

К (идеальному образу), сводятся здесь к полезной диагностической информации θ_0 об индивидах и информационным помехам.

Математическую процедуру определения принадлежности к таксону на основе материала обучения принято называть распознаванием образов. Видовой таксон в отличие от таксона (кластера), выделяемого путем обычной феноменологической типизации и распознавания, определяется с помощью модели воспроизведения индивидов в видовой популяции. После выделения и обоснования видовых или квазивидовых таксонов (выяснения идеализированных их образов) отнесение к ним новых индивидов может во многих случаях выполняться тоже путем феноменологической типизации и распознавания с учетом связей между признаками. Иногда оказываются достаточными отдельные признаки. Но обычно при определении видовой принадлежности должна учитываться уже отмеченная выше разная существенность признаков в системе воспроизведения.

Литература

1. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. М.: Мир, 1982. 488 с.
2. Белоусов А.Ф. Неоднородность распределения составов в ассоциациях изверженных горных пород и представление о породных группах // Геология и геофизика. 1967. № 5. С.26-24.
3. Белоусов А.Ф. Популяционная модель в исследовании ассоциаций магматических пород // Геология и геофизика. 1979. № 1. С.35-45.
4. Белоусов А.Ф. Системный подход и некоторые методологические проблемы исследования геолого-географических формаций // Методологические и философские проблемы геологии. Новосибирск, 1979. С.276-294.
5. Белоусов А.Ф., Кривенко А.П., Полякова З.Г. Вулканические формации. Новосибирск: Наука, 1982. 281 с.
6. Белоусов А.Ф., Кривенко А.П. Магмогенез вулканических формаций. Новосибирск: Наука, 1983. 167 с.
7. Мейен С.В., Шрейдер Ю.А. Методологические аспекты теории классификации // Вопросы философии. 1976. № 12. С.67-79.
8. Розова С.С. Классификационная проблема в современной науке. Новосибирск: Наука, 1985. 223 с.

9. Фон Нейман Дж. Теория самовоспроизводящихся автоматов. М.: Мир, 1971. 382 с.

10. Шаратов И.П. Логический анализ некоторых проблем геологии. М.: Недра, 1977. 144 с.

11. Эшби У.Р. Теоретико-множественный подход к механизму и гомеостазису // Исследования по общей теории системы. М., 1969. С.398-441.

12. Яблонский А.И. Конференция советских ученых по системным исследованиям: методологические аспекты // Вопросы философии. 1985. № 12. С.140-146.

А.Н.Дмитриев, В.Н.Кандыба

ВЫЯВЛЕНИЕ ТАКСОНОМИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ В ЗАДАЧАХ ДИАГНОЗА

Сравнительное изучение множества геологических объектов, представленных их описаниями, зачастую сводится к анализу концептуальной модели этого множества. В самом общем смысле концептуальная модель считается заданной, если составляющие ее элементы (объекты) охарактеризованы так, что допускают сравнительное изучение по каждому элементу характеристики (признаку). В нашем случае рассматриваются концептуальные модели исходных данных, в которых строкам соответствуют объекты, а столбцам - признаки.

Длительная практика работы с концептуальными моделями геологических объектов показала, что в основе организации той или иной концептуальной модели лежит явно или неявно формулируемое требование родства объектов организуемого множества. Это родство на ранних этапах работы было выявлено и названо "критерием общности" исследуемых объектов /1,6/. По критерию общности объекты, вошедшие в данную концептуальную модель, неразличимы. Эта неразличимость, базирующаяся на выявленной группе признаков, принимающих одинаковое значение для всех объектов модели, является обоснованием исходного группирования объектов. Но она ничего не говорит о "распределении родства" между объектами (строками) моде-

ли (таблицы). На этот вопрос отвечают те или иные таксономические процедуры, которые выявляют степень родства объектов по общему основанию анализа (сходства-различия), т.е. по всему набору признаков без системы предпочтения. Именно эта непредвзятость и делает таксономию не только существенным, но и первоочередным этапом решения задач сравнительного изучения геологических объектов. Существует много подходов и алгоритмо-программных решений задач таксономии, но мы рассмотрели возможности метода согласованных оценок /3,5/ в таксономической роли. Ведь цели сравнительного изучения объектов очень часто адресуются вопросам сцепленности, взаимосвязи объектов внутри заданной таблицы. Вот почему весьма важны таксономические процедуры.

Выявление таксономических закономерностей в задачах диагноза тесно сопряжено с общей проблематикой таксономии. Применяя метод согласованных оценок в качестве основополагающего алгоритма для обнаружения таксономических закономерностей, мы модифицируем ряд процедур алгоритма и нарастим его некоторыми дополнительными математическими операциями. Содержательно под таксономической закономерностью понимается выявление такой группы признаков из общего пространства, которая на исходных данных устанавливает родственные группы объектов без систем предпочтения.

I. Вычислительные приемы

Пусть согласованные оценки строк (объектов) и столбцов (признаков) таблиц концептуальных моделей геологических объектов представляют собой неподвижную точку некоторого оператора. Для нахождения этого оператора применяется итерационная процедура/4/, в соответствии с которой способ вычисления оценок строк идет через оценки столбцов $\lambda \cdot \pi_i$, а оценки столбцов вычисляются через оценки строк $\mu \cdot \omega_j$.

$$\lambda \cdot \pi_i = \sum_{j=1}^n t_{ij} \cdot \omega_j; \quad \mu \cdot \omega_j = \sum_{i=1}^m t_{ij} \cdot \pi_i. \quad (I)$$

Нагрузки строк (π_i) и столбцов (ω_j) представляют собой первую пару собственных векторов, соответствующую максимальному собственному числу в сингулярном разложении таблицы исходных данных.

