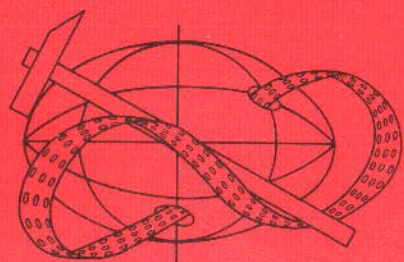


550

Л 694

АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ГЕОЛОГИИ И ГЕОФИЗИКИ

ЛОГИКО-ИНФОРМАЦИОННЫЕ
ИССЛЕДОВАНИЯ
В ГЕОЛОГИИ



НОВОСИБИРСК—1977

550 + 51

Л 694

А Н Н О Т А Ц И Я

Данный сборник тематически специализирован, поскольку в нем рассматриваются математические вопросы и практическое применение тестового подхода в исследовании геологической информации.

В первой части сборника приводятся работы теоретического характера, во второй даны примеры решения конкретных задач в новых методических направлениях, продолжающих тестовый подход. Основная идеология математических разработок связывается с задачами сравнительного изучения описаний геологических объектов.

Работа предназначена для специалистов по математической обработке информации. Она будет полезна тем, кто изучает математические методы в геологии в связи с решением задач прогнозно-поискового профиля. Кроме того, особенно в своей теоретической части, сборник поможет геологам ориентироваться в логической структуре сравнительного изучения объектов широкого профиля.

Редакционная коллегия

Федосеев Г.С., Красавчиков В.О., Штанова Т.И.

Ответственный редактор А.Н.Дмитриев



Печатается по решению секции
стратиграфии, литологии и осадочных
полезных ископаемых Ученого Совета
Института геологии и геофизики

СО АН СССР

© Институт геологии
и геофизики СО АН СССР
1977 г.

I. ТЕОРЕТИЧЕСКИЕ ВОПРОСЫ

А.Н.Дмитриев, В.О.Красавчиков

ТЕСТОВЫЙ ПОДХОД В РЕШЕНИИ ПРОБЛЕМ ОБРАБОТКИ
ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Введение

По мере возрастания сложности и скорости решения геологических проблем (в областях ресурсного обеспечения, подземного и надземного строительства, теоретического и экспериментального исследования общепланетарных систем и процессов) геологические исследования становятся неизбежно комплексными. Особенно острне обстановки возникают в геологии в местах, где требуется точное и оперативное обращение с большими информационными массивами. Именно в качестве средства преодоления этих специфических трудностей (возникающих в связи с увеличением информационного потока от технических средств исследования геосистем) и созданы новые способы обработки геологической информации разнообразной природы. К настоящему времени геология (отечественная и мировая) располагает довольно обширным арсеналом математических методов для автоматической обработки данных по довольно широкому перечню целей исследования. В данной работе мы освещаем один из таких подходов в работе с информацией, который возник и развивается на грани геологии и теоретической кибернетики, вернее, в русле идей кибернетики, дающих начало прикладным математическим разработкам в сфере естествознания.

Цель работы многозначна. С одной стороны она суммирует проделанные этапы работы, с другой, освещает новые направления развития и иллюстрирует ряд новых результатов по тестовому подходу. Кроме того, приведена достаточно строгая систематизация типичных постановок в классе решаемых геологических задач и алгоритмов их решения. В целом изложение подчинено нуждам геологов, в связи с чем авторы стремились показать разнообразие возможностей тестовых методов при сравнительном изучении геологических объектов. Этими целями вызвана специфика языка изложения материала и характер подбора литературы.

§ I. Основная характеристика области применения тестового подхода в геологии

Неотъемлемой частью развития современной геологии как в ее производственных, так и в научно-исследовательских сечениях, является глобализация и тотализация изучения геологических тел и процессов. Именно поэтому большая часть геологических проблем преобразуется в информационные проблемы. Причем эти проблемы могут возникать и формулироваться как в недрах классических геологических постановок (например, задачи ритмокомплексов или построение новых классификационных подразделений в палеонтологии), так и в проблемах, возникающих в геологии сегодня (например, вопросы космогеологии или острые геоэкологические задачи, требующие точного и оперативного решения). Сузим задачу нашего рассмотрения геологических проблем, сводимых к информационным проблемам, до задач прогнозно-поискового профиля.

Указанный ракурс "поисковых геологических проблем" является иллюстрацией того, что информационный подход по своему существу способен преодолеть стандарт целей геологического поиска и поставить перед геологической наукой совершенно новый набор целей, подлежащих реализации. То есть геологический уровень информации уже содержит зачатки к выходу на новый вид геологического поиска, а именно, поиск геологических последствий результатов промышленно ориентированных геологических разработок полезных ископаемых [58] .

Действительно, рассмотрение геологического поиска, как информационной проблемы обладает рядом постановочных преимуществ, среди которых отметим:

- 1) естественную широту охвата поисковых проблем, формулируемых в сообщениях о фактах общепланетарного масштаба;
- 2) возможность количественной оценки совокупности еще неоткрытых месторождений перспективных веществ;
- 3) общепланетарную оценку количества того или иного вещества и установление нормы изъятия этого вещества согласованием

темпов и жизненной целесообразности его промышленной переработки;

4) постановка задач чисто геологического плана о характере распределения вещественных аномалий планеты;

5) выяснение геологической роли рудных и углеводородных скоплений, а также их пространственно-временная локализация на эволюционной траектории Земли;

6) отыскание зон поврежденности общепланетарного равновесия и естественное уравнивание промышленно и научно стимулируемых геологических поисков.

С учетом вышеизложенного, геологический поиск в его информационных смыслах содержательно нами трактуется довольно широко.

§ 2. Некоторые вопросы состояния информационного направления исследований в геологии

Теоретические поиски и организация практики при геологических исследованиях движимы как экологическими стимулами развития, так и внутренними нуждами развития широкого перечня наук о Земле. Независимо от конкретного стимула каждая последующая фаза возрастания геологического знания зависит от общей интеллектуальной и инструментальной оснащенности человеческой деятельности. Новые качественные и количественные возможности развития геологии возникли в связи с инструментальной вооруженностью разума. В последнее время эта вооруженность разума, воплотившаяся в видовое разнообразие вычислительных машин, обеспечила своеобразную индустриализацию почти всех отраслей деятельности, включая и практическую геологию (с ее экономическими целями).

Естественность и необходимость внедрения автоматических средств получения и исследования геологической информации выявилась в середине пятидесятых годов. Потребовалось еще десятилетие для реализации этой необходимости в конкретных практических и теоретических задачах некоторых разделов геологии [57] . По мере получения первых обнадеживающих результатов решения задач и совершенствования автоматических средств получения и исследования информации возрастала глубина задач, подлежащих решению, и увеличивалось их геологическое разнообразие. Начали соз-

ревать некоторые предпосылки для попыток создания "теоретической геологии". Эти предпосылки были замечены и претворены в конкретные исследования. В одних случаях база "теоретической геологии" проецировалась на содержательной геологической основе и, с той или иной долей уместности, обрамлялась формализованными средствами. В других же случаях [8, 57] "теоретическая геология" базировалась на формальных построениях и с той же долей уместности, как и для первого случая, формальная схема "подтверждалась геологическими примерами". Обе стороны быстро обнаружили слабые места в построениях "теоретической геологии" и со взаимной готовностью подняли открытую дискуссию [60].

В то же время, параллельно усилиям "создания теоретической геологии", шел интенсивный поиск "хороших геологических задач" и создавались приемлемые для их решения алгоритмы [4, 14]. Так, в интервале нескольких лет, обозначились геологические разделы задач, для которых были построены довольно разнообразные комплексы программ [5, 23, 35]. Возникали производственные и научные подразделения, специализированные по математическому профилю решения геологических задач.

Геофизические направления, по своему содержанию будучи существенно математическими, менее болезненно вошли в режим широкого использования вычислительных машин. В "чисто" геологических разделах возникли не только новые исследовательские трудности (как со стороны геологии, так и со стороны математики), но и чисто психологические. К настоящему времени психологические затруднения выведены в скучную область монотонных рассуждений, а более резко и трезво обрисовались трудности принципиального характера [4, 8, 27]. Причем эти трудности выросли не только в связи с проводкой математических средств к действительно крупным геологическим проблемам, но и в связи с возрастанием уровня общей требовательности математических средств, нацеленных на решение геологических задач [12, 18, 30-32, 51].

Кроме того, серьезные содержательные и формализованные вопросы возникают в связи с поступлением в геологическую отрасль знания большого количества сообщений, не предусмотренных предыдущими ступенями развития геологии. Эти сообщения связываются с результатами исследования земли из космоса (спутники) и новым потоком фактов о ложе мирового океана (бурение). Если учесть тех-

ническую оснащенность геологов-практиков и инструментальную оснащенность геологов-исследователей, а также возросшее в связи с этим производство информации, то становится очевидной та острота необходимости развития автоматических средств исследования информации, которая резко выявилась в задачах прогнозно-поискового профиля [18, 21, 34].

Действительно, несмотря на применимость математических средств для широкого класса геологических задач, связанных с исследованием информации, своеобразным эпицентром попыток и разработок методов математического изучения объектов и событий является отрасль геологии о поиске и разведке полезных ископаемых. Дальнейшие разделы работы посвящены содержательной трактовке математических средств анализа геологической информации.

§ 3. Возникновение тестового подхода

Возникновение тестового подхода в практике восходит к работам С.В. Яблонского для области технических приложений. Период с 1958 года по 1964 год можно характеризовать как период внутреннего становления основного направления подхода в области теории и поиска областей приложения (новых и в новых сферах). Практическое применение тестов на этом этапе сосредотачивалось в решении задач поиска неисправностей в сложных технических системах: счетно-решающих устройств, устройств типа управляющих систем и др. Поскольку эти системы представляют собой своеобразную реализацию функций алгебры логики, то естественной была и попытка осуществить логический контроль неисправностей контактных систем [59].

Начиная с 1964 года произошло интенсивное внедрение тестового подхода в область естествознания. Первые применения тестов в задачах геологического профиля относятся к 1965 году, по инициативе Ю.И. Журавлева. Конкретная попытка решить задачу прогнозирования комплексного оруденения с помощью построения всех тупиковых тестов, по предложению Ф.П. Кренделева, начата была в 1966 г. [12]. К 1970 году тестовый подход, развиваясь и адаптируясь к задачам естествознания, был применен в довольно широком классе задач геологии прогнозно-поискового профиля [5, 14, 21, 28]. Методами тестов в классе задач, сводимых к задачам распознавания, были

получены результаты большой практической ценности [9, 22, 32, 46]. Одновременно был создан алгоритмический комплекс для многоцелевой обработки таблиц бинарных символов и создан небольшой альбом программ (в основном для машин М-220, ЭЭСМ-6) [5, 12, 27]. В этот же период были обнаружены некоторые математические особенности подхода в работе геологической информацией, которые дали основание к дальнейшему развитию теории и расширению практических приложений. Именно в интервале этого времени подход подвергся критике [8] и сравнительному изучению с другими направлениями [6, 8, II].

В период с 1970 по 1975 год тестовый подход вступает в этап математической и прикладной зрелости. Разработки в этом направлении интенсивно велись в математическом отношении в Москве [27, 30-35, 52, 61], Новосибирске [13, 16, 23, 24], Алма-Ате [4, 5], Владивостоке [47, 48, 50], Риге [6]. География практических приложений гораздо шире. Резко возросло и количество сфер применения тестового подхода; были решены задачи: медицинского профиля [6, 15, 19], географического направления, в социологии, в рыбном хозяйстве, в промышленности и др. Среди практического направления основным руслом применения тестов была геология. Именно в геологии размах разнообразия задач, подлежащих решению данным подходом, оказался довольно большим. Несмотря на новизну подхода, появляются довольно обширные изложения правил и рабочих процедур на конкретных примерах решения теоретически и производственно важных задач.

§ 4. Современные особенности подхода

Дальнейшее развитие подхода, начиная с 1975 года, характеризуется стабилизацией теоретических и практических "границ" и переходом к проблемам качества и обоснования подхода в целом. При этом в этом периоде (захватывающим настоящее время, ддящимся в будущее), с его задачей удовлетворить внутренние нужды подхода, обнаруживается тенденция к установлению связей между отдельными методами, с одной стороны, и детализацией особенностей методов, с другой. На базе поиска "макросвязей" вскрылась возможность наращивать подход новыми методами, нацеленными на обнаружение закономерностей, фиксированных в неявном виде, в естественно-науч-

ных описаниях [36, 38]. На основе анализа и детализации тестовых процедур открылась возможность рассмотреть некоторые количественные стороны и природу разнообразия, содержащегося в естественнонаучных описаниях [17, 37, 54]. Одновременно с теоретическими разработками подхода успешно продолжается (а в некоторых случаях наращивается) практическое применение в различных сферах геологии, особенно в области задач прогнозно-поискового направления.

Настоящая обстановка в области приложений характеризуется широкой комплексацией тестовых программ не только со "смежными" алгоритмами (эвристическими) [41, 52], но и с алгоритмами, например, статистического направления и алгоритмами, построенными на итерационной основе [7]. Характерна также ступенчатость в решении многоцелевых и информоемких задач. В задачах этого типа тестовым методам принадлежат наиболее трудоемкие места по обнаружению детализаций при исследовании сходств и различий объектов, заданных естественно-научными описаниями [12, 13, 17]. Продолжается поиск алгоритмов построения всех тупиковых тестов с целью выхода на большеформатные таблицы [35, 47]. Появляются работы обобщающего характера и профилирующие дальнейшие исследования [17, 30, 41] и обоснования подхода [5, 27, 61] в сравнительном изучении с алгоритмами иных направлений (в частности, комплекс алгоритмов "Кора") [5, 8, 41]. Исходя из общей идеологии тестового подхода на прямом и косвенном его продолжении развития, были созданы новые алгоритмы и построены обеспечивающие их программы различного предназначения [19, 23, 41, 50]. В целом текущий период характерен отчетливой тенденцией к синтезу довольно широкого набора решающих правил и вычислительных операций в компактную, содержательно интерпретируемую совокупность алгоритмов многоцелевого предназначения. Причем эта комплексация возникает не только из самой природы подхода, допускающей синтезирование своих частей, но и из природы задач, подлежащих решению. Глубокое внедрение в специфику информационных исследований в геологии находится в зрелой фазе.

§ 5. Структура изложения

Дальнейшее изложение материала, характеризующего возможности

и формализованную основу тестового подхода, подчинено двум целям следующего содержания. Во первых, приводится последовательное освещение методов, составляющих тестовый подход, в заданном перечне математических идей и средств их математического описания. Во вторых, в местах их естественного приложения и возникновения, указывается на предназначенность каждого характеризующего метода в области его применения. Причем это применение методов представлено конкретными указаниями на то, какие задачи предпочтительно решать тем или иным методом.

Данная нацеленность работы и обуславливает подразделение на части, вернее, на определенные главы, членение которых на параграфы произведено по различным мотивам. В первой главе (содержание которой профилировано идеей сравнительного изучения объектов исследования с помощью математических средств) последовательность и количество параграфов устанавливается разнообразием моделей тестовой обработки. Во второй главе параграфы построены с учетом не только разнообразия моделей обработки информации, но и с учетом видов решающих функций. Трехчленное подразделение моделей и решающих функций имеет различное происхождение. Подразделение моделей исходит из требований (или операций) сравнительного изучения объектов в координатах исследования: различий, сходств и согласованности (сходство-различие). Этим координатам исследования в содержательном срезе в формализованных схемах тестового подхода соответствуют Р-тесты, Q-тесты и H-тесты [17, 54].

Происхождение подразделения видов решающих функций обязано специфике конкретных целей исследования информации. Кроме того, сам вид постановок геологических задач (в интересующем нас классе) обуславливает уместность и достаточность таких видов решающих функций, как:

- а) "взвешенные расстояния" - для нужд таксономии и подразделения объектов на классы [18, 41];
- б) "веса строк" - для решения задач упорядочения исследуемых объектов по комплексу характеристических признаков [8, II];
- в) "голосования" - для задач, сводимых к распознаванию образов, во всех случаях, когда требуется произвести сортировку множества объектов (проб) на заранее выделенных классах (эталонах) [17, 43].

Естественно допустить, что эти модели и виды решающих функций позволяют обрабатывать информацию, содержащуюся в естественно-научном описании объектов, по довольно широкому списку целей, имеющих практическое и теоретическое значение в геологии.

По мере практического и теоретического расширения тестового подхода, оказалось уместным применить ряд новых понятий и определений таких, например, как вводимая ниже модель обработки, согласованность и др.

Для большинства из приводимых алгоритмов даны программные реализации и проведено широкое и всестороннее апробирование при решении конкретных задач.

В заключении указаны некоторые дальнейшие возможности развития и применения тестового подхода.

ОСНОВНЫЕ МОДЕЛИ ТЕСТОВОЙ ОБРАБОТКИ ИНФОРМАЦИИ

§ I. Понятие тестовой модели

Применению тестовых алгоритмов при математической обработке информации предшествует этап доалгоритмической обработки. Конечным шагом этого этапа, предопределяющим порядок проведения обработки и выбор конкретных алгоритмов, является задание модели тестовой обработки. Эта модель обработки обычно тесно увязана с имеющейся у исследователя моделью системы исследуемых им объектов и явлений. И если модель тестовой обработки имеет формальное определение (см. ниже), то модель системы (по крайней мере, в геологических исследованиях), содержательная роль которой весьма велика, не поддается формализации. Поэтому модель тестовой обработки представляет собой удобную грань, за которой начинается обычная математическая работа.

Пусть задана информация, представляющая собой перечень описаний объектов S_1, S_2, S_3, \dots , подлежащих исследованию. Характеристические признаки объектов бинарные и образуют последовательность x_1, \dots, x_n (с естественной полнотой описывающую объекты). Таким путем организованная информация и подлежит тестовой обработке.

Чтобы задать модель тестовой обработки, следует прежде всего указать систему целевых отношений между объектами.

Целевыми отношениями (в плане данной системы исследований) являются отношения следующего вида:

- а) в соответствии с целью обработки информации объекты S_i и S_k сходны;
- б) в соответствии с целью обработки информации объекты S_i и S_k различны;
- в) в соответствии с целью обработки информации объекты S_i и S_k сходно-различны;
- г) согласно цели обработки учетной информации объекты S_i и S_k индифферентны (т.е. для них не имеет место ни одно из отношений (а); (б); (в)).

Пример I. Исходная совокупность исследуемых объектов S_1, S_2, S_3, \dots представляет собой перечень рудосодержащих площадей, для которых даны оценки запасов. Эти оценки подразделяют перечень объектов на месторождения и рудопроявления. Известно также – месторождения промышленно значимые. Признаковое пространство x_1, x_2, \dots, x_n характеризует объекты двумя наборами: признаки геологического содержания и признаки геофизического профиля. Задана следующая цель обработки информации – путем сравнительного изучения месторождений и рудопроявлений (представленных соответствующими таблицами описаний) следует построить решающую функцию $h(s) = h(x_1(s), \dots, x_n(s))$, пригодную для разбраковки вновь обнаруживаемых рудоносных площадей на месторождения и рудопроявления.

Система отношений для объектов указанного примера будет следующей:

- 1) для всех пар месторождений имеет место отношение (а); для всех пар рудопроявлений имеет место отношение (а);
- 2) для всех пар (S_i, S_k) объектов, составленных из одного месторождения и одного рудопроявления, имеет место отношение (б);

Вслед за указанием целевых отношений исследуемых объектов осуществляется этап выделения классов, т.е. для модели устанавливается, какие из подмножеств исходной совокупности $S_1, S_2, S_3, \dots, S_m$ считать классами объектов, выделенными соответственно цели обработки информации. Предполагается, что эти классы попарно не пересекаются. В примере I выделяются два попарно не пересекающиеся класса: класс месторождений и класс рудопроявлений. Отметим, что иногда целесообразно первые два этапа (выяснение системы целевых отношений и классовое образование) поменять местами.

Следующий, третий этап, заключается в переходе от системы целевых отношений к системе отношений модели, которая получается некоторым видоизменением системы целевых отношений. Необходимость и смысл этого этапа были выяснены в результате решения многих практических задач прогнозно-поискового профиля.

Основной смысл этого этапа заключается в учете того факта, что цель обработки информации и порождаемая ею система целевых отношений являются внешними по отношению к исследуемой совокупности объектов и поэтому могут не иметь достаточно четкого отражения в обрабатываемых описаниях. С другой стороны, в исходной

совокупности S_1, S_2, \dots может присутствовать система естественных отношений между объектами, не зависящая от поставленных нами целей и отражающая некоторые закономерности природных процессов, управляющих возникновением и развитием этих объектов. Опыт решения задач в тестовом подходе показывает, что учет естественных отношений между объектами улучшает качество решения - например, уменьшает число ошибок при сортировке рудосодержащих площадей на месторождения и рудопроявления. Такой учет и приводит к системе отношений модели, представляющей собой результат своего рода адаптации наших целей к реальной природной ситуации.

Следует признать, конечно, что приведенные выше общие рассуждения не приводят к очевидным практическим указаниям - как поступать в том или ином конкретном случае, да и в теоретическом плане нуждаются в определенных уточнениях. Поэтому в публикуемой работе авторы ограничиваются изложением простых частных случаев, где указанные соображения представляются вполне очевидными. Однако целесообразность таких общих рассуждений о структуре обработки информации для нас представляется несомненной, тем более, что этот вопрос является малоизученным.

Пример 2. В качестве иллюстрации вышеизложенного несколько усложним ситуацию примера 1. Предположим, что некоторые из месторождений - например S_1 и S_2 находятся в несколько иной геолого-структурной обстановке и образование рудных тел было тесно сопряжено с условиями, заметно отличающимися от условий, в которых происходило образование рудных тел у остальных месторождений, для которых генетические условия рудообразования были существенно подобны (довольно частый случай в природе). Тогда основная масса месторождений образует, как бы, как иногда говорят, ядро класса месторождений $[9, II]$, а месторождения S_1, S_2 образуют периферию класса. Такая периферичность объектов (в данном случае S_1, S_2) обычно существенно проявляется в их описаниях, резко отличая их от остальных объектов данного класса. Учитывая это, внесем следующие изменения в систему целевых отношений: для S_1, S_2 удалим все пары типа (а), в которые входят S_1 или S_2 , т.е. откажемся от условий сходства S_1, S_2 с остальными месторождениями. При этом отношения различия S_1, S_2 с классом рудопроявлений полностью сохраняются.

Далее, как уже отмечалось в [30, 32] причины, препятствовавшие

образованию крупных месторождений могут оказаться весьма разнообразными, что и проявляется в существенном несхождении их описаний. Описания месторождений со сходной причиной образования в среднем более похожи друг на друга (скажем, по хэммингову расстоянию). Предположим, что имеет место именно такая ситуация разнообразия причин. Учитывая это, удалим из системы целевых отношений все отношения сходства между мелкими рудопроявлениями. Полученная в результате всех описанных преобразований система отношений и будет системой отношений модели. В данном случае она имеет следующий вид:

- а) если $i, j \leq m_1, i \neq 1, 2, j \neq 1, 2$, то S_i сходно с S_j ;
- б) если $i \leq m_1, k > m_1$, то S_i, S_k различны.

Построенная система отношений модели однозначным образом определяет множество τ некоторых сочетаний исходных признаков. По этим сочетаниям в данной модели и проводится сравнительное изучение объектов и вычисляются значения решающих функций (см. подробнее гл.3), по которым на основе решающих правил делается определенное суждение об исследуемом объекте, результирующее проведенную работу с информацией. Обычно - это суждение об отнесе - нии к одному из классов модели или о месте объекта в ряду по степени проявления некоторого целевого признака либо оценка типичности объекта для того или иного класса.

Поскольку информация сосредоточена в таблицах Т описаний (где столбцы отвечают признакам, а строки объектам), то обычно говорят не о сочетаниях признаков, а о наборах столбцов таблицы. Поэтому и множество τ обычно определяется как множество наборов столбцов, обладающих определенными свойствами. Используя введенный в [34] термин, будем называть τ множеством информационных единиц модели, тогда как составляющие его наборы - информационными единицами.

В основу выделения множества τ информационных единиц в тестовом подходе положен следующий излагаемый ниже принцип.

Пусть задана система отношений модели D и таблица описаний Т. Ее строки $S_i = (t_{i1}, \dots, t_{in})$ - описание объектов S_i признаками x_1, \dots, x_n .

Определение 1. Строки S_{i_1}, S_{i_2} называются различными, если существует такой столбец $x_j, j = 1, \dots, n$, что $t_{i_1 j} \neq t_{i_2 j}$; строки S_{i_1}, S_{i_2} сходны, если существует такой столбец x_j , что

$t_{i,j} = t_{i_2,j}$; строки s_{i_1}, s_{i_2} будем называть сходно-различными, или согласованными, если они одновременно сходны и различны.

Определение 2. Подтаблица $T' \subseteq T$, полученная из T удалением некоторых столбцов, называется D - допустимой, и набор ее столбцов называется D - тестом¹⁾, если она удовлетворяет следующему условию соответствия. Если объекты s_i, s_k , согласно системе D отношений модели, сходны, различны или согласованы, то и отвечающие им строки s_i, s_k таблицы T являются соответственно сходными, различными или согласованными.

Определение 3. Набор столбцов $t = (x_{j_1}, \dots, x_{j_k})$, образующий подтаблицу T' таблицы T , называется информационной единицей модели или тупиковым D - тестом, если никакая другая подтаблица $T'' \subset T'$ не отвечает D - тесту.

Таким образом, в основу выделения информационной единицы модели положены 2 условия:

- 1) условие соответствия (см. определение 2);
- 2) условие избыточности (см. определение 3).

Суммируя вышеизложенное, формулируем

Определение 4. Тестовой моделью обработки называется четверка $\langle T, K, D, \tau \rangle$, где T - таблица описаний, K - список классов модели; $K = \{K_1, \dots, K_m\}$, D - система отношений модели, τ - множество информационных единиц модели.

Описанная здесь в общих чертах схема будет детально изложена в последующих параграфах при описании важнейших типов таких моделей, наиболее часто используемых на практике - P -моделей, Q -моделей и H -моделей.

§ 2. P - модели

Следя терминологии данной работы, можно сказать, что применение тестового подхода к решению практических задач началось с реализации P -моделей тестовой обработки.

Определение 1. Модель тестовой обработки $\langle T, K, D, \tau \rangle$ называется моделью типа P или P -моделью, если

1) См. более детальное рассмотрение этого вопроса в работе Е.А.Смертина в данном сборнике.

1) объекты исходной совокупности относятся к одному классу либо разбиты на несколько непересекающихся классов;

2) в случае, когда все объекты относятся к одному классу, система D состоит из отношений различия для любой пары объектов;

3) в случае нескольких классов система D состоит из отношений различия для всех пар объектов, лежащих в разных классах, и отношений индифферентности для всех пар объектов, относящихся к одному и тому же классу.

Если в модели выделен только один класс, то информационная единица такой P -модели называется тупиковым тестом [12]. Понятие тупикового теста впервые введено С.В.Яблонским в 1958 г. в работе [59]. Если в P -модели выделено несколько классов, то ее информационная единица называется тупиковым тестором [12]. Следует отметить, что понятие тупикового тестора является частным случаем обобщенного понятия теста, приведенного в [59].

Формулируем понятия тупикового теста и тестора в табличных терминах, как это обычно и делается в литературе [12, 56].

Определение 2. Набор столбцов x_{j_1}, \dots, x_{j_k} называется тестом, если при удалении из T всех столбцов, кроме x_{j_1}, \dots, x_{j_k} , в полученной таблице все строки будут попарно различны.

Определение 3. Тест называется тупиковым, если при удалении из него любого столбца полученный набор не является тестом.

Определение 4. Разобьем таблицу $T = (t_{ij})_{m \times n}$ на k подтаблиц T_1, \dots, T_k , где $T_1 = (t_{ij}^1)_{m_1 \times n}, \dots, T_k = (t_{ij}^k)_{m_k \times n}$. Таблица T_1 образуется первыми m_1 строками T , таблица T_2 - строками с номерами $m_1 + 1, m_1 + 2, \dots, m_1 + m_2$, таблица T_3 - следующими m_3 строками и так далее (см. рис.1). Набор столбцов x_{j_1}, \dots, x_{j_k} таблицы T называется тестором для T_1, \dots, T_k , если в таблице, полученной из T удалением всех столбцов, кроме x_{j_1}, \dots, x_{j_k} все строки, относящиеся к разным подтаблицам T_p, T_q попарно различны.

$$T = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_k \end{bmatrix}$$

Рис. 1



Определение 5. Тестор называется тупиковым, если при удалении из него любого столбца полученный набор не является тестором.

Пример 1. Рассмотрим таблицу T следующего вида [12].

	x_1	x_2	x_3	x_4	x_5
$\$ 1$	1	0	0	1	1
$\$ 2$	1	1	0	1	1
$\$ 3$	1	0	1	1	0
$\$ 4$	1	1	1	0	0
$\$ 5$	0	1	1	0	1
$\$ 6$	0	1	1	1	0

Тупиковые тесты таблицы T :

(x_1, x_2, x_5) , (x_2, x_4, x_5) , (x_1, x_2, x_3, x_4) .

Тем самым множество \mathcal{T} состоит в данном случае из трех тестов.

Пример 2. Для иллюстрации понятия тупикового тестора разобьем таблицу T

	x_1	x_2	x_3	x_4	x_5
$\$ 1$	0	1	0	1	0
$\$ 2$	1	0	1	1	1
$\$ 3$	1	0	1	1	0
$\$ 4$	1	0	0	0	1
$\$ 5$	0	1	0	0	0
$\$ 6$	1	1	1	1	0

на две подтаблицы T_1, T_2 :

$$T_1 = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix} \quad T_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Тупиковыми тесторами здесь являются наборы (x_1, x_2, x_4) и (x_2, x_3, x_4) . Множество \mathcal{T} состоит из двух тупиковых тесторов.

Следует отметить, что понятия тупикового теста и тестора тесно взаимосвязаны. Именно, если таблицу $T = (t_{ij})_{m \times n}$ разбить на m подтаблиц T_1, \dots, T_m так, что каждая из подтаблиц будет состоять из одной строки, то тупиковыми тесторами для T_1, \dots, T_m будут в точности тупиковые тесты T . Таким образом,

тупиковый тест можно рассматривать как частную реализацию тупикового тестора.

С точки зрения обработки на ЭВМ, весьма важно бывает оценить, исходя из размерности матрицы T , минимальное и максимальное число столбцов, могущих составить тупиковый тестор. Число столбцов, составляющих информационную единицу t (тест, тестор и т.д.), принято называть длиной этой единицы. Имеют место следующие оценки.

Предложение 1 [13]. Если t - тупиковый тест T , то

$$\lceil \log_2 m \rceil \leq \ell(t) \leq m - 1$$

Предложение 2. Если t - тупиковый тестор для T_1, \dots, T_k , то

$$\lceil \log_2 k \rceil \leq \ell(t) \leq m - 1$$

Последнее утверждение непосредственно вытекает из предложения 1.

Пример 3. Таблица T разбита на подтаблицы

$$T_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad T_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Имеем при $k = 2$, $m = 5$, $\lceil \log_2 2 \rceil = 1$, $m - 1 = 4$. Нетрудно проверить, что набор (x_5) будет тупиковым тестором минимальной длины, а набор (x_1, x_2, x_3, x_4) - максимальной. Других тупиковых тесторов в таблице нет.

Для общего числа тупиковых тестов таблицы T известна верхняя оценка.

Предложение 3 [56].

Пусть $\mathcal{Z}(T_n)$ - число тупиковых тестов таблицы T с n столбцами. Через $R(n)$ обозначим величину $\max \mathcal{Z}(T_n)$, где максимум берется по всем таблицам с n столбцами. Имеют место оценки:

$$0,28 C_n^{\frac{5}{2}} \leq R(n) \leq C_n^{\frac{1}{2}},$$

причем при $n \geq 10$, $R(n) < C_n^{\frac{2}{3}}$.

При математической обработке информации существенную роль играют так называемые оценки информативности признаков [51, 47 и др.]. Они используются как для целей распознавания и упорядочения объектов [51, 12], так и для минимизации исходной совокупности признаков [55, 2], а также играют важную роль при содержательной интерпретации результатов обработки.

