

## НОВЫЕ ТЕСТОВЫЕ РАЗРАБОТКИ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ РУДОНОСНОСТИ (НА ПРИМЕРЕ ТРАППОВЫХ ИНТРУЗИЙ)

Геологическая отрасль знания и производства на современном этапе развития оснащается средствами автоматического преобразования полученной и учтенной информации. Это преобразование в основе своей опирается на готовность геологических исследований становиться все более точной наукой и на инструментальную готовность математики, преодолевающей в ряде своих разделов сугубую абстракцию. Плодотворная встреча системы математических процедур с множеством сообщений, скомпонованных в геологическую постановку задачи, осуществляется на новой территории научного познания, которая еще ожидает своего наименования. Оснащение геологической науки математическими средствами исследования — процесс интенсивный и не всегда достаточно определенный, особенно на этапах постановок задач и результативного принятия решений. С нашей точки зрения, эта напряженность процесса подтверждает его полезность и значимость. В самом общем смысле прогрессирующее проникновение математики в геологию (и обратно) осуществляется в наиболее развитых и точных разделах геологии (мы не имеем при этом в виду геофизику, внутренняя математичность которой всегда очевидна).

Взаимопроникновение математики и геологии должно быть обоюдоплодотворным, т.е. вслед за признанием полезности математики для геологии следует ожидать обнаружения полезности геологии для создания новых разделов прикладной математики.

### НЕКОТОРЫЕ СВОЙСТВА СРАВНИТЕЛЬНОГО ИЗУЧЕНИЯ ОБЪЕКТОВ

Широко развитый в геологии метод сравнительного изучения объектов и явлений можно представить в качестве предмета математического исследования. Эта возможность основывается на том, что сравнение результатов геологических наблюдений, регистрируемых описаниями объектов и явлений, элементаризируется на отдельные логические процедуры (Бугаец и др., 1968; Васильев и др., 1974). Как правило, эти процедуры в конечном итоге сводятся к выяснению двух рядов сообщений: ряд сообщений о сходстве объектов и ряд сообщений об их различии (Родионов, Серых, 1968; Бабич и др., 1974). После установления существенности данного сообщения в характеристике объектов и выяснении его роли (отождествляет их или различает) производится со-

тавление нового списка признаков, который содержит уже информацию не только о данном объекте, но и об отношениях (сходны, различны) сравниваемых объектов. Признаки (или свойства объектов), по которым осуществляется процедура сравнительного изучения, являются как бы внешними, выведенными в плоскость прямого и обратного взаимодействия объектов исследования в сфере их информационного существования (Константинов и др., 1970; Москаленко, Николаев, 1972).

Объект, информационные характеристики которого не соприкасаются с характеристиками каких-либо других объектов, называется информационно-замкнутым объектом (неоткрытое месторождение совершенно нового типа). Объект, информационные характеристики которого слабо соприкасаются с характеристиками каких-либо других объектов, называется информационно ограниченным объектом (например, открытое месторождение нового типа). Объект, информационные характеристики которого образуют характеристические сходства с характеристиками каких-либо других объектов, называется информационно открытым объектом (например, открытое месторождение широко известного типа).

Совокупность информационно замкнутых объектов представляет собой проблематичное множество; совокупность информационно ограниченных — уникальное множество объектов (несколько десятков). Совокупность информационно открытых — развитое множество объектов (сотни и тысячи объектов).

В настоящее время трудно что-либо определенное сказать в отношении объектов проблематичного множества. Сравнительное изучение в области уникального множества объектов в геологии развито очень широко. Причины повышенного увлечения этим множеством лежат не только в содержательном интересе, но и в практической существенности объектов, составляющих это множество (например, комплексные месторождения золота и урана в докембрийских конгломератах, медно-никелевые месторождения дифференцированных трапповых интрузий). Однако эпицентром сравнительного изучения геологических объектов в области металлогении все же, видимо, будет развитое множество объектов. Рассмотрим "ручные" возможности работы в сфере сравнительного изучения объектов уникального и развитых множеств.

Сравнительное изучение уникальных объектов производится с помощью логических и (или) измерительных операций на каждом из исследуемых объектов. Причем собственно сравнение характеристик объектов осуществляется на базе знаний геолога и его практического опыта. Небольшое число объектов изучения не позволяет "сшивать непрерывно" сравниваемые результаты, поскольку в классе таких объектов каждый объект это класс. Чтобы типизовать уникальное множество, геолог вынужден обращаться к деталям, что резко увеличивает число необходимых процедур сравнения и их целевое разнообразие (Бишаев, 1973).