Учитывание π_i и ω_j , как показала практика, приводит к выявлению ряда структурных особенностей таблиц данных. Это выявление заложено в характере встречного пересчета столбцов и строк, который устанавливает структурные соотношения строк и столбцов. При этом пересчете цель обработки таблиц имеет руководящую роль в процедурах на последующих этапах преобразования структур, т.е. $\lambda \cdot \pi_i$ и $\mu \cdot \omega_j$ вычисляются на концептуально заданной таблице. Результаты нагрузок столбцов и строк подвергаются оценке их пригодности при построении структурной модели геологических объектов:

а) нагрузка признака x_{j1} и нагрузка объекта S_{i1} существенны в структурном отношении для концептуальной модели - гипотеза A_1 ;

б) нагрузка признака x_{j2} и нагрузка объекта S_{i2} не существенны в структурном отношении для концептуальной модели - гипотеза A_2 .

Существенность столбца или строки для преобразования исходной структуры в целеориентированную устанавливается простым соотношением нагрузок. Столбец (строка) является существенным в структурном отношении, если величина его (ее) нагрузки не ниже некоторого заданного порога значений. Причем этот порог выбирается с учетом разброса значений нагрузок (например, больше среднего значения или меньше). Выбор порога обусловлен следующим предположением.

Предположение I. Максимальное значение нагрузок строк (π_i) и столбцов (ω_j), полученное на исходной структуре табличных данных, означает максимальную связность (согласованность) строк и столбцов между собой и с другими строками и столбцами.

Следовательно, нулевые значения π_i и ω_j можно считать совпадающими с гипотезой A_2 , а если $\pi_i \neq 0$ и $\omega_j \neq 0$, то - под гипотезу A_1 . Объекты и признаки с максимальными значениями нагрузок являются носителями силы взаимосвязи объектов и признаков.

Конкретизируем предположение I. Формулы (I), если сделать их подстановки друг в друга, приводят к следующим соотношениям:

$$\lambda \cdot \mu \cdot \pi_i = \sum_{i=1}^m h_{\alpha i} \cdot \pi_{\alpha}; \quad \lambda \cdot \mu \cdot \omega_j = \sum_{j=1}^n q_{\beta j} \cdot \omega_{\beta}, \quad (2)$$

где $h_{\alpha i}$, $q_{\beta j}$ - коэффициенты, отражающие собой силу взаимосвязи строк S_{α} , S_i и соответственно столбцов x_{β} , x_j между собой.

Как видно из формул (2), чем больше эта сила взаимосвязи, тем с большим коэффициентом вес соответствующей строки (или столбца) вносит вклад в вес вычисляемой строки (столбца).

Формулы (1)-(2), как и предположение I по своему назначению очень подробно ревизуют концептуальную модель на предмет содержания в ней скрытых взаимосвязей не только между объектами, но и между признаками. Причем этот опрос синхронный по строкам и столбцам.

Но метод может быть применяем и для поиска несогласованности строк и столбцов, особенностей неструктурности исследуемых свойств. Для этой цели можно употребить следующие формулы вычисления нагрузок строк и столбцов:

$$\lambda \cdot \pi_i = \sum_{j=1}^n (1 - t_{ij}) \cdot \omega_j; \quad \mu \cdot \omega_j = \sum_{i=1}^m (1 - t_{ij}) \cdot \pi_i. \quad (3)$$

Может иметь значение также отыскание связи между особенностями структурной выраженности и невыраженности (слабой согласованности объектов и признаков), которое проводится по формуле:

$$\lambda \cdot \pi_i = \sum_{j=1}^n t_{ij} \cdot \omega_j; \quad \mu \cdot \omega_j = \sum_{i=1}^m (1 - t_{ij}) \cdot \pi_i. \quad (4)$$

Исследуемые таксономические взаимосвязи между объектами и их признаками на одной таблице исходных данных можно распространить на случай решения задач с двумя таблицами.

Пусть таблице исходных данных T_1 поставлен в соответствие естественно выделенный класс объектов (в нашем случае список месторождений нефти, представленный в табличном виде) из m_1 строк и n столбцов. Второй класс объектов (месторождения газа) представлен таблицей T_2 с m_2 строк. Тогда для изучения различий и сходств заданных классов потребуется:

Предположение II. Максимальные нагрузки столбцов и строк составляют характеристику меры согласованности объектов и признаков исследуемой таблицы; меры согласованности различных таблиц — не сходны, они выявляют взаимосвязи строк и столбцов исследуемых таблиц.

Для исследования различий отношений взаимосвязей объектов двух классов необходимо на представляющих их таблицах оценить нагрузки столбцов и строк по (I). Выявленные значения нагрузок

позволяют провести сравнение мер согласованности классов, для этого потребуется:

Определение I. Столбцы, имеющие максимальное значение нагрузок (по заданному порогу) для I класса (T_1) и минимальное значение для II класса (T_2) называются контрастными. Столбцы, имеющие средние значения нагрузок (выходящие за порог контрастности), для I и II классов называются неконтрастными. Столбцы, разница нагрузок которых по I и II классам близка к нулю, называются равновесными по нагрузкам.

Тогда с учетом предположения II и определения I, для решения задачи различения объектов двух классов (представленных T_1 и T_2) формулируется.