Подобные оценки информативности были введены на самых первых шагах и в тестовом подходе. К настоящему моменту предложен целый ряд таких оценок [12, 13, 15, 61, 47, 52]. Некоторые из них будут подробно рассмотрены в главе III. Здесь отметим только, что первые оценки информативности были сформулированы для тестов и тесторов в [12], а именно, информационный вес признака P_j и разделяющий вес признака R_j . Информационный вес признака P_j предназначен для P -моделей с одним выделенным классом, а разделяющий вес R_j — для P -моделей с несколькими классами. Остальные оценки, в основном, являются модификациями этих исходных. Как P_j , так и R_j , определяются по следующей схеме. Пусть τ — множество информационных единиц модели, в которые входит j -ый столбец, отвечающий признаку x_j . Тогда как P_j , так и R_j , равняются

$$\frac{|\tau_j|}{|\tau|}$$

и, тем самым, представляют собой частоту вхождения признака x_j в информационные единицы модели. В работе [12], стр. 10, высказывалось предположение, что эти величины оценивают важность признака x_j при изучении класса объектов, эталонными представителями которого являются S_1, S_2, S_3, \dots . Если расположить признаки x_j в последовательность по убыванию P_j (или R_j соответственно), то признаки, получившие меньший номер, оказываются более важными. К настоящему времени сохранился достаточно обширный эмпирический материал [9, 46, 25], показывающий, что такого рода оценки действительно могут быть полезны при обработке экспериментальных данных в естественных науках и, по-видимому, отражают существенные, с точки зрения цели обработки, характеристики признаков. Однако исчерпывающего истолкования содержательной природы этой существенности, на наш взгляд, пока не найдено. Кроме того, существует и неполная математическая разработка системы приводимых соотношений. Именно поэтому в некоторых направлениях недостаточно четко очерчен круг задач, где эти

оценки целесообразно использовать. Работы в этом направлении ведутся (например [34] и др.) довольно широко, а полученные практические (в геологии, географии, медицине и др.) и теоретические результаты самоутверждаются актуальностью и своевременностью решаемых задач.

§ 3. Q — модели

Если в P -моделях система отношений модели фиксирует различия между членами исходной совокупности S_1, S_2, \dots , то Q -модели оперируют с информационными единицами, отвечающими условиям сходства между объектами.

Определение 1. Модель $\langle T, K, D, \tau \rangle$ называется моделью типа Q или Q -моделью, если:

- 1) все объекты относятся к одному классу;
- 2) система D состоит из отношений сходства для любой пары объектов исходной совокупности.

Информационная единица Q -модели, впервые введенная Е.А. Смергиным в [54], называется тупиковым Q -тестом. Приведем соответствующие определения.

Определение 1. Набор столбцов x_{j_1}, \dots, x_{j_e} называется Q -тестом, если при удалении из T всех столбцов, кроме x_{j_1}, \dots, x_{j_e} , в полученной таблице все строки будут попарно сходны.

Определение 2. Q -тест называется тупиковым, если при удалении из него любого столбца полученный набор не является Q -тестом.

Пример 1. Таблица T из примера I § 2 имеет в точности один тупиковый Q -тест (x_1, x_3, x_4, x_5) .

Пример 2. Таблица T вида

$$\begin{bmatrix} I & 0 & I & 0 & I \\ 0 & I & 0 & I & I \\ 0 & I & 0 & I & 0 \end{bmatrix}$$

не имеет тупиковых Q -тестов, $\tau = \emptyset$, поскольку первая и третья строки различны по всем столбцам.

Точно также, как и в тестовом случае, вводится Q -тестовая мера важности признака [54]:

$$Q_j = \frac{|\tau_j|}{|\tau|}$$

Как уже отмечалось, система D в случае Q -модели состоит исключительно из условий сходства для всех объектов. С содержательной точки зрения система D выражает условие однородности исходной совокупности S_1, \dots, S_m . Информационная единица - тупиковый Q -тест - представляет собой неизбыточную совокупность признаков, на которой эта однородность проявляется в сходстве описаний объектов.

§ 4. Н - модели

Впервые Н-модели были применены на практике в [3,17] для обработки информации по дифференцированным трапповым интрузиям^{ж)}.

Определение 1. Модель $\langle T, K, D, \tau \rangle$ называется моделью типа Н или Н-моделью, если:

1) объекты относятся к одному или нескольким непересекающимся классам;

2) в случае одного класса система D состоит из отношений сходства - различия (согласования) для любой пары объектов;

3) в случае нескольких классов система D состоит из отношений различия для объектов, лежащих в разных классах. Кроме того, для некоторых из выделенных классов (возможно всех) A_1, \dots, A_2 D содержит отношения сходства для объектов, лежащих в одном и том же классе.

Сформулируем в табличных терминах определение соответствующих информационных единиц.

Определение 2. Набор столбцов x_{j_1}, \dots, x_{j_e} таблицы T называется Н-тестом, если при удалении из T всех столбцов, кроме x_{j_1}, \dots, x_{j_e} , в полученной таблице все строки будут попарно сходны и различны.

Определение 3. Н-тест называется тупиковым, если при удалении из него любого столбца полученный набор не является Н-тестом.

ж) Рассматривалась частная реализация Н-модели, информационная единица которой была названа К-тестором.

Пример 1. Таблица T из примера 1 §2 представляет собой тупиковый Н-тест.

Предположим, что исходная совокупность S_1, S_2, \dots разбита на классы A_1, \dots, A_p , где $A_1 = \{S_1, \dots, S_{m_1}\}$, $A_2 = \{S_{m_1+1}, \dots, S_{m_1+m_2}\}, \dots$. Для классов A_{z_1}, \dots, A_{z_q} система D содержит условия сходства принадлежащих им членов. Классам A_1, \dots, A_p отвечают подтаблицы T_1, \dots, T_p таблицы T , где

$$T = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_p \end{bmatrix}$$

Определение 4. Набор столбцов x_{j_1}, \dots, x_{j_e} называется Н-тестором, если при удалении из T всех столбцов, кроме x_{j_1}, \dots, x_{j_e} , получается таблица, в которой:

1) все строки, отвечающие разным подтаблицам T_i, T_j , различны;

2) любая пара строк, отвечающих одной и той же подтаблице T_{z_j} , $j = 1, \dots, q$, имеет сходство.

Определение 5. Н-тестор называется тупиковым, если при удалении из него любого столбца полученная таблица не является Н-тестором.

Пример 2.

$$T_1 = \begin{matrix} & x_1 & x_2 & x_3 & x_4 & x_5 \\ \begin{bmatrix} I & 0 & I & 0 & I \\ I & I & 0 & 0 & I \\ 0 & I & I & I & 0 \\ I & I & I & 0 & 0 \end{bmatrix} \end{matrix}$$

$$T_2 = \begin{matrix} & x_1 & x_2 & x_3 & x_4 & x_5 \\ \begin{bmatrix} I & 0 & 0 & I & 0 \\ 0 & 0 & 0 & I & I \\ 0 & I & 0 & 0 & I \\ 0 & 0 & I & 0 & 0 \end{bmatrix} \end{matrix}$$

Предположим, что условие сходства - различия заданы для обоих классов. Тогда множество τ тупиковых Н-тестором состоит из единственного набора столбцов: (x_1, x_2, x_3) , тогда как множес-

тво тупиковых тесторов состоит из наборов (x_1, x_2, x_3) , (x_1, x_2, x_4) , (x_1, x_2, x_5) , (x_1, x_3, x_4) , (x_1, x_4, x_5) . Как и в предыдущих случаях, для информационных единиц N -модели можно определить вес признака:

$$H_j = \frac{|T_j|}{|T|}$$

На рис.2 приведены веса H_j и R_j для рассматриваемых в настоящем примере таблиц.

№	I	II	III	IV	V
H_j	I	I	I	0	0
R_j	I	0.6	0.4	0.6	0.4

Рис. 2

Нетрудно видеть, что различия в оценке признаков по H_j и R_j довольно существенны.

Дополнительные условия сходства-различия для членов классов A_{2_1}, \dots, A_{2_4} представляют собой условия однородности объектов, относящихся к одному и тому же классу. Таким образом, при построении N -модели необходимо указать, какие из классов модели принимаются за однородные.

§ I. Виды решающих функций

В дальнейшем будем считать, что задано некоторое множество объектов \mathcal{J} , члены которого называются объектами обработки. Исходная совокупность S_1, S_2, S_3, \dots может совпадать с \mathcal{J} или быть его собственным подмножеством. В последнем случае множество \mathcal{J} иногда называют генеральной совокупностью, а его подмножество $\{S_1, S_2, \dots\}$ - обучающей выборкой объектов из \mathcal{J} [5I]. В литературе по распознаванию рассматривается как весьма важная проблема представительности обучающей выборки (см. [5I] и др.), однако рамки настоящей статьи не позволяют обсудить этот вопрос применительно к тестовому подходу. Мы его оставляем для дальнейших публикаций.

Совпадение или несовпадение \mathcal{J} с обучающей выборкой зависит от цели исследования. Например, если поставлена задача оценки важности признаков x_j при упорядочении объектов исходной совокупности S_1, S_2, \dots по степени проявления некоторого целевого признака X_{n+1} (например, масштаба запасов месторождений), то \mathcal{J} совпадает с обучающей выборкой.

Для наглядности изложения ограничимся рассмотрением приложения тестового подхода к решению следующих двух важных задач обработки информации, часто встречающихся на практике: задачи распознавания и задачи упорядочения [5I, I4].

Задача распознавания может быть сформулирована следующим образом. Генеральная совокупность \mathcal{J} разбита на p классов, из эталонных представителей которых составлена обучающая выборка

S_1, S_2, \dots . Объекты S_1, S_2, \dots охарактеризованы бинарными признаками x_1, \dots, x_n . Требуется для любого объекта $S \in \mathcal{J}$ на основании сравнения его описания признаками x_1, \dots, x_n с описаниями объектов обучения распознать, к какому из выделенных классов он относится.

В задаче упорядочения \mathcal{J} представляет собой один класс, для

объектов которого определен целевой признак x_{n+1} . Требуется для произвольного $s \in \mathcal{J}$ найти его место в упорядоченной по убыванию степени проявления признака x_{n+1} последовательности объектов обучающей выборки. При этом искомое место может быть то же, что и у какого-либо объекта обучения, либо может располагаться между двумя соседними объектами.

Для решения обеих задач в тестовом подходе используются решающие функции h вида $h: \mathcal{J} \rightarrow R$, где R - множество всех вещественных чисел. Для объекта $s \in \mathcal{J}$, подлежащего распознаванию или упорядочению, вычисляется значение $h(s)$. По вычисленному значению $h(s)$ на основе заданного решающего правила производится отнесение s к одному из классов либо указание места в последовательности объектов. Решающая функция h , вообще говоря, в тестовом подходе зависит не только от s . В общем виде ее можно записать как $h = h(s, T, \mathcal{e})$, где T -таблица описаний объектов обучения и \mathcal{e} - некоторая дополнительная информация об объектах обучения, не содержащаяся в таблице. В \mathcal{e} может включаться, например, информация о разбиении объектов на классы, оценка их "важности" (например, запасы месторождений) или "типичности" для своих классов [15].

Решающие функции, наиболее часто применяемые в практических задачах, можно подразделить на 3 основных типа:

- 1) решающие функции типа взвешенных расстояний [12, 14];
- 2) решающие функции типа весов строк [34, 12, 61];
- 3) решающие функции типа числа голосов [43, 16].

Опишем общий вид и основные свойства решающих функций каждого типа, оставляя конкретные примеры таких функций и отвечающих им решающих правил для последующих параграфов.

1. Решающие функции типа взвешенных расстояний

Эти решающие функции применяются в задачах распознавания. В соответствии с [12] введем операцию Δ над элементами множества O, I при помощи таблицы

B	A	O	I
O		I	O
I		O	I

Предположим, что исходная совокупность S_1, S_2, \dots разбита на непересекающиеся классы A_1, \dots, A_p , которым отвечают подтаблицы T_1, \dots, T_p . Определим для $s \in \mathcal{J}$ и A_k расстояние S до A_k . Пусть описание s объекта имеет вид $s = (\delta_1, \delta_2, \dots, \delta_n)$

$$T_k = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{m_k 1} & \alpha_{m_k 2} & \dots & \alpha_{m_k n} \end{bmatrix}$$

и для признаков x_1, \dots, x_n заданы неотрицательные веса ρ_1, \dots, ρ_n . Тогда искомое расстояние может быть определено двумя способами:

а) как среднее расстояние до объектов A_k (см. [12])

$$\rho^1 = \rho^1(s, A_k) = \frac{\sum_{i=1}^{m_k} \sum_{j=1}^n (\alpha_{ij} \circ \delta_j) \rho_j}{m_k}$$

(черта над $(\alpha_{ij} \circ \delta_j)$ означает булево отрицание);

б) как расстояние до ближайшего объекта

$$\rho^2 = \rho^2(s, A_k) = \max_{i=1, \dots, m_k} \sum_{j=1}^n (\alpha_{ij} \circ \delta_j) \rho_j$$

Сравнивая расстояния S до классов, можно произвести отнесение s к ближайшему из них. Конкретные способы такого отнесения - решающие правила - будут изложены в последующих параграфах применительно к рассмотренным моделям обработки.

Какой именно из двух видов оценки расстояния предпочтительнее избрать, обычно определяется в каждом конкретном случае путем сравнения результатов экзамена системы распознавания [14], а также содержательной интерпретации полученных по ρ^1 и ρ^2 расстояний. Следует отметить [14], что ρ^1 более устойчива к малым изменениям обучающей выборки, а с помощью ρ^2 всегда безошибочно распознаются объекты обучающей выборки (что не гарантируется при применении ρ^1).

2. Решающие функции типа весов строк

Решающие функции этого типа применяются как для задач распознавания [34], так и для задач упорядочения [14]. Предположим, что заданы веса признаков ρ_1, \dots, ρ_n и отображения

$\varphi_j: \{0,1\} \rightarrow \{0,1\}$, $j = 1, \dots, n$. Весом $I(s)$ строки $s = (\alpha_1, \dots, \alpha_n)$, являющейся описанием объекта $s \in \mathcal{J}$, называется величина $I(s) = \sum_{j=1}^n \varphi_j(\alpha_j) \beta_j + C$, где C - произвольная постоянная.

Образование φ_j называется отображением ориентации. Легко видеть, что φ_j является либо тождественным отображением $\varphi_j(\alpha_j) = \alpha_j$, либо булевым отрицанием α_j , т.е. $\varphi_j(\alpha_j) = \bar{\alpha}_j$. В этом случае говорят об инверсии признака x_j . Если все признаки x_k , претерпевшие инверсию, заменить их отрицаниями, то $I(s) = \sum_{j=1}^n \alpha'_j \beta_j + C$ (*), где $s' = (\alpha'_1, \dots, \alpha'_n)$ - соответственно видоизмененные описания s . 0 способами задания φ_j см. ниже, а также [18]. Обычно перед началом обработки такая замена (если она необходима) производится, и для расчетов используется формула (*). В этом виде $I(s)$ представляет собой обычную линейную функцию. Отметим, что выбор ориентации признаков зачастую является достаточно трудной задачей и производится либо специалистами-поставщиками информации, хорошо владеющими обрабатываемым материалом [16], либо с помощью ЭВМ [17]. Веса β_j могут быть отрицательными.

3. Решающие функции типа числа голосов

Эти решающие функции применяются в задачах распознавания. Их применение в тестовом подходе было обосновано в [43]. Дальнейшее развитие методов голосования, проведенное в [27, 28 и др.] привело к созданию классов алгоритмов, основанных на методе вычисления оценок [27]. Поэтому эти решающие функции и основанные на них алгоритмы распознавания можно рассматривать как некоторые реализации общей схемы метода вычисления оценок.

Пусть задана система опорных множеств Ω , т.е. некоторое множество подмножеств множества признаков. В тестовом подходе это всегда множество \mathcal{T} информационных единиц соответствующей модели либо его часть. Для $s \in \mathcal{J}$ подмножество (набор) признаков $t = (x_{j_1}, \dots, x_{j_e})$ голосует за объект обучения s_i , если на s и s_i признаки x_{j_1}, \dots, x_{j_e} принимают одинаковые значения. Предположим, что исходная совокупность разбита на классы A_1, \dots, A_p . Для $s \in \mathcal{J}$ обозначим через $\Gamma(s, s_i)$ общее число голосов, поданных за s_i наборами из Ω . С помощью

$\Gamma(s, s_i)$ можно оценить голоса, поданные за класс A_k , следующими двумя способами [16]:

а) как среднее число голосов за объекты данного класса

$$\Gamma_k^I(s) = \frac{\sum_{i=1}^{m_k} \Gamma(s, s_i^k)}{m_k | \Omega |},$$

где $A_k = \{s_i^k, \dots, s_{m_k}^k\}$;

б) как максимальное число голосов за объект данного класса

$$\Gamma_k^2(s) = \max_{i=1, \dots, m_k} \Gamma(s, s_i^k) / | \Omega |$$

Как и в случае взвешенных расстояний, выбор Γ_k^I либо Γ_k^2 определяется по результатам экзамена системы распознавания или из практических соображений. В литературе [5] отмечаются примеры сходства результатов распознавания по взвешенным расстояниям и оценкам числа голосов.

§ 2. Обработка информации для P-модели

1. Задача упорядочения

Применительно к P-модели рассмотрим приложение тестового подхода к решению задачи упорядочения. Предполагается, что для объектов обучения s_1, \dots, s_m известны значения целевого признака - порядкового или количественного - $\gamma_1, \dots, \gamma_m$. Это могут быть, например, масштабы запасов месторождений или их ранги. Будем считать, что $\gamma_1 > \gamma_2 > \dots > \gamma_m$. Предполагается также, что, с точки зрения цели обработки, все величины $\gamma_1, \dots, \gamma_m$ существенно различны и объекты s_1, \dots, s_m составляют один класс. Задан список объектов экзамена и проб $s_{m+1}, \dots, s_q \in \mathcal{J}$. Требуется найти их место в последовательности s_1, \dots, s_m .

Для решения этой задачи используются решающие функции типа весов строк. Общая схема обработки информации заключается в следующем. Вначале производится построение множества \mathcal{T} всех тупиковых тестов T (или некоторого его подмножества) и вычисляются тестовые информационные веса признаков P_1, \dots, P_n . Затем по выбранной решающей функции $I(s)$ (см. ниже) производится вычисление $I(s_1), \dots, I(s_m)$. Если $I(s_1) > I(s_2) > \dots > I(s_m)$, то система распознавания считается допустимой и вычис-

ляются значения $I(S_k)$ для объектов экзамена $S_k, k > m$. Если эти объекты с известным значением r_k занимают места в последовательности $I(S_1), I(S_2), \dots, I(S_m)$ в достаточно хорошем соответствии с их положением в последовательности r_1, \dots, r_m то производится упорядочение проб.

Опишем решающие функции, употребляемые для этой цели.

1) Информационные веса строк $J(S)$.

$$J(S) = \sum_{j=1}^n \alpha_j P_j$$

Практические применения $J(S)$ см. в [14, 10, 49].

2) Веса $y(S)$ [61, 34].

Обозначим через $I_{j\alpha}$ множество объектов исходной совокупности, у которых значение признака x_j равно $\alpha, \alpha \in \{0, 1\}$. Определим вес признака x_j по значению α :

$$P_j(\alpha) = \left(\frac{\sum_{S_k \in I_{j\alpha}} r_k}{\sum_{i=1}^m r_i} \right) P_j.$$

Весом $y(S)$ для $S \in J$ назовем величину $y(S) = \sum_{j=1}^n P_j(\alpha_j)$.

Полагая $(\sum_{S_k \in I_{j1}} r_k / \sum_{i=1}^m r_i) = h_{j1}, (\sum_{S_k \in I_{j0}} r_k / \sum_{i=1}^m r_i) = h_{j0}$, получаем

$$y(S) = \sum_{j=1}^n P_j(\alpha_j) = \sum_{\alpha_j=1} P_j(1) + \sum_{\alpha_j=0} P_j(0) =$$

$$= \sum_{\alpha_j=1} h_{j1} P_j + \sum_{\alpha_j=0} h_{j0} P_j = \sum_{j=1}^n \alpha_j h_{j1} P_j + \sum_{j=1}^n (1 - \alpha_j) h_{j0} P_j =$$

$$= \sum_{j=1}^n \alpha_j h_{j1} P_j + \sum_{j=1}^n h_{j0} P_j - \sum_{j=1}^n \alpha_j h_{j0} P_j =$$

$$= \sum_{j=1}^n \alpha_j [(h_{j1} - h_{j0}) P_j] + \sum_{j=1}^n h_{j0} P_j$$

Полагая $(h_{j1} - h_{j0}) P_j = \tilde{P}_j$ и учитывая, что при фиксированных r_1, \dots, r_m и T сумма $\sum_{j=1}^n h_{j0} P_j$ величина постоянная, получаем

$$y(S) = \sum_{j=1}^n \alpha_j \tilde{P}_j + C,$$

т.е. $y(S)$ отвечает общему виду $I(S)$ из § I.

Преимуществом этой решающей функции является то, что она не зависит от выбора ориентации признаков. С другой стороны, резуль-

таты упорядочения существенно зависят от точности измерения r_1, \dots, r_m , что ограничивает область применения $y(S)$. Учитывая этот факт, в [52] предложена модификация $y(S)$ для случая, когда целевой признак x_{n+1} измерен в шкале порядка. Используя шкалу порядка, можно с большей надежностью оценить взаимоотношения между объектами обучения по целевому признаку x_{n+1} , хотя при этом и происходит некоторое "загрубление" данных. Указанная модификация использует так называемый смешанный треугольник Паскаля - комбинаторную конструкцию, применяемую в исследованиях по теории информации [1]. Для подробного ознакомления с модифицированной решающей функцией отсылаем читателя к [52].

2. Задача распознавания

В этом случае объекты обучения разбиты на p классов A_1, \dots, A_p . Информационной единицей модели является тупиковый тестор. Используются решающие функции всех трех основных типов.

Для удобства изложения ограничимся случаем, когда $p = 2$, т.е. в задаче выделены два класса, и $A_1 = \{S_1^1, \dots, S_{m_1}^1\}$, $A_2 = \{S_1^2, \dots, S_{m_2}^2\}$, $m_1 + m_2 = m$.

I. Решающие функции типа взвешенных расстояний.

Вычисляются разделяющие веса признаков R_1, \dots, R_n . Затем, одним из двух описанных выше способов находятся расстояния объекта $S \in J$ до классов A_1, A_2 :

$$a) \rho_1^1 = \sum_{i=1}^{m_1} \sum_{j=1}^n (\alpha_{ij} \circ r_i) R_j / m_1, \rho_2^1 = \sum_{i=1}^{m_2} \sum_{j=1}^n (\beta_{ij} \circ r_i) R_j / m_2;$$

$$b) \rho_1^2 = \max_{i=1, \dots, m_1} \sum_{j=1}^n (\alpha_{ij} \circ r_i) R_j, \rho_2^2 = \max_{i=1, \dots, m_2} \sum_{j=1}^n (\beta_{ij} \circ r_i) R_j,$$

где $T_1 = (\alpha_{ij})_{m_1 \times n}$, $T_2 = (\beta_{ij})_{m_2 \times n}$ - таблицы описаний, отвечающие A_1, A_2 ,

$$T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix},$$

и описание S признаками x_1, \dots, x_n имеет вид (r_1, \dots, r_n) . Пусть для распознавания выбран один из вариантов оценки расстояний, и расстояния S до классов равны ρ_1, ρ_2 . Тогда используется следующее решающее правило:

- при $\rho_1 - \rho_2 \geq \epsilon$ объект S относится к первому классу;
- при $\rho_2 - \rho_1 \geq \epsilon$ объект S относится ко второму классу;
- при $|\rho_1 - \rho_2| < \epsilon$ объект S не распознается.

Здесь $\epsilon > 0$ - порог распознавания, выбираемый на основе экзамена системы распознавания или из эмпирических соображений.

В работе [48] взвешенные расстояния используются при тестовой обработке количественных данных. Для этого предложен оригинальный способ выбора порогов различимости значений признака, основанный на идее избыточности, наподобие таковой же в тестовом подходе.

II. Решающие функции типа весов строк.

Для задачи распознавания в [61, 34, 52] использована решающая функция вида $y(S)$, но с тесторными разделяющими весами R_j , $j = 1, \dots, n$.

Заданы $\delta_1^1, \dots, \delta_{m_1}^1, \delta_1^2, \dots, \delta_{m_2}^2$ - дополнительные характеристики объектов обучения (например, оценки запасов месторождений). При этом $\delta_i^1 > \delta_i^2$ при всех i, κ . Такая ситуация возникает, в частности, тогда, когда A_1 представляет класс "крупных" месторождений, а A_2 - "мелких".

Вычисляются веса признаков для $\alpha \in \{0, 1\}$:

$$R_j(\alpha) = \left(\sum_{S_k \in I_{j\alpha}} \delta_k / \sum_{i=1}^m \delta_i \right) \cdot R_j$$

и веса $y(S)$ объектов $y(S) = \sum_{j=1}^n R_j(\alpha_j)$.

Затем определяются числа (пороги) C_0, C_1, C_2 так, чтобы для максимального числа объектов обучения выполнялось условие: если S относится к классу A_K , то $C_{K-1} > y(S_i) > C_K$. Распознавание ведется по следующему решающему правилу:

- при $C_0 > y(S) > C_1$ объект S относится к первому классу;
- при $C_1 > y(S) > C_2$ объект S относится ко второму классу;
- при $y(S) > C_0, y(S) = C_1, y(S) < C_2$ объект S не распознается.

III. Решающие функции типа числа голосов.

Производится построение множества τ тупиковых тесторов и для объекта $S \in \mathcal{J}$, подлежащего распознаванию, одним из двух описанных способов вычисляются оценки числа голосов за каждый из классов. Либо $\Gamma_1^1(S) = \sum_{i=1}^{m_1} \Gamma(S, S_i^1) / |\tau| m_1$, $\Gamma_2^1(S) = \sum_{i=1}^{m_2} \Gamma(S, S_i^2) / |\tau| m_2$, либо $\Gamma_1^2(S) = \max_{i=1, \dots, m_1} \Gamma(S, S_i^1) / |\tau|$, $\Gamma_2^2(S) = \max_{i=1, \dots, m_2} \Gamma(S, S_i^2) / |\tau|$.

Затем по выбранному способу оценивания голосов проводится распознавание по следующему решающему правилу:

- при $\Gamma_1^k(S) - \Gamma_2^k(S) \geq \epsilon$ объект S относится к первому классу;
- при $\Gamma_2^k(S) - \Gamma_1^k(S) \geq \epsilon$ объект S относится ко второму классу;
- при $|\Gamma_1^k(S) - \Gamma_2^k(S)| < \epsilon$ объект S не распознается.

Как уже отмечалось, в ряде практических задач отмечено сходство между результатами распознавания по взвешенным расстояниям и по оценкам числа голосов [5]. Кроме того, описанное решающее правило может иметь следующее дополнение [16]: если $\max_{i=1,2} \Gamma_i^k(S) < \delta$, то S не распознается. Здесь $\delta > 0$ - порог, оценивающий близость к материалу обучения. Описаны также случаи [13, 3], когда в качестве Ω бралось не все множество τ , а его часть $\tau^{\ell} = \{t \in \tau | \ell(t) \leq \ell\}$, т.е. множество тупиковых тесторов длины, не превосходящей ℓ . Выбор ℓ может быть оптимизирован относительно результатов экзамена системы распознавания, что приводит к заметному улучшению результатов экзамена. В проведенных экспериментах было обнаружено, что оптимальное значение ℓ обычно ниже среднего, равного $\ell_{cp} = \frac{\sum_{j=1}^n |\tau_j|}{|\tau|}$ и приближается к минимально возможному для данной таблицы.

Может быть использовано обобщение метода голосования по тупиковым тесторам на случай таблиц количественных данных. Для этого используются, как и в схеме взвешенных расстояний, пороги различимости значений признаков.

Сравнение результатов практического применения алгоритма голосования по тупиковым тесторам с голосованием по выборкам фиксированной длины рассмотрено в [34].

§ 3. Обработка информации для Q -модели

I. Задача распознавания

Общая схема применения Q -модели обработки в излагаемом ниже алгоритме распознавания заключена в следующем. Рассматриваются по отдельности Q -модели для обоих классов: $M_1 = \langle T_1, \{1, \dots, m_1\}, Q_1, \tau_1 \rangle$, $M_2 = \langle T_2, \{1, \dots, m_2\}, Q_2, \tau_2 \rangle$. Для объекта $S \in \mathcal{J}$, подлежащего распознаванию, на основе сопоставления с объектами класса A_K по множеству τ_K , где $K=1, 2$,

оценивается "степень принадлежности" $E_K(S)$ к классу, представленному объектами $S_1^k, \dots, S_{m_k}^k$. Затем ведется распознавание по заданному решающему правилу:

- при $\theta = \max_{k=1,2} E_K(S) < \delta$ объект S не распознается;
- при $\theta > \delta$ и $E_I(S) - E_2(S) \geq \epsilon$ объект S относится к I кл.;
- при $\theta > \delta$ и $E_2(S) - E_I(S) \geq \epsilon$ объект S относится к II кл.;
- при $|E_I(S) - E_2(S)| < \epsilon$ объект S не распознается.

Оценка $E_K(S)$ определяется таким образом: наряду с $T_I = (\alpha_{ij})_{m \times n}$, $T_2 = (\beta_{ij})_{m_2 \times n}$ рассмотрим таблицы $\bar{T}_I = (\bar{\alpha}_{ij})_{m \times n}$, $\bar{T}_2 = (\bar{\beta}_{ij})_{m_2 \times n}$.

Нетрудно заметить, что тупиковыми тесторами для таблицы

$$\hat{T}_K = \begin{bmatrix} T_K \\ \bar{T}_K \end{bmatrix}$$

при ее разбиении на подтаблицы T_K, \bar{T}_K будут в точности тупиковые Q -тесты таблицы T_K .

Обозначим через $\bar{T}_K(S)$ оценку числа голосов за \bar{T}_K ,

$$\bar{T}_K(S) = \sum_{i=1}^{m_k} \Gamma(S, \bar{S}_i^k) / m_k |r_k|,$$

где $\bar{S}_i^k = (\bar{\alpha}_{i1}, \dots, \bar{\alpha}_{in})$, $\bar{S}_i^k = (\bar{\beta}_{i1}, \dots, \bar{\beta}_{in})$, r_k - множество тупиковых Q -тестов таблицы T_K .

Полагаем

$$E_I(S) = I - \bar{T}_I(S), \quad E_2(S) = I - \bar{T}_2(S).$$

Выбор такого способа оценки $E_K(S)$ можно пояснить, исходя из следующих соображений. Величина $\Gamma(S, \bar{S}_i^k)$ равняется числу тупиковых Q -тестов $t = (x_{j_1}, \dots, x_{j_t})$ таблицы T_K , удовлетворяющих условию: ни по одному из признаков x_{j_1}, \dots, x_{j_t} объекты S и S_i^k не сходны. С другой стороны, т.к. t является Q -тестом таблицы T_K , все объекты класса A_K являются сходными по крайней мере по одному из признаков x_{j_1}, \dots, x_{j_t} . Таким образом, $\bar{T}_K(S)$ представляет собой нормированное число отклонений от условий сходства между объектами класса A_K . Легко видеть, что если описание S совпадает с одной из строк таблицы T_K , то $\bar{T}_K(S) = 0$ и $E_K(S) = I$. Кроме того, для любого $S \in J$ имеет место неравенство:

$$0 \leq \bar{T}_K(S) \leq I,$$

откуда

$$0 \leq E_K(S) \leq I.$$

Этот алгоритм распознавания апробирован при обработке геологической информации по рудным месторождениям.

2. Оценка степени однородности класса

При решении задачи распознавания иногда важно оценить степень однородности классов A_1, A_2 . Потребность в анализе однородности обычно возникает тогда, когда не удается удовлетворительным образом разделить классы. Такая ситуация может быть вызвана различными причинами. В частности, она может быть вызвана неоднородностью классов A_1, A_2 в связи с присутствием в них объектов, имеющих слабое сходство с остальными членами данного класса. Использование Q -тестовой модели позволяет оценить степень однородности класса и выявить объекты, уменьшающие степень однородности. О других приемах анализа обучающей выборки в тестовом подходе см., например, работу [50].

Упомянутые оценки могут быть получены следующим образом: поочередно удаляем из таблицы T_K описание объектов $S_1^k, \dots, S_{m_k}^k$, обозначив полученные таблицы $T_K^1, \dots, T_K^i, \dots, T_K^{m_k}$ соответственно. Вычисляем степени принадлежности $E_K^i(S_i^k)$, $i=1, \dots, m_k$ объектов S_i^k к классам, представленным совокупностями $A_K \setminus \{S_i^k\}$. Определим степень однородности E_K класса A_K по формуле:

$$E_K = \sum_{i=1}^{m_k} E_K^i(S_i^k) / m_k.$$

Величины E_K^i , определяющие E_K , оценивают влияние объектов S_i^k класса A_K на степень однородности A_K . Объекты, имеющие минимальное значение E_K^i , подвергаются содержательному анализу на предмет правомерности их присутствия в классе A_K (и вообще в обучающей выборке) и наличия ошибок в их описаниях.

§ 4. Обработка информации для N-модели

I. Задача распознавания

Имеются экспериментальные данные, показывающие, что совмест-

ный учет сходства-различия в определенных ситуациях приводит к улучшению результатов распознавания. В частности, при использовании решающей функции типа взвешенных расстояний с N -тесторными весами признаков H_j результаты распознавания оказались лучше, чем с разделяющими тесторными весами R_j . Это относится как к проценту ошибок на экзамене системы распознавания, так и к общей картине распределения расстояний до классов у испытуемых объектов, которая в N -тесторном случае более отчетливо выявляла характер принадлежности к классам.

