976. 242 с.

64. *Реми Г.* Курс неорганической химии. М.: Изд-во иностр. лит., 1963. 920 с.

65. *Родионов Д.А.* Статистические методы разграничения геологических объектов по комплексу признаков. М.: Наука, 1968. 158 с.

66. *Родионов Д.А.* Статистические решения в геологии. М.: Недра, 1981. 231 с.

67. *Ронов А.Б.* Строение и геохимическая история осадочной оболочки Земли//Геохимия. Минералогия. М.: Наука, 1980. С. 3–14.

68. *Сворень И.М.* К вопросу о воспроизводимости газовой среды минералообразования//Углерод и его соединения в эндогенных процессах минералообразования. Львов: Изд-во Львов. ун-та, 1975. С. 26–28.

69. *Смирнов Б.И.* Статистические методы в геологии. Львов: Высш. шк., 1975. Ч. 2. 120 с.

70. *Соколов В.А.* Процессы образования нефти и газа//Происхождение нефти

и газа и формирование их месторождений: Материалы Всесоюз. совещ. по генезису нефти и газа. М.: Недра, 1977. С. 16–39.

71. *Соколов В.Л., Симоненко В.Ф., Гуляева Н.Д.* Гумусовое органическое вещество – источник генерации углеводородов: (Эксперим. исслед.)//Происхождение нефти и газа и закономерности образования и размещения их залежей: Тез. докл. Львов: Изд-во Львов. ун-та, 1977. С. 16–18.

72. Статистические методы при геохимических поисках месторождений//Под ред. Д.А. Минаева. М.: ИМГРЭ, 1973. 122 с.

73. *Суворов Ю.П.* Теории исследования операций и принятия решений. Возможность их применения в создании АСУ-Геология: Обзор. Сер. II. ВИЭМС, 1973. 35 с.

74. Тепловой режим недр СССР/Под ред. Ф.А. Макаренко, Б.Г. Поляка. М.: Наука, 1970. 224 с.

75. *Тимофеева Н.В.* К уточнению сортности руд в полиметаллических месторождениях с помощью статистических методов разграничения//Тр. симпоз. "Горнорудный Пршибрам в науке и технике". Пршибрам, 1977. С. 456–465.

76. *Фишер Ф.* Новейшие взгляды на происхождение нефти//Нефт. хоз-во. 1930. № 11. С. 1073–1077.

77. *Чернов Г., Мозес Л.* Элементарная теория статистических решений. М.: Сов. радио, 1962. 406 с.

78. *Чупров А.А.* Основные проблемы теории корреляции. М.: Планхозгиз, 1926. 165 с.

79. *Шарапов И.П.* Логический анализ некоторых проблем геологии. М.: Недра, 1977. 144 с.

80. *Шейдина И.Л.* Методы экспертных оценок в практике США//США – экономика, политика, идеология. 1971. № 9. С. 113–121.

81. *Энглер К.* К вопросу о происхождении нефти//Горн. журн. 1888. № 9. С. 306–315.

82. *Юрг И.У., Эйсма Э.* Образование углеводородов из жирных кислот//Органическая геохимия: (Успехи в познании природы нефти и нефтематеринских веществ). М.: Недра, 1967. Вып. 1. С. 141–143.

83. *Якушкин Б.В.* Классификация//СЭ. 2-е изд. М., 1973. Т. 12. С. 793–794.

84. *Янч Э.* Прогнозирование научно-технического прогресса. М.: Прогресс, 1970. 568 с.

ОГЛАВЛЕНИЕ

Введение	3
<i>Глава 1</i>	
Теоретические основы объективного принятия решений	5
1. О характере исходных геологических данных и уровне их неопределенности	5
2. Краткий обзор существующих методов решения сложных, трудно формализуемых проблем	10
<i>Глава 2</i>	
Постановка задачи	29
1. Математическая модель объекта исследования	30
2. Математическая модель принятия решения	31
3. Проблема генезиса нефти и газа	31
<i>Глава 3</i>	
Формирование массива исходных данных	35
1. Методика бесконтактного опроса специалистов	35
2. Характеристика массива исходных данных "Литература"	46
<i>Глава 4</i>	
Обработка информации	47
1. Систематизация исходных данных	47
2. Формализация смысловой информации	58
3. Характеристика массива данных "Индивидуальные высказывания"	64
4. Свертывание информации	72
<i>Глава 5</i>	
Поиск объективных высказываний	94
1. Типичные методы и критерии принятия решений	95
2. Назначение (выбор) параметров эффективности	99
3. Алгоритм сопоставления оценок, полученных по разным критериям оптимальности	101
<i>Глава 6</i>	
Интерпретация результатов исследования	105
1. Обзор наилучших гипотез происхождения нефтяных углеводородов	105
2. Выводы о современном состоянии проблемы генезиса нефти и газа	109
Заключение	113
Литература	115