Предположение III. Лучшей разделяющей способностью для объектов двух классов обладают контрастные по нагрузкам признаки; лучшей способностью к выявлению подобия для объектов двух классов обладают равновесные по нагрузкам признаки.

Подчеркнем, что выявление различия и подобия исследуемых объектов во многом исчерпывает цели сравнительного изучения геологических объектов. На основании изложенных определений и предположений проведено нижеследующее конкретное решение вопросов по выявлению таксономических закономерностей в задачах диагноза.

2. Алгоритм выявления таксономических закономерностей

Выдвигаемый алгоритм анализа табличных данных нацелен на решение таксономических и диагностических задач. Напомним, что мы рассматриваем таксономию как разновидность группирования объектов по полному списку характеристических признаков, т.е. без особо выделенных классифицирующих признаков. Эта трактовка таксономии позволяет нам наряду с решением общих задач анализа данных заниматься вопросами таксономического характера. Причем зачастую таксономический результат в методе согласованных оценок можно получить попутно. Если же в геологической постановке задачи содержится вопрос о группировании исследуемой совокупности объектов по полному основанию (по всему списку признаков), то задача выделения "таксонов" сводится к выявлению групп объектов, весовые нагрузки которых не превосходят заданного порога. Порог ус-

танавливается на основе количественных оценок разброса весовых нагрузок. Это и естественно, поскольку метод не только выявляет характер взаимной сцепленности объектов (в зависимости от структуры исходных табличных данных), но и группирует их по близости нагрузок. Эти процедуры и интерпретация не противоречат общей идеологии таксономии. Нижеследующий алгоритм ориентирован как для нужд таксономии, так и для задач распознавания в двухклассовой постановке.

2.1. Алгоритм решения

Остановимся более детально на вопросах таксономического порядка. Причем несмотря на двухклассовую постановку задачи, сосредоточимся на процедурах группирования объектов внутри данного класса. Учитывая особенность метода согласованных оценок подчеркнем, что в данном случае таксономия осуществляется "явным" образом, т.е. исследователь знает, какие конкретно столбцы (признаки) играют решающую роль в группировании объектов. Разработана такая алгоритмическая схема.

1. Вычисляются нагрузки строк (π_i и столбцов (ω_j) для каждой из таблиц T_I и T_{II} по методу согласованных оценок. Причем, T_I - объекты первого класса, T_{II} - объекты второго класса, T_I и T_{II} - эталоны. В соответствии с требованием задачи вычисляются разности нагрузок столбцов между нагрузками для первого и второго классов. Эта разность выявляет структурные различия классов:

$$|\Delta I_1^I| = \omega_1^I - \omega_1^{II}, \dots, |\Delta I_n^I| = \omega_n^I - \omega_n^{II}$$

$$|\Delta I_1^{II}| = \omega_1^{II} - \omega_1^I, \dots, |\Delta I_n^{II}| = \omega_n^{II} - \omega_n^I.$$

Далее ΔI_1^I и ΔI_1^{II} упорядочиваются по величине

$$\Delta I_k^I > \Delta I_1^I > \Delta I_m^I; \Delta I_k^{II} > \Delta I_1^{II} > \Delta I_m^{II}$$

причем, происходит одновременно и сортировка признаков по их способности различать внутреннее сходство - различие объектов исследуемых классов. Для этого задается порог величины разностей нагрузок столбцов. Поскольку $0 \leq \pi_i \leq 1$, то целесообразно этот порог задать как $\Delta I_q^I = 0,5$ и $\Delta I_q^{II} = 0,5$, что не противоречит и широкому численному эксперименту. В случаях же (практически ред-

ких) если значения π_i лежат, например, в интервале $0 \leq \pi_i \leq 0,5$, то принимается простейшее решение: $\Delta I_q^I = 0,25$ и $\Delta I_q^{II} = 0,25$. Тогда все столбцы T_I и T_{II} , для которых разность $\Delta I_i > 0,5$, образуют разделяющую группу признаков, а совокупность признаков, для которой разность меньше 0,5, образует группу признаков, отождествляющую объекты двух классов.

При необходимости в процессе решения задачи распознавания формируются подтаблицы таблиц T_I и T_{II} :

T_I^P и T_{II}^P - таблицы, составленные из упорядоченных по разности столбцов, которым соответствует различающая группа признаков, для I и II класса соответственно;

T_I^O и T_{II}^O - таблицы, составленные из упорядоченных по величине разности столбцов, которым соответствует отождествляющая группа признаков, для I и II классов соответственно.

Выделенные таким образом подтаблицы имеют другую структуру табличных данных, а именно подтаблицы T_I^P и T_{II}^P собрали признаки, максимально характеризующие индивидуальность первого и второго классов, а T_I^O и T_{II}^O отражают общность структур первого и второго классов.

2. Производится диагностика на T_I и T_{II} (с внутренним экзаменом /2/) с записью результатов распознавания на материале обучения. Далее производится распознавание проб (объектов третьего класса T_3) на общем исходном пространстве признаков с записью результатов распознавания и вычислением процента распознавания. Вслед за этим осуществляется процедура распознавания на подтаблицах T_I^P и T_{II}^P и сравниваются результаты распознавания.