По-видимому, это можно объяснить тем, что N -тестор, помимо своей способности различать объекты разных классов, является своеобразной "единицей памяти" о принадлежности объектов к одному и тому же классу. Поэтому закономерности, проявляющиеся во взаимосвязях между составляющими его признаками, имеют больше шансов оказаться отражением действительных, объективных закономерностей, присущих изучаемым объектам и явлениям, чем таковые же у "обычного" тестора. Причем этот фактор имеет тем большее значение, чем меньше объем обучающей выборки. Дело в том, что на малых выборках, как показывают наши наблюдения, могут возникать сочетания признаков, являющиеся информационными единицами модели (например, тупиковыми тесторами), но не поддающиеся содержательной интерпретации и имеющие чисто комбинаторное происхождение.

Введение дополнительных условий на вид информационной единицы (как в случае N -тестора), по-видимому, может производить селекцию информационных шумов [52] такого типа.

Применительно к N -модели опишем алгоритмы распознавания, использующие решающие функции типа взвешенных расстояний и типа числа голосов. Для изложения понадобятся следующие предварительные замечания. Чтобы однозначно определить информационную единицу N -модели, необходимо указать, для какого из классов A_1, A_2 (либо для них обоих) система D содержит условия сходства (гл. II) для объектов одного и того же класса. Может возникнуть ситуация, когда такие условия задаются только для одного из классов. Поэтому возникает вопрос, каким образом определяется задание условий сходства-различия в N -модели. В дополнение к сказанному об этом в § 4 гл. I отметим, что условие однородности задается только для одного из классов обычно в том случае, когда один из классов по содержательным соображениям оценивается как существ-

венно более однородный, чем другой. Оценка степени однородности классов может быть также проведена с помощью приемов, описанных в § 3, либо на основе выделения типичного объекта ("реплики" [15]) каждого класса с использованием меры типичности объектов, рассмотренной в [18].

2. Решающие функции типа взвешенных расстояний

В описанном алгоритме вначале вычисляются N -тесторные веса признаков H_1, \dots, H_n . Затем для каждого из объектов S , подлежащих распознаванию, вычисляются расстояния до классов. Как и в случае P -модели, возможны два способа оценки расстояний.

$$1) \rho_1^1(s) = \frac{\sum_{i=1}^{m_1} \sum_{j=1}^n (S_i \circ S_i^j) H_j}{m_1}, \quad \rho_2^1(s) = \frac{\sum_{i=1}^{m_2} \sum_{j=1}^n (S_i \circ S_i^j) H_j}{m_2}$$

$$2) \rho_1^2(s) = \max_{i=1, \dots, m_1} \sum_{j=1}^n (S_i \circ S_i^j) H_j, \quad \rho_2^2(s) = \max_{i=1, \dots, m_2} \sum_{j=1}^n (S_i \circ S_i^j) H_j$$

Пусть $\rho_1(s), \rho_2(s)$ - расстояния до классов по выбранному способу оценки. Тогда распознавание ведется по следующему правилу:

при $\rho_1(s) - \rho_2(s) \geq \epsilon$ объект S относится к I классу

при $\rho_2(s) - \rho_1(s) \geq \epsilon$ объект S относится ко II классу;

при $|\rho_1(s) - \rho_2(s)| < \epsilon$ объект S не распознается.

3. Решающие функции типа числа голосов

Здесь могут быть использованы два алгоритма распознавания. Первый из них совпадает с методом голосования по тупиковым тесторам, и поэтому его описание мы опустим*).

Схема распознавания, реализуемая вторым алгоритмом, существенно использует специфику N -тестора. В этом алгоритме используется метод голосования, имеющий сходные черты как с таковым для P -модели, так и методом голосования, описанным для Q -модели. Для простоты изложения ограничимся случаем, когда условия

*) Совпадение общей схемы распознавания, конечно, не означает совпадения результатов распознавания, т.к. меняется множество \approx информационных единиц.

сходства-различия заданы для обоих классов A_1, A_2 . Описание алгоритма на случай сходства-различия только для одного класса легко усматривается из последующего изложения.

Заметим, что если для $S \in \mathcal{J}$ набор $t = (x_{j_1}, \dots, x_{j_e})$, принадлежащий множеству τ тупиковых N -тесторов, голосует за объект S_i^k , $k = 1, 2$, то описание S признаками x_{j_1}, \dots, x_{j_e} имеет сходство со всеми описаниями объектов из A_k по данному набору признаков. Таким образом, если описание S присоединить к T_k , то набор t в расширенной таким образом таблице T останется тупиковым N -тестором. Это значит, что расширенная совокупность S_1^k, \dots, S_m^k удовлетворяет условию однородности на наборе $t = (x_{j_1}, \dots, x_{j_e})$. Этот факт представляет собой косвенное свидетельство в пользу принадлежности S к классу, представленному S_1^k, \dots, S_m^k .

Однако свидетельство такого рода мы имеем и тогда, когда t не голосует ни за один из объектов обучения но, тем не менее, S имеет сходство по данному набору признаков со всеми объектами из A_k . В этой ситуации присоединение S к T_k также сохраняет t тупиковым N -тестором. Будем говорить, что t голосует за класс A_k "в целом", если S имеет сходство по признакам x_{j_1}, \dots, x_{j_e} со всеми объектами из A_k , не голосуя за объекты другого класса. Ясно, что если t голосует за S_i^1 , то t голосует и за A_k "в целом". Обратное, вообще говоря, неверно.

Обозначим $\Delta_1(S), \Delta_2(S)$ - числа голосов "в целом" за классы A_1, A_2 . Предположим, что $\Gamma_1(S), \Gamma_2(S)$ - обычные оценки числа голосов за классы по выбранному способу (одному из двух, описанных ранее). Тогда распознавание ведется по следующему решающему правилу: при $\Delta_1(S) - \Delta_2(S) \geq \delta^1, \Gamma_1(S) - \Gamma_2(S) \geq \epsilon$ объект S относится к первому классу;

при $\Delta_2(S) - \Delta_1(S) \geq \delta^2, \Gamma_2(S) - \Gamma_1(S) \geq \epsilon$ объект относится ко второму классу;

во всех остальных случаях S не распознается.

Пороги $\delta^1 > 0, \delta^2 > 0$ выбираются из соображений оптимизации результатов экзамена системы распознавания.

В случае, когда условия сходства-различия заданы только для объектов одного класса (например, A_1), указанное решающее правило имеет следующую модификацию:

при $\Delta_1(S) \geq \delta^1, \Gamma_1(S) - \Gamma_2(S) \geq \epsilon$ объект S относится к первому классу;

при $\Delta_1(S) < \delta^1, \Gamma_2(S) - \Gamma_1(S) \geq \epsilon$ объект S относится ко второму классу;

во всех остальных случаях S не распознается.

Заключение

В публикуемой статье авторы не ставили своей целью дать полный обзор результатов в области тестового подхода. В стороне остались исследования математического аппарата теории тестов. Не обсуждаются и конкретные приложения подхода, хотя обобщение этих результатов и их критический анализ принесли бы несомненную пользу. Основное внимание уделяется описанию моделей тестовой обработки информации, расширяющих рамки тестового подхода и дающих возможность более полного соответствия с целью обработки информации. Вводимые модели позволяют предложить новые алгоритмы распознавания и упорядочения. Эти алгоритмы, описываемые в работе, апробированы на геологическом материале - в основном, по рудным месторождениям.

Такое рассмотрение тестового подхода, предпринятое авторами, совершенно не случайно. Дело в том, что за последние годы интенсивная разработка тестового подхода, с одной стороны, и развитие геологической науки, с другой, привели к возникновению новых целей и новых объектов исследования, что с неизбежностью повлекло за собой значительное возрастание разнообразия постановок задач обработки информации. В частности, возникли и получили свое развитие представления о геологическом поиске как информационной проблеме. Укажем на такие сферы геологического поиска:

- 1) поиск месторождений полезных ископаемых, стимулируемый всей системой промышленных реализаций на всем отрезке времени человеческой деятельности;
- 2) поиск соответствующих геологических обстановок и временных условий для строительства наземных, подземных и гидросооружений;
- 3) поиск геологических последствий, возникающих в условиях выравнивания масштабов промышленных и геологических процессов.

К настоящему времени представляется несомненным, что логико-информационный подход дает вполне подходящий математический

аппарат для решения этих задач геологического поиска. Такой вывод основывается на анализе специфики информации, которую приходится учитывать в процессе их решения. Тестовый же подход является одной из наиболее развитых частей логико-комбинаторного аппарата обработки информации.

Отметим также, что новое разнообразие целей и постановок привело к "отпочкованию" от тестового подхода новых логико-комбинаторных методов. Таковы, в частности, метод целевой итерационной классификации ("Цикл" [2]), метод согласованных оценок ("Качели" [7]) и метод Т-свойств [36]. Метод согласованных оценок дает оценки признаков и объектов из предположения наличия прямой и обратной связи между существенностью признаков и "важностью" объектов, на которых эти признаки выполняются. Указанные соотношения описываются системой уравнений, связывающих оценки признаков и объектов. Искомые оценки являются решением этой системы. Для их нахождения предложен [7] итерационный процесс ("Качели"), доказана его сходимость и проведен анализ скорости сходимости.

В методе "Цикл" по заданной системе целевых расстояний между объектами обучения производится отбор подсистемы признаков x_{j_1}, \dots, x_{j_k} из исходного признакового пространства x_1, \dots, x_n . Выбранная подсистема признаков располагает исследуемые объекты на соответствующем расстоянии. Этот отбор (а также оценка информативности признаков) производится в ходе содержательно интерпретируемого вычислительного итерационного процесса.

Метод Т-свойств нацелен на поиск определенного вида закономерностей, присущих классам объектов обучения. На основе вводимого понятия тушиковости [36, 38] производится отбор эмпирических закономерностей, не сводимых к более "простым" (в определенном смысле) закономерностям и на их основе производится анализ материала обучения и последующее распознавание проб. В отличие от тестового подхода (в его настоящем виде) метод Т-свойств занимается непосредственно взаимосвязями между признаками, а не оперирует такими обобщенными характеристиками, как веса признаков и объектов. Проведенные исследования показали, что метод Т-свойств удачно дополняет тестовый подход при логико-комбинаторном анализе информации. Частная реализация метода Т-свойств - метод пакетов - используется в настоящем сборнике для обработки данных рентгено-структурного анализа образцов браннерита (см. статью

Н.А.Кулик, В.О.Красавчикова, Т.И.Штатновой).

Важно отметить, что два метода ("Цикл" и "Т-свойства") могут работать не только с бинарными признаками, но и с данными, измеренными в любом из четырех общераспространенных типов шкал. Метод согласованных оценок работает с логическими признаками и признаками, измеренными в шкале интервалов.

В целом, на вопрос о полезности и своевременности развития и применения тестового подхода в области решения геологических задач авторы отвечают положительно.

Мы также благодарим В.Б.Кудрявцева за уместную стимуляцию наших усилий в работе; Ю.Л.Васильеву глубоко признательны за строгость и объективность критики. Все недочеты работы следует отнести за счет авторов.

Л и т е р а т у р а

1. Б а б к и н В.Ф. Метод универсального кодирования источника независимых сообщений неэкспоненциальной трудоемкости. - В кн.: Проблемы передачи информации, т.7, М., "Наука", 1971.
2. Б и ш а е в А.А. Итерационный способ нахождения информативной системы признаков для целевой классификации объектов. - В кн.: III Всесоюзная конференция по проблемам теоретической кибернетики. (Тезисы докладов). Новосибирск, 1974, с.185-187.
3. Б и ш а е в А.А., К а р б ы ш е в В.Д. Методика решения задач классификации геологических объектов с помощью ЭВМ в геологии. Алма-Ата, 1974, с.192-196.
4. Б у г а е ц А.К., Д в о р н и ч е н к о Г.К., М а ц а к А.П., С е р о в а Л.Л. Алгоритмы и программы решения геологических задач на ЭВМ "Минск-22" и "ЕЭСМ-3М", вып.2, Алма-Ата, 1969.
5. Б у г а е ц А.Н., Д у д е н к о Л.Н. Математические методы при прогнозировании месторождений полезных ископаемых. Л., "Недра", 1976, 272 с.
6. Б у м е й с т е р В.К., М а р к о в и ч З.П. Применение непараметрических методов распознавания для дифференциальной диагностики тиреотоксикоза и вегетоневроза. - "Изв. АН Латвийской ССР", 1970, № II, с.104-117.
7. В а с и л ь е в Ю.Л., Д м и т р и е в А.Н. Спектраль-

ный подход к сравнению объектов, охарактеризованных набором признаков. - "Докл. АН СССР", 1972, т.206, № 6, с.1309-1312.

8. Геология и математика. Задачи диагноза и распознавания в геологии, геохимии и геофизике. Новосибирск, "Наука", 1970, с.224. - Авт. Воронин Ю.А., Ионина Н.А., Каратаева Г.Н. и др.

9. Вышемирский В.С., Дмитриев А.Н., Трофимук А.А. Поисковые признаки гигантских нефтяных месторождений. М., 1971, 16 с. (Спец. докл. к УШ Мировому нефтяному конгрессу).

10. Васильев Ю.Р., Дмитриев А.Н., Золотухин В.В. Распознавание и оценка никеленосных дифференцированных трапловых интрузий Севера Сибирской платформы. - "Геология и геофизика", 1973, № 1, с.13-23.

11. Васильев Ю.Р., Велинский В.В., Дмитриев А.Н. Количественная оценка различий химических составов гипербазитов с помощью логико-дискретного анализа. - "Геология и геофизика", 1971, № 6, с.57-63.

12. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. О математических принципах классификации предметов и явлений. - В кн.: Дискретный анализ. Вып.7, Новосибирск, "Наука", 1966, с.3-15.

13. Дмитриев А.Н. Использование длин тупиковых тестов при обработке таблиц. - В кн.: Дискретный анализ. Вып.14, Новосибирск, "Наука", 1970, с.18-21.

14. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. Об одном принципе классификации и прогноза геологических объектов и явлений. - "Геология и геофизика", 1968, № 5, с.50-64.

15. Дифференциальная диагностика шизофрении и органических заболеваний головного мозга логико-дискретным методом. - В кн.: Проблемы моделирования психической деятельности. Вып.2, Новосибирск, 1968, с.307-322. - Авт. Волков П.П., Дмитриев А.Н., Нагаева Г.М., Пантелеев В.И.

16. Организация и обработка геологической информации с помощью ЭВМ на основе построения тупиковых тестов. - В кн.: Логико-информационные решения геологических задач. М., "Наука", 1975,

с.83-128. - Авт.: Дмитриев А.Н., Кренделев Ф.П., Бишаев А.А. и др.

17. Дмитриев А.Н. Новые тестовые разработки в задачах прогнозирования рудоносности (на примере трапловых интрузий). - В кн.: Математические методы при прогнозе рудоносности. М., "Наука", 1977, с.104-163.

18. Дмитриев А.Н., Красавчиков В.О. Процедура математической обработки описаний нефтяных месторождений. - "Геология и геофизика", 1976, № II, с.86-96.

19. Алгоритмы и программы вычислительной диагностики психических заболеваний. Новосибирск, 1969, 146 с. - Авт.: Волков П.П., Дмитриев А.Н., Киселев А.С. и др.

20. Дмитриев А.Н. Некоторые табличные числа. - В кн.: Дискретный анализ. Вып. 12, Новосибирск, "Наука", 1968, с.22-26.

21. Дмитриев А.Н., Васильев Ю.Р., Золотухин В.В. Логико-математическая обработка информации при выявлении перспективности сульфидного оруденения в некоторых трапловых интрузиях Севера Сибирской платформы. - "Геология и геофизика", 1968, № II, с.95-101.

22. Дмитриев А.Н., Золотухин В.В., Васильев Ю.Р. Опыт применения дискретной математической обработки информации по дифференцированным рудоносным трапловым интрузиям Северо-Запада Сибирской платформы. - "Сов. геология", 1968, № 12, с.98-108.

23. Логико-математическая обработка геологической информации. Программы к ЭВМ для логико-математической обработки геологической информации. (Оперативно-информационный материал). Новосибирск, 1975, 189 с. - Авт.: Дмитриев А.Н., Бабич В.В., Федосеев Г.С.

24. Программы "тестовых оценок" (тесты, тесторы). - В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.36-45. - Авт.: Дмитриев А.Н., Слущкая Т.Л., Смертин Е.А., Штатнова Т.И.

25. Опыт крупномасштабного прогнозирования редкометального оруденения в вулканогенных формациях с применением ЭВМ. - В кн.: Математические методы при прогнозе рудоносности. М., "Наука", 1977, с.164-192. - Авт.: Еремеев А.Н., Модников И.С., Писаревский В.М. и др.

26. Принципы крупномасштабного прогнозирования редкометалло-го оруденения в вулканоплутонических формациях (с применением ЭВМ).—*"Сов.геология"*, 1972, № I, — Авт.: Еремеев А.Н., Модников И.С., Писаревский В.М., Чесноков Л.В.
27. Журавлев Ю.И., Никифоров В.В. Алгоритмы распознавания, основанные на выполнении оценок.—*"Кибернетика"*, 1971, № 3.
28. Журавлев Ю.И., Туляганов Ш.Е. Измерение важности признака.—В кн.: *Вопросы кибернетики*. Вып.33, Ташкент, 1970.
29. Загоруйко Н.Г. Методы распознавания и их применение. М., *"Сов. радио"*, 1972.
30. Константинов Р.М. Математические методы в металлогенетических исследованиях рудных районов.—В кн.: *Металлогенетический анализ рудоуправляющих факторов в рудных районах*. М., *"Недра"*, 1972.
31. Логическое планирование экспериментов по выращиванию окрашенного кварца.—*Докл.АН СССР*, 1974, т.218, № I.— Авт.: Константинов Р.М., Белякова Ю.А., Сиротинская С.В., Хетчиков Л.Н.
32. Константинов Р.М., Дмитриев А.Н. Использование математических методов для анализа геологических факторов, влияющих на масштабы оруденения.—*"Геология рудных месторождений"*, 1970, № 2, с.56—64.
33. Константинов Р.М., Дмитриев А.Н. Методы обработки качественной геологической информации для определения формационного типа месторождений.—*"Изв. АН СССР, сер. геол."*, 1971, № 8, с.3—15.
34. Константинов Р.М., Королева З.Е., Кудрявцев В.Б. О комбинаторно-логическом подходе к задачам прогноза рудоносности.— В кн.: *Проблемы кибернетики*. М., *"Наука"*, 1976, с.5—34.
35. Королева З.Е. О некоторых характеристиках тестовых алгоритмов распознавания.—В кн.: *Тезисы III Всесоюзной конференции по проблемам теоретической кибернетики*. Новосибирск, 1974.
36. Красавчиков В.О. Модификация тестового подхода к анализу таблиц описаний на основе понятия пакета.—В кн.:

Дискретный анализ. Вып.26. Новосибирск, *"Наука"*, 1974, с.36—60.

37. Красавчиков В.О. Комплекс алгоритмов для программ распознавания по Т-свойствам.—В кн.: *Программные комплексы для целевой обработки информации*. Новосибирск, 1977, с.144—162.

38. Красавчиков В.О. Непрерывная Д.Н.Ф. и ее приложение к распознаванию.—В кн.: *Тезисы IV Всесоюзной конференции по теоретической кибернетике*. Новосибирск, 1977.

39. Кренделев Ф.П., Дмитриев А.Н., Журавлев Ю.И. Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики.—*Докл. АН СССР*, 1967, т.173, с.1149—1152.

40. Кренделев Ф.П., Дмитриев А.Н. Применение дискретной математики для выбора районов и направления поисково-разведочных работ с целью выявления крупных месторождений типа Витватерсранд.—В кн.: *Проблема металлоносности древних конгломератов на территории СССР*. М., *"Наука"*, 1969, с.64—94.

41. Кренделев Ф.П., Кренделев С.Ф. Эвристические методы в геологии (математические основы). М., *"Наука"*, 1977, 152 с.

42. Математические методы при прогнозе рудоносности. М., *"Наука"*, 1977, 276 с.

43. Мацак А.П. Проблемы обучения по малым выборкам при геологическом прогнозировании.—*"Изв. Каз.ССР, сер. физ.-мат."*, 1969, № 5, с.29—34.

44. Математические методы в геологии и геологическая информация. Международный геологический конгресс, XXIV сессия, доклады советских геологов. Проблема 16, симпозиумы 104, 105. М., *"Наука"*, 1972.

45. Миронов Ю.П. Теоретико-множественные модели графитовидов. М., *"Наука"*, 1975.

46. Оценка масштаба редкометаллового оруденения, локализованного в вулканических аппаратах (с помощью ЭВМ).—*"Сов.геология"*, 1969, № II, с.100—109,—Авт.: Модников И.С., Дмитриев А.Н., Еремеев А.Н. и др.

47. Москаленко Ю.С., Николаев В.В. О количественной оценке информативности признаков.—В кн.: *Информационные методы в системах управления, измерений и контроля*. Владивосток, 1972.

48. Москаленко Ю.С. О способах представления исходной информации в задачах распознавания образов.-В кн.: Управление и информация. Вып.14, Владивосток, 1974, с.132-136.
49. Нестеренко Г.В., Дмитриев А.Н., Бишаев А.А., Штатнова Т.И. Генетический анализ золотосносных россыпей с помощью математических средств.-В кн.: Минералогия и геохимия рудных месторождений Сибири. Новосибирск, "Наука", 1977, с.100-114. (Труды ИГиГ СО АН СССР, вып.370).
50. Николаев В.В. О получении тупиковых тестов при изменении диагностических таблиц.-В кн.: Управление и информация, вып.14, Владивосток, 1974.
51. Родионов Д.А. Статистические методы разграничения геологических объектов по комплексу признаков. М., "Недра", 1968.
52. Сиротинская С.В. Метод вариационных рядов и его применение к исследованию некоторых геологических особенностей оловянно-вольфрамовых месторождений.-В кн.: Логико-информационные решения геологических задач. М., "Наука", 1975, с.5-82.
53. Слущкая Т.Л. Алгоритм вычисления информационных весов признаков.-В кн.: Дискретный анализ. Вып.12, Новосибирск, "Наука", 1968, с.75-90.
54. Смертин Е.А. Ц-тесты в задачах тестового распознавания.-В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. (Краткие тезисы докладов к конференции 12-14 февраля 1973 г.). Новосибирск, 1973, с. 65-66.
55. Соколов А.Д. Оптимальное бинарное кодирование признаков, подсчет строчечных нагрузок и минимизация.-В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.151-158.
56. Соловьев Н.А. О максимальном числе тупиковых тестов.-"Кибернетика", 1972, № 1, с.28-30.
57. Усманов Ф.А. Основы математического анализа геологических структур. Ташкент, "ФАН", 1977, 206 с.
58. Федосеев Г.С. Предварительный анализ данных в логико-математических исследованиях.-В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1976, с.22-42.
59. Чегис И.А., Яблонский С.В. Логические способы контроля работы электрических схем.-"Труды Матем.ин-та В.А. Стеклова", 1958, т.51, с.270-339.
60. Шаратов И.П. К истории математической геологии.- В кн.: Математические методы в геологии. Львов, 1973, (Вестник Львовского гос.ун-та)
61. Тестовый подход к количественной оценке геолого-структурных факторов и масштабов оруденения.-"Геология рудных месторождений", 1971, № 2, - Авт.: Яблонский С.В., Демидова Н.Г., Константинов Р.М. и др.

ВОПРОСЫ ТЕОРИИ И АЛГОРИТМЫ НА БАЗЕ ПОСТРОЕНИЯ \mathcal{D} -ТЕСТОВ

В предыдущей работе сборника дана общая идеология \mathcal{D} -тестового подхода, здесь же мы рассматриваем детально некоторые его вопросы и доводим это рассмотрение до конкретных разработок, нацеленных в перспективе на класс задач практического значения. За быстрое воплощение работы в печать автор благодарит А.Н.Дмитриева. Автор признателен Ю.Л.Васильеву и В.Н.Носкову, беседы с которыми помогли ему расширить математический кругозор.

Цель настоящей работы - исследовать несколько основных моментов построения тестовых процедур [1-4] решения задачи сравнительного изучения объектов и дать на базе этого исследования соответствующие выражения их обоснования, дополнения, объяснения или обобщения. Так, употребляемое обычно отношение различия двух строк дополняется еще тремя отношениями: сходства, согласования и отсутствия сравнения. Определения теста и тестора обобщаются в определении \mathcal{D} -теста. Объясняется целевой смысл оценки столбца. Обосновывается, чем вызвана необходимость \mathcal{D} -тестовых оценок столбцов в задачах с наличием цели. Приводится общий алгоритм построения всех тупиковых \mathcal{D} -тестов. Вводятся дополнительные числовые меры для строк. Производится их классификация на математические и геологические. Процедура голосования по тупиковым \mathcal{D} -тестам дополнена с учетом голосов "против". Дается обобщенная формула поощрения длин тупиковых \mathcal{D} -тестов при оценке столбцов таблиц. Изложенное выше - краткое содержание работы, посвященной исследованию бинарных таблиц, т.е. матриц $T = [t_{ij}]$, в которых элемент $t_{ij} \in \{0, 1\}$ стоит на пересечении строки S_i , $i = 1, 2, \dots, m_2$ и столбца x_j , $j = 1, 2, \dots, n$; m - число строк, а n - число столбцов таблицы T .

Строки бинарных таблиц соответствуют геологическим или другим объектам, а столбцы - признакам объектов. Объекты будем обозначать символами S_i , а признаки - x_j , так же, как и соответствующие им строки и столбцы таблиц. На объектах S_i известны также значения $t_{i, n+1}$ целевого признака x_{n+1} , изменяющиеся в промежутке $0 \leq t_{i, n+1} < \infty$, $i = 1, 2, \dots, m$. В из-

ложении § I они не будут учитываться, но уже в § 2 будут приняты во внимание. Перейдем к изложению § I.

§ I. Определение наличных отношений между строками бинарных таблиц

Первый момент исследования - определение различимости двух строк таблицы. Элементарно она сводится к наличию в таблице такого столбца, по которому соответствующие 2 элемента строк различны.

Вообще, если проверить все C_m^2 сочетаний по 2 элемента бинарного столбца, то обнаружится, что каждое такое сочетание имеет вид

$$\begin{aligned} &\text{либо } (a, \bar{a}), \\ &\text{либо } (a, a), \text{ где } a \in \{0, 1\} \end{aligned}$$

и \bar{a} - булево отрицание a . Сочетания вида (a, \bar{a}) выражают противоположность, а сочетания вида (a, a) - равенство соответствующих элементов столбца; для целой бинарной таблицы они отражают различие и сходство тех строк, в которых стоят эти элементы. Причем те строки, которые различимы по каждому столбцу - несходны, а те, которые сходны по каждому столбцу - неразличимы. Остальные же строки бинарных таблиц одновременно и сходны, и различимы.

Эти отношения строк целесообразно оформить в следующее определение.

Определение I. Строки S_{i_1} , S_{i_2} называются различимыми по столбцу x_j , если $t_{i_1 j} \neq t_{i_2 j}$; строки S_{i_1} , S_{i_2} называются сходными по столбцу x_j , если $t_{i_1 j} = t_{i_2 j}$. Строки S_{i_1} , S_{i_2} называются несходными, если они различимы по каждому столбцу x_j , $j = 1, 2, \dots, n$; строки S_{i_1} , S_{i_2} называются неразличимыми, если они сходны по каждому столбцу x_j , $j = 1, 2, \dots, n$ и строки называются сходно-различимыми или согласованными, если они одновременно и сходны, и различимы.

Итак, строго говоря, любые две строки S_{i_1} , S_{i_2} таблицы T могут быть в одном из следующих трех отношений:

- 1) S_{i_1} и S_{i_2} неразличимы;
- 2) S_{i_1} и S_{i_2} несходны;
- 3) S_{i_1} и S_{i_2} согласованы.

Отношения же сходства и различимости представляют собой не-

полное выражение отношений неразличимости и несходства или неполное выражение отношения согласования двух строк.

Однако на установление согласования двух строк необходимо проверить 2 столбца, а на установление того, что строки неразличимы или несходны, необходимо проверить все столбцы таблицы. Поэтому на практике обычно ограничиваются установлением сходства или различия строк S_{i_1} , S_{i_2} , для чего достаточно проверить 1 столбец.

Кроме того, о тех парах строк S_{i_1} , S_{i_2} , для которых проверка не проводилась, мы еще не знаем, в каком из указанных отношений они находятся друг к другу. Мы можем сказать о таких строках

4) S_{i_1} и S_{i_2} не сравнивались.

Последнее высказывание выражает пустое отношение между строками S_{i_1} и S_{i_2} .

Перечисленные возможные отношения между строками таблицы T , после проверки некоторых пар строк, составят перечень \mathcal{N} — различных отношений между парами строк таблицы T , в котором для каждой из C_m^2 пар строк (S_{i_1}, S_{i_2}) указывается, что

- либо S_{i_1} , S_{i_2} неразличимы,
- либо S_{i_1} , S_{i_2} несходны,
- либо S_{i_1} , S_{i_2} согласованы,
- либо S_{i_1} , S_{i_2} не сравнивались.

Перечень \mathcal{N} показывает в общем, какого характера информацию о строках можно извлечь из описания таблицы T , а какого нельзя, т.е. служит основанием для принятия или непринятия того или иного перечня целевых требований к строкам таблицы. В следующем параграфе говорится о составлении целевого перечня.

§ 2. Построение перечня целевых требований и определение тупиковых \mathcal{D} -тестов

Перечень \mathcal{D} -целевых требований к строкам S_{i_1} , S_{i_2} таблицы T составляется, исходя из значений $t_{i_1, n+1}$ целевого признака применительно к строкам бинарных таблиц, по следующим правилам:

а) строки S_{i_1} и S_{i_2} не подлежат сравнению, если разница значений $t_{i_1, n+1}$, $t_{i_2, n+1}$ между собой для нас не играет роли;

- б) строки S_{i_1} и S_{i_2} должны быть различимы, если разница между $t_{i_1, n+1}$ и $t_{i_2, n+1}$ должна быть учитываема;
- в) строки S_{i_1} и S_{i_2} должны быть сходны, если из значений $t_{i_1, n+1}$ и $t_{i_2, n+1}$ делается вывод, что следует учесть их близость;
- г) строки S_{i_1} и S_{i_2} должны быть согласованы, если значения $t_{i_1, n+1}$ и $t_{i_2, n+1}$ таковы, что надо указать как на сходство, так и на различие их;
- д) строки S_{i_1} и S_{i_2} должны быть несходны, если разница между $t_{i_1, n+1}$ и $t_{i_2, n+1}$ велика;
- е) строки S_{i_1} и S_{i_2} должны быть неразличимы, если $t_{i_1, n+1} = t_{i_2, n+1}$.
- Перебрав все C_m^2 паросочетаний $t_{i_1, n+1}$, $t_{i_2, n+1}$ и установив K каждой из пар S_{i_1} , S_{i_2} одноцелевое условие из числа а) — е), получим перечень \mathcal{D} -целевых требований к строкам таблицы T . Поскольку перечень \mathcal{D} составляется независимо от перечня \mathcal{N} , может получиться так, что эти 2 перечня противоречивы. Так, если из \mathcal{N} известно, что строки S_{i_1} и S_{i_2} неразличимы, а согласно \mathcal{D} они должны быть различимы, то перечни \mathcal{D} и \mathcal{N} противоречивы. В этом случае таблица T не удовлетворяет целевым условиям признака x_{n+1} . Такие таблицы в дальнейшем изложении не рассматриваются. В том случае, если \mathcal{D} и \mathcal{N} непротиворечивы, таблица T называется \mathcal{D} -допустимой. Всякую таблицу T' , которая получается из T удалением некоторых столбцов, и в которой строки удовлетворяют \mathcal{D} , будем называть \mathcal{D} -допустимой частью таблицы T . Заметим, что условия д), е) выполняются для любой части \mathcal{D} -допустимой таблицы T и поэтому строки тех пар строк (S_{i_1}, S_{i_2}) таблицы T , на которые наложены условия д), е) перечня \mathcal{D} , можно не сравнивать между собой при проверке на \mathcal{D} -допустимость частей таблицы T . Будем считать, что они не подлежат сравнению. Так все разнообразие целевых требований к строкам частей таблицы T сводится к следующим четырем:
- а) строки S_{i_1} , S_{i_2} не подлежат сравнению;
- б) строки S_{i_1} , S_{i_2} должны различаться;
- в) строки S_{i_1} , S_{i_2} должны быть сходны;
- г) строки S_{i_1} , S_{i_2} должны быть согласованы;
- преимущество \mathcal{D} -допустимых частей перед таблицей T выражается в том, что, удовлетворяя всем требованиям целевого перечня \mathcal{D} , они более экономны по числу столбцов.