Виктор Юрьевич Раевский

МАТЕМАТИЧЕСКИЕ
РЕШЕНИЯ
ТРУДНО
ФОРМАЛИЗУЕМЫХ
ГЕОЛОГИЧЕСКИХ
ЗАДАЧ

*Утверждено к печати
Всесоюзным
научно-исследовательским
институтом экономики
минерального сырья
и геологоразведочных работ*

Редактор издательства

Ю.А. Юдина

Художник

С.А. Резников

Художественный редактор

Н.Н. Власик

Технический редактор

Н.А. Торгашова

Корректор

О.А. Пахомова

Набор выполнен в издательстве
на наборно-печатающих автоматах

ИБ № 35287

Подписано к печати 18.12.86. Т — 22842

Формат 60×90 1/16

Бумага офсетная № 1

Гарнитура Пресс-Роман

Печать офсетная. Усл.печ.л. 7,5

Усл.кр.-отт. 7,9. Уч.-изд.л. 8,5

Тираж 750 экз. Тип. зак. 1014

Цена 1 р. 30 к.

Ордена Трудового Красного Знамени

издательство "Наука"

117864 ГСП-7, Москва В-485

Профсоюзная ул., д. 90

Ордена Трудового Красного Знамени

1-я типография издательства "Наука"

199034, Ленинград В-34, 9-я линия, 12

В ИЗДАТЕЛЬСТВЕ "НАУКА"

готовится к печати:

Лукин А.А. Опыт разработки методики морфоструктурно-гидрогеологического анализа. 10 л.

В книге рассматривается новая методика гидрогеодинамического и морфоструктурно-гидрогеологического районирования, которая является методическим ядром морфоструктурно-гидрогеологического анализа и позволяет решать на уровне достоверных прогнозов многие задачи гидрогеологии.

Заказы просим направлять по одному из перечисленных адресов магазинов "Книга-почтой" "Академкнига":

Магазины "Книга-почтой."

252030 Киев, ул. Пирогова, 4; 197345 Ленинград, Петрозаводская, 7
117192 Москва, Мичуринский пр-т, 12

Магазины "Академкнига" с указанием отделов "Книга-почтой":

480091 Алма-Ата, ул. Фурманова, 91/97 /"Книга-почтой"/; 37005 Баку, ул. Коммунистическая, 51 /"Книга-почтой"/; 232600 Вильнюс, ул. Университето, 4; 690088 Владивосток, Океанский пр-т, 140 /"Книга-почтой"/; 320093 Днепропетровск, пр-т Гагарина, 24 /"Книга-почтой"/; 734001 Душанбе, пр-т Ленина, 95 /"Книга-почтой"/; 375002 Ереван, ул. Туманяна, 31; 664033 Иркутск, ул. Лермонтова, 289 /"Книга-почтой"/; 420043 Казань, ул. Достоевского, 53 /"Книга-почтой"/; 252030 Киев, ул. Ленина, 42; 252142 Киев, пр-т Вернадского, 79; 252030 Киев, ул. Пирогова, 2; 277012 Кишинев, пр-т Ленина, 148 /"Книга-почтой"/; 343900 Краматорск, Донецкой обл., ул. Марата, 1 /"Книга-почтой"/; 660049 Красноярск, пр-т Мира, 84; 443002 Куйбышев, пр-т Ленина, 2 /"Книга-почтой"/; 191104 Ленинград, Литейный пр-т, 57; 199164 Ленинград, Таможенный пер., 2; 196034 Ленинград, В/О, 9 линия, 16; 220012 Минск, Ленинский пр-т, 72 /"Книга-почтой"/; 103009 Москва, ул. Горького, 19-а; 117312 Москва, ул. Вавилова, 55/7; 630076 Новосибирск, Красный пр-т, 51; 630090 Новосибирск, Морской пр-т, 22 /"Книга-почтой"/; 142284 Протвино, Московской обл., ул. Победы, 8; 142292 Пушино, Московской обл., МР, "В", 1; 620161 Свердловск, ул. Мамина-Сибиряка, 137 /"Книга-почтой"/; 700000 Ташкент, ул. Ю. Фучика, 1; 700029 Ташкент, ул. Ленина, 73; 700070 Ташкент, ул. Ш. Руставели, 43; 700185 Ташкент, ул. Дружбы народов, 6 /"Книга-почтой"/; 634050 Томск, наб. реки Ушайки, 18; 634050 Томск, Академический пр-т, 5; 450059 Уфа, ул. Р.Зорге, 10 /"Книга-почтой"/; 450025 Уфа, ул. Коммунистическая, 49; 720000 Фрунзе, бульвар Дзержинского, 42 /"Книга-почтой"/; 310078 Харьков, ул. Чернышевского, 87 /"Книга-почтой"/

1 р. 30 к.

4798