В случае распознавания по T_I^P и T_{II}^P появляются исходы:

Полное (безошибочное) распознавание, при этом следует перейти к минимизации пространства признаков путем удаления последнего признака в упорядоченном ряду столбцов по величине разности Δ_i . Разделяющее пространство признаков считается минимальным, если оно характеризуется безошибочным распознаванием и не содержит избыточных признаков (убирание хотя бы одного признака приводит к ошибке распознавания). Неизбыточное безошибочное распознающее пространство признаков является базой для построения узких таблиц исследуемых классов, в которых сконцентрированы максимально разделяющие признаки; узкие таблицы классов T_I и T_{II} являются структурно не сходными, различными.

Распознавание с ошибкой на материале обучения производится в сторону наращивания пространства признаков путем добавления первого из оставшихся признаков отождествляющей группы. Добавляемые признаки видоизменяют структуру таблиц и если происходит ошибочное распознавание, то такое пространство признаков (полученное расширением различающей группы признаков) называется широким разделяющим пространством признаков.

В случае когда безошибочного распознавания не происходит, то устанавливается группа признаков, по которой получен наилучший результат и такая группа называется оптимальной группой признаков.

3. Распознавание по оптимальной группе признаков производится с коррекцией объектов в классах, т.е. из классов, представленных таблицами T_I и T_{II} , удаляются не распознающиеся объекты. Как правило, такие объекты представляют собой отдельную, характеристически обособленную группу. Вывод этой группы приводит к исправлению априорной классификации исследуемых объектов. Минимизация T_I и T_{II} по объектам производится до безошибочной диагностики. Эта коррекция обучающего материала оказывается практически полезной и имеет, как правило, интересные интерпретации. Кроме того, пообъектная минимизация "гибридизирует" таксономию и диагностику, именно здесь появляется возможность говорить о выявлении таксономической закономерности. Эта закономерность представляет собой некоторое "ядро" объектов таблицы, причем устойчивость этого ядра поддерживается максимально таксономизирующими признаками, своеобразная "двумерная таксономия".

После достижения стопроцентного распознавания на материале обучения производится сортировка объектов, представленных таблицей проб T_{III} . При этом класс проб может состоять из объектов, принадлежность которых к классу заранее известна. В этом случае производится решение задачи в режиме "внешнего" экзамена. Если объекты в T_{III} заданы так, что их принадлежность к классу заранее неизвестна, то решается задача прогнозирования, т.е. производится диагностика проб на материале обучения, заданном T_I и T_{II} . В ряде случаев полезно диагностику объектов в прогнозных целях производить после процедур внешнего экзамена. Это особенно желательно, если число объектов в обучающей совокупности объектов достаточно велико ($m > 10$, m - число объектов в классе; тогда

оставшиеся объекты можно пустить на внешний экзамен). Естественно, что организация классов обучения должна учитывать требование типичности объектов для данного класса.

Экзамен проб может осуществляться по заранее заданной схеме диагноза, составляемой с учетом характера этапа обучения (включая и экзамен). Программа решения учитывает все возможные варианты диагноза, а решение ведется по той или иной схеме, т.е. прогнозный диагноз может вестись по:

- а) узкому разделяющему пространству признаков;
- б) широкому разделяющему пространству признаков;
- в) узкому разделяющему пространству признаков на скорректированном составе объектов в T_I^K и T_{II}^K ;
- г) широкому разделяющему пространству признаков на скорректированном составе объектов в T_I^K и T_{II}^K .

В связи с задачами обнаружения закономерностей и оценки избыточности (или недостатка) исходных данных по сравнению с новыми целевыми структурами таблиц, полезно проводить сравнение полученных результатов (с минимизацией столбцов и строк) с результатами таблиц по исходным данным.

2.2. Выявление области неопределенности

Совокупность признаков, отождествляющих объекты первого и второго классов по существу представляет собой общую часть структур исследуемых классов. Эта общая структурная часть T_I и T_{II} является областью неопределенности в целях диагноза.

Рассматривается совокупность смешивающих признаков, выявленная анализом структуры таблиц T_I и T_{II} путем оценки \mathcal{F}_i и ω_j . Цель этого рассмотрения состоит в том, чтобы обнаружить необходимое и достаточное количество столбцов, по которым заданные классы различаются не лучше случайного уровня, т.е. смешивающие признаки, содержащиеся в исходных данных в неявном виде, выделяются, и таблицы T_I и T_{II} сливаются в один класс. Естественно предположить, что каждый признак, который является носителем неопределенности в отношении его способности разделять классы, сосредотачивает в себе информацию другого порядка внецелелеполющенную. Эта неопределенность выявляется методом согласованных оце-

нок путем обнаружения межтабличных связей (T_I и T_{II}) и оценкой структурной существенности каждого признака.

Математические вопросы обнаружения и оценки неопределенности освещены в /4/, здесь же мы укажем на практическую часть исследования:

а) решается задача по оценке различия класса эталонов и класса проб путем количественной оценки неопределенности для каждой пробы к сводной таблице эталонов и проб; эта оценка ранжирует все пробы по степени их "непохожести" на эталоны;

б) полагая, что таблица эталонов, представляющая собой концептуальную модель геологических объектов, содержит в себе скрытые закономерности взаимосвязей объектов, примем, что получившая максимальную оценку на несходство с эталонами, будет одновременно и максимально неопределенной в отношении ее принадлежности к эталонам.

Ранее разработанный и примененный подход к решению задачи неопределенности путем восстановления фиктивных пропусков для объектов /9/ дополняется, таким образом, более общей моделью выявления областей неопределенности в таблицах исходных данных. Кроме того, насыщение признаков с максимальной неопределенностью делает более сложной и менее надежной задачу таксономии. Выявление таксономических закономерностей для таблиц с неопределенностями становится трудоемким делом, но наличие такого факта говорит о необходимости анализа исходной концептуальной модели.