Проиллюстрируем сказанное выше на примере.

Пример I. Пусть таблица T и целевой признак имеют следующий вид:

$$T = \begin{array}{l} S_1 \begin{bmatrix} I & I \\ I & 0 \\ 0 & I \\ I & I \end{bmatrix} \begin{array}{l} 0,80 \\ 0,30 \\ 0,75 \\ 0,70 \end{array} \\ x_1 \quad x_2 \quad x_3 \end{array}$$

Тогда перечень M для таблицы T выглядит так (допустим, что S₃ и S₄ еще не сравнивались):

- 1) S₁ и S₄ неразличимы;
- 2) S₂ и S₃ несходны;
- 3) S₁ и S₂, S₁ и S₃, S₂ и S₄ согласованы;
- 4) S₃ и S₄ не сравнивались;

целевой перечень D возьмем в такой форме:

- а) S₁ и S₄, S₃ и S₄ не подлежат сравнению;
- б) строки S₁ и S₂, S₂ и S₃ должны быть различимы;
- в) строки S₁ и S₃ должны быть сходны;
- г) строки S₂ и S₄ должны быть согласованы;

сравнивая перечни D и M, видим, что они не противоречат друг другу. Следовательно, таблица T является D-допустимой. Если из T удалить любой столбец - x₁ или x₂, то оставшийся столбец не удовлетворяет условиям D, так как строки S₂ и S₄ не будут согласованы. Поэтому таблица T не имеет D-допустимых частей и ее набор столбцов образует тупиковый D-тест (x₁, x₂) согласно нижеследующему определению.

Определение 2. Набор столбцов (x_{j₁}, x_{j₂}, ..., x_{j_ε}) называется D-тестом для таблицы T, если таблица, образованная столбцами x_{j₁}, x_{j₂}, ..., x_{j_ε}, является D-допустимой частью таблицы T. D-тест называется тупиковым, если при удалении из него любого столбца он перестает быть D-тестом.

В зависимости от того или иного вида перечня D множество тупиковых D-тестов для таблицы T меняется. Отметим следующие 4 важных случая, когда D целиком состоит из условий только одного вида: а), б), в) или г). Тогда в качестве тупиковых D-тестов для T выступают:

- а) - столбцы таблицы T;

в случае б) - тупиковые тесты для T [I];

в) - тупиковые G-тесты [3];

г) - тупиковые H-тесты [4].

Таким образом, определение D-теста имеет большую общность и вид D-теста существенно зависит от способа задания целевого перечня D. Например, тестор для таблиц T₁, T₂, ..., T_μ [2, 4] есть частный случай D-теста для таблицы T, которая состоит из всех строк таблиц T₁, T₂, ..., T_μ, размещенных так, что строки таблиц с меньшими номерами располагаются выше, чем строки таблиц с большими номерами. Здесь перечень D составляется так, что каждая пара строк из T должна удовлетворять условию:

- а) если она лежит целиком в одной из T₁, T₂, ..., T_μ;
- б) если строки пары лежат в разных таблицах.

В заключение пункта отметим тот факт, что наличие цели попарного сравнения строк в условиях б), в), г) и ее отсутствие в условиях а) выражаются в том, что в качестве тупиковых D-тестов в первых случаях предстают сцепленные в коллективы наборы признаков, а во втором - изолированные, отдельные столбцы.

Перейдем к определению оценки столбцов с помощью тупиковых D-тестов.

§ 3. Определение, объяснение и обоснование D-тестовой оценки столбцов

D-тестовую оценку столбца x_j таблицы T определим по аналогии с тестовой оценкой P_j [2] следующим образом. Пусть K^D - число всех тупиковых D-тестов таблицы T, а K_j^D - число тех ее D-тестов, в которых содержится столбец x_j. Тогда величина

$$D = \frac{K_j^D}{K^D},$$

представляющая собой частоту участия x_j в тупиковых D-тестах, принимается в качестве такой оценки.

Объясним целевой смысл этой оценки столбца. Заметим, что если t = (x_{j₁}, x_{j₂}, ..., x_{j_ε}) - тупиковый D-тест, то никакой столбец не может быть удален из t без нарушения каких-либо условий, входящих в D. Таким образом, внутри фиксированного D-теста t все столбцы являются в равной степени необходимыми

для выполнения условий из \mathcal{D} .

Далее, число $K_{\mathcal{D}}$ всех тупиковых \mathcal{D} -тестов представляет собой количество неизбыточных наборов признаков, удовлетворяющих всем условиям из \mathcal{D} или элементарных вариантов. В таких вариантах, обеспечивающих выполнение целевого отношения \mathcal{D} , число K_j тупиковых \mathcal{D} -тестов с участием j -го столбца характеризует частоту необходимости j -го столбца. Но, в силу сказанного выше, естественно предположить, что все эти участия имеют равную значимость. Таким образом, \mathcal{D}_j представляет собой частоту необходимости участия x_j в равнозначных элементарных вариантах, обеспечивающих выполнение всех целевых условий из \mathcal{D} . Поэтому основной тезис, принятый в настоящем пункте, заключается в том, что чем выше \mathcal{D}_j , тем существеннее признак x_j для реализации поставленной цели обработки таблицы. Таким образом, \mathcal{D}_j представляет собой оценку важности признака x_j для сравнительного изучения объектов в тех рамках, которые определяются целеуказанием. Цель определяет все множество просматриваемых элементарных вариантов (\mathcal{D} -тестов) и тем самым задает дифференциацию признаков по их существенности.

Для того, чтобы обосновать необходимость \mathcal{D} -тестовых оценок столбцов в задачах с наличием цели попарного сравнения объектов, заметим, во-первых, что если \mathcal{D} состоит только из условий вида а), то все столбцы получают один и тот же вес $\mathcal{D}_j = \frac{1}{n}$, $j = 1, 2, \dots, n$. Поэтому, если никакая пара строк не подлежит целевому условию различимости, сходства или согласования ее строк между собой, то все столбцы получают одинаковые веса, т.е. равноправны для исследования. Таким образом, тот случай, когда все признаки равноправны для исследователя — тоже частный случай \mathcal{D} -тестового подхода.

Во-вторых, допущение об индивидуальности столбцов и их равноправном участии в удовлетворении условий перечня \mathcal{D} относительно строк справедливо лишь тогда, когда \mathcal{D} состоит только из условий а), но в общем случае оно недостаточно соответствует действительности. Так как обычно никакой отдельный столбец не отражает полного удовлетворения всем условиям перечня \mathcal{D} , а лишь частичное, то его участие зависит от участия остальных столбцов. Фактор неравноправия столбцов опирается на возможность обойтись без участия того или иного столбца при достижении выполнения

всех условий перечня \mathcal{D} .

Так, необходимость \mathcal{D} -тестовых оценок столбцов обосновывается их предпочтительностью перед оценками тривиальными, не учитывающими целевых требований к строкам таблицы T .

В следующем пункте приводится изложение алгоритма построения всех тупиковых \mathcal{D} -тестов.

§ 4. Алгоритм построения всех тупиковых \mathcal{D} -тестов

Алгоритм построения всех тупиковых \mathcal{D} -тестов для таблицы T включает этап построения таблицы сравнений T^* , которая строится из таблицы T с учетом целевых требований перечня \mathcal{D} .

Каждая строка таблицы T^* соответствует некоторой паре строк из T , связанных условиями б), в) или г) перечня \mathcal{D} . Никакой паре строк из T , связанных условием а), в таблице T^* не соответствует ни одной строки.

Перебирая все пары строк (S_{i_1}, S_{i_2}) таблицы T , мы включаем в таблицу T^* строку вида

$S_{i_1} \oplus S_{i_2}$, если S_{i_1} и S_{i_2} связаны условием вида б);

$S_{i_1} \odot S_{i_2}$, если S_{i_1} и S_{i_2} связаны условием в);

если же S_{i_1} и S_{i_2} связаны требованием г), то в T^* включаются обе строки — $S_{i_1} \oplus S_{i_2}$ и $S_{i_1} \odot S_{i_2}$; символом $S_{i_1} \oplus S_{i_2}$ обозначается такая строка $\tilde{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n)$, в которой $\sigma_j = t_{i_1 j} \oplus t_{i_2 j}$, а символом $S_{i_1} \odot S_{i_2}$ — такая строка $\tilde{\pi} = (\pi_1, \pi_2, \dots, \pi_n)$, в которой $\pi_j = t_{i_1 j} \odot t_{i_2 j}$; операции \oplus , \odot над элементами $t_{i_1 j}, t_{i_2 j}$ алфавита $\{0, I\}$ определяются по такой таблице

$t_{i_1 j}$	$t_{i_2 j}$	\oplus	\odot
0	0	0	I
0	I	I	0
I	0	I	0
I	I	0	I

Из определения таблицы T^* следует, что, если набор $t = (x_{j_1}, x_{j_2}, \dots, x_{j_e})$ столбцов таблицы T — тупиковый \mathcal{D} -тест для T , то набор $t^* = (x_{j_1}^*, x_{j_2}^*, \dots, x_{j_e}^*)$ столбцов T^* с теми же номерами имеет следующее свойство \mathcal{D}^* :

во-первых, не имеет нулевых строк,

во-вторых, при удалении хотя бы одного столбца приобретает нулевые строки.

Справедливо также то, что если набор $t^* = (x_{j_1}^*, \dots, x_{j_c}^*)$ таблицы T^* имеет свойство \mathcal{D}^* , то соответствующий ему набор столбцов $t = (x_{j_1}, \dots, x_{j_c})$ является тупиковым \mathcal{D} -тестом для таблицы T .

Таким образом, между тупиковыми \mathcal{D} -тестами для таблицы T и наборами столбцов из T^* , имеющими свойство \mathcal{D}^* (т.е. тупиковыми пакетами для T^* [5]), имеет место взаимнооднозначное соответствие и задача построения всех тупиковых \mathcal{D} -тестов для таблицы T сводится к задаче построения всех наборов столбцов из T^* , имеющих свойство \mathcal{D}^* .

Для того, чтобы не пропустить ни одного набора столбцов, их как бы нумеруют указанием номеров столбцов входящих в набор t^* . Это производится построением строки $\tilde{r} = (r_1, r_2, \dots, r_n)$, в которой

$$r_j = \begin{cases} 1, & \text{если } x_j^* \text{ входит в } t^*, \\ 0, & \text{если } x_j^* \text{ не входит в } t^*, \end{cases} \quad j = 1, 2, \dots, n.$$

Последовательная проверка всех наборов столбцов из T^* (на обладание свойством \mathcal{D}^*), соответствующих строкам r от $(0, 0, \dots, 0)$ до $(1, 1, \dots, 1)$ позволяет выявить и построить все тупиковые \mathcal{D} -тесты для таблицы T .

Описанный выше алгоритм может быть улучшен (если из таблицы T^* удалить строки - расширения, переставить в ней столбцы и вести проверку не всех наборов из T^* , а таких, которые удовлетворяют соответствующим условиям [6]) в смысле уместности для программной реализации на ЭВМ, но мы не будем останавливаться на этом подробнее здесь, а приступим к описанию числовых оценок строк как без учета целевых требований вида б), в), г), так и с учетом их.

§ 5. Числовые оценки строк в отсутствии целевых требований сравнения

Определим числовые характеристики строки $s = (t_1, t_2, \dots, t_n)$ в отсутствии целевых требований вида б), в), г) на строки, т.е. в том предположении, что все столбцы индивидуальны для сравнения строк.

Оценку различия между строками пары (s, s_i) можно получить из формулы

$$\rho(s, s_i) = \frac{1}{n} \sum_{j=1}^n |t_j - t_{ij}|, \quad 1 \leq i \leq m \quad (I)$$

для нормированного числа различий между соответственными компонентами строк s, s_i . Заметив, что в рассматриваемом случае коэффициент $\frac{1}{n}$ равен величинам \mathcal{D}_j , $j = 1, 2, \dots, n$, запишем формулу (I) в более общем виде

$$\rho(s, s_i) = \sum_{j=1}^n |t_j - t_{ij}| \cdot \mathcal{D}_j = \sum_{j=1}^n (t_j \oplus t_{ij}) \cdot \mathcal{D}_j$$

Оценка сходства строк пары (s, s_i) дается следующим выражением

$$q(s, s_i) = \frac{1}{n} \sum_{j=1}^n (1 - |t_j - t_{ij}|), \quad 1 \leq i \leq m, \quad (2)$$

нормированного числа совпадений одноименных компонент строк s, s_i и может быть представлена в виде

$$q(s, s_i) = \sum_{j=1}^n (1 - |t_j - t_{ij}|) \mathcal{D}_j = \sum_{j=1}^n (t_j \circ t_{ij}) \cdot \mathcal{D}_j$$

Из формул (I), (2) усматривается равенство $\rho(s, s_i) + q(s, s_i) = 1$.

Оценка согласования строк пары (s, s_i) выражается более сложно:

$$h(s, s_i) = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n (t_j \oplus t_{ij}) \cdot (t_k \circ t_{ik}), \quad (3)$$

а в общем виде выглядит так:

$$h(s, s_i) = \sum_{j=1}^n \sum_{k=1}^n (t_j \oplus t_{ij}) \cdot \mathcal{D}_j \cdot (t_k \circ t_{ik}) \cdot \mathcal{D}_k$$

Она равна $h(s, s_i) = \rho(s, s_i) \cdot q(s, s_i)$

Величины (I)-(3) построены в том естественном предположении, что значения "1" и "0" - равноправные значения алфавита $\{0, 1\}$. Другая точка зрения на значения "1" и "0" такова, что "1" лучше "0" в целевом отношении. Поэтому оценки строк индивидуальные можно получить двумя путями:

во-первых, считая, что "1" и "0" равноправные значения алфавита $\{0, 1\}$, суммированием парных оценок тех пар строк, в которые входит оцениваемая строка;

во-вторых, считая, что в геологическом отношении "1" значимее, чем "0", оцениваем то, насколько оцениваемая строка различима с

нулевой (т.е. "наихудшей") или сходна с единичной.

Так, индивидуальную оценку строки S по различию можно получить, суммируя парные оценки ее различия со всеми остальными, что дает

$$\pi(S) = \sum_{i=1}^m f_i(S, s_i) = \frac{m}{n} \sum_{j=1}^n [t_j(1 - \mathcal{P}_j) + \mathcal{P}_j(1 - t_j)], \quad (4)$$

где \mathcal{P}_j - частота встречаемости единиц в столбце x_j ; видно, что $\pi(S)$ отражает суммарное различие между строкой S и строками S_1, S_2, \dots, S_m , если "I" и "0" равноправные значения.

Еще одну индивидуальную оценку строки S_i по различию получим, считая, что "I" лучше "0"; оцениваем то, насколько S лучше нулевой строки (которая может и не быть строкой из T) по формуле

$$f_i(S) = \frac{1}{n} \sum_{j=1}^n t_j = f_i(S, \tilde{0}) \quad (5)$$

и в качестве оценки строки S выступает число единиц в ней.

Величина $f_i(S)$ может быть записана и в более общем виде следующим образом $f_i(S) = \sum_{j=1}^n t_j \mathcal{D}_j$, откуда видно, что число единиц в строке S - частный случай \mathcal{D} - тестовой оценки строки S .

Пример 2. Вычислим значения $f_i(i_1, i_2)$, $\pi(i)$, $f(i)$ для следующей таблицы T .

$$T = \begin{bmatrix} 0 & 0 & I & I & I \\ 0 & 0 & 0 & 0 & 0 \\ 0 & I & I & I & 0 \\ I & 0 & I & 0 & I \end{bmatrix}$$

Получим $f_i(I, 2) = \frac{3}{5}$, $f_i(I, 3) = f_i(I, 4) = \frac{2}{5}$, $f_i(2, 3) = f_i(2, 4) = \frac{3}{5}$, $f_i(3, 4) = \frac{4}{5}$, $\pi(I) = \frac{7}{5}$, $\pi(2) = \pi(3) = \pi(4) = \frac{9}{5}$, $f_i(I) = \frac{3}{5}$, $f_i(2) = 0$, $f_i(3) = f_i(4) = \frac{3}{5}$. Заметим, что значение $\pi(2)$ оценки строки S_2 принадлежит к разряду максимальных среди $\pi(I) - \pi(4)$, а значение $f_i(2)$ - наименьшее в ряду $f_i(I) - f_i(4)$. Это замечание наглядно показывает различие между способом оценки строк с учетом предпочтения "I" над "0" и способом оценки строк с предположением равноправия "I" и "0".

Индивидуальная оценка строки S по сходству в первом случае (равноправия "0" и "I") получается суммированием парных оце-

нок ее сходства со всеми и равна

$$h(S) = \sum_{i=1}^m q(S, s_i) = \frac{m}{n} \sum_{j=1}^n [t_j \mathcal{P}_j + (1 - t_j) \cdot (1 - \mathcal{P}_j)] \quad (6)$$

Поэтому $n(S) + \pi(S) = m$.

Во втором случае ("I" лучше "0") оценивается сходство S с единичной ("наилучшей") строкой и в качестве индивидуальной оценки по сходству выступает число

$$q(S) = \frac{1}{n} \sum_{j=1}^n t_j = q(S, \tilde{I}) = f_i(S), \quad (7)$$

единиц в строке S .

Пример 3. Для таблицы T , следующего вида

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & I & I \\ 0 & I & 0 & 0 & I \\ I & 0 & 0 & I & 0 \end{bmatrix}$$

вычислим значения величин $q(i_1, i_2)$, $n(i)$, $f(i)$, извлекаемых из понятия сходства. Получим такие результаты:

$$\begin{aligned} q(1, 2) = \frac{2}{5}, \quad q(1, 3) = q(1, 4) = \frac{3}{5}, \quad q(2, 3) = q(2, 4) = \frac{2}{5}, \quad q(3, 4) = \frac{4}{5}; \\ n(1) = \frac{13}{5}, \quad n(2) = n(3) = n(4) = \frac{11}{5}; \\ q(1) = 0, \quad q(2) = \frac{3}{5}, \quad q(3) = q(4) = \frac{2}{5}. \end{aligned}$$

Принципиальное отличие оценок $n(S)$ и $q(S)$ между собой наглядно видно на оценках строки S_1 , для которой $n(I)$ представляет собой максимальное из значений $n(I) - n(4)$, а $q(I)$ - минимальное из значений $q(I) - q(4)$.

Индивидуальная оценка строки S по согласованию в случае равноправия "0" и "I" есть сумма парных оценок ее согласования с остальными:

$$b(S) = \sum_{i=1}^m h(S, s_i) = \sum_{i=1}^m f_i(S, s_i) \cdot q(S, s_i). \quad (8)$$

Индивидуальная оценка S по согласованию при учете того, что "I" лучше "0" совпадает с парными оценками ее согласования с нулевой и с единичной строками. Она равна

$$h(S) = \frac{1}{n} \sum_{j=1}^n \sum_{k=i}^n t_j \cdot (1 - t_k) = h(S, \tilde{0}) = h(S, \tilde{I}). \quad (9)$$

Пример 4. Для таблицы T примера 3 вычислим значения величин $h(i_1, i_2)$, $b(i)$, $f(i)$, построенных на основе

понятия согласования. Придем к следующим значениям:

$$h(1,2) = h(1,3) = h(1,4) = h(2,3) = h(2,4) = \frac{3}{5}, \quad h(3,4) = \frac{2}{5};$$

$$b(1) = b(2) = \frac{2}{5}, \quad b(3) = b(4) = \frac{3}{5};$$

$$h(1) = 0, \quad h(2) = h(3) = h(4) = \frac{3}{5}$$

Отличие оценок $b(i)$ от оценок $h(i)$ заметно на примере оценок строки S_1 . Для нее $b(1)$ входит в число максимальных значений в ряду $b(1) - b(4)$, а $h(1)$ является минимальным значением из $h(1) - h(4)$.

По различиям получены такие оценки:

$$f(1,2) = \frac{3}{5}, \quad f(1,3) = f(1,4) = \frac{2}{5}, \quad f(2,3) = f(2,4) = \frac{3}{5}, \quad f(3,4) = \frac{4}{5};$$

$$\pi(1) = \frac{7}{5}, \quad \pi(2) = \pi(3) = \pi(4) = \frac{9}{5}; \quad h(1) = 0, \quad h(2) = \frac{3}{5}, \quad h(3) = h(4) = \frac{4}{5}$$

Оценки строк, заданные формулами (1) - (9), построены в том предположении, что все столбцы таблицы равноправны для сравнения ее строк. Эти оценки следует преобразовать так, чтобы новые оценки были справедливы в более общем предположении на столбцы, ибо допущение об индивидуальности столбцов и их равноправном участии в удовлетворении условий перечня \mathcal{D} относительно строк справедливо лишь тогда, когда \mathcal{D} состоит из условий вида а), но в общем случае оно недостаточно соответствует действительности. Так как обычно никакой отдельный столбец не отражает полного удовлетворения всем условиям вида б), в) или г) перечня \mathcal{D} , а лишь частичное, то его участие зависит от участия остальных столбцов.

Фактор неравноправия столбцов опирается на возможность обойтись без участия того или иного столбца при достижении выполнения всех условий перечня \mathcal{D} . Учитывая сказанное, будем считать столбцы сцепленными, а в качестве меры участия столбца x_j установим частоту привлечения его для получения несжимаемого выполнения всех условий перечня \mathcal{D} , то есть частоту его участия в тупиковых \mathcal{D} -тестах

$$D_j = \frac{\kappa_j}{\kappa_0}, \quad j = 1, 2, \dots, n$$

С учетом этих новых уточненных весов преобразуем введенные ранее оценки строк (1)-(9).

§ 6. Числовые оценки строк при наличии целевых требований сравнения строк

Преобразование оценок (1)-(9) в новые оценки производится так, чтобы эти новые оценки совпадали с оценками (1)-(9) в том случае, если \mathcal{D} сводится к одним только условиям вида а). В новых оценках каждое различие, сходство или согласование между строками учитывается со своим коэффициентом, равным частоте его участия в тупиковых \mathcal{D} -тестах.

Так, общая оценка различия строк $S = (t_1, t_2, \dots, t_n)$ и S_i и аналогичная (1) принимает вид

$$\rho(S, S_i) = \sum_{j=1}^n (t_j \oplus t_{ij}) \cdot D_j, \quad (1')$$

а соответствующая (2) общая оценка сходства строк S , S_i выглядит так:

$$\alpha(S, S_i) = \sum_{j=1}^n (t_j \circ t_{ij}) \cdot D_j. \quad (2')$$

Выше, в § 5, эти формулы были получены из других соображений.

Для оценки согласования строк S , S_i введем в учет величины

$$D_{j_1, j_2} = \frac{\kappa_{j_1, j_2}}{\kappa_0}, \quad 1 \leq j_1 < j_2 \leq n,$$

где κ_{j_1, j_2} - число тупиковых \mathcal{D} -тестов с участием пары столбцов (x_{j_1}, x_{j_2}) ; величины D_{j_1, j_2} представляют собой соответствующие тестовые характеристики, введенные в работе [7], а здесь распространенные на \mathcal{D} -тесты. Теперь оценка согласования, построенная по подобию оценки (3), представляется следующей формулой:

$$h(S, S_i) = \sum_{j=1}^n \sum_{k=1}^n (t_j \oplus t_{ij}) \cdot (t_k \circ t_{ik}) \cdot D_{j,k} \quad (3')$$

Два пути получения индивидуальных оценок строк S по различию приводят к следующим выражениям:

$$\Pi(S) = \sum_{i=1}^m \rho(S, S_i) = m \sum_{j=1}^n \{ [t_j (1 - p_j) + p_j (1 - t_j)] \cdot D_j \} \quad (4')$$

- в случае равноправия "0" и "1" - и

$$P(S) = \sum_{j=1}^n t_j \cdot D_j \quad (5')$$

- в случае предпочтения "1" перед "0", которые похожи на (4) и (5).

В качестве индивидуальных оценок строк по сходству проявля-

ются, с одной стороны, оценки, сходные с оценками формул 6)

$$N(s) = \sum_{i=1}^m Q(s, s_i) = m \sum_{j=1}^n \left\{ [t_j \cdot \mathcal{P}_j + (1-t_j) \cdot (1-\mathcal{P}_j)] \cdot \mathcal{D}_j \right\}, \quad (6')$$

а с другой стороны, оценки, сходные с оценками (7)

$$Q(s) = \sum_{j=1}^n t_j \cdot \mathcal{D}_j = P(s). \quad (7')$$

Наконец, воспользовавшись формулами (8) и (9), получаем индивидуальные оценки строк S_i по согласованию в случае равноправия "I" и "0"

$$B(s) = \sum_{i=1}^m H(s, s_i), \quad (8')$$

а в случае предпочтения "I" перед "0"

$$H(s) = \sum_{j=1}^n \sum_{k=1}^n [t_j \cdot (1-t_k) \mathcal{D}_{j,k}] \quad (9')$$

Оценки (I)-(9), (I')-(9'), полученные выше, можно классифицировать по тем или иным признакам и давать им соответствующие наименования. Ниже производится такая классификация.

§ 7. Классификация числовых оценок строк

Первое подразделение величин (I)-(9), (I')-(9') произведем по признаку необходимости построения тупиковых \mathcal{D} -тестов для их подсчета. Величины (I)-(9) не нуждаются в таком построении и называются априорными, а характеристики (I')-(9') называются апостериорными; подразделение на априорные и апостериорные характеристики следует работе [8], продолжающей [9].

Второе подразделение характеристик строк проведем по числу строк, описываемых ими. Величины (I)-(3), (I')-(3') характеризуют по две строки S и s_i и называются парными, а величины (4)-(9), (4')-(9'), характеризующие по одной строке, называются индивидуальными.

По характеру учета сравнений величины (I)-(9), (I')-(9') подразделяются на три класса:

оценки (1), (4), (5), (1'), (4'), (5') - величины, построенные на учете различий;

оценки (2), (6), (7), (2'), (6'), (7') - величины, учитывающие сходства;

характеристики (3), (8), (9), (3'), (8'), (9') - величины, построен-

ные для учета согласований.

Далее, оценки вида (I)-(4), (6), (8), (I')-(4'), (6'), (8') - называются симметричными или замкнутыми, относительно инверсии "0" и "I". Это означает, что замена любых столбцов таблиц на симметричные [10] не влияет на значения этих величин. Величины (5), (7), (9), (5'), (7'), (9') - несимметричные или открытые для такой инверсии, так как она приводит к изменению их значений.

Индивидуальные оценки (4)-(9), (4')-(9') по способу их получения подразделяются на банальные (4), (6), (8), (4'), (6'), (8') и оригинальные (5), (7), (9), (5'), (7'), (9'). Банальные оценки получают усреднением суммы парных оценок, а оригинальные имеют "начало отсчета", т.е. нулевую строку с нулевым весом.

Проведенная классификация величин (I)-(9), (I')-(9') не претендует на полноту, но и в таком виде может быть полезна как для содержательного ознакомления с ними, так и для дальнейшего размышления. Например, из рассмотрения этой классификации становится видно то, что в банальных оценках, вследствие усреднения, геологическое основание несколько заглушено, а в априорных оценках недораскрыто математически целевое соотношение между строками. Потому банальные оценки можно отнести к более математическим оценкам, а априорные - оригинальные к более геологическим. Они сами по себе являются односторонними. Недостатки этих оценок преодолены в величинах (5'), (7'), (9'), которые сочетают в себе математический учет целевой направленности и геологического начального основания и представляют собой наглядное изображение совместной работы геологии и математики.

Выше были рассмотрены \mathcal{D} -тестовые оценки столбцов и строк бинарных таблиц, получаемые в том предположении, что участие столбца в тупиковых \mathcal{D} -тестах различной длины имеет одну и ту же целевую нагрузку, независимо от их длины. В ряде задач, однако, [11], учет длин оказался существенным для их решения. Следующий пункт посвящен изложению одного способа обобщения такого учета.

§ 8. Учет длин тупиковых \mathcal{D} -тестов при оценке столбцов

Привлечение учета длин тупиковых \mathcal{D} -тестов для оценки столбцов бинарных таблиц представляет собой следствие того

взгляда на столбцы тупиковых \mathcal{D} -тестов, согласно которому столбцы более коротких тупиковых \mathcal{D} -тестов важнее, чем столбцы длинных в отношении перечня \mathcal{D} целевых условий. Один из возможных способов учета длин - следующий. Строится перечень $\{l_1, l_2, \dots, l_s\}$ всех наличных длин тупиковых \mathcal{D} -тестов таблицы с одновременным подсчетом чисел κ^{α} , $\alpha = 1, 2, \dots, s$, всех тупиковых \mathcal{D} -тестов длины l_{α} , чисел κ_j^{α} всех тупиковых \mathcal{D} -тестов с участием столбца x_j (и имеющих длину l_{α}). Затем подсчитываются величины

$$D_j^f = \frac{\sum_{\alpha=1}^s \kappa_j^{\alpha} \cdot f(l_{\alpha})}{\sum_{\alpha=1}^s \kappa^{\alpha} \cdot f(l_{\alpha})}, \quad j = 1, 2, \dots, n, \quad (10)$$

которые и представляют характеристики столбцов x_j , учитывающие длины тупиковых \mathcal{D} -тестов посредством новой функции f . Функция f называется функцией поощрения длин. Вид функции f зависит от цели исследования и задается исследователем. Благодаря этому последнему обстоятельству веса столбцов таблицы можно варьировать в некоторых пределах, пользуясь той или иной функцией поощрения.

Представляет интерес задача выяснения того, каковы эти пределы. Не останавливаясь на этом в настоящей работе, приведем примеры использования различных функций поощрения:

- $f = \text{const}$; в данном случае $D_j^f = D_j$, т.е. величины D_j^f представляют собой обычные веса D_j ;
- $f(l) = l$; получаемые веса D_j^f в том случае представляют собой обобщение весов из работы [II];
- $f(l) = \begin{cases} 1, & \text{если } l = l_{\alpha} \\ 0, & \text{если } l \neq l_{\alpha} \end{cases}$; в рассматриваемом случае веса D_j^f есть веса, построенные по тупиковым \mathcal{D} -тестам длины l_{α} .

Приведенные примеры показывают, что формула (10) обладает большой общностью. С употреблением весов D_j^f можно дать соответствующие оценки строк, подобные (I)-(9), подставив вместо D_j значения D_j^f . Не будем выписывать их здесь, а изложим процедуру голосования по тупиковым \mathcal{D} -тестам, в которой оценки строк подсчитываются без подсчета оценок столбцов; излагаемая процедура есть обобщение процедур голосования по тестам и по тесторам [I2, I3].

§ 9. Голосование по тупиковым \mathcal{D} -тестам

Голосование по тупиковым \mathcal{D} -тестам производится следующим образом. Строится множество δ всех тупиковых \mathcal{D} -тестов для таблицы T и одновременно вычисляется, сколько голосов "за" и "против" получает объект, подлежащий оценке. Считается, что тупиковый \mathcal{D} -тест t подает I голос либо а1) за отнесение строки s к i -й строке - s_i , если s сходна со строкой s_i по всем компонентам из t ;

а2) против отнесения s к строке s_i , если s различима от s_i по каждой компоненте из t ;

либо воздерживается от голосования за или против отнесения строки s к той или иной строке s_i , если ни для какой из s_i не выполнены ни а1), ни а2).

Перебрав все тупиковые \mathcal{D} -тесты, получим, что из них: за i -ю строку проголосуют $\kappa^+(s, s_i)$, $i = 1, 2, \dots, m$, против i -й строки проголосуют $\kappa^-(s, s_i) = 1, 2, \dots, m$, а остальные, числом $\kappa^0(s, s_i)$ не голосуют ни за, ни против i -й строки.

Решение об отнесении строки s к таблице T выносится на основании величины

$$\max_{1 \leq i \leq m} [\kappa^+(s, s_i) - \kappa^-(s, s_i)]$$

Если она положительна, то s относится к T , а в противном случае s не относится к T .

Если строка s относится к T , то она может быть отнесена по степени родства к той или иной строке из T на основании следующего алгоритма.

- Алгоритм отнесения строки s к строкам таблицы T s относится к той строке s_{i_0} , для которой

$$\kappa^+(s, s_{i_0}) - \kappa^-(s, s_{i_0}) = \max_{1 \leq i \leq m} [\kappa^+(s, s_i) - \kappa^-(s, s_i)].$$

Если s не относится к T , алгоритм б) не употребляется.

В данной работе мы затронули далеко не все вопросы теории \mathcal{D} -тестов. И даже в тех, которые освещались, некоторых коснулись только вскользь. Так обстоит дело с изложением алгоритма построения всех тупиковых \mathcal{D} -тестов, классификации числовых оценок строк, алгоритмов распознавания. Одной из причин такой схематичности было стремление представить себе и понять целост-

ную, единую картину разнообразных приемов и построений на базе \mathcal{D} -тестов. Поэтому акцент изложения делался на те построения, которые выясняли суть и содержательный смысл тех или иных приемов, в ущерб тем построениям, которые были бы лишь делом техники.