3. Задача оценки различий нефтяных и газовых месторождений

Цель этого раздела состоит в иллюстрировании конкретного решения задачи, тесно связанной с вопросами количественного анализа взаимосвязей строк и столбцов в заданных таблицах. Отметим, что схема решения задачи имеет явно гибридную природу таксономического и диагностического характера. Таксономический раздел содержится в части уточнения априорной классификации по общему основанию (по всему набору признаков), а задача распознавания решается в режиме адаптации (минимизация признаков и объектов, выявление неопределенности). В связи с методическим характером ра-

боты геологическая часть интерпретации результатов вынесена за пределы этой статьи.

Важность и трудоемкость задач отдельного прогноза месторождений нефти и газа подчеркивалась многими авторами /6,8/. Часть этих задач решается с помощью математических методов и вычислительных машин. В данной работе частная задача отдельного прогноза решается методом согласованных оценок /1,3/. Эта работа методически и содержательно примыкает к предыдущим исследованиям.

Задача решалась в геологической постановке и информационном обеспечении В.С.Вышемирским /2,7/. Таблица исходных данных состояла из 36 месторождений (строк) и 182 характеристических признаков (столбцов). Выборка объектов подразделена на два класса: I - нефтяные месторождения ($m^I = 21$) и II - газовые месторождения ($m^{II} = 15$).

Пространство признаков подразделялось на 5 подпространств по аналогии с /6/, но в формализованной постановке задачи это во внимание не принималось. Элементы t_{ij} таблицы, стоящие в пересечении i -й строки и j -го столбца, соответствуют объекту S_i и признаку x_j и равны

$$t_{ij} = \begin{cases} 1, & \text{если } x_j \text{ выполнен на } S_i \\ 0, & \text{если } x_j \text{ не выполнен для } S_i. \end{cases}$$

Ставилась задача полного (100 %) разделения нефтяных и газовых месторождений по минимально возможному количеству признаков. Разделение по всем 182 параметрам считалось неэффективным, поскольку трудоемкость обеспечения информацией объектов прогноза по полному списку признаков крайне велика. Для этого потребовалось бы изучить прогнозируемые объекты весьма подробно, при этом задача "прогноза на ранней стадии изученности объектов" вырождается в поисково-разведочную.

В плане специфики постановки задачи и таксономических возможностей метода в общем ходе решения предусматривалась коррекция априорной классификации. Эта процедура выявления зачастую сложного группирования объектов естественна для таксономии и наряду с широко практикуемой минимизацией признаков составляет основу таксономического этапа. Поскольку метод выявляет сцепленность всех признаков со всеми объектами, то характер близости оценок строк (объектов) и столбцов (признаков) вводит элемент

(таксономии) и позволяет корректировать естественную (априорную) классификацию. Именно в связи с двойной минимизацией (объектов и признаков) возникла необходимость разделения решения задачи на два взаимосвязанных этапа: первый – решение задачи с минимизацией количества признаков; второй – решение задачи с минимизацией количества объектов.

3.1. Первый этап решения

Решение для обоих этапов проводилось с помощью нецентрированного варианта метода согласованных оценок. На первом этапе и шаге решения проводилась оценка нагрузок объектов и признаков по полной таблице исходных данных ($m \times n = 36 \times 182$). При этом в первом классе имелось 21 месторождение нефти, во втором – 15 месторождений газа. Исследуемые классы объектов (нефтяные и газовые месторождения) представляли собой не только эталонную выборку, но и принимались в качестве объектов для внутреннего экзамена.

В результате распознавания на процедурах внутреннего экзамена разделение объектов двух классов составило 72 % (по всему списку признаков, $n = 182$). Правильно к своим классам отнеслись объекты 1–12, 14–17 (I класс) и 22–27, 29, 30, 32, 33 (II класс). Остальные объекты попали в область пересечения.

Согласно принятой схеме решения на первом этапе (рис.1) проводились последующие шаги решения с целью достижения 100 % разделения нефтяных и газовых месторождений. В связи с предположением о том, что контрастность признаков (max нагрузка по I и min по II классам и наоборот) играет существенную роль в оценке расстояния между классами, была проведена сортировка признаков ($n = 182$) по характеру нагрузок как внутри, так и между классами. При этом выделились группы (табл.1):

а) все контрастные признаки (у которых разница между нагрузками по I и II классам превышала 0,4, т.е. треть максимального веса признака), число таких признаков составило 51;

б) все равновесные признаки, числом в 52 признака (у которых разница весов составила до 0,15);

в) все неконтрастные признаки (не вошедшие в состав контрастных и равновесных) – 79 признаков.

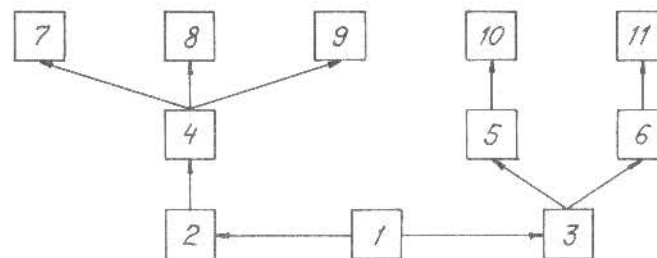


Рис.1. Схема первого этапа решения:

1 – решение с минимизацией состава признаков; 2 – общее подразделение состава признаков; 3 – контрастные по преимуществу класса; 4 – общий список признаков; 5 – преимущество по первому классу; 6 – преимущество по второму классу; 7 – все контрастные признаки; 8 – все равновесные признаки; 9 – все неконтрастные признаки; 10 – контрастные по первому классу; 11 – контрастные по второму классу.

По каждой выделенной группе признаков проведено решение, при этом выявилась их различная диагностическая существенность. Контрастная группа признаков дала наиболее высокий процент разделения нефтяных и газовых месторождений (70 %). Минимальный процент разделения (28 %) дала группа равновесных признаков. Группа неконтрастных признаков дала средний процент разделения (53 %).

Таким образом, предположение о существенности признаков, имеющих максимальный вес для данного класса, подтвердилось практически в задаче распознавания. Следовательно, для характеристики одного класса целесообразно пользоваться признаками с максимальным весом, а при разделении объектов на два класса следует применять максимально контрастные признаки. Похожесть объектов двух классов следует выявлять группой равновесных признаков.

По мере исследования разделяющей способности группы признаков и в поисках максимально различающей группы признаков потребовалось изучить влияние на разделение преимуществ весов признака одного класса над другим. Для этого контрастные признаки были, в свою очередь, подразделены на признаки с преимуществом

Таблица 1

Номер л/п	Содержание таблиц исходных данных	Количество объектов		Количество признаков		Распознавание, %		Коррекция априорной классифи- кации
		I класс	II класс	Число	Процент	общее	на I признак	
1	Общий список признаков	21	15	182	100	72	0,4	По признакам
2	Все контрастные признаки	21	15	51	28	70	1,4	
3	Все равновесные признаки	21	15	52	29	28	0,5	
4	Оставшиеся признаки	21	15	79	44	53	0,7	
5	Контрастные признаки по преимуществу I кл.	21	15	26	14	58	2,2	
6	Контрастные признаки по преимуществу II кл.	21	15	25	14	64	2,6	
7	Все признаки по преиму- ществу I кл.	21	15	89	48	61	0,7	
8	Все признаки по преиму- ществу II кл.	21	15	91	50	56	0,6	
9	Минимизация объектов по I классу	11	15	182	100	88	0,5	По объектам
10	Минимизация объектов по I и II классам	11	10	182	100	90	0,5	
11	Решение без первых II объектов I кл.	10	15	182	100	92	0,5	
12	Все возрастные признаки по I классу	11	15	38	20	92	1,6	
13	Контрастные признаки без первых II объектов I кл.	10	15	37	20	100	2,7	
14	Контрастные признаки по I классу	11	15	22	12	62	0,6	
15	Контрастные признаки по II классу	11	15	89	49	58	0,5	

по весу: I - I класса - 26 признаков; 2 - II класса - 25 признаков. В полном признаковом пространстве также были выделены признаки с преимуществом по весу: 3 - по I классу - 87 признаков; 4 - по II классу - 91 признак. По группам признаков (I - 4) разделение следующее (%) (см. рис.1): I - 58; 2 - 64; 3 - 61 и 4 - 56.

Здесь уместно обратить внимание на более высокую разделяющую силу контрастных признаков. Так, индивидуальный вклад в диагностику каждого контрастного признака составил 1,2%, а каждый признак по преимуществу веса по классу дал вклад всего 0,3%. Рассматривая общие результаты решения данного этапа, можно отметить следующее:

1. Выявлена совокупность контрастных признаков ($n = 51$), которая пригодна для изучения различий нефтяных и газовых месторождений.

2. Выявлена совокупность равновесных признаков ($n = 52$), которая пригодна для изучения сходств нефтяных и газовых месторождений.

3. Подтвердилось положение I о разделяющей способности контрастных признаков и роднящей способности контрастных признаков и роднящей способности равновесных признаков.

3.2. Второй этап решения

Этот этап решения оказался целесообразным по двум причинам: во-первых, согласно целеуказанию, требуется провести полную разделимость нефтяных и газовых месторождений по исходным данным; во-вторых, в соответствии с таксономическими способностями метода, оказалось возможным симметризовать процедуру минимизаций признаков минимизацией объектов.

Анализ состава объектов обоих классов, по нагрузкам строк на первом этапе, показал тот же характер подразделения, что и для столбцов. При этом обнаружилось два основных вида объектов: контрастные - хорошо различающиеся между собой месторождения нефти и газа (с I-II номер в составе I класса и 22-27 номер для II класса) и равновесные месторождения, попадающие на пересечение значений нагрузок для обоих классов. Характерно, что это подразделение объектов в изучаемых классах весьма устойчиво и подт-

верждается всеми шагами решения первого этапа. Именно поэтому минимизация объектов началась с удаления контрастных (хорошо разделяющихся) объектов.

На втором этапе (см. рис.2) решение состояло из шести шагов. На первом шаге минимизации объектов было принято решение удалить из таблицы первые 10 из 11 родственных объектов Аравийской платформы, оставив наиболее типичный. Объекты второго класса не изымались, т.е. таблица решения состояла из 11-21 объектов I класса и 22-36 II класса.