В то же время рассмотрения доводятся до конкретных разработок и некоторые представляют собой конкретизацию идей статьи [14].

Л и т е р а т у р а

1. Чегис И.А., Яблонский С.В. Логические способы контроля работы электрических схем.-"Труды МИАН", т.51, 1958, с.270-360.

2. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. О математических принципах классификации предметов и явлений.-В кн.: Дискретный анализ, вып.7, Новосибирск, "Наука", 1966, с.3-15.

3. Смертин Е.А. \mathcal{D} -тесты в задачах тестового распознавания.-В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск, "Наука", 1973, с.65-66.

4. Дмитриев А.Н. Новые тестовые разработки в задачах прогнозирования рудоносности (на примере трапсовых интрузий).- В кн.: Математические методы при прогнозе рудоносности. М., "Наука", 1977, с.104-163.

5. Красавчиков В.О. Модификация тестового подхода к анализу таблиц описаний на основе понятия пакета.-В кн.: Дискретный анализ, вып.26, Новосибирск, "Наука", 1974, с.36-60.

6. Бугаец А.Н., Дуденко Л.Н. Математические методы при прогнозировании месторождений полезных ископаемых. Л., "Недра", 1976, с.72.

7. Кудрявцев В.Б., Кудрявцев Вит.Б. О перспективности населенных пунктов.-В кн.: Исследование операций, вып.3, М., 1972, с.34.

8. Дмитриев А.Н., Кренделев Ф.П., Бишаев А.А., Смертин Е.А., Штатнова Т.И. Организация и обработка геологической информации с помощью ЭВМ на основе построения тупиковых тестов.-В кн.: Логико-информационные

решения геологических задач. М., "Наука", 1975, с.83-128.

9. Дмитриев А.Н. Некоторые табличные числа.-В кн.: Дискретный анализ, вып.12, Новосибирск, "Наука", 1968, с.22-26.

10. Дмитриев А.Н., Смертин Е.А. Алгоритм вычисления тестовых параметров бинарных таблиц в задачах распознавания.-В кн.: Алгоритмы и программы решения геологических задач на ЭЦВМ "Минск-2" и "БЭСМ-3М", вып.3, Алма-Ата, 1970, с.119-133.

11. Дмитриев А.Н. Использование длин тупиковых тестов при обработке таблиц.-В кн.: Дискретный анализ, вып.17, Новосибирск, "Наука", 1970, с.18-21.

12. Мацак А.П. Проблемы обучения по малым выборкам при геологическом прогнозировании.-"Изв. АН Каз.ССР. Сер.Физ.-мат.", 1969, № 5, с.29-34.

13. Смертин Е.А., Дмитриев А.Н. Дополнение к алгоритму распознавания "голосованием" по тестам и тесторам.- В кн.: Алгоритмы и программы решения геологических задач на ЭЦВМ "Минск-2" и "БЭСМ-3М". Вып.3, Алма-Ата, 1970, с.117-119.

14. Дмитриев А.Н. Вопросы формализованных постановок геологических задач прогнозно-пожского профиля.-В кн.: Логико-математическая обработка геологической информации, Новосибирск, 1976, с.3-22.

II. ВОПРОСЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ

Гуваков А.И., Дмитриев А.Н., Кандыба В.Н.

ОЦЕНКА ПЕРСПЕКТИВНОСТИ ОЛОВОРУДНЫХ РАЙОНОВ ПРИМОРЬЯ

Оценка перспективности оловорудных районов является частью задачи сравнительного изучения оловорудных месторождений Приморья при проведении региональных исследований по проблеме генезиса и рудоконтролирующих факторов на указанной территории. Работа по изучению месторождений олова осуществляется по инициативе Р.М.Константинова. Первые шаги в постановке и решении задачи проводились в тесном сотрудничестве с группой информационных исследований в геологии (ИГЕМ АН СССР) и геологами Приморского геологического Управления. В данной статье излагаются некоторые результаты решения задачи по прогнозу перспективных объектов путем разбора их на заранее выделенных классах. Излагаются также схема решения и перечень алгоритмов, привлеченных для сравнительного анализа и приведены количественные оценки исследуемых объектов.

Исходная информация

Исходная информация, подлежащая исследованию, была подготовлена коллективом научных сотрудников и геологов под руководством Р.М.Константинова (ИГЕМ АН СССР) и П.Н.Антонова (ШГУ). Первоначально информация представляла собой таблицу, содержащую 62 строки (объекта) и 167 столбцов (признаков, которыми характеризуются объекты).

Общее количество объектов (по качественным значениям целевого признака) было подразделено на три класса: крупные месторождения, мелкие месторождения, рудопроявления, а также был задан перечень объектов, подлежащих распознаванию проб. Кроме этого подразделения, было задано подразделение месторождений на три класса не по запасам, а по морфологическому критерию. В результате предварительного анализа исходной информации число целевого подразделе

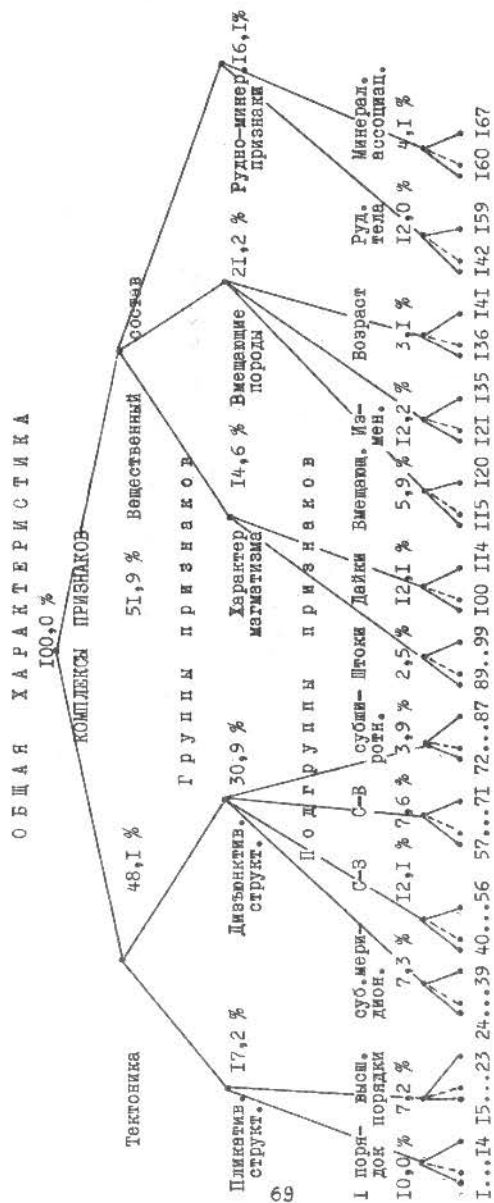


Рис. 1. Информационный граф оловорудных месторождений и проб.

ния классов по запасам было сокращено до двух, поскольку оказалось целесообразным объединить крупные и мелкие месторождения в один класс. Таким образом, в качестве классов эталонных [I] объектов принят класс месторождений (9 объектов) и класс рудопроявлений (18 объектов). Класс проб (в количестве 35) в результате предварительного анализа никаких изменений не претерпел. Согласно предварительным геологическим интерпретациям, часть из предъявленных проб более предпочтительна в отношении оруденения, но конкретный перечень и возможный масштаб этих оруденений задан в очень широких пределах.

Совокупность месторождений, рудопроявлений и проб охарактеризована списком в 167 признаков. В целом, общий список признаков подразделен на два блока:

- 1) освещающий тектонику района и объектов исследования;
- 2) характеризующий вещественный состав исследуемых объектов.

Более дробное подразделение признаков на родственно сжатые группы иллюстрируются на рис.1. Надо отметить высокую насыщенность пространства признаков тектоническими характеристиками. Так, число признаков, по существу характеризующих геометрию объектов, равно 104. Количество признаков вещественного характера значительно меньше, и этим признакам присуща меньшая информационная детальность.

В исходном признаковом пространстве представлены качественные, порядковые и количественные признаки [2,3]. Первоначально каждая градация количественных признаков была возведена в ранг самостоятельного признака, что в отношении распознавания снижало диагностическую роль некоторых из них. Для того, чтобы в схемах распознавания это не происходило, было проведено перекодирование общего списка признаков.

Постановка задачи

В геологическом отношении задача имеет как чисто производственный аспект, так и аспект вопросов, связанных с проблемой рудогенеза и рудного контроля. Однако вопросы, возникающие для разрешения как в практике, так и в теории, по существу своему взаимодополняющие.

- I. Подразделение пространства признаков на группы, характери-

зующие тот или иной продуктивный класс объектов;

2. Обнаружение комплекса разделяющих признаков для выделенных классов эталонных объектов;

3. Выделение группы признаков, наиболее тесно коррелирующей с запасами месторождений;

4. Выделение группы признаков для проведения сортировки объектов экзамена на классах месторождений;

5. Указание наиболее перспективных объектов в качестве первоочередных для детальной разведки из числа представленных проб.

В основу решения этих довольно сложных вопросов положены некоторые первоначальные предположения, которые содержат в себе гипотезу геолога. Так, в соответствии с содержательной геологической постановкой Р.М.Константинова, считается полезным в плане получения теоретических результатов подразделить общую совокупность признаков на три группы:

- 1) тектоническая группа признаков;
- 2) магматическая группа признаков;
- 3) ведущие минеральные ассоциации.

Для более детального рассмотрения сцепленности геологических процессов, имевших место на изучаемой территории, требуется (рис.1) выяснение "веса" каждой из характеристических подгрупп признаков.

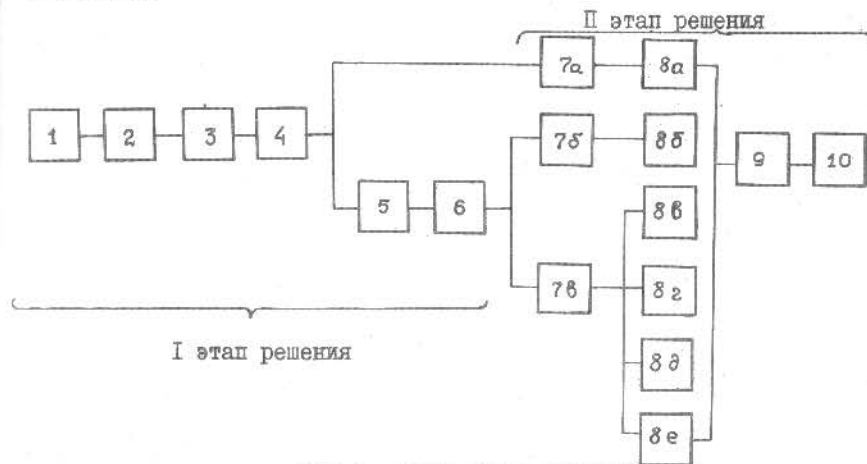


Рис.2. Блок-схема решения

I - постановка задачи; 2 - сбор исходной информации; 3 - бинарное кодирование; 4 - доалгоритмический анализ исходного материала; 5 - перекодирование исходной информации; 6 - анализ однородности эталонной выборки; 7 - минимизация признакового пространства: а, б) - по качельному алгоритму; в) - по алгоритму "Каскад"; 8 - распознавание проб по алгоритмам: а, б) - "Качели", в) - "Тесторы"; г) - "Q-тесты", д) - "Пакеты", е) - "Каскад"; 9 - обсуждение результатов распознавания; 10 - принятие решения. III этап решения по морфологии рудных тел на схеме не приведен.

В формализованном срезе задача сравнительного изучения месторождений была подразделена на три этапа (см. блок-схему решения рис.2).

На первом этапе было проведено изучение исходной таблицы в первоначальном коде характеристических признаков [4]. При этом основным средством исследования учтенной информации был метод согласованных оценок [5].

На втором этапе, посвященном в основном уточнению схемы распознавания и большей строгости процедур копирования исходной информации, применялся тестовый комплекс методов [6]. Полученные результаты, имеющие практический характер, были переданы для принятия решений в Приморское геологическое управление, они оказались согласованными с практическими результатами, полученными в ИГЕМе СССР.

На третьем этапе решалась задача об исследовании специфики сходств и различий месторождений и рудопроявлений, подразделенных по морфологическому критерию. В этом случае применялись, в основном, тестовые методы (такие, как P- и Q-тесты) [7], и полученные результаты во многом имеют теоретическую трактовку в области рудогенеза и рудоконтроля.

Кратко изложим схемы решения и основные результаты по двум первым этапам.

Решение по исходной таблице

Для данного этапа решения задачи характерен довольно большой объем исходной таблицы. Поэтому она и обрабатывалась методом согласованных оценок, позволившим охватить таблицу в целом. Полученные результаты по этой таблице, с одной стороны, дали оценки

объектов и признаков, которые представили интерес с содержательной точки зрения сами по себе, с другой стороны, эти оценки сориентировали ход дальнейшего анализа.

В частности, упорядочение строк матрицы по значениям их нагрузок (вычисленных по всем 167 признакам) показало, что в пределах значений нагрузок строк первого ранга локализовались все объекты первого класса, в следующий ранг попали объекты второго класса (рудопроявления). Эта первичная проверка качества информации и уместности алгоритма оказалась вполне приемлемой и по результатам оценок признаков, из общей совокупности которых для дальнейших диагностических целей было отобрано (по максимальным весам) 32 признака.

Для выяснения некоторых теоретических вопросов и, в частности, проблемы выделения рудоконтролирующего комплекса признаков путем суммирования "качельных" весов по подгруппам, группам и комплексам признаков в граф исходной информационной таблицы были внесены соответствующие количественные оценки (рис.1). Полученные величины, взятые в процентах от общей суммы информационных весов, помещены в вершинах графа. Значительного преимущества это построение не дало ни одному из комплексов, так, "вещественный состав" получил 51,9%, а "тектоника" - 48,1%, но на уровне групп обнаружено значительное информационное превосходство признаков, характеризующих "дизъюнктивные структуры" - 30,9% над остальными.

При переходе к более подробным подразделениям признакового пространства получена следующая (табл.1) последовательность подгрупп:

Таблица № 1

№ п/п	Название подгруппы признаков	Число признаков в подгруппе	Суммарное значение нагрузки в подгруппе в %
1.	Изменения вмещающих пород	15	12,2
2.	Северо-Западные нарушения	17	12,1
3.	Дайки кислого состава	15	12,1
4.	Элементы залегания рудных тел	18	12,0
5.	Пликативные структуры I порядка	14	10,0
6.	Северо-Восточные нарушения	14	7,6
7.	Субмеридиональные нарушения (крупные)	16	7,3

8.	Высшие порядки пликативных структур	9	7,2
9.	Вмещающие породы	6	5,9
10.	Ведущие минеральные ассоциации	8	4,1
11.	Субширотные нарушения	16	3,9
12.	Возраст	6	3,1
13.	Штоки, штокообразные интрузии	11	2,5
		Итого: 100%	

Кроме того, для выявления информационной существенности выделенных подгрупп, групп и комплексов признаков была проведена осредненная количественная оценка наиболее общих единиц подразделения признаков — тектонического и вещественного комплексов. Величины, характеризующие эти комплексы, оказались, по существу, равными: комплекс "тектонические признаки" имеет средний вес 0,325, а комплекс "вещественный состав" — 0,324. Неожиданным оказался факт "информационного превосходства" (в смысле поиска черт сходства сравниваемых объектов) группы признаков "пликативные структуры" ($\bar{\omega} = 0,407$) над группой признаков "дизъюнктивные структуры" ($\bar{\omega} = 0,244$).

На материале исходной таблицы, на базе метода согласованных оценок и понятия "реплика" в три шага была проведена минимизация признаков и распознавание проб.

Для проверки влияния процедуры минимизации на результаты распознавания предварительно было проведено распознавание объектов на полном признаковом пространстве. В качестве объектов экзамена случайным образом были выбраны 8 объектов из II-го класса. При распознавании правильно были рассортированы только три из восьми рудопроявлений (5 попали в класс месторождений).

Затем из полного пространства были отобраны 44 признака, обладающих наибольшими информационными "начальными" весами. Отбор производился следующим образом. Все признаки были упорядочены по убыванию информативности, и минимизация производилась по резкому перепаду значений в весах признаков. Результаты экзамена при этом значительно улучшились: лишь один объект был ошибочно отнесен к I классу, в двух случаях объекты попали в зону неопределенности, а пять объектов были распознаны правильно.

Так как в упорядоченном ряду, составленном из 44 признаков, наблюдается еще один перепад в информационных весах, признаковое

пространство было сокращено до 32. На этом признаковом пространстве были получены следующие результаты внешнего экзамена: 7 объектов были распознаны правильно и лишь один объект попал в зону неопределенности.

Таким образом, можно сделать вывод о том, что процедура минимизации признакового пространства приводит к улучшению качества распознавания, хотя вопрос об оптимальности минимизации в данном случае не решался. На третьем же шаге решения (прогнозом) на 35 проб четыре пробы были отнесены к объектам высокого промышленного значения (табл.7).

Таким образом, в задаче изучения совокупности оловорудных месторождений Приморья и разбраковки площадей на их промышленную перспективность на первом этапе решения были получены следующие результаты практического и теоретического значения:

1. Выделены существенные группы признаков и обнаружен следующий порядок их значимости в рудоконтроле: вмещающие породы — высшие порядки пликативных структур — вторичные изменения — дайки кислого состава.

2. Выделена группа из 32 признаков, которая пригодна для процедуры сортировки объектов по эталонным классам.

3. Указаны объекты, перспективные на обнаружение в них значительных скоплений оловянных руд.

В дальнейшем, в тесном контакте с геологами, задача была уточнена и в некоторых своих разделах подвергалась перестановке. Эти уточнения и модификация задачи в более жесткий "прогнозный вид" вызвали необходимость в преобразовании исходной информации.

Преобразование исходной таблицы

Преобразование исходной таблицы, обусловленное необходимостью повышения надежности прогноза, осуществлено путем перекодирования характеристических признаков, и в плане отбора однородной совокупности объектов для последующего распознавания.

Перекодированию подвергались порядковые и количественные признаки, обладающие двумя и более градациями. Эта процедура производилась путем объединения градационных делений в один признак, новая граница кодирования которого обеспечила максимально возможное число различий между выделенными классами. В результате

Т а б л и ц а 2
Таблица признаков и принятых кодов

№ признака в таблице	Название признаков	Код	Расшифровка кода
I	2	3	4
I. Тектонические и рудно-минералогические признаки			
I. Складчатость II порядка			
I	Складки II порядка	I 0	синклинали антиклинали
2	Угол падения крыла	I 0	выше 45° до 45°
3	Простирание складки	I 0	северо-восточное остальные направления
4	Амплитуда складки	I 0	до 5 км более 5 км
5	Флексуры	I 0	имеются отсутствуют
2. Складчатость высших порядков			
6	Ядро антиклинали	I 0	есть нет
7	Ядро синклинали	I 0	есть нет
8	Крылья синклинали	I 0	есть нет
9	Простирания складки	I 0	северо-восточное остальные направления
10	Число складок	I 0	> 2 < 2
3. Субмеридиональные нарушения (крупные)			
11	Отклонение простирания нарушения	I 0	есть нет
12	Падение западное	I 0	угол = 60° угол > 60°
13	Падение восточное	I 0	угол > 60° угол = 60°
14	Длина нарушения	I 0	длина > 200 м длина ≤ 200 м

Продолжение таблицы 2

I	2	3	4
15	Мощность нарушения	I 0	мощность ≥ I м мощность < I м
16	Число нарушений	I 0	> 3 ≤ 3
4. Северо-западные нарушения (крупные)			
17	Отклонение простирания нарушения	I 0	есть нет
18	Падение северо-западное	I 0	угол < 75° угол ≥ 75°
19	Падение юго-восточное	I 0	угол ≥ 60° угол < 60°
20	Длина нарушения	I 0	> 1000 м ≤ 1000 м
21	Мощность нарушения	I 0	> 10 м ≤ 10 м
22	Число нарушений	I 0	> 3 ≤ 3
5. Северо-восточные нарушения (крупные)			
23	Отклонение простирания нарушения	I 0	есть нет
24	Падение северо-западное	I 0	угол ≥ 60° угол < 60°
25	Падение юго-восточное	I 0	угол ≥ 60° угол < 60°
26	Длина нарушения	I 0	> 500 м ≤ 500 м
27	Мощность нарушения	I 0	> I м ≤ I м
28	Число нарушений	I 0	> 3 ≤ 3
6. Субширотные нарушения (крупные)			
29	Отклонение простирания нарушения	I 0	есть нет
30	Падение северное	I 0	угол < 75° угол ≥ 75°
31	Падение южное	I 0	угол ≥ 60° угол < 60°
32	Длина нарушения	I 0	> 1000 м ≤ 1000 м

Продолжение таблицы 2

1	2	3	4
33	Мощность нарушения	I 0	\geq I м $<$ I м
34	Нарушения	I 0	есть нет
7. <u>Размеры и элементы залегания рудных тел</u>			
35	Простираание рудных тел в одном или нескольких направлениях	I 0	в двух и более в одном направлении
36	Угол падения	I 0	угол $> 75^{\circ}$ угол $\leq 75^{\circ}$
37	Длина рудных тел	I 0	> 500 м ≤ 500 м
38	Средняя мощность	I 0	$> I$ м $\leq I$ м
39	Количество рудных тел	I 0	> 10 м ≤ 10 м
II. Петролого-литологические признаки			
Магматические признаки			
8. <u>Дайки</u>			
40	Дайки кислого состава	I 0	есть нет
41	Дайки среднего состава	I 0	есть нет
42	Протяженность дайки	I 0	> 1000 м ≤ 1000 м
43	Мощность даек	I 0	> 10 м ≤ 10 м
44	Число даек	I 0	> 10 ≤ 10
45	Взаимотношение даек	I 0	пересекающееся параллельное
9. <u>Рудовмещающие породы</u>			
46	Песчаники	I 0	есть нет
47	Алевролиты и сланцы	I 0	есть нет

Продолжение таблицы 2

1	2	3	4
48	Флиш	I 0	есть нет
49	Кислые магматические породы	I 0	есть нет
10. <u>Изменения вмещающих пород предрудный метасоматоз</u>			
50	Биотизация	I 0	есть нет
51	Пропилитизация	I 0	есть нет
52	Серицитизация	I 0	есть нет
53	Хлоритизация	I 0	есть нет
54	Турмалинизация	I 0	есть нет
Околожильные изменения			
55	Окварцевание	I 0	есть нет
56	Серицитизация	I 0	есть нет
57	Хлоритизация	I 0	есть нет
58	Турмалинизация	I 0	есть нет
59	Сульфидизация	I 0	есть нет
60	Карбонатизация	I 0	есть нет
II. <u>Возраст вмещающих пород</u>			
61	Возраст пород	I 0	нижний мел все остальное
III Рудно-формационные признаки			
12. <u>Ведущие минеральные ассоциации</u>			
62	Кв.-касс. (пир.+а/пир.+вольфр.) главная	I 0	есть нет
63	Колчеданная главная и второстепенная	I 0	есть нет
64	Гал.-сфал.-х/пир. главная и второстепенная	I 0	есть нет
65	Карбонатная второстепенная	I 0	есть нет

перекодирования, исходное число признаков сократилось до 65 и был получен отождествляющий блок (А) признаков [9] для первого класса (5 признаков). Содержание признаков и кодовые обозначения приведены в таблице 2, в таблицах 3 и 4 приведены классы эталонов, а таблица 5 иллюстрирует общий перечень проб в новом кодовом обозначении.

Информационный граф, по расчетам может быть построен для класса месторождений тем же способом, что и граф на рис. I. Он хорошо выявляет для этого класса большую значимость тектонического комплекса признаков - 61%, по сравнению с комплексом вещественного состава - 39%, а среди групп признаков - группу, характеризующую дизъюнктивные структуры - 43,3%, которая по значимости более чем вдвое превосходит ближайшую к ней группу - "пликативные структуры" - 17,7%. Оценка существенности подгрупп признаков выявила их очередность, представленную в таблице 6.

Таблица № 6

№ п/п	Название подгруппы признаков	Суммарное значение подгруппы в %
1.	Субмеридиональные нарушения	12,4
2.	Дайки кислого состава	12,1
3.	Северо-Западные нарушения	11,0
4.	Пликативные структуры высшего порядка	10,4
5.	Субширотные нарушения	10,1
6.	Изменения вмещающих пород	9,8
7.	Северо-Восточные нарушения	9,8
8.	Элементы залегания рудных тел	8,5
9.	Пликативные структуры I порядка	7,3
10.	Ведущие минеральные ассоциации	4,7
11.	Вмещающие породы	3,2
12.	Возраст пород	0,7

Анализ однородности совокупности описаний объектов становится особенно настоятельным во всех видах задач, сводимых к проблеме распознавания образов. В данном случае анализ однородности не заканчивается обнаружением факта однородности или неоднородности

объектов, а продолжается в задачу организации однородности выделением наиболее родственных объектов (по общему списку значений признаков) в соответствующие группы. Такую организацию однородности можно осуществить удалением малотипичных объектов или привлечением типичных в ту или иную совокупность описаний объектов. Очевидно, что на этапе составления исходной таблицы к объектам предъявляются требования естественно-содержательной однородности, а при математической обработке таблиц описаний требования однородности становятся более жесткими, и их соблюдение выделяется анализом конкретных величин значений каждого признака для каждого объекта. Помимо естественных черт неоднородности, неоднородность описаний, вскрываемая математически, может происходить от неполноты и неточности сведений, внесенных в описания объектов. Именно поэтому и возникает необходимость в алгоритмической обработке информации, нацеленной на ее фильтрацию [10].

Эта фильтрация может заключаться в том, что производится поиск тех признаков, которые, являясь наиболее общими для объектов, служат как бы связующими элементами сочленов класса. Например, отождествляющие признаки или признаки с фиксированным порогом частот встречаемости данного значения (для бинарных символов $> 0,5$)

При выяснении однородности каждого класса по отдельности уместно в исходном признаковом пространстве задаться соответствующей метрикой и установить меру значимости каждого описания объекта в данном классе. Раздельная работа с каждым классом, заканчивающаяся вычислением мер каждого объекта в классе, позволяет, в первом приближении, судить о компактности классов. Подчеркнем, что хотя и много информации заключено в индивидуальном описании объекта, при исследовании однородности и построении компактного класса учитывается информация, относящаяся к организации структуры всей таблицы в целом.

Анализ однородности объектов

Анализ однородности исследуемой совокупности объектов производится с целью более точного получения результата процедурами распознавания. Минимизация признакового пространства осуществлялась в связи с поиском рудоконтролирующих признаков комплексов

с последующим их применением для разбраковки объектов из класса проб. Основные результаты исследования однородности объектов и организации компактных классов были получены с помощью метода согласованных оценок строк и столбцов матриц [5]. По вычисленным величинам нагрузок строк (важности месторождений по заданному критерию) четко выделилась однородная группа из семи объектов (с диапазоном значений строчечных нагрузок, в условных ненормированных единицах: 31,7 - 37,5). Два объекта - № 8 и № 5 оказались малотипичными для класса месторождений в данном пространстве признаков (величины строчечных нагрузок составили 27,0 - 27,8 соответственно).

Далее был произведен анализ близости объектов второго класса к объектам первого. Процедура анализа заключалась в подсчете аналогичных нагрузок строк испытуемых объектов с использованием информационных весов признаков, полученных при обработке первого класса. В результате 5 объектов: а именно, № 25, № 20, № 13, № 24, № 27 попали в диапазон строчечных нагрузок объектов первого класса, причем первый из них расположился в пределах однородной группы. Остальные объекты получили более низкие строчечные нагрузки. Таким образом, объекты второго класса в целом образуют компактную группу и лишь пять объектов оказываются малотипичными для данной выборки в принятом признаковом пространстве.

С содержательной позиции малотипичность какого-либо объекта может объясняться либо его ошибочным включением в анализируемый класс (по целевому признаку), либо его обособленной геологической характеристикой.

На основании результатов проведенного анализа однородности из материала обучения были удалены 4 объекта - № 25, № 20-П, № 24 и № 13. При решении задачи другими алгоритмами они использованы в качестве объектов экзамена. Это позволило привести исходный материал к более компактному и, следовательно, более пригодному для распознавания проб виду.

Минимизация признакового пространства

Сокращение числа характеристических признаков для целей улучшения рабочей схемы распознавания (минимизация признакового пространства) осуществлено двумя методами:

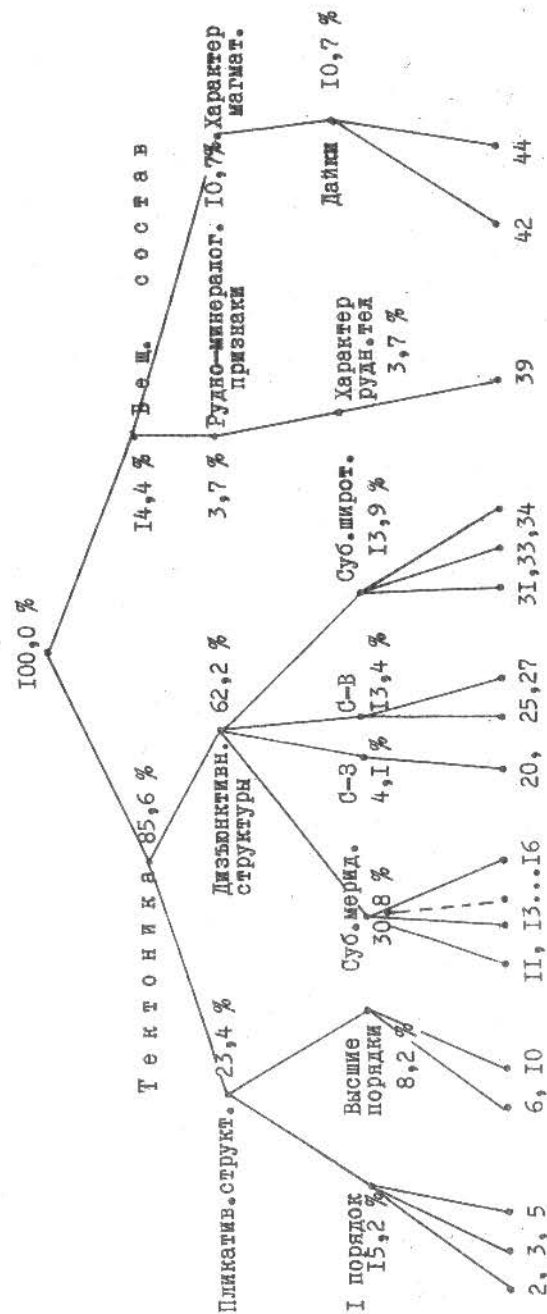


Рис. 3. Информационный граф различий промышленного и непромышленного классов.

1. Минимизация преобразованного признакового пространства на основе метода согласованных оценок проводилась для класса месторождений (в количестве 9) и рудопроявлений (в количестве 14) в соответствии с ранее описанной схемой. Для данного случая порог отбора признаков (в минимизированное пространство) по строчечным нагрузкам составил значение - $\omega_i > 0,5$. Количество признаков после минимизации составило 58.

2. Минимизация признаков по методу суммарного учета мер приуроченности и согласования ("Каскад") [9] производилась простейшим способом. При обработке информации для каждого признака, вошедшего в преобразованную таблицу, была получена оценка информативности, характеризующая его межклассовую способность. Граница минимизированного пространства определялась по наиболее резкому перепаду величины P_j^v (0,397 и 0,198) в упорядоченном ряду признаков. Полученное таким образом пространство включает 19 признаков.

На данном шаге общего решения задачи оказалось целесообразным произвести оценку разделяющей способности не только каждого признака в отдельности, но и подгрупп, групп и комплексов признаков. Для этого разделяющие способности признаков, входящих в соответствующие подразделения, суммировались сначала по подгруппам, группам, а затем и комплексам признаков. Интегральная разделяющая способность минимизированного пространства была принята за 100%, что позволило разделяющие способности всех подразделений признаков также выразить в процентах. Этот анализ позволил построить граф разделяющей информационной способности пространства, изображенный на рис.3, из которого следует, что наиболее контрастно месторождения и рудопроявления различаются по группе признаков "дизъюнктивные структуры", содержащей 62,2% различий.