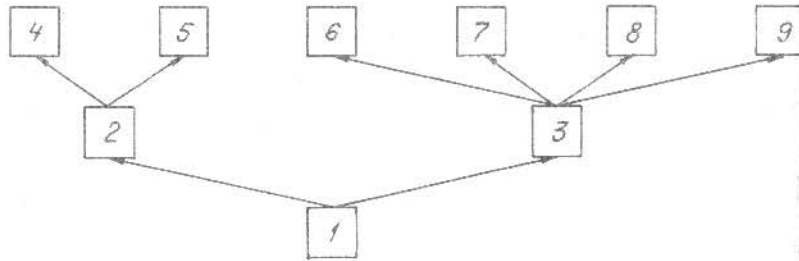


Рис.2. Схема второго этапа решения:

1 - решение с минимизацией состава объектов; 2 - коррекция априорной классификации объектов в полном признаковом пространстве; 3 - решение скорректированной классификации по контрастным признакам; 4 - минимизация объектов по первому классу; 5 - минимизация объектов по второму классу; 6 - все контрастные признаки; 7 - без объекта № 11 первого класса; 8 - признаки, контрастные по первому классу; 9 - признаки, контрастные по второму классу.

Для проверки работы признакового пространства решение проведено по полному составу признаков (n = 182). При этом было достигнуто разделение 88 %, т.е. большая часть трудно различимых объектов при исправлении априорной классификации распознана верно (решение на первом этапе по полному составу признаков составило 72 %). В I класс правильно попали объекты 12-17, 19-21, во второй - 22-33,36. Если же учесть двухступенчатость схем распознавания, то суммарное разделение (по первому и второму

Номер решения	Процент распознавания	Количество признаков по группам				
		наднефтяная свита	продуктивная толща	поднефтяная свита	структурная ловушка	геотектоническая обстановка
1	72	28/100	50/100	32/100	27/100	45/100
2	70	10/36	12/24	8/25	1/4	20/44
3	28	5/18	17/34	5/16	13/48	11/24
4	53	13/46	21/42	19/59	13/48	13/29
5	58	4/14	0/0	5/16	0/0	17/38
6	64	6/21	12/24	3/9	1/4	3/7
7	61	16/57	15/30	19/59	8/30	31/69
8	56	12/43	35/70	13/41	18/67	13/29
9	88	28/100	50/100	32/100	27/100	45/100
10	90	28/100	50/100	32/100	27/100	45/100
11	92	28/100	50/100	32/100	27/100	45/100
12	92	12/43	12/24	4/13	4/15	6/13
13	100	6/21	14/28	6/19	4/15	7/16
14	62	8/29	5/10	0/0	4/15	5/11
15	58	14/50	33/66	17/53	12/44	13/29

этапу этого шага) составило 92 %. Причем улучшение получено за счет минимизации объектов по I классу. Дальнейшее улучшение распознавания следовало наращивать за счет исправления состава II класса. Для этого из состава II класса были изъяты объекты 22-26.

На очередном шаге решения таблица была задана объектами 11-21 по I классу и объектами 27-36 по II классу. Решение дало 90 % разделения месторождений по полному составу признаков (n = 182). Область пересечения составили 11 объект первого класса (оставшийся представитель родственной группы месторождений Аравийской платформы) и 34 объект II класса. На третьем шаге второго этапа решения задачи таблица состояла из 12-21 объектов I класса и из 22-36 объектов II класса. Процент разделения по всему составу признаков составил 92 % (табл.2).

В связи с достигнутым насыщением процента распознавания (минимизацией объектов) по полному составу признаков (n = 182), ока-

залось целесообразным провести решение таблиц в соответствии с введенным подразделением признаков. Согласно цели исследования для нас важной оказалась группа контрастных признаков. Поэтому на четвертом шаге решения второго этапа были взяты контрастные признаки ($n = 38$) и объекты, вошедшие в состав первого шага решения данного этапа. Получено 92 % разделения объектов, что на 4 % выше, чем для первого шага. Таким образом, наше предположение о наличии "шумящих" признаков в общем составе признаков ($n = 182$) снова подтвердилось.

На пятом шаге решалась таблица с объектами третьего шага второго этапа (без объекта II) числом контрастных признаков, равным 37. На этом шаге получено 100 % разделения как по результату второй ступени распознавания, так и по сумме двух ступеней распознавания (см. табл.2).

Для проверки сделанных предположений проведено два следующих решения. Шестой шаг, как и четвертый, имеет тот же состав эталонов, но решение проводилось не по всем контрастным признакам, а по тем из них ($n = 22$), которые имеют более высокий вес по I классу. Полученное разделение (62 %) оказалось гораздо ниже, чем по всем контрастным признакам. На седьмом шаге решалась таблица с тем же составом объектов, что и на предыдущем шаге, но по тем контрастным признакам, которые по II классу имели более высокий вес ($n = 89$). Процент правильного распознавания оказался еще меньше - 58.

Таким образом, предположение о высокой диагностической сущности всех контрастных признаков подтверждено конкретными расчетами и на практической задаче.

Если в случае разделяющих признаков, каждый из них является нагруженным в смысле различения классов, то смешивающие признаки выявляют общие места в структуре таблиц, т.е. отождествляют классы, очерчивая таким образом наиболее неразличимые области таблиц. Эта способность выявлять области различия и сходства таблиц содержится в алгоритме и базируется на поэлементном учете структуры таблиц (пример дифференциации признаков по разности нагрузок приведен в табл.3).

Ранжированные смешивающие признаки могут употребляться для выяснения подпространства минимально различающих признаков. Таким путем можно выявить область максимальной неопределенности.