Поскольку опыт решения прикладных прогнозных задач, а также результаты экзамена на минимизированном исходном материале говорят об улучшении качества распознавания при минимизации признакового пространства, внешний экзамен в данном случае не проводился. Сравнение результатов минимизации перекодированной информации с применением программы "Качели" и программы "Каскад" показывает, что при выполнении этой операции разными математическими приемами получаются различные по объему признаковые пространства (58 и 19 признаков). Тем не менее все 19 признаков, полученные по "Каскаду", входят в состав 58, полученных по "Качелям";

Таблица № 7

Название процедуры	Качели	Частота встреч	Качели	Тестеры	Шагеты	Q-тесты		Каскад	Каскад
						Каскад	Каскад		
Кол-во признаков									
Номера проб	167	65	65	19	19	19	19	56	19
№ 41	I	I	I	I	I	I	I	I	I,00
№ 30	0	I	I	I	I	I	I	I	0,88
№ 38	0	I	I	I	0,5	0,5	0,5	2	0,69
№ 58	0	0	I	I	I	I	I	0,5	0,62
№ 46	I	0	0	I	I	I	0,5	0,5	0,56
№ 49	I	I	I	I	0,5	0,5	0	0	0,56
№ 28	I	0	0	I	0,5	0	0	0,5	0,44
№ 45	I	I	I	0	0	0	0	0	0,38
№ 47	I	0	0	I	0,5	0,5	0	0	0,38
№ 50	0	I	I	I	0	0	0	0	0,38
№ 34	I	0	I	0,5	0	0	0	0	0,31
№ 51	0	0	0	0,5	0,5	1	0	0	0,25
№ 29	0	0	0	0,5	I	0,5	0	0	0,25
№ 52	0	0	I	0	0	0	0	0	0,12
№ 42	0	0	0	0	0,5	0	0	0,5	0,19
	0,44	0,76	0,91	I,00	0,87	0,70	0,44	0,64	

I - первый класс; 0,5 - неопределенность; 0 - II класс.

Сопоставление этих пространств с результатами минимизации на исходном материале весьма затруднено, т.к. при перекодировании информации претерпела значительные изменения как в количественном, так и в содержательном отношении.

Прогнозирование объектов

В соответствии с принятой схемой решения общей задачи (рис.2) распознавания проб и прогнозная оценка каждой пробы производилась как на исходном признаковом пространстве, так и на перекодированном.

1. Распознавание проб на таблицах исходного кодирования значений признаков проводилось по сокращенному списку признаков (насчитывающему 32 признака) с помощью программы, построенной на базе метода согласованных оценок и понятия "реплики". Для выявления перспективных промышленных объектов распознаванию подверглись все 35 проб. Причем, ввиду отсутствия в признаковом пространстве отождествляющего блока -А, анализ проб на допустимость их к распознаванию не производился. В результате счетных процедур к классу месторождений с различной степенью "интенсивности" было отнесено 7 объектов: № 28, 34, 41, 45, 46, 47 и 49 остальные 28 проб отнесены к классу рудопоявлений.

2. Распознавание проб на перекодированных значениях признаков было проведено по комплексу программ: "Качели", "Каскад", "Q-тесты" [13], "Пакеты" [12], "Тесторн" (табл.7).

Распознавание в пространстве 58 бинарных признаков по программе "Качели" привело к следующим результатам: 10 проб отнесены к месторождениям (табл.7).

По другим программам сортировка проб по классам проводилась на минимизированном пространстве, содержащем 19 признаков.

По применению тесторного метода (программа "Тесторн") к классу месторождений были отнесены 10 проб (№ 28, 30, 38, 41, 46, 47, 49, 58 и 60), к классу рудопоявлений - 22 пробы и три (№ 29, 34, 51) попали в зону неопределенности.

По тестовому методу (программа "Q-тесты") к месторождениям были отнесены 4 пробы (№ 30, 41, 51 и 58), 5 проб попали в зону неопределенности (№ 29, 38, 46, 47 и 60), остальные 26 объектов были отнесены к рудопоявлениям.

Близкие результаты распознавания были получены по программе "Пакеты" [12]: 5 проб попали в класс месторождений (№ 29, 30, 41, 46 и 58), 7 проб - в зону неопределенности (№ 28, 38, 42, 47, 49, 51 и 60) и 23 пробы были отнесены к рудопоявлениям.

При обработке информации по алгоритму суммарных мер приуроченности и согласования (программа "Каскад") были получены следующие результаты: 2 пробы распознавались как месторождения (№ 30 и 81), 6 проб попали в зону неопределенности (№ 28, 38, 42, 45, 58 и 60) и 27 проб были отнесены к классу рудопоявлений.

Обсуждение результатов распознавания

Сравнительный анализ результатов распознавания проб различными алгоритмами представляется весьма полезным, т.к. с одной стороны, позволяет более критически подойти к принятию окончательного решения о перспективности, а с другой стороны, помогает выявить некоторые возможности привлекаемых к обработке алгоритмов, а также их специфичность.

Прежде всего, следует отметить, что по всем указанным выше алгоритмам подавляющее большинство проб классифицируются как рудопоявления и лишь незначительная их часть относится к месторождениям.

При более детальном анализе можно усмотреть и некоторые различия между использованными алгоритмами по результатам распознавания. Так, по числу проб, отнесенных к классу месторождений, алгоритмы можно расположить в следующем порядке: "Тесторн" (10), "Качели" (10), "Пакеты" (5), "Q-тесты" (4), и "Каскад" (2). По числу проб, отнесенных к рудопоявлениям, алгоритмы располагаются практически в обратном порядке: "Каскад" (27), "Q-тесты" (26), "Качели" (25), "Пакеты" (23), "Тесторн" (22). Закономерность нарушается лишь переменной мест "Качельного" алгоритма и "Пакетов". Это, возможно, вызвано тем, что распознавание по "Качелям" велось на ином, по сравнению с другими алгоритмами, признаковом пространстве. По количеству проб, попавших в зону неопределенности, отмеченной закономерности в порядке алгоритмов не наблюдается "Пакеты" (7), "Каскад" (6), "Q-тесты" (5), "Тесторн" (3), "Качели" (0).

Таким образом, количественный учет проб, отнесенных к тому

или иному классу, позволяет условно разделить примененные алгоритмы на две группы. К первой можно отнести "Тесторы", "Пакеты" и "Качели", обладающие наибольшей "щедростью" при отнесении проб к месторождениям и наибольшей "скупостью" при отнесении проб к рудопроявлениям. Другую группу образуют "Каскад" и "Q-тесты", характеризующиеся противоположной тенденцией в определении классовой принадлежности проб (табл.8). "Качельный" метод, по результатам распознавания проб на пространстве из 32 признаков, не находит определенного места в отмеченном ряду, поскольку обработка велась на совершенно ином, не перекодированном материале.

Естественно, что полученные выводы ни в коей мере не могут отождествляться с оценкой качества распознавания перечисленными алгоритмами.

Для более обоснованного принятия решения о перспективности распознаваемых проб имеет смысл провести анализ результатов несколько в ином аспекте. Вычислим для каждой пробы усредненный результат распознавания по всем алгоритмам, причем эта величина условная и вводится для более удобного использования результатов. Для этого обозначим через "1" отнесение пробы к классу месторождений, через "0" - отнесение к рудопроявлениям, а через "0,5" - попадание пробы в зону неопределенности. Просуммировав эти значения по строкам, поделим полученные величины на число примененных алгоритмов распознавания. Окончательные результаты подсчета внесены в табл.7. Естественно, если эта величина больше 0,5, то это значит, что данная проба чаще относится к I классу, чем ко II классу; если она меньше 0,5, то это свидетельствует об обратном соотношении; если же усредненный результат равен 0,5, то вопрос о классовой принадлежности пробы остается открытым. Как видно из табл. № 7, пробы (№ 30, 38, 41, 46, 49, 58 и 60) получили усредненную оценку результатов распознавания больше 0,5; 7 проб (№ 29, 34, 42, 45, 50, 51 и 52) - меньше 0,5 и 2 пробы (№ 28 и 47) - около 0,5. Таким образом, в целом проявляется единая направленность в определении классовой принадлежности проб, поскольку лишь в двух случаях были получены равновероятные результаты распознавания.

Из числа проб, классифицируемых как перспективные, следует особо отметить объекты: № 41, который по всем без исключения алгоритмам относится к классу месторождений и № 30, который также

практически всеми (исключение составляет результат по "Качелям" в пространстве 32 признаков) алгоритмами отнесен к I классу. Близкие к этим результаты распознавания были получены также по пробам № 46 и 58.

Таблица № 8

Число проб (в % от общей суммы), отнесенных к I классу	Число проб (в % от общей суммы), отнесенных ко II классу	Число проб (в % от общей суммы), попавших в зону неопределенности
"Тесторы" - 28,5(19)	"Тесторы" - 63(19)	"Пакеты" - 20(19)
"Качели" - 28,5(58)	"Пакеты" - 65,7(19)	"Каскад" - 17,1(19)
"Пакеты" - 14,3(19)	"Качели" - 71,5(58)	"Q-тесты" - 14,3(19)
"Каскад" - 5,7(19)	"Каскад" - 27,2(19)	"Качели" - 0(58)
"Качели" - 20(32)	"Качели" - 80(32)	"Качели" - 0(32)

Примечание: В скобках указан объем признакового пространства, на котором производилось распознавание.

По результатам решения задачи по оценке перспективности некоторых районов Приморья на оловооруденение можно сформулировать следующие основные выводы.

1. Материал обучения в целом представлен достаточно "компактными" классами, пригодными для распознавания проб. "Малотипичными" в своих классах оказались 2 объекта из I класса и 5 объектов из II класса.

2. Процедуры минимизации признакового пространства, применяемые при решении задачи, улучшают результаты распознавания.

3. По результатам распознавания в качестве перспективных на оловооруденение могут рассматриваться следующие объекты: № 41, № 30, № 58, № 46, № 38, № 49 и № 60.

Л и т е р а т у р а

Г. Кренделев Ф.П., Дмитриев А.Н., Журавлев Ю.И. Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики. "ДАН СССР", 1967, т.173, № 5, с.1149-1152.

2. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. О математических принципах классификации предметов и явлений.-В кн.: Дискретный анализ. Вып.7, Новосибирск, "Наука", 1966, с.3-15.
3. Бишаев А.А., Карбышев В.Д. Методика решения задач классификации геологических объектов с помощью ЭВМ (на примере дифференцированных трапловых интрузий).-В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, 1974, с.192-196.
4. Константинов Р.М., Дмитриев А.Н. Использование математических методов анализа геологических факторов, влияющих на масштабы оруденения (на примере месторождений касситерит-сульфидной формации).- "Геология рудных месторождений", 1970, № 2, с.56-64.
5. Васильев Ю.Л., Дмитриев А.Н. Спектральный подход к сравнению объектов, охарактеризованных набором признаков.- "ДАН СССР", 1972, т.206, № 6, с.1309-1312.
6. Логико-математическая обработка геологической информации. Под редакцией Дмитриева А.Н. Новосибирск, 1975, 189 с.
7. Смертин Е.А. Q - тесты в задачах тестового распознавания.-В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск, 1973, с.65-66.
8. Волков П.П., Дмитриев А.Н., Нагаева Г.М., Пантелеев В.И. Дифференциальная диагностика шизофрении и органических заболеваний головного мозга логико-дискретным методом.-В кн.: Проблемы моделирования психической деятельности. Вып.2, Новосибирск, 1968, с.307-322.
9. Бабич В.В., Федосеев Г.С. Типизация прогнозно-поисковых задач и некоторые подходы к их решению.-В кн.: Состояние и направление исследований по металлогении траплов. Красноярск, 1974, с.118-119.
10. Вышемирский В.С., Дмитриев А.Н., Трофимук А.А. Поисковые признаки гигантских нефтяных месторождений.-В кн.: УШ Мировой нефтяной конгресс. М., 1971, с.3-16.
11. Модников И.С., Еремеев А.Н., Писаревский В.И., Россман Г.И., Дмитриев А.П., Кренделев Ф.П., Резник Т.Л. Оценка масштаба редкометального оруденения, локализованного в вулканических аппара-

тах (с помощью ЭВМ).- "Сов. геология", 1969, № II, с.100-110.

12. Красавчиков В.О. Модификация тестового подхода к анализу таблиц описаний на основе понятия пакета.-В кн.: Дискретный анализ. Вып.26, 1974, с.36-60.

13. Смертин Е.А., Дмитриев А.Н. Дополнение к алгоритму распознавания "голосованием" по тестам и тесторам. - В кн.: Алгоритмы и программы. Решения геологических задач на ЭЦВМ "Минск-2" и "БЭСМ-3М". Вып.3, Алма-Ата, 1970, с.117-119.

ПРОГНОЗНАЯ ОЦЕНКА МАГНИТНЫХ АНОМАЛИЙ ХОЛЗУНО-ИНСКОГО РАЙОНА
(ПО ГЕОФИЗИЧЕСКИМ ДАННЫМ)

Холзуно-Инской район, располагающийся в области сочленения каледонид Горного Алтая и герцинид Рудного Алтая, представляет собой сочетание двух структурно-формационных зон - Чарыско-Талицкой и Коргонской. В строении этих зон принимают участие разнообразные породы протерозойского, кембрийского, ордовикского, силурийского, девонского, раннекаменноугольного и неогенового возраста. Среди магматических образований, относящихся к герцинскому циклу, выделяются три среднедевонских и одна нижнекаменноугольная группы. К первой относятся субвулканические интрузивы переменного состава. Вторая группа существенно гранитовая, сопровождающаяся широко развитой дайково-жильной фацией и значительными контактово-метаморфическими изменениями. Мелкие тела основного состава составляют третью группу. И, наконец, четвертая группа представлена гранитоидами батолитового типа, с которыми ассоциируют гибридные породы гранодиорит-габбрового ряда.

Математической обработке было подвергнуто 115 объектов Холзуно-Инского района (рис.1), разбитых по профессиональным соображениям* на три класса - месторождения (запасы более 50 млн. тонн), рудопроявления (запасы менее 50 млн. тонн) и пробы (запасы неизвестны).

В данной задаче под объектами понимаются участки земной коры, горизонтальные размеры и конфигурация которых определяются проекцией магнитных аномалий, выявленных при съемке масштаба 1:25000. Вертикальный размер определяется величиной естественных эрозионных врезов. В геологическом строении объектов первого класса принимают участие главным образом вулканогенно-осадочные отложения среднего девона (туфы кислого состава с прослоями туффилов, алевролитов и известняков). Исключением является Белорецкое месторождение, располагающееся среди

* Исходная информация представлена геологами и геофизиками Алтайской геофизической экспедиции ЗСГУ Ю.В.Загайновым, Н.В.Истоминным и др.

терригенных и карбонатных пород силура. Обычно в пределах рудных участков устанавливаются субвулканические тела альбитофиоров, кварцевых альбитофиоров и граносиенитов, с которыми связаны явления мраморизации, ороговикования и скарнирования. Мощности рудных и скарново-рудных зон меняются от нескольких десятков до 100-130 метров при горизонтальной протяженности 2-6 км и изученной глубине 600-800 метров. Гематитовое и гематит-магнетитовое оруденение имеет вкрапленный, прожилковый, линзовидный и пластообразный характер. Общая мощность пластообразных рудных тел - 10-30 м с увеличением до 80-100 м и уменьшением до 3 м. На магнитометрических планах рудные тела обычно оконтуриваются изолиниями в 1000-10000 гамм, а максимальные значения аномалий достигают 60000-100000 гамм.

Геологическое строение представителей второго класса отличается большим разнообразием и значительно меньшей степенью изученности. В целом, также как и объекты первого класса, они располагаются среди среднедевонских и нижнесилурийских отложений. При этом одни из них локализируются в пределах сильно метаморфизованных, рассланцеванных и брекчированных туфов, другие - в экзоконтактной зоне массива кварцевых диоритов и гранодиоритов, прорывающих окварцеванные порфириды и, наконец, третьи - в зонах смятия, среди песчаников, известняков и глинисто-кремнистых сланцев. В последнем случае в пределах участков устанавливаются интрузивные тела кварцевых порфиров, гранит-порфиров и фельзитов, к висячим контактам которых приурочены скарновне и скарново-рудные зоны протяженностью до 1,5 км и мощностью до 50 м. В пределах рудных зон насчитывается от одного до пяти рудных тел. Они имеют в основном линзообразную, реже столбообразную форму и обычно выклиниваются на глубине. Руды преимущественно магнетитовые с подчиненным развитием гематит-магнетитовых и марганцевосодержащих магнетитовых. Объекты в большинстве своем характеризуются слабыми площадными аномалиями (до 300 гамм), в пределах которых максимальная интенсивность достигает 11000-20000 гамм.

К пробам отнесены участки и площади с различной степенью изученности. Преимущественно это коренные обнажения руд и оруденелых пород, в меньшей степени - делювий, пролювий и аллювий

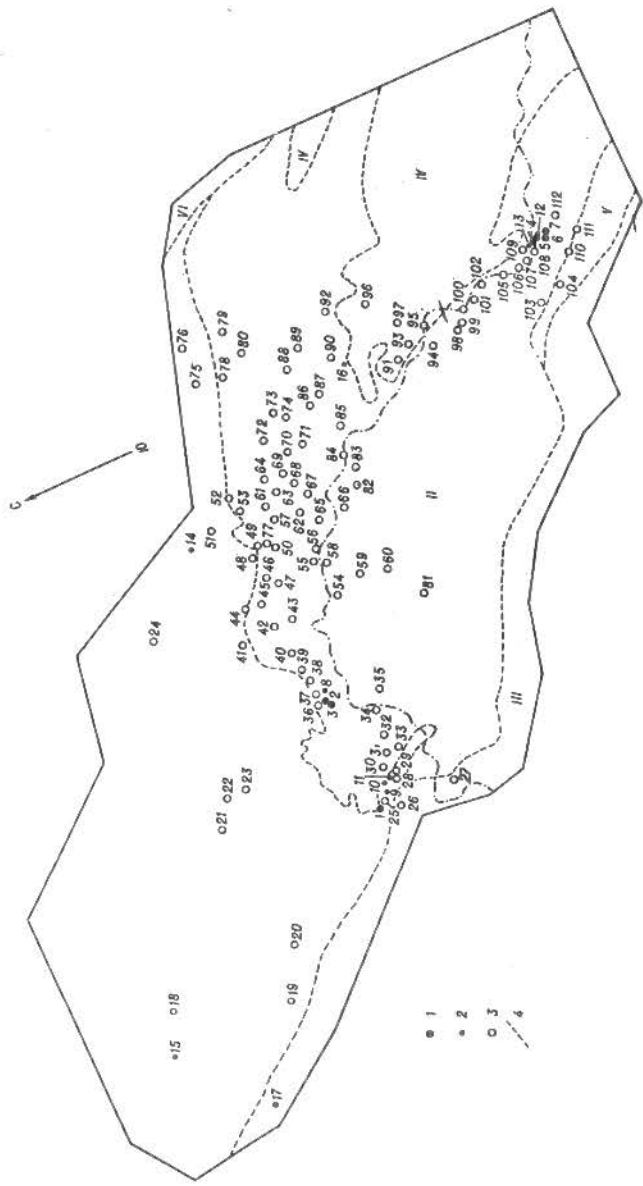


Рис. 1. Схема размещения объектов Холмско-инского района (по Б.В. Сагайнову)

[I] - М Е С Т О Р О Ж Д Е Н И Я : I - Белоречное; 2 - Инское, Центр. Уч.-к.3 - Инское, Юго-Зап. Уч.-к.; 4 - Холзуновское, перевальный Уч.-к. ан 7-9; 5 - Холзуновское, Перев. Уч.-к. ан-6; 7 - Холзуновское, Тургу. Уч.-к. ан 1-2. [2] - Р У Д О П Р О Я В Л Е Н И Я : 8 - Инское, Вост. Уч.-к.; 9 - Лефовичинское; 10 - Пономаревское I; II - Пономаревское II; 12 - Холзуновское, Перев. Уч.-к. ан-5; 13 - Холзуновское, Перев. Уч.-к. ан-10; 14 - Чесноковское; 15 - Новознаменское; 16 - Верхне-Кедровское; 17 - Кузнецовское. [3] - П Р О Я В (Участка) : 18 - Кузнецкий; 19 - Ручьевский; 20 - Локтевский; 21 - Пупчихинский; 22 - Хангаринский; 23 - Листвяный; 24 - Чалинский; 25 - Балтихинский; 26 - Правый Борт Уч. Белой; 27 - Боровский; 28 - Крохалинский юв; 29 - Крохалинский св; 30 - Верхне-Иркутский; 31 - Нижне-Иркутский I; 32 - Нижне-Иркутский II; 33 - Нижне-Иркутский III; 34 - Промежуточный I; 35 - Промежуточный II; 36 - Северный I; 37 - Северный II; 38 - Северный III; 39 - Сергеевский I; 40 - Сергеевский II; 41 - Рубежный; 42 - Таларка I; 43 - Таларка II; 44 - Сечеленский; 45 - Горький I; 46 - Горький II; 47 - Горький III; 48 - Моховой I; 49 - Моховой II; 50 - Епишинский I; 51 - Верхне-Логовушинский; 52 - Подбельный I; 53 - Подбельный II; 54 - Верхне-Шумишинский; 55 - Кристальный I; 56 - Кристальный II; 57 - Королевский; 58 - Верх-Коровихинский; 59 - Коровихинский; 60 - Поливановский; 61 - Белолозов-Кортонский; 62 - Кортонский; 63 - Викторовский; 64 - Усть-Антонов-Кортонский; 65 - ПримоКортонский; 66 - Убинский; 67 - Верх-Кортонский; 68 - Осинь-Кортонский; 69 - Кортонский; 70 - Верхне-Куртеньковский; 71 - Верхне-Полетное; 72 - Куртеньковский; 73 - Шелкинский I; 74 - Шелкинский II; 75 - Чуглихинский; 76 - Талицкий; 77 - Кипишинский II; 78 - Тимофеевский; 79 - Нижне-Кумирский; 80 - Левачий; 81 - Теремовский; 82 - Раскатинский I; 83 - Раскатинский II; 84 - Верхне-Раскатинский; 85 - Соболевый; 86 - Катменский I; 87 - Катменский II; 88 - Кумирский I; 89 - Кумирский II; 90 - Средне-Кедровский; 91 - Южная Комса; 92 - К. Деганка; 93 - Стамовой; 94 - Мено-Стамовой I; 95 - Мено-Стамовой II; 96 - Ночной I; 97 - Ночной II; 98 - Косиновский III; 99 - Косиновский IV; 100 - Косиновский V; 101 - Косиновский VI; 102 - Косиновский VII; 103 - Барсуковский I; 104 - Барсуковский II; 105 - Динейно-Кульдинский I; 106 - Динейно-Кульдинский II; 107 - Кульдинский; 108 - Кульдино-Кедровый I; 109 - Кульдино-Кедровый II; 110 - Западный I; 111 - Западный II; 112 - Тургушунский. [4] - Г Р А В И И И Ч А Г Ы С К О - Т А Л И С К И (I), Кортонской (II), Рудно-Атлаской (III), Холзуно-Чушской (IV), Кено-Атлаской (V) и Анрейско-Чушской (VI) структурно-формационных зон

гематит - и магнетит-содержащих пород. Известные рудные зоны располагаются среди скарнированных вмещающих пород в пределах экзоконтактов гранитоидных массивов и среди вулканогенно-осадочных толщ среднего девона. В первом случае руды в основном магнетитовые, немногочисленные рудные тела мощностью до 15-25 м и протяженностью до 200-300 м ассоциируют с эпидот-гранатовыми, пироксен-амфиболовыми скарнами. В пределах скарно-во-рудных зон широко развиты ороговикованные породы. Во втором случае гематит-магнетитовые и гематитовые. Рудные тела пластообразной и линзовидной формы залегают согласно с пачками эффузивно-пирокластических пород. Магнитометрические данные свидетельствуют, в основном, о сложном распределении магнитовозмущающих масс. На отдельных участках имеется множество аномалий с интенсивностью до 3000-5000 гамм.

С содержательных позиций решаемую задачу можно сформулировать следующим образом: используя информацию по разведанным месторождениям и рудопроявлениям железа, требуется расклассифицировать в меньшей степени изученные объекты Холзуно-Инского района на перспективные и неперспективные с точки зрения обнаружения в них промышленного магнетитового или магнетит-гематитового оруденения. Следовательно, в соответствии с принятой видовой классификацией [1], эта задача является задачей классифицирования в двухклассовой постановке, что и определяет схему ее решения по алгоритму "Каскад-1".

Гравиметрические и магнитометрические данные, положенные в основу математической обработки, сведены в паспорт, фрагмент которого приводится в таблице I. В первоначальном варианте число геофизических параметров достигало 48, но после логического анализа для математической обработки было принято пространство из 35 признаков, 4 из которых являются качественными, 2 - порядковыми, 19 - количественными и 10 - точечными [2]. Для каждого введенного в паспорт признака указано количество и значение градаций, а также тип. В соответствии с видом решаемой задачи критерием оптимальности при бинарном кодировании исходных данных являлась оценка информативности R_j^y . Нахождение оптимальных вариантов производилось на ЭВМ М-220 по программам, реализующим алгоритмы кодирования "перебор" (для точечных при-

Таблица I

Фрагмент паспорта АЗ-73

Информационный массив 02		Геофизические признаки		
Номер признака	Содержание	Градация	Значение Тип ¹⁾	
I	2	3	4	5
А. РЕГИОНАЛЬНАЯ ГРАВИМЕТРИЯ (I: 500000)				
43	Характер регионального поля	1	Положительное	
		2	Отрицательное	4
		3	Ступень	
Б. РЕГИОНАЛЬНАЯ МАГНИТОМЕТРИЯ (I: 500000)				
44	Характер регионального поля	1	Положительное	
		2	Отрицательное	4
		3	Ступень	
45	Связь объекта с региональной зоной ³⁾	1	С единой зоной	
		2	С фрагментарной зоной	4
46	Характер региональной зоны ³⁾	1	Упорядоченный	4
		2	Неупорядоченный	
50	Ориентировка региональной зоны	1	СЗ	4
		2	СВ	
В. СРЕДНЕМАСШТАБНАЯ МАГНИТОМЕТРИЯ (I: 200000)				
47	Характер поля	1	Положительное	
		2	Отрицательное	4
		3	Ступень	
48	Связь объекта с зоной ²⁾	1	С единой зоной	
		2	С фрагментарной зоной	4
49	Характер зоны ³⁾	1	Упорядоченный	4
		2	Неупорядоченный	
50	Ориентировка зоны	1	СЗ	
		2	СВ	4
		3	Широтная	
		4	Меридиональная	

I	2	3	4	5
Г. КРУПНОМАСШТАБНАЯ МАГНИТОМЕТРИЯ (I:25000)				
52	Контрастность аномалии на фоне окружающего поля в радиусе 10 км ⁴⁾	1	Сильная	2
		2	Средняя	
		3	Нет	
		4	Нет	
53	Ориентировка длинной оси аномалии	1	СВ	4
		2	СВ	
		3	Широтная	
		4	Меридиональная	
54	Интенсивность намагничения (ед. СГС)			3
55	Глубина до возмущающего объекта (м)			3
56	Максимальное значение ΔT_a (МЭ)			3
57	Максимальное среднее значение ΔT_a (МЭ)			3
58	Минимальное значение ΔT_a (МЭ)			3
59	Среднее минимальное значение ΔT_a (МЭ)			3
60	Максимальный горизонтальный градиент (γ° /метр)			3
61	Осредненный горизонтальный градиент (γ° /метр)			3
62	Однородность по градиенту	1	Есть	1
		2	Нет	
63	Однородность конфигурации ⁵⁾	1	Есть	1
		2	Нет	
64	Длина аномалии L (в м)			3
65	Ширина аномалии B (в м)			3
66	Вытянутость аномалии L/B			3
67	Количество уровней	1	Один	2
		2	Два	
68	Амплитуда I-го уровня (МЭ)			3
69	Радиус I-го уровня (в м)			3
Д. ХАРАКТЕРИСТИКА КРУТЯЗНЫ АНОМАЛИИ				
72	d_1^1/d_2^1 ^{6/}			3
73	d_1^1/d_3^1			3
74	d_2^1/d_3^1			3

I	2	3	4	5
75	d_1''/d_2''			3
76	d_1''/d_3''			
77	d_2''/d_3''			
Е. ПРОЧИЕ ХАРАКТЕРИСТИКИ				
78	Наличие других аномалий на расстоянии 10 км от объекта	1	Есть	1
		2	Нет	
79	Наличие в связи с объектом отрицательных магнитных аномалий	1	Есть	1
		2	Нет	

Примечания:

1). Для типов признаков приняты следующие обозначения:
1 - качественные, 2 - полуколичественные (порядковые), 3 - количественные, 4 - точечные;

2). В признаках 45 и 48 под фрагментарной зоной понимается региональная или среднемасштабная магнитная аномалия, в пределах которой выделяются аномалии более высоких порядков, в противном случае аномалия считается единой;

3). В признаках 46 и 49 упорядоченной считается фрагментарная зона, в которой аномалии более высоких порядков обладают близкой пространственной ориентацией, в противном случае - зона неупорядоченная;

4). В признаке 52 под контрастностью аномалии понимается степень непохожести по конфигурации на соседние аномалии;

5). В признаке 63 под однородностью конфигурации понимается единство аномалии, под неоднородностью конфигурации - ее фрагментарность;

6). В признаках 72-77 d_1 - полурадиус аномалии по изолинии, равной $2/3 \Delta T_a$; d_2 - полурадиус аномалии по изолинии, равной $1/2 \Delta T_a$; d_3 - полурадиус аномалии по изолинии, равной $1/3 \Delta T_a$; признаки 72-74 характеризуют пологую ветвь аномалии, признаки 75-77 - крутую.

знаков) и "скользящая" граница (для остальных признаков) [3,4]. На основании полученных результатов была составлена рабочая информационная таблица 2.

Оценка качества распознавания на полном признаковом пространстве (минимизация не производилась) осуществлялась по процедуре внешнего экзамена путем поочередного удаления из информационной таблицы каждого из 17 объектов обучения с последующим перекодированием информации и пересчетом оценок информативности. Классифицирование удаленных объектов выполнялось по результатам подсчета строчечных нагрузок (таблица 3). При экзамене правильно распознано 12 объектов и имело место 5 ошибок, т.е. качество распознавания достигает 70,6% (рис.2).

Для выяснения соотношения реальных и "идеальных" классов при полной обучающей выборке был проведен внутренний экзамен. Эталонные объекты получили при этом следующие оценки строчечных нагрузок $J[\tilde{P}_j^y]$: № 1 - 644, № 2 - 788, № 3 - 721, № 4 - 715, № 5 - 578, № 6 - 653, № 7 - 799 (I класс), № 8 - 301, № 9 - 176, № 10 - 217, № 11 - 139, № 12 - 592, № 13 - 194, № 14 - 503, № 15 - 76, № 16 - 135, № 17 - 182 (II класс); $R_I = 851$, $R_{II} = 0$ ("идеальные объекты"). Величины строчечных нагрузок позволяют сделать вывод о том, что объекты I класса образуют "компактную" группу. Основная масса объектов II класса также достаточно "компактна", но два объекта, № 14 и особенно № 12, вносят элемент неоднородности (рис.3). В этой связи эталонные классы имеют область пересечения "в", ширина которой определяется величиной $\rho = J[\tilde{P}_j^y]_{min}^I - J[\tilde{P}_j^y]_{max}^{II}$ и равна $(578 - 592) = -14$. Следует отметить, что при внешнем экзамене объект № 12 (также, как и № 14) дает ошибку классифицирования и во всех случаях определяет величину $J[\tilde{P}_j^y]_{max}$, снижая тем самым оценку качества распознавания. С формальной точки зрения эти особенности Перевальского участка АН-5 Холзунского месторождения (№ 12) можно объяснить либо нетипичностью (по сравнению с остальными рудопроявлениями) данного объекта, отразившейся в его геофизической характеристике, либо его ошибочным отнесением к классу рудопроявлений

Логический анализ информационной таблицы, проведенный с целью выделения допустимых к распознаванию на данном материале обучения проб, показал, что критерий общности (блок А)

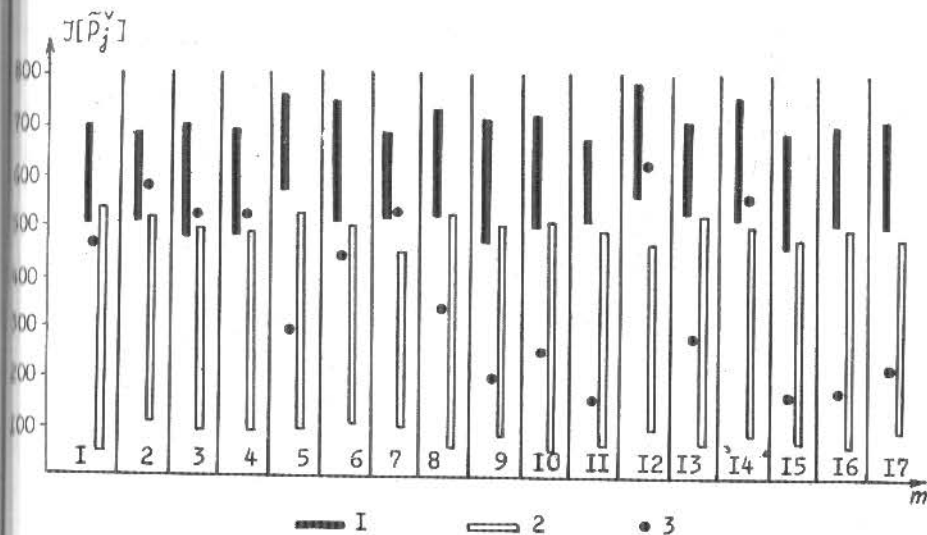


Рис. 2. Классифицирование объектов при внешнем экзамене. I - первый класс, 2 - второй класс, 3 - экзаменуемые объекты. $J[\tilde{P}_j^y]$ - строчечные нагрузки, m - номера экзаменуемых объектов.

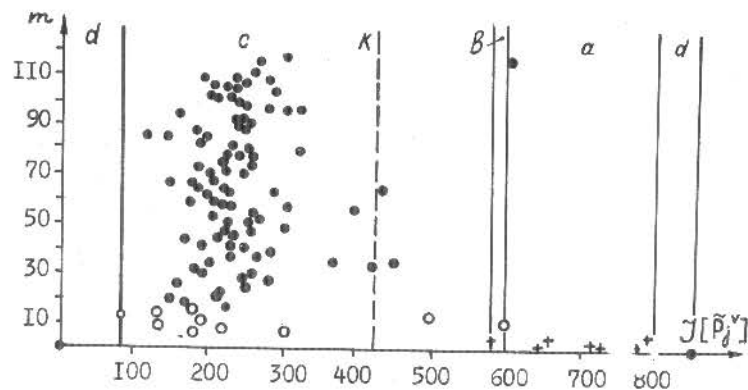


Рис. 3. Результаты внутреннего экзамена и распознавания проб. "4" - эталонные объекты I класса; "o" - эталонные объекты II класса; K - граница "идеальных" классов. a, b, c, d - диагностические зоны.

Таблица 3

Строчечные нагрузки объектов при внешнем экзамене

Номер объекта	Номера экзаменов																
	I	2	3	4	5	6	7	8	9	10	II	I2	I3	I4	I5	I6	I7
1	584	554	570	620	590	588	564	587	558	531	601	535	622	562	549	580	
2	693	614	692	754	748	666	733	684	696	687	746	706	758	658	695	705	
3	648	644	642	698	702	684	700	639	658	629	679	673	697	614	634	651	
4	506	512	-	682	634	602	644	613	632	618	695	662	675	600	616	633	
5	556	604	480	576	510	522	520	465	500	508	554	527	512	459	505	491	
6	694	658	642	764	658	518	584	571	566	551	642	544	570	527	535	528	
7							714	711	720	672	789	698	732	690	697	676	
8	292	232	218	254	264	254	-	253	255	239	293	308	236	222	239	225	
9	112	194	158	140	114	138	152	-	148	140	152	112	155	145	124	148	
10	174	226	158	162	180	176	178	181	125	194	185	178	185	139	216	195	
11	136	108	94	88	128	102	111	123	106	118	118	111	121	96	116	121	
12	536	520	494	488	448	448	524	504	506	489	-	522	503	472	489	475	
13	222	196	168	158	148	202	237	149	177	168	213	432	140	143	161	147	
14	436	426	426	416	416	446	446	450	447	435	465	60	80	435	453	470	
15	56	132	96	100	136	98	62	94	51	60	93	432	80	-	56	80	
16	88	134	92	94	132	166	99	80	139	106	106	106	114	65	-	114	
17	118	180	156	126	166	194	155	159	157	145	149	156	180	144	163	-	
[ГР]	472	580	520	524	296	440	532	366	198	247	156	620	275	558	161	166	208

Примечание: 1) Прочерки в столбцах строчечных нагрузок соответствуют экзаменуемым объектам;
2) [ГР] - величина строчечных нагрузок экзаменуемых объектов.

имеет минимальную возможную длину. Это не позволяет осуществить строгий контроль за пробами при допуске их к классифицированию: только три пробы (№ 19, 26, 27) не были допущены к распознаванию по этому критерию. Оба подблока в блоке С не являются тесторами и не обладают самостоятельной классифицирующей способностью. Из-за большого количества прочерков (от 22 до 25) из дальнейшей обработки были исключены пробы № 21, 24, 41, 81 и 92. Классифицирование оставшихся 90 проб осуществлялось с привлечением признаков из блока Д.

Результаты классифицирования отражены в таблице 4 и на рисунке 3. Отсутствие фигуративных точек в периферических зонах

Результаты классифицирования проб

Таблица 4

Проба	Проба	Проба	Проба	Проба	Проба	Проба	Проба	Проба	Проба
18	226	42	229	60	213	78	238	98	327
20	162	43	241	61	172	79	226	99	249
22	145	44	196	62	205	80	327	100	244
23	213	45	165	63	234	82	253	101	233
25	249	46	216	64	288	83	237	102	214
28	160	47	238	65	219	84	181	103	207
29	279	48	217	66	180	85	200	104	288
30	242	49	259	67	433	86	143	105	236
31	245	50	302	68	174	87	120	106	229
32	197	51	232	69	148	88	249	107	208
33	188	52	259	70	207	89	174	108	246
34	425	53	268	71	206	90	242	109	276
35	203	54	265	72	246	91	257	110	236
36	363	55	213	73	226	93	246	111	190
37	454	56	394	74	172	94	234	112	263
38	263	57	311	75	259	95	162	113	609
39	233	58	236	76	220	96	310	114	260
40	288	59	222	77	262	97	275	115	309

Примечание: в таблицу включены также результаты распознавания трех проб, информация по которым была получена позднее: 113-Белоуобинский, 114-Западно-Церевальный и 115-Юго-Восточный Тургутинский.

свидетельствуют о том, что привлеченные к распознаванию пробы соответствуют материалу обучения. Практически все пробы располагаются в зоне второго класса, но можно выделить несколько участков - Северный I (№ 36), Кристальный II (№ 56), Северный II (№ 37), Верх-Коргончиковый (№ 67), Промежуточный I (№ 34), Девичий (№ 80) и Коксинский III (№ 98), обладающих некоторой особенностью и характеризующихся большей относительной близостью, по сравнению с остальными пробами, к объектам первого класса. Лишь единственный объект - Белоубинский участок (№ II3), расположен в зоне первого класса. Он и все выше перечисленные объекты были рекомендованы для более детального изучения.

Л и т е р а т у р а

1. БАБИЧ В.В., ФЕДОСЕЕВ Г.С. Метод целевого классифицирования и упорядочения объектов ("Каскад-I"). - В кн.: Логико-математическая обработка геологической информации (сборник научных трудов). Новосибирск, 1976, с.42-70.
2. ФЕДОСЕЕВ Г.С. Предварительный анализ данных в логико-математических исследованиях. - В кн.: Логико-математическая обработка геологической информации (сборник научных трудов). Новосибирск, 1976, с.22-42.
3. БАБИЧ В.В. Программа П4 "Оптимальное бинарное кодирование признаков". - В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.143-150.
4. БАБИЧ В.В. Программа П3 "Подсчет строчечных нагрузок по заданной информативности для бинарных таблиц". - В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.135-142.

Н.А. Кулик, В.О. Красавчиков, Т.И. Шатнова

ОПЫТ ИСПОЛЬЗОВАНИЯ ЛОГИКО-ДИСКРЕТНОГО ПОДХОДА ПРИ РЕНТГЕНОВСКОЙ ДИАГНОСТИКЕ БРАННЕРИТА

Рентгеновская характеристика является неотъемлемой и наиболее убедительной частью диагностики минералов вообще и в особенности минералов сложного состава; когда же малые количества материала просто не позволяют произвести необходимых химических определений, рентгеновская характеристика нередко становится единственным надежным критерием при идентификации минерала среди близких по составу и физическим свойствам минеральных видов. Сказанное в полной мере относится к сложным окислам и титанатам, которые, обладая широкими возможностями изоморфных замещений вследствие сложности состава, вдобавок ко всему часто метамиктны.

Настоящая работа посвящена рассмотрению рентгенографического материала по браннериту - одному из титанатов урана, для которого весьма остро встает вопрос - с чем сравнивать испытуемый материал при рентгеновской его характеристике? В литературе опубликовано в общей сложности около 25 рентгеновских анализов образцов, относимых к браннериту. Сравнение химизма некоторых из них и сопоставление их рентгеновских данных, проведенное

М.М. Повилайтис [20], демонстрирует весьма широкий спектр химических составов минерала, начиная от собственно урановых разновидностей до существенно ториевых – торутитов, которые М.М. Повилайтис предлагает рассматривать как конечные члены торий-уранового изоморфного ряда. Имеются данные [23] и о богатых кальцием разновидностях браннерита. Такое многообразие порождает вопрос – который же из образцов в рентгеновском отношении наиболее представительен для всего минерального вида браннерит. Как показывает даже беглое сравнение опубликованных рентгенограмм (табл. I, рис. I), колебания в интенсивностях соответствующих линий нередко не позволяют сравнивать их, а колебания межплоскостных расстояний для каждой линии таковы, что нередко соседние линии "перекрывают" друг друга либо же расстояния между ними меньше, чем пределы колебаний d/n линий. Очевидно, что при таком положении утверждение о "хорошей" или "плохой" сходимости рентгенограммы испытуемого образца с рентгенограммами уже имеющимися становится проблематичным: что называть "хорошей сходимостью" (как, например, отмечается для лодочникита и торутита), если рентгенограммы, используемые для сравнения, сами обнаруживают такие расхождения? Ссылки на сложность и непостоянство состава минерала хотя и объясняют эти расхождения, однако не помогают исследователю при необходимости диагностики, и, волею неволей, приходится удовлетворяться весьма приблизительным соответствием рентгенограмм, не имея "эталонного" в рентгеновском отношении образца.

При этом наличие рентгенограмм искусственного браннерита мало меняет положение, поскольку, как выяснилось, искусственное соединение отнюдь не идентично по рентгеновской характеристике природным образцам. В то же время несомненно, что для браннерита как минерального вида при всем многообразии природных объектов должны существовать характерные черты общности и в рентгеновском отношении. Это побудило авторов провести анализ накопленного материала по рентгенограммам браннерита по двум направлениям:

1. Выявление на дебаграммах таких бесспорно браннеритовых линий, которые являются характерными для этого минерального вида и совокупность которых надежно отличает его от других сходных по составу минералов;

2. Выявление по имеющимся рентгенограммам наиболее выразительных представителей для всего минерального вида – т.е. установление эталонных представителей для браннерита.

В настоящей работе сделана попытка применить для решения поставленных задач логико-дискретный подход [6], включающий логическое кодирование информации, составление информационных таблиц и выявление общих закономерностей на их основе. На этапе кодирования был применен аппарат функций алгебры логики [24].

Характеристика материала

В работе использованы опубликованные данные по рентгенограммам пригодного метамиктного браннерита, по искусственным соединениям и неметамиктным образцам и результаты анализа образцов одного из авторов. Из этого материала в табл. I выделены данные по браннеритовым рентгенограммам, наиболее часто используемым в литературе для сравнения, т.е. по существу выполняющим роль эталонных. Для них характерно следующее.

1. Рентгенограммы представляют отожженные образцы, причем условия отжига не идентичны. Температура отжига колеблется от 600–750°C (рентгенограммы 5,6) до 1000°C. Как правило, отжиг метамиктного материала проводился на воздухе, образец 22 отожжен в вакууме.

2. Рентгеновская съемка проведена в неодинаковых условиях – на различном излучении, с различной экспозицией, в камерах разного диаметра.

3. Иногда опубликованные данные не исчерпывают всей рентгеновской характеристики порошкограммы, некоторые исследователи опускают слабые линии, особенно в области больших углов отражения. Так, Пабст [29] после линии с межплоскостным расстоянием 1,230 указывает: "Плюс еще около 20 линий, по большей части широких и диффузных", значения которых не приводятся. Именно этим, по-видимому, объясняется обрыв значений d/n и в рентгенограммах № 9, 10, 23.

4. Исходные данные в различных работах приводятся с различной точностью значений d/n . Так, если для образца 9, судя

Описание рентгенограмм отожженного браннерита

№ линии	2		3		4		5		6		7		9		10		12	
	д	d/n	д	d/n	д	d/n	д	d/n	д	d/n	д	d/n	д	d/n	д	d/n	д	d/n
1							5	6,52	2	6,54								
2	3	6,07	4	6,05	5	5,88					2	5,89	26	6,00				
3							3	5,24	3	5,27								
4							1	5,14			1	4,88						
5							1	4,76					4	4,810				
6							4	4,73	3	4,73			82	4,720				
7	5	4,70	6	4,73	5	4,70					4	4,656						
8							1	4,53	1	4,56			3	4,580				
9							1	4,46	1	4,40								
10							1	4,35	1	4,34			22	4,310				
11	2	4,31	3	4,32														
12					2	4,29												
13							2	4,19										
14							5	4,14	3	4,15	1	4,064					1	4,04
15							3	3,85	1	3,85							0,5	3,88
15 ^a							2	3,76										
15 ^o											0,5	3,61						
16	10	3,42	10	3,41	10	3,41	10	3,44	10	3,43	7	3,48 ^x	100	3,418	ср.	3,43	10	3,40
17	6	3,32	6	3,32	6	3,29	4	3,33	3	3,34	5	3,322	60	3,308	ср.	3,32	8	3,31
18							3	3,27	3	3,27								
19							3	3,24										
20							4	3,18										
21													13	3,243				
22	5	3,02	3+	3,02					5	3,04			7	3,190	сл.	3,192	6	3,115
23					3	2,98							7	3,040				
24													28	3,001			5	3,000
24 ^a	5	2,91	5	2,92	5	2,90			7	2,911	5	2,884	45	2,905	сл.	2,916	4	2,89
25													3	2,870				
26	4	2,76	4	2,75	4	2,74			6	2,761	4	2,788	27	2,755			4	2,75
26 ^a							5	2,734					5	2,720				
27							8	2,617	9	2,635	0,5	2,607	2	2,620			4	2,63
27 ^a													5+	2,550				
28	4	2,511	4	2,508	4	2,491	3дв	2,522	5	2,523	4	2,499	22	2,504				
29	7	2,455	6	2,462	7	2,449			5	2,456	5	2,448	29	2,455	ср.	2,461	7	2,45
30	2	2,426	1	2,432	2-	2,429	2дв	2,430	1	2,433			14	2,429				
30 ^a							1	2,297					3+	2,298				
31													24	2,275	сл.	2,285	7	2,26
32	7	2,276	6	2,275	6	2,270	8	2,278	8	2,278	6	2,258	1	2,250				
32 ^a																		
33							1	2,185	2	2,187								
34			2	2,161	2	2,158			3	2,166			4	2,160				
34 ^a	2	2,151					2в	2,151			3	2,148	5	2,151				

по оценке интенсивностей в 100-бальной шкале, использовано фотометрическое измерение расстояний между линиями и значимым приводится третий десятичный знак, начиная уже с линии с $d/n = 4,720$, то в анализах 24-27 предпочтается точность лишь второго десятичного знака вплоть до $d/n = 1,000 \text{ \AA}$. По-видимому, это связано и с качеством съемки, и с качеством анализируемого материала, и с профессиональными навыками каждого исследователя. Необходимо принимать во внимание и время опубликования работ, и соответствующий ему уровень развития рентгеновского анализа.

Материал таблицы наиболее наглядно охватывается в виде графика (см. вклейку), на котором линии с межплоскостными расстояниями до 2,30 нанесены в масштабе $1 \text{ мм} = 0,01 \text{ \AA}$, а с меньшими значениями в масштабе $1 \text{ мм} = 0,001 \text{ \AA}$. Такой порог масштабности выбран для того, чтобы иметь возможность сравнения данных по эталонной подборке природных объектов с данными по искусственным браннеритам.

Поскольку анализируемые рентгенограммы представляют отожженный материал, а при отжиге браннерита не исключено появление примесных не-браннеритовых фаз, необходимым этапом перед дальнейшим рассмотрением приведенного в табл. I материала явилась оценка степени его "рентгеновской чистоты". Известно, что для браннерита при отжиге возможно появление, по крайней мере, пяти "примесных" фаз. Это три структурные модификации TiO_2 , UO_2 , U_3O_8 - наиболее вероятные при принятых температурах отжига. Поэтому на графике показаны также наиболее сильные (с интенсивностью не ниже 5) линии брукита, рутила, анатаза, уранинита и ураноксида, возможность присутствия которых в каждом образце рассматривается в специальном разделе "Исключение примесей". Особо следует остановиться на так называемой "фазе-X", образование которой в интервале температур $800-1000^\circ\text{C}$ связывается с окислением U^{+4} до U^{+6} [14]. Сравнение рентгенограммы браннерита, отожженного в вакууме (что исключает возможность образования фазы-X), и рентгенограмм образцов, отожженных при разных температурах - в том числе таких, когда образование фазы-X должно быть максимальным, и в условиях, когда образование ее исклю-

Продолжение I табл. I

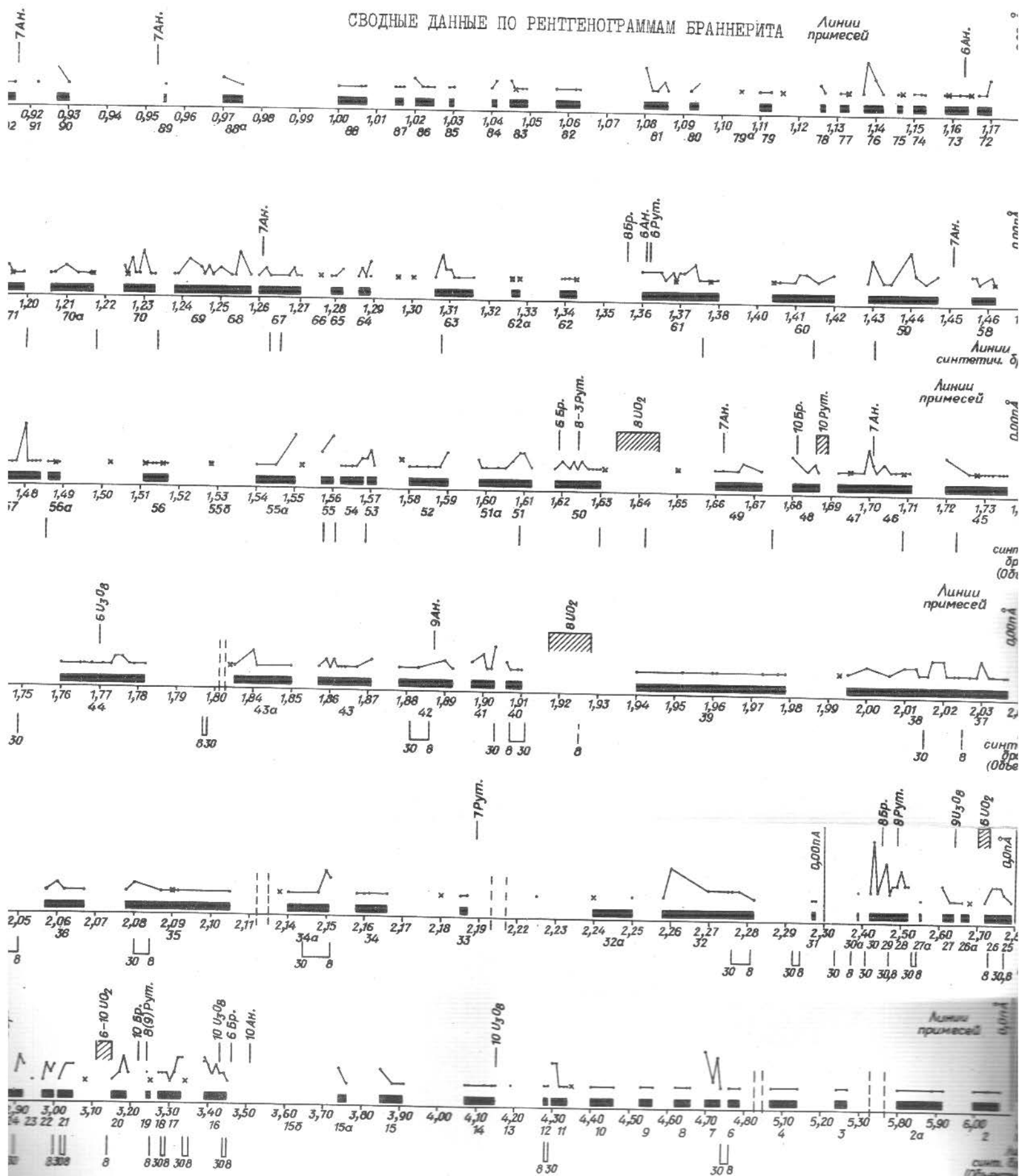
№ п/п	21		22		19		23		24		25		26		27		28		29		
	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	
1																					
2			28	6,04			I 5,92				I 5,8	2 5,8	I 5,8								
3							I 5,07 ^{A)}														
4																					
5																					
6																					
7	4	4,73	59	4,74				5 4,65	5 4,70	4 4,7	4 4,7	4 4,7	4 4,7			I 4,719				I 4,78	
8																					
9																					
10																					
11	2	4,3I	I9	4,305					I 4,30	I 4,30	I 4,30	I 4,30	I 4,30							I-2	4,35
12																					
13																					
14																					
15	2	3,9I																			
15 ^{A)}	2	3,74					5 3,74 ^{A)}														
15 ^{B)}																2	3,744				
16	7	3,42	100	3,44	7	3,45	10 3,38	10 3,40	10 3,39	10 3,39	10 3,40	10 3,40	10 3,40								
17	5	3,33	59	3,33	6	3,33														I-2	3,34
18																					
19																					
20																					
21	3	3,0I	28	3,02	10	3,17	2 3,18 ^{A)}				4 3,17	10 3,18								?	3,17
22																					
23																					
24	5	2,92	54	2,9I	3	2,90	3 2,88	4 2,89	4 2,88	4 2,89	4 2,89	5 2,89									
24 ^{A)}																					
25	5	2,77	4I	2,77	3	2,74		3 2,74													
26																					
26 ^{A)}																					
27	I	2,6I				I 2,6I															
27 ^{A)}																					
28	5	2,50	45	2,52	8	2,5I		3 2,49	3 2,48	3 2,48	3 2,50	3 2,50			2	2,552					
29	7	2,45	45	2,465	8	2,46	4 2,475 ^{B)}	4 2,44	4 2,43	4 2,43	4 2,44	4 2,44			2	2,390	0,5-I	2,42			
30																					
30 ^{A)}																					
31																					
32	7	2,27	38	2,28	6	2,275	6 2,26	4 2,26	5 2,26	5 2,25	5 2,26	5 2,26									
32 ^{A)}																					
33																					
34																					
34 ^{A)}	2	2,15						I 2,15	I 2,15	I 2,14	I 2,15	I 2,15	I 2,15								

Продолжение 2 табл. I

№ п/п	2		3		4		5		6		7		9		10		12				
	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н	г	д/н			
35	I	2,080	I+	2,078			III	2,087	I	2,104	III	2,09I	7	2,075							
36							I	2,067	3	2,062	4	2,058									
37	4	2,029	4	2,032	4	2,02I	3дв	2,036	4	2,037	4ш	2,023	16	2,027	о.сл	2,024			3	2,02	
38	2	2,017	2	2,014	I-	2,00	2дв	2,006	2	2,013	I	1,995	I2	2,010					3	1,94	
39							Iдв	1,973	2	1,977											
40									8	1,906											
41	8	1,903	8	1,903	7	1,903	6	1,903			9	1,902	22	1,906	сл	1,909			2	1,90	
42																					
43	6	1,86I	6	1,864	5	1,857	6	1,862	5	1,86I	7	1,859	15	1,863	о.сл	1,87I			I	1,87	
43 ^{B)}																					
44	I	1,776	I-	1,778	I-	1,779	I	1,775	10	1,768	2	1,782	3-	1,774	о.сл	1,776			5	1,76	
44 ^{A)}																					
45	3	1,729	3	1,735	3	1,73I							3	1,726	5+	1,734					
46									3	1,711											
47	4	1,700	3	1,702	2	1,704	3	1,70I	5	1,70I	4	1,704	4	1,699	I2	1,692					
48									2	1,684	3	1,687	4	1,686							
49									III	1,672				2	1,662						
50	6	1,623	6	1,625	7	1,618	7	1,624	9	1,625	10ш	1,623	17	1,620	сл	1,666			5	1,66	
51	4	1,609	2	1,609			3	1,605	4	1,609										5	1,63
51 ^{B)}																					
52	2	1,585	I-	1,588					I	1,582											
53	5	1,569	5	1,570																	
54									6	1,565	8	1,567	7 дв	1,563							
55	3	1,557	2	1,557					2	1,560				10ш	1,557						
55 ^{A)}																					
55 ^{B)}									2	1,545											
56													7дв	1,517	2	1,513					
56 ^{A)}																					
57	I	1,489																			
57 ^{A)}	2	1,48I	2дв	1,484	3дв	1,477	7	1,480	3	1,480	4	1,482	5+	1,486							
58	2-	1,46I	2	1,458	2-	1,457	2ш	1,456	3	1,457				4	1,456						
59	2+	1,440	3	1,444	2	1,44I	I	1,44I	3ш	1,447	5	1,433	7	1,440							
60	I-	1,41I	I+	1,413	2-	1,410	I	1,420			4	1,406	I	1,41I							
61	4дв	1,37I	5дв	1,370	4дв	1,369	3ш	1,375	5	1,374	7	1,368	4	1,37I	сл	1,374			4	1,38	
62																					
62 ^{A)}																					
63	3-	1,308	2	1,311	3	1,308	III	1,316	3р	1,309	3	1,306	3	1,308							
64	Iдв	1,289	2дв	1,287	2	1,286	I	1,287	2	1,289				I	1,289						
65																					
66																					
67	2-	1,268	I	1,267	I+	1,263															
68	3	1,255	3	1,253	3	1,247	3	1,254													
69	I-	1,242																			

СВОДНЫЕ ДАННЫЕ ПО РЕНТГЕНОГРАММАМ БРАННЕРИТА

Линии примесей



чено λ , показывает очень слабое различие между сравниваемыми объектами. Это позволяет авторам настоящей работы считать, что присутствие фазы X не вызывает необходимости исключения ее линий из рентгенограмм отожженных браннеритов, в отличие от остальных примесей.

Исключение примесей

Р у т и л. Наиболее сильные линии рутила, которые могли бы сказаться на рентгенограмме браннерита в случае его присутствия — это 3; 24 ($\dot{\lambda} = 8-9$; 2,49 ($\dot{\lambda} = 8-4$); 2,189 ($\dot{\lambda} = 7-3$); I,686-I,689 ($\dot{\lambda} = 10$); I,624-I,628 ($\dot{\lambda} = 8-3$); I,362 ($\dot{\lambda} = 6$) (Михеев, 1958). Наиболее важной в диагностическом отношении для рутила является линия 3,24, отсутствие которой практически означает отсутствие примеси рутила в рентгеновски ощутимом количестве в исследуемом материале. В образцах табл. I линии с таким значением d/λ (линии I9) имеются лишь у образцов 5,9 и 28, что избавляет от необходимости проверки на присутствие рутила в других образцах. Отметим, что интенсивность этой главной рутиловой линии в объектах 5,9 и 28 очень низкая, так что можно с уверенностью считать, что даже при наличии примеси рутила в этих образцах ее влияние не может быть существенным и может сказаться лишь на появлении (или наложении) наиболее сильных его линий. Действительно, линия 48, соответствующая наиболее сильной и постоянной по интенсивности линии рутила I,686-I,689, присутствует только у одного из отмеченных образцов — № 5. Отсутствие ее у обр. 9 и 28 исключает возможность проявления в них более слабых линий рутила, и, следовательно, линия 28 у этих образцов не является рутиловой, несмотря на близость значения d/λ с линией 2,49 рутила. Тем более, что линия 2,49 у рутила непостоянна по интенсивности, которая может снижаться до 4, а аналогичная ей по интенсивности линия 2,189 у обоих образцов отсутствует. На таком же основании линии 50 и 6I для образ-

№/ Образцы 5,6 — температура отжига 600-750°C; обр. № 24 и 25 и их отожженные при 1300°C и в атмосфере азота аналоги (обр. 24^a и 25^a, табл. 5,6).

цов 9 и 28 нельзя принять как рутиловые, несмотря на близость их по d/n к линиям рутила I,624 и I,362, тем более, что эти линии хорошо выражены и у образцов, для которых примесь рутила исключена отсутствием его более сильных линий.

Таким образом, как проявление примеси рутила можно считать линии 19, 33, 48 у обр. 5. Сдвигивание линии 28 у этого образца позволяет предполагать наложение линии рутила на собственно браннеритовую линию с таким же значением d/n , а нормальная, по сравнению с образцами, не содержащими примеси рутила, интенсивность линии 50 у обр. № 5 свидетельствует о том, что если здесь и имеет место наложение рутиловой линии на линию браннерита, то влияние этого наложения очень незначительно.

А н а т а з. Наиболее важной линией анатаза является линия 3,50 ($\dot{J} = 10$). То, что у всех анализируемых рентгенограмм в этой области углов отражения линии отсутствует (фиг. I), надежно свидетельствует об отсутствии примеси анатаза в сравниваемых образцах и позволяет не принимать во внимание возможное совпадение значений d/n линий I,887, I,662 и I,261, тем более, что такие сильные линии анатаза, как I,452 ($\dot{J} = 7$) и I,335 также не имеют себе аналогов в приведенных рентгенограммах браннеритов.

Б р у к и т. Несовпадение линий брукита 3,22 ($\dot{J} = 10$) и I,356 ($\dot{J} = 8$) (Михеев, 1958) даже при "перекрытии" линии брукита 2,45 значениями d/n линии 29 и отсутствие повышения интенсивности линий 48 и 50 (совпадающих с линиями брукита I,681 ($\dot{J} = 10$) и I,619 ($\dot{J} = 6$), дает основание считать, что брукит как рентгеновски значимая примесь в анализируемом материале отсутствует. Это подтверждается несовпадением линий 3,46 и 2,87 брукита с линиями рентгенограмм табл. I, несмотря на большой разброс значений d/n для линий браннерита в этой части рентгенограмм.

У р а н и н и т. Даже при наличии колебаний в значениях d/n уранинита [18], попадание линий последнего I,917-I,928 ($\dot{J} = 8$) и I,633-I,645 ($\dot{J} = 8$) в "пустое" от линий испытуемых образцов пространство (фиг. I), надежно свидетельствует об отсутствии влияния UO_2 в рентгенограммах табл. I. Не перекрываются и значения линий 3,11-3,15 ($\dot{J} = 6-10$) и 2,70-2,73 ($\dot{J} = 5-6$); все это позволяет с уверенностью исключить возможность наложения линий уранинита из рассмотрения.

У р а н о к с и д U_3O_8 . У этого соединения 4 сильные линии, которые в случае его присутствия могли бы проявиться в узоре рассматриваемых рентгенограмм. Так, с линией U_3O_8 4,15 ($\dot{J} = 10$) совпадает линия I4 у объектов 5 и 6; к линии 3,43 ($\dot{J} = 10$) близки значения линии I6 у тех же объектов; с линией 2,64 ($\dot{J} = 9$) - значения d/n линии 27 у объектов № 5, 6, 12; к линии I,77 ($\dot{J} = 6$) близка линия 44 у объектов № 6, 19, 23. Рассмотрим имеющиеся совпадения. Повышенная интенсивность линии I4 у объектов 5 и 6 по сравнению с другими объектами и ее большее межплоскостное расстояние служит первым сигналом о возможном присутствии U_3O_8 в этих образцах. Подтверждением служит несопоставимо высокая интенсивность линии 27 у этих образцов с интенсивностью той же линии в остальных случаях табл. I. Наконец, $\dot{J} = 10$ линии 44 у образцов № 5 и 6 по сравнению с интенсивностью этой широко представленной у многих образцов линии позволяет уже уверенно говорить о наложении линий ураноксида и собственно браннерита. В то же время высокая интенсивность линии I6 почти у всех образцов не позволяет в этом случае усмотреть аналогичное наложение, хотя и не исключает его. На основании рассмотренного можно считать, что проявление линий I4 и 27 в рентгенограммах обр. № 5 и 6 если не целиком, то в большей мере обязано присутствию примеси ураноксида, в то время как в остальных случаях возможно наложение его линий на линии собственно браннерита. Возможно, влияние примеси ураноксида сказывается также и в рентгенограмме обр. № 12, у которого линия 27 имеет интенсивность 4, а линия 44 - интенсивность 5; но худшее совпадение значений d/n линии I4 и ее низкая интенсивность свидетельствует о том, что это влияние меньше, чем в случае обр. № 5 и 6. Отсутствие линии I4 у образцов 19 и 23 не позволяет говорить о сколько-нибудь существенном проявлении примеси ураноксида в этих образцах даже при близости значений d/n линии 44 этих образцов к линии ураноксида.

Таким образом, для рассмотренных примесей вероятным является присутствие рутила и ураноксида в обр. № 5, ураноксида - в обр. № 6 и 12, рутила - в обр. № 9 и 28, причем в последнем случае - в очень незначительном количестве. В целом же материал, представленный в табл. I, характеризуется высокой рентге-

новской чистотой. Каким образом присутствие примесей учитывалось в дальнейшей обработке материала, будет показано ниже.