Номер п/п	К л а с с		Разность $ \Delta $
	I(ω_1^I)	II(ω_1^{II})	
I	0,09	0,27	0,18
2	0,24	0,34	0,10
3*	0,48	0,46	0,02
4	0,52	0,07	0,45
5**	<u>0,01</u>	<u>0,83</u>	<u>0,82</u>
6	0,04	0,35	0,31
7	<u>0,15</u>	<u>0,69</u>	<u>0,54</u>
8	<u>0,18</u>	0,21	0,03
9	0,27	0,56	0,29
10	<u>0,73</u>	<u>0,11</u>	<u>0,62</u>
<u>11</u>	<u>0,11</u>	<u>0,75</u>	<u>0,64</u>
12	0,08	0,10	0,02
13	0,14	0,18	0,04
14	0,43	0,89	0,46
<u>15</u>	<u>0,94</u>	<u>0,14</u>	<u>0,80</u>

* Признаки: 3,8,12,13 - максимально смешивающие (отождествляющие).

** Признаки: 5,7,10,11,15 - максимально разделяющие.

Она устанавливается группой столбцов максимально неразличающих объекты двух классов, представленных T_1 и T_2 . Для этого производится поиск группы признаков, которая максимизирует ошибку распознавания, т.е. требует найти избыточное смешивающее пространство признаков. Оно формируется из смешивающих признаков, которые не улучшают и не ухудшают распознавания (достигшего случайного уровня).

Заключение

I. Применение метода согласованных оценок с целью выявления таксономических закономерностей в задачах диагноза нарацаива-

ет значение таксономии при сравнительном изучении множества геологических объектов. Это наращивание идет в сторону выявления таксономических "ядер", которые и концентрируют в себе закономерности взаимосвязей объектов в исследуемом классе. Устойчивость таксономического ядра обеспечивается не только максимальной схожестью объектов по заданным характеристикам, но и выделением группы признаков, которые максимизируют эту схожесть.

2. В рамках предложенного метода оказалось, что таксономическая подготовка материала обучения к диагностическим процедурам на своих заключительных этапах совершенно органично переходит к процедурам распознавания. Таксономический этап переорганизовывает таблицы исходных данных в сторону повышения их закономерности, путем исключения из состава таблиц "шумящих" строк (объектов) и столбцов (признаков). Причем это продвижение исходных концептуальных моделей к целеориентированным сопровождается наращиванием именно таксономических закономерностей.

3. Осуществлено конкретное решение задачи по сравнительному изучению нефтяных и газовых месторождений. Обнаружена слабая степень родства объектов в обучающих классах и избыточность пространства признаков для целей разделения классов. Последовательными шагами решения выявлены таксономические "ядра" исследуемых классов, на базе которых получено полное разделение классов. Эта таксономическая подготовка таблиц исходных данных по заданным классам позволила не только осуществить требуемую диагностику объектов и установить область неопределенности, т.е. выявить группу признаков смешивающих классы, что резко повышает интерпретационные возможности по составу объектов и их признаков.

Нам приятно поблагодарить профессора В.С.Вышемирского за постановку геологической задачи и любезно предоставленный материал обучения и экзамена, а также математиков С.В.Макарова и Е.А.Смертина за ряд своевременных замечаний и помощь в работе.

Литература

1. Васильев Ю.Я., Дмитриев А.Н. Спектральный подход к сравнению объектов, охарактеризованных набором признаков // Докл. АН СССР. 1972. Т.206, № 6. С.1309-1312.

2. Вышемирский В.С., Соколов А.Д. Оценка различий нефтяных и газовых гигантских месторождений логико-математическими методами // Геология и геофизика. 1986. № 2. С.51-57.

3. Дмитриев А.Н., Смертин Е.А. Закономерности, исследуемые методом согласованных оценок // Методы и программное обеспечение обработки информации и прикладного статистического анализа на ЭВМ. Минск, 1985. С.18-19.

4. Макаров С.В. Восстановление одиночного пропуски в таблице "объект-признак" на основе сингулярного разложения данных // Логико-математические методы в геологических исследованиях. Новосибирск, 1986. С.25-29.

5. Метод согласованных оценок / А.Н.Дмитриев, С.В.Макаров, Е.А.Смертин и др. Новосибирск; ИГиГ СО АН СССР, 1982. 133 с.

6. Трофимук А.А., Вышемирский В.С., Дмитриев А.Н. и др. Распознавание образов гигантских нефтяных месторождений // Проблемы нефтегазовости Сибири. Новосибирск, 1971. С.34-50.

7. Трофимук А.А., Бишаев А.А., Вышемирский В.С. и др. Диагностика нефтяных и газовых месторождений логико-математическими методами // Условия раздельного формирования зон нефте- и газонакопления в земной коре. Тюмень, 1978. С.20-28.

8. Условия раздельного формирования зон нефте- и газонакопления в земной коре / А.А.Трофимук, А.А.Бишаев, В.С.Вышемирский и др. Тюмень, 1978. 152 с.

9. Федосеев Г.С. Восстановление пропусков в таблицах данных // Логико-математические методы в геологических исследованиях. Новосибирск, 1986. С.29-42.

В.В.Бабич, В.И.Лебедев, Г.С.Федосеев

ТАКСОНОМИЧЕСКИЙ АНАЛИЗ КЛАССИФИКАЦИИ
КОБАЛЬТОНОСНЫХ ОБЪЕКТОВ ЦЕНТРАЛЬНОЙ АЗИИ

В практику металлогенических исследований прочно вошел рудно-формационный анализ, нацеленный на выявление важных типов месторождений, установление закономерностей их размещения и ло-