Формирование и характеристика признакового пространства Анализ закодированной информации

С точки зрения математической обработки, табл. I представляет собой совокупность описаний объектов S_1, \dots, S_m признаками x_{j1}, \dots, x_{jn} , причем каждая рентгенограмма, представляющая образец, является объектом, а ее линии — признаками. Каждая линия имеет две характеристики: количественную — величину d/n и качественную — интенсивность проявления. Первая характеристика, получаемая инструментальным измерением, более объективна в том смысле, что точность ее в большей мере зависит не от квалификации исполнителя-расшифровщика, а от положения линии: дисперсия значений d/n в области малых углов отражения (линии с малыми порядковыми номерами) увеличивается, но в области средних и больших углов отражения (средняя и нижняя часть (табл. I) они могут быть измерены весьма точно. Интенсивность же линий, как правило, определяется на глаз, и оценка ее сильно зависит как от квалификации специалиста, так и от качества рентгенограммы. Это не позволяет проводить раздельное рассмотрение признаков по каждому параметру, а заставляет лишь фиксировать наличие или отсутствие данной линии (признака) у каждого объекта.

Закодируем признаки табл. I следующим образом: наличие линии (признака) у объекта пометим I (единицей). Полученная таким образом табл. 2 содержит 109 признаков и 19 объектов. Рассмотрение материала табл. I и табл. 2 позволяет сделать следующие замечания.

I. Появление линии I2 со значением d/n , очень близким к таковому линии II, лишь в двух случаях (у объектов 6 и 4) позволяет предполагать линию I2 результатом расщепления K_α ; при столь малых различиях в значениях d/n между II и I2 линиями появление лишь одной из них не позволяет сопоставить ее с той или другой порознь, поэтому предлагается рассматривать II и I2

линии в единой совокупности как один сложный признак — пакет линий x_{II-I2} . Точно также непрерывность имевшихся в табл. I значений d/n для линий I7 и I8 не позволяет различать принадлежность к той или другой линии проявление лишь одной из них с пограничным для обеих значением d/n . Поэтому и в данном случае линии I7 и I8 предлагается рассматривать как единый признак — пакет из двух линий $x_{I7-I8} = x_{I7} \vee x_{I8}$. Присутствие одновременно линий 2I и 22 у объекта № 9 позволяет и в наборе всех значений d/n , соответствующих этим линиям, иметь в виду возможность различения их друг от друга при проявлении лишь одной из них, как в предыдущих случаях, заставляет рассматривать их как единый сложный признак-пакет. Та же самая мотивировка используется и при объединении в пакет линий 25, 26, 28; 29 и 30; 37 и 38; 46 и 47; 5I и 5Ia; 53, 54 и 55; 56a и 57; 67, 68 и 69. При этом в пакет объединялись линии не только с очень близкими значениями d/n , не позволяющими различить их друг от друга в случае неполного проявления всех линий, но и при условии, что наилучшее проявление их позволяет уверенно различать количество линий в пакете.

2. На основании изложенного в разделе "Исключение примесей" из дальнейшего рассмотрения исключены линии I9 и 33 (как проявление примеси рутила) и I4 в объектах № 5 и 6 как обусловленную в сильнейшей мере проявлением примеси ураноксида. В то же время, в случае наложения линий примеси на собственные линии браннерита, они оставлены для дальнейшей работы. Это относится к линиям 27, 44 (объекты № 5 и 6) и линии 29 в объектах 5, 28 и 9, к линиям 48 и 50 в обр. 5.

3. Наличие у объекта 9 линий 34 и 34a заставляет различать обе эти линии и у других объектов.

4. Очень велик разброс значений d/n у линии 39. Для нее, однако, невозможно наложение примесей, в том числе и фазы-X (по наличию этой линии у образцов с низкой температурой отжига, ког-

Ж/ Таким образом, в терминах функций алгебры логики этот сложный признак представляет собой дизъюнкцию [24] признаков x_{II} и x_{I2} : $x_{II-I2} = x_{II} \vee x_{I2}$. Отметим, что в работе [12] рассмотрены только те пакеты, которые выполняются для всех объектов S_1, \dots, S_m таблиц описаний T.

таблица 4

Таблица эталонов

№ отжи- га	№ об- ектов	17	11-12	16	17-18	120	121-22	124	125-26	127	128-30	132	134a	135	136	137-38	139	141	143	143a	144	145	146-47	148	149	150	151-51a	152	153-55	155a	156a-57	158	159	160	161	163	164	167-69	170											
600°	5	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I								
750°	6	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I						
800°	12	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I				
900°	7	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I			
	21	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I			
	23	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	
	9	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	
	3	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	
	4	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	
	26	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	
	25	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	
	27	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	28	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	19	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
	24	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
1000°		I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I

разных объектов) не позволяет относить их к отождествляющим признакам. Такие линии также будут весьма характеристичными для минерального вида в целом. Это в первую очередь признаки-пакеты линий I7-I8, 2I-22, 53-55, признак 24.

Проверка результатов анализа

Для того, чтобы оценить пригодность выделенной совокупности отождествляющих и характеристичных признаков (линий) в целях практической диагностики, необходимо установить, насколько присущи они другим порошкограммам браннерита и как надежно отличаются они браннерит от других сходных по составу и свойствам минералов, вызывающих наибольшие затруднения при диагностике. Для этой цели использован фактический материал табл. 5 и 5а, в которых собраны описания рентгенограмм браннерита, найденные в литературе уже в процессе обработки табл. I и потому не вошедшие в нее, а также рентгенограммы образцов, отождествленных в условиях, сильно отличных от наиболее распространенных, либо снятые с неотожженных образцов. Наиболее полно проявленные из них могут служить своеобразными "экзаменаторами" на правильность выделения характеристичных и отождествляющих линий. В качестве проб для примера практического распознавания в табл. 5 включены образцы одного из авторов.

Для отобранных объектов выделены и закодированы линии, ранее определенные как достаточно проявленные (признаковое пространство табл. 4), в результате чего получилась табл. 6, в которой образцы также упорядочены по температурам отжига, и в которую включены для сравнения объекты 22 и 28-29, ранее исключенные из рассмотрения при формировании табл. 3. Наличие у объекта 22е линий с $d/n = I,449$ и $I,433$, отмеченных в табл. 5 как линии 59 и 59а, показывает, что значения линии 59 в табл. I соответствуют двум линиям, столь близко расположенным друг к другу, что значения их межплоскостных расстояний перекрывают друг друга - то есть в этом случае мы имеем фактически дело с пакетом линий. Подтверждением этому является наличие двух максимумов в интервале линий 59 (см. вклейку). Поэтому присутствие линий 59

Продолжение 2 табл. 5.

№ п/п	13		14		15		16		17		18		20		30		31		32		32а		32б	
	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н
49																								
50	30	1,628	2	1,62																				
51	15	1,595	5	1,59																				
52	20	1,567																						
54																								
55	20	1,540	2	1,54																				
56																								
57	10	1,480	0,5	1,48																				
58	58	1,456																						
59	5	1,442	2м	1,44																				
60	5	1,410	6	1,41																				
61	51	1,369	1	1,36																				
62																								
62а																								
63	5	1,313																						
64	5	1,287																						
65																								
67																								
68	5	1,248	3	1,25																				
70	10	1,230	3	1,22																				
70а																								
71																								
74																								
75																								
76																								
78																								
79																								
80																								
81																								
82																								
85																								

Продолжение 3 табл. 5

№ п/п	22а		22б		22в		22г		22д		22е		22ж		22а		22а		
	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	д/н	
2																			
5																			
6																			
7																			
8	Сил.	4,6	Ср.	4,6															
11																			
12																			
13																			
14																			
15																			
16																			
17	Ср.	3,3	Ср.	3,3															
18																			
19																			
20																			
21																			
22	Ср.	2,99	Сил.	3,2 ^{XX}															
23																			
24	Сил.	2,88	Сил.	2,95															
24а																			
25																			
26	Сил.	2,72	Сл.	2,72															
27																			
27а																			
28																			
29	Сил.	2,47	Сил.	2,48															
30																			
30а																			
31																			
32																			
32а	Ср.	2,25	Сл.	2,28															
33																			
34а																			
35																			
36																			
37																			
38	Сл.	2,00	Сл.	2,00															
39																			
39а																			
41																			
42	Ср.	1,89	Ср.	1,88															
43	Ср.	1,85																	
43а																			
43б																			
44	О.сл.	1,78	Сл.	1,77 ^X															
44а																			
45	Сл.	1,73																	
46																			
47																			
48																			
49																			
50																			
51	Сил.	1,61	Сл.	1,61															
52																			
53																			
54	Сил.	1,56	Сл.	1,56															

№ линии	22a		22б		22в		22г		22д		22е		22ж		24а		25а	
	g	d/n	g	d/n	g	d/n	g	d/n	g	d/n	g	d/n	g	d/n	g	d/n	g	d/n
55а							23ш	1,546	18	1,550	54	1,555	13	1,541	2	1,55	2	1,54
56а																		
57					10	1,475	33ш	1,478	25	1,469	17	1,470	17ш	1,482	1	1,47	2	1,48
58																		
59																		
59а																		
60							30	1,418 ^x										
61	Сл.	1,36	Сл.	1,36			20	1,364	20ш	1,368	34ш	1,377	5	1,392	2в	1,365	2в	1,365
62							15	1,345	17	1,342			12	1,340				
62а							16	1,325 ^x					12	1,316 ^x				
63							30	1,310	20	1,308	31	1,308	8	1,311	1	1,310	1	1,310
64											17	1,285						
65							20	1,283 ^x	20	1,282	15	1,279						
66							35	1,248	40ш	1,247	28	1,248	37	1,255	10	1,245	3	1,249
69													6	1,237				
70	о.сл.	1,23					35	1,228	31	1,228	34	1,233			4	1,227	3	1,229
70а							21ш	1,215			29	1,214						
71							17	1,202	14	1,206	14	1,207	17	1,208				
72							17ш	1,193	15	1,195	24	1,193	9	1,182				
73							25ш	1,163	15	1,160	30ш	1,170	12	1,158	1в	1,160	1в	1,160
75											30ш	1,147						
76							25ш	1,139	22	1,138					3в	1,137	3в	1,140
78																		
80																		
81	Ср.	1,08					37ш	1,081	22	1,081	50ш	1,087						
82							25ш	1,061			20ш	1,061	9	1,063				
83							25ш	1,046	20	1,047	11	1,050						
84																		
86							21ш	1,021	15	1,021	13ш	1,025			1в	1,020	1в	1,020
89																		
90																		
91																		
92																		

Анализ 13 [25]. Материал отожжен при 800°-900°C. Условия съемки: Cu(Ni), V = 35кв, A = 12ма. Анализ 14, 16, 17, 18, 20 - образцы Кулик в различных пунктах Сибири, материал отожжен при 800°C на воздухе в течение 1 часа. Ре-антикатод, V = 35кв, A = 10, экспозиция от 12 до 28 час. Анализ 15. Тот же исходный материал, что и в ан. 14, но отожженный при 650°C. Остальные условия отжига и съемки те же. Анализ 8 [28]. Синтетический (1000°C) браннерит. Анализ 30 [30]. Синтетический браннерит. Анализ 24а, 25а [27]. Ан. 24а - тот же исходный материал, что и для ан. 24 (табл. I), но отожженный в атмосфере азота при 1300°C в течение часа. Автором отмечается присутствие помимо браннерита кубической фазы C и еще одной фазы "...", которые, однако, в рентгенограмме не выделяются. Ан. 25а - тот же исходный материал, что и для ан. 25 (табл. I), но отожженный при 1300°C на воздухе. Помимо браннерита автор указывает присутствие фазы "...", фазы не идентифицированной и не выделенной в рентгенограмме. Анализ 22а, 22б, 22в, 22г, 22д, 22е, 22ж [13, 14]. Исходный материал тот же, что и для ан. 22 (табл. I), но отожженный в различных режимах температуры на воздухе: 22а-600°C, 22б-700°C, 22в-800°C, 22г-1000°C, 22д-1100°C, 22е-1200°C, 22ж-900°C. Пробы 22а и 22б рассматриваются автором как чистый браннерит, остальные - как смесь фаз: 22б - браннерит + U₃O₈ + рутил, 22в - фаза X + U₃O₈ + рутил, 22г - браннерит + фаза X, 22д - браннерит + U₃O₈ + рутил, 22ж - чистая фаза X. Линии, по мнению автора принадлежащие фазе U₃O₈ отмечены в рентгенограммах X-линии рутила (xx), фаза X - (a). Анализ 31 [31]. Материал отожжен при 1000°C в течение 36 часов. Обозначения интенсивностей: и - средняя, v - сильная, w - слабая, vw - очень слабая, ww - очень, очень слабая. Рентгенограммы не исчерпываются приведенными линиями, указывается наличие других с меньшим межплоскостным расстоянием, которые обозначены в соответствии с рентгенограммой Пабста, взятой автором для сравнения. Анализ 32, 32а, 32б, 32в - [14]. Ан. 32 - материал отожжен при 800°C, помимо браннерита содержит фазу U₃O₈; Ан. 32а - тот же материал, отожженный при 900°C - идентифицирован автором как браннерит + фаза X + U₃O₈, 32б - отожжен при 1000°C, браннерит + фаза X; 32в - отожжен при 1100°C браннерит + U₃O₈. Обозначение фаз, как в ан. 32а и следующих того же автора. В продолжениях 3 и 4 табл. 5 указаны отсутствующие у объектов 22а-25е линии 1,4,6 и др., точно также, как имеющиеся здесь линии 14,24а,32а и др. указаны в предшествующей части таблицы.

№ линии	I		3а		II		26а	
	g	d/n	g	d/n	g	d/n	g	d/n
16					3	3,45	I	3,40
21	4	3,10	I+	3,11	10	3,02	5	3,08
	4-	(3,11)	2+	(3,08)				
22					I	2,98		
24							1в	2,90
26а	I	2,69	I-	2,68	6	2,69	2	2,65
	I-	(2,70)		(2,68)				
35					2	2,10	2	2,08
40	3	1,91	I	1,92				
	3-	(1,91)	2	(1,91)				
41					I	1,90		
42							3	1,89
43б					I	1,79		
50	3-	1,626	I	1,626				
	2+	(1,627)	2+	(1,630)				
51							3	1,61
54					I	1,56		
69	I-	1,238						
70							1в	1,23
70а	I-	1,213						
71							1в	1,20
79	I-	1,104						
80							1в	1,09
84	I-	1,057					1в	1,035

Анализ 1,3а - [29]. Тот же материал, что и для анализов 2 и 3 соответственно (табл. I), но без прокалывания. В скобках-d/n тех же линий после оджига при 520° C (анализ I) и 600°C (анализ 3). Анализ II [15]. Тот же материал, что и для анализа 12 (табл. I). Анализ 26а неотожженный аналог обр. 26 (табл. I).

и 59а в объекте 22е фиксировано в табл. 6 единицей.

Сравнение табл. 6 и 4 показывает следующее. Выделенные в качестве объектов-"экзаменаторов" рентгенограммы № 30, 24а, 25а, 22е, 22д, 22г, 31, 13, 22, 22в обнаруживают присутствие всех 8 отождествляющих признаков в 8 случаях у объекта 22 отсутствует признак-пакет 56а-57, у объекта 22в - признаки 32 и 59. Поскольку объект 22 представляет собой рентгенограмму образца, отожженного в вакууме, можно было бы думать, что отсутствие отождествляющего признака 56а-57 вызвано невозможностью окисления четырехвалентного урана до шестивалентного, что вероятно при отжиге на воздухе. Однако этот признак есть у образца, отожженного в атмосфере азота (объект 24а), где окисление также невозможно и потому отсутствие ее пока не находит объяснения, как и в случае образца 22в.

В двух случаях из десяти у объектов-"экзаменаторов" отсутствует по одному характеристичному признаку, и лишь у того же объекта 22в нет сразу трех из них.

Таким образом, наиболее полно отождествляющие и характеристичные линии реализованы у 8 объектов из десяти. Это вполне подтверждает правильность отбора совокупности линий-признаков, сделанного на табл. 4, а объекты, сгруппированные в ней, позволяют рассматривать как "эталонный" материал, представительный для всего минерального вида "браннерит".

Посмотрим, как проявляется выделенная совокупность линий у остальных объектов табл. 6.

1. Группа объектов (8, 32, 32а, 32в, 32б) характеризуется неполнотой рентгенограм и сравнение не может быть проведено по всей совокупности. Однако в пределах сравниваемой части объектов присутствуют все отождествляющие признаки и лишь в одном случае нет одного из трех характеристичных.

2. Объект 22ж - "фаза-х". Отсутствует один отождествляющий и 3 из четырех характеристичных признаков браннерита.

3. Из образцов-проб наиболее близки к группе эталонных браннеритов объекты 18 и 20, у которых отсутствует по одному отождествляющему и одному характеристичному признаку, и объект 14, у которого нет двух отождествляющих и одной характеристичной линий-признаков. Для объектов 16 и 17 расхождение очень ве-

тип объекта	т-отж	№ объекта	8	30	24а	25а	22е	22д	22г	31	13	22	22в	22б	14	16	17	18	20	22б	15	22а	11	26а	13а	28	29
экс.	иск.	8																									
экс.	1300	24а																									
экс.	1300	25а																									
экс.	1200	22е																									
экс.	1100	22д																									
экс.	1100	32в																									
экс.	1000	22г																									
экс.	1000	31																									
экс.	000	13																									
экс.	000	22а																									
экс.	000	22ж																									
экс.	800	22в																									
проба	800	14																									
проба	800	16																									
проба	800	17																									
проба	800	18																									
проба	800	20																									
	700	22б																									
	600	15																									
	неотожженные	22а																									
	неотожженные	11																									
	1000	26а																									
	1000	13а																									
	1000	28																									
	1000	29																									

ливо: нет соответственно 4 и 6 отождествляющих, 2 и 3 характеристичных признаков.

4. В рентгенограммах неотожженных и отоженных при низкой температуре образцов:

а) Обр. 22б - нет 5 отождествляющих и 2 характеристичных признаков; в обр. 22а - нет четырех отождествляющих признаков. Между тем обе рентгенограммы получены на том же материале, что и остальные рентгенограммы обр. 22, но при иных условиях отжига. На примере этих объектов видно, как меняется информационная значимость рентгенограммы и ее возможности для диагностики минерала в зависимости от температуры отжига материала.

б) То же самое относится к объектам I, 3а, II, I5, 26а, которые характеризуются очень низкой проявленностью отождествляющих и характеристичных линий, хотя тоже являются браннеритами, как это показывают рентгенограммы тех же образцов после отжига - № 2, 3, I2, I4, 26. Приведенное сравнение позволяет утверждать, что выделенная совокупность отождествляющих и характеристичных признаков-линий не соответствует наиболее устойчивым плоским сеткам структуры браннерита, которые дольше всех сохранились бы при метамиктном распаде, и что при имеющемся скудном фактическом материале по неметамиктным образцам такие устойчивые сетки выделить нельзя.

5. Наконец, сравнение лодочникита (объект 28) и торутита (объект 29) обнаруживает отсутствие у обоих по 2 отождествляющих и 3 (из 4-х) характеристичных признака. Здесь также, как при рассмотрении проб, встает вопрос, какую проявленность отождествляющих и характеристичных признаков следует считать достаточной для отнесения объекта к минеральному виду браннерит при выбранном подходе к диагностике. Этот вопрос будет рассмотрен после проведения сравнения эталонных браннеритов с типичными не-браннеритами.

Из не-браннеритов для сравнения выбраны давидит, эвксенит и ряд минералов, сходных по внешнему виду и физическим свойствам с браннеритом и также зачастую метамиктных [I6, I7, I8, 2I, 22]. Закодированные в признаковом пространстве таблицы 4 данные по этим минералам составляют табл. 7. Рассмотрение этой таблицы и ее сравнение с табл. 4 ("эталонные браннериты") показывает сле-

дующее. Группа минералов-бетафит - викиит в признаковом пространстве браннерита резко отличается малым количеством признаков и их расположением. Рентгенограммы этих минералов содержат, однако, в интервале тех же значений d/n линии, не совпадающие с линиями браннерита. Очевидно, что при практической диагностике такие образцы легко отличимы от браннерита и потому могут быть исключены из дальнейшего рассмотрения. Сравнение оставшейся части табл. 7 с табл. 4 показывает, что в признаковом пространстве браннерита давидит, эвксенит и остальные близкие минералы обнаруживают большое количество признаков. Так, по числу выполняемых признаков некоторые давидиты (обр. 40) и эвксениты (обр. 36) оказываются ближе к эталонным браннеритам, чем некоторые из собственно браннеритов (обр. 22в), причем выполняется большое число отождествляющих (до 6-7) и значительная часть характеристичных признаков браннерита. Тем самым выявляется, что значительная часть отождествляющих и все характеристичные линии, свойственные браннериту и объединяющие объекты внутри минерального вида, в то же время не являются различающими этот вид от других, поскольку присущи и другим титанатам и ниобатам. Это показывает, что задача различения браннеритов и этих минералов действительно является актуальной.

Рассмотрение табл. 7 показывает, однако, что среди отождествляющих линий браннерита имеются такие, которые устойчиво отсутствуют в рентгенограммах остальных сравниваемых минералов. Это линия I6, которая не встречается у эвксенита и близких к нему минералов, и пакет линий 67-69, надежно отличающий браннерит от давидита. Эти признаки присутствуют во всех просмотренных объектах браннерита, т.е. как раз характеризуются высокой устойчивостью проявления. Именно они, таким образом, среди отождествляющих линий браннерита являются одновременно различающими по отношению к наиболее трудно отличимым минералам.

Близкими к различающим оказываются также признак 32, отличающий браннерит от эвксенита, и пакет линий 56а-57, редко встречаемый у давидита. Появление линий в пределах этого пакета у давидита может быть объяснено специфичным присутствием в качестве примесей давидита минералов ильменита и гематита. Чтобы показать это, в табл. 7 для давидитов был проведен анализ на

или невозможным даже на заведомо браннеритовых образцах в случае малого количества линий — как это видно из сравнения объектов 22б, 15, 22а, 11, 26а и других неотожженных или слабо отожженных образцов с их хорошо отожженными аналогами. Исходные данные по браннеритам и минералам, взятым для сравнения позволяют считать хорошей проявленностью, достаточной для успешного распознавания, присутствие на рентгенограмме более 20 пар линий.

2. Для распознавания браннерита в первую очередь оказывается важным присутствие в рентгенограмме линий, соответствующих признакам I6 и 67-69. Как было показано на табл. 7, отсутствие хотя бы одного из них свойственно только не-браннеритам.

3. Поскольку, однако, область проявления линии I6 совпадает с одной из наиболее сильных линий ураноксида — примеси, образование которой не исключено при отжиге и других ураносодержащих титанатов, (хотя как видно из табл. 7, у не-браннеритов ее образование не наблюдается) — для надежности диагностики необходимо быть уверенным, что присутствие линии I6 не вызвано только примесью ураноксида. Предлагается для проверки использовать различающие признаки 7, II-I2, 20, 43, 60, свойственные главным образом браннериту, интенсивность которых меньше, чем у линии I6, и в случае их проявления можно быть уверенным, что проявлена и линия I6, как собственно браннеритовая.

4. Как видно из табл. 4 и 6, для всех достаточно полно проявленных рентгенограмм браннеритов свойственно одновременное присутствие по крайней мере 6 из 8 отождествляющих линий (признаков). У не-браннеритов выполнение такого количества отождествляющих признаков крайне редко.

Схема диагностики

На основании изложенного можно предложить следующую схему диагностики.

A. Проверка полноты проявленности линий в рентгенограмме.

1. Проявленность хорошая (в рентгенограмме свыше 20 пар линий).
2. Проявленность недостаточная (менее 20 пар линий).

а) Все линии свойственны браннеритам — необходимо улучшение проявленности рентгенограммы (дополнительный отжиг, изменение экспозиции съемки и т.п.)

б) Присутствуют заведомо не-браннеритовые линии — не-браннерит.

B. В случае выполнения A-I: Проверка присутствия в рентгенограмме признаков I6 и 67-69.

I. Нет одного из признаков — не браннерит.

2. Оба признака присутствуют —> B

B. В случае B-2: Проверка наличия хотя бы одного из различающих признаков (7, II-I2, 20, 43, 60).

I. Нет ни одного из указанных признаков —> Г

2. Есть хотя бы один из указанных признаков —> Д

Г. В случае B-I: Проверка наличия примесей U_3O_8 , брукита, анатаза.

I. Указанные примеси не проявлены в области признака I6 —> Д

2. Указанные примеси проявлены в области признака I6

а) Проявление признака I6 обусловлено только примесью — не браннерит (см. B-I)

б) Возможно наложение линий примеси и браннеритовых —> Д

Д. В случаях B-2, Г-I и Г-2-б: Установление числа отождествляющих признаков (I6, 28-30, 32, 37-38, 56а-57, 59, 61, 67-69), присутствующих в рентгенограмме.

I. Число отождествляющих признаков ≥ 6 — браннерит

2. Число отождествляющих менее 6 — диагностика ненадежна;

необходимо проверить качество рентгенограммы и полноту ее проявленности, можно воспользоваться проверкой на присутствие дополнительных различающих признаков — линий 52, 34а, 63. Следует провести сопоставление с другими минералами (не браннеритами).

В случае Д-I для подтверждения принадлежности к браннеритам следует провести дополнительную проверку на присутствие характеристичных признаков браннерита: I7-I8, 21-22, 24, 53-55.

В соответствии с приведенной схемой лодочникит может быть отнесен к минеральному виду "браннерит", тогда как для торутита структурные отличия оказываются слишком велики.

Выводы

Использование аппарата функций алгебры логики и логико-дискретного подхода к обработке данных рентгено-структурного анализа привело к следующим результатам:

1. Установлена группа дебаграмм браннерита, которая может служить сводным эталонным объектом. Сравнение с ним позволяет надежно диагностировать браннерит.

2. Установлено, что при диагностике нет необходимости проводить сравнение по всем возможным для браннерита линиям и выделены наиболее важные в диагностическом отношении линии. Эти линии не всегда являются наиболее сильными, что, однако, несущественно, поскольку для браннерита вообще характерна большая изменчивость интенсивности для одних и тех же линий.

3. Предложена схема диагностики, проверенная авторами на всех имевшихся образцах.

В этой схеме наряду с важными в диагностическом отношении отдельными линиями (признаками) существенную роль играют пакеты признаков, выполняющиеся для всех эталонных рентгенограмм и представляющие собой логические дизъюнкции исходных признаков.

Литература

1. Александров В.Б., Пятенко Ю.А. Рентгенометрическое исследование некоторых метамиктных титанониобатов. - "Докл. АН СССР", 1959, т.124, № 1, с.179-182.

2. Андропова Е.П. Физико-химическое исследование браннерита. - В кн.: Петрология и минералогия некоторых рудных районов СССР. М., Госгеолгиздат, 1951, с.81-102.

3. Гецева Р.В., Савельева К.Т. Руководство по определению урановых минералов. М., Госгеолгиздат, 1956, с.51-52.

4. Готман Я.Д. Лодочникит - новый минерал из группы титанатов урана и тория. - "Записки Всесоюз.минералогич.об-ва", 1958, ч.87, вып.2, с.197-200.

5. Готман Я.Д., Хапаев И.А. Торутит - новый минерал из группы титанатов тория. - "Записки Всесоюз.минералогич.об-ва", 1958, ч.87, вып.2, с.201-203.

6. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. О математических принципах классификации предметов и явлений. - В кн.: Дискретный анализ. Вып.7, Новосибирск, "Наука", 1966, с.3-15.

7. Дмитриев А.Н. Некоторые табличные числа. - В кн.: Дискретный анализ. Вып.12, Новосибирск, "Наука", 1968, с.22-26.

8. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. Об одном принципе классификации и прогнозе геологических объектов и явлений. - "Геология и геофизика", 1968, № 5, с.50-64.

9. Жабин А.Г., Александров В.Б., Казакова М.Е. Об эшните гидротермального генезиса из Вишневых гор. - "Труды ИМГРЭ АН СССР", 1961, вып.7, с.108-112.

10. Жабин А.Г., Воронкова А.А., Халезова Е.Б., Казакова М.Е. Новые данные об акцессорном давидите из Вишневых Гор на Урале и о так называемых уфертите и ферутите. - "Труды ИМГРЭ АН СССР", 1963, вып.15, с.110-120.

11. Заварзин А.В. Об условиях нахождения браннерита в кварц-пирит-серицитовых метасоматических телах. - "Геология рудных месторождений", 1961, № 6, с.54-58.

12. Красавчиков В.О. Модификация тестового подхода к анализу таблиц описаний на основе понятия пакета. - В кн.: Дискретный анализ. Вып.26, Новосибирск, "Наука", 1974, с.36-60.

13. Кривоконова Г.К. О поведении метамиктного браннерита в процессе термообработки. - "Записки Всесоюз.минералогич.об-ва", 1972, № 5, с.254-267.

14. Кривоконова Г.К. Фазовые превращения браннерита в процессе отжига. - В кн.: Краткие сообщения по минералогии и геохимии за 1968 г. М., "Наука", 1970, с.77-82 (Труды ВИМС, вып.1).

15. Кулик Н.А. Совместный одновременный рост уранинита и браннерита и некоторые данные об условиях образования урановых минералов. - В кн.: Геохимия и минералогия радиоактивных элементов Сибири. Новосибирск, "Наука", 1970, с.17-53.

16. Ложникова О.Н., Яковлева С.Я. Рентгенометрический справочник-определитель минералов, содержащих редкоземельные элементы. М., "Наука", 1961, 224 с.

17. "Минералы". Справочник, т.2, вып.3, М., "Наука", 1967, 675 с.

18. Михеев В.И. Рентгенометрический определитель минералов. М., Госгеолтехиздат, 1957, 867 с.

19. Повилайтис М.М. По поводу новых минералов - лодочникита, абсита и торутита. - "Записки Всесоюзн.минералогич. об-ва", ч.92, вып. I, 1963, с.118-123.

20. Повилайтис М.М., Яковлевская Г.А., Князева Д.Н., Беляева И.Д. О свойствах браннерита и возможности определения абсолютного возраста по нему. - "Записки Всесоюзн.минералогич.об-ва", 1968, ч.97, вып.2, с.150-161.

21. Сидоренко Г.А. Рентгенометрический определитель урановых и урансодержащих минералов. М., Госгеолтехиздат, 1960, 116 с.

22. Соболева М.В., Пудовкина И.А. Минералы урана. М., Госгеолтехиздат, 1957, 408 с.

23. Чернышева Л.В., Куприянова И.И., Сидоренко Г.А. Браннерит из Средней Азии. - В кн.: Краткие сообщения по минералогии и геохимии за 1968 г. М., "Наука", 1970, с.83-88. (Труды ВИМС, вып. I).

24. Яблонский С.В. Функциональные построения в К-значной логике. М., "Наука", 1958, с.3-142 (Труды математического института им. Стеклова, т.51).

25. Ishihara S., Kavachi Y. Uraniferous Ore Deposit at Futago-jima, Kagoshima Prefecture. - Report geol. surv. Japan, N° 190, 1961.

26. Karkhanavala M.D. X-Ray & electron diffraction study of Brannerit from Rajasthan. J. Scient and Industr. Res. (Ind.), BC 18 N° 1, B 7-B10, 1959.

27. Lima-deFaria J. Identification of metamict minerals. 1964 (Portug.).

28. Oztunali Önder. Über die Struktur von Brannerit. - Neus Jahrb. Mineralogie, Monatsheft N°8, 1959.

29. PABST A. Brannerite from California. - Amer. Mineralogist, 39, N° 1-2, 1954.

30. PATCHETT J.E., NUFFILD E.W. Studies of radioactive compounds X. The synthesis and crystallography of brannerite. - Canad. Mineral, 1960, Vol. 6, pt 4.

31. UMAMESHVARARAO G.V., KRISHNASVAMY R. Brannerit from Rajasthan. - J. Scient and Industr. Res. (Ind.) A-15, N° 9, 1956.

32. Whittle A.W.G. The nature of davidite. - Economic Geology 1959, V. 54, N° 1.

СОДЕРЖАНИЕ

I. ТЕОРЕТИЧЕСКИЕ ВОПРОСЫ

1. Дмитриев А.Н., Красавчиков В.О. Тестовый подход в решении проблем обработки геологической информации 3
2. Смертин Е.А. Вопросы теории и алгоритмы на базе построения D-тестов 48

II. ВОПРОСЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ

3. Гуваков А.И., Дмитриев А.Н., Кандыба В.Н. Оценка перспективности оловорудных районов Приморья 68
4. Бабич В.В., Федосеев Г.С. Прогнозная оценка магнитных аномалий Холзуно-Инского района по геофизическим данным 94
5. Кулик Н.А., Красавчиков В.О., Штатнова Т.И. Опыт использования логико-дискретного подхода при рентгеновской диагностике браннерита 109