Работа в области развитого множества объектов становится почти невозможной для ручного исполнения начиная с процедур мобилизации информации.

Целевая подготовка информации по задаче развитого множества (например, мобилизация информации по золоторудным месторождениям Забайкалья заняла три года работы хорошего специалиста). И "ручное" сравнительное изучение в пространстве 100 объектов и 200 признаков практически уже не выполнимо.

Возможное увеличение и уточнение количества логических и вычислительных операций лежит на путях автоматизации ряда звеньев процесса сравнительного изучения. Кибернетический подход и является именно тем подходом, который на базе математических разработок ряда логических и вычислительных операций позволяет автоматизировать трудоемкие участки исследовательской работы этого направления (Родионов, 1968; Константинов, 1973).

### ВОПРОСЫ ВОЗНИКНОВЕНИЯ ТЕСТОВОГО ПОДХОДА

Как было указано выше, широко развитый в геологии метод сравнительного изучения объектов сводится к выяснению черт их сходства и различия. Если совокупность объектов (в нашем случае месторождений) представлена перечнем характеристических признаков так, что мобилизованная по цели исследования информация может быть задана таблицей из нулей и единиц, то для облегчения процедур сравнительного изучения может быть применен тестовый подход.

Математическое понятие "тест" впервые введено С.В. Яблонским (Чегис, Яблонский, 1958) в связи с задачей отыскания экономичных способов обнаружения неисправности в сложных системах, таких, как счетно-решающие устройства, управляющие системы. Эти системы, как правило, являются материальными реализациями одной или нескольких функций алгебры логики, а удобное их описание дают контактные схемы. Мысль о возможном применении тестов в сфере естествознания в беседе с автором высказал Ю.И. Журавлев в 1964 г. Совместная попытка адаптировать техническую схему тестового подхода применительно к геологическим задачам начата в 1965 г. После преодоления первой трудности организации пространства признаков и кодирования логической информации геологических описаний месторождений в начале 1966 г. была рассмотрена задача о сравнительном изучении зарубежных месторождений докембрийских конгломератов в постановке Ф.П. Кренделева. Существенные отличия геологических задач от технических в тестовом рассмотрении были обнаружены на первых же шагах; среди них отметим следующие: а) большой по сравнению с таблицами технических объектов объем геологических таблиц и как результат этого большое число тестов (десятки тысяч), исключающее оценку диагностического значения каждого теста; б) реальные таблицы геологических объектов содержали отождествляющие столбцы (для всех объектов значения одинаковы — 1 или 0) и равные строки, а также прочерки — значения неизвестны; в) задачи в геологической постановке редко обхо-

дятся одним классом объектов; как правило, приходится изучать два, три и более классов, что выдвинуло проблему многотабличных исследований.

Способы, которыми преодолевались трудности, встретившие тестовый подход в применении к геологическим задачам, преобразились в реальные алгоритмы и программы для ЭВМ. Так, например, суммирование числа вхождения каждого признака в общее число тестов позволило оценить существенность характеристических признаков, сумма существенностей признаков оценила важность строки (объекта, т.е. месторождения). Введение табличных чисел позволило привести реальную таблицу геологического материала к допустимой в тестовом смысле без особой потери полезной информации (Дмитриев, 1968). Многотабличные исследования в тестовом подходе были начаты введением в свое время понятия тестов (Дмитриев, и др., 1966).

Дальнейшее развитие тестового подхода в естествознании выразилось в подразделении на два основных направления: поиск новых точек приложения разработанных алгоритмов и программ (Бугаец и др., 1968; Вышемирский и др., 1971; Петровская и др., 1971, Трофимук и др., 1971; Яблонский и др., 1971; Константинов, 1973; Гольдман и др., 1974) и дальнейшее развитие самого подхода (Нестеренко и др., 1969; Смертин, Дмитриев, 1970; Дмитриев, 1973; Дмитриев и др., 1973; Николаев, 1974). Второе направление в свою очередь подразделяется на разработки усовершенствования и новые разработки. Разработки усовершенствования проецируются в область исследований, где уточняются и расширяются имеющиеся тестовые процедуры, а новые разработки проецируются в область исследований, где на естественном продолжении тестового подхода строятся новые процедуры в соответствии с теоретической и практической необходимостью. Так, процедуры построения тестов для большеобъемных таблиц вероятностным методом, различные виды целевого голосования по тестам и тесторам, использование длин тупиковых тестов и тесторов являются разработками использования (Красавчиков, 1973; Николаев, 1974). Введение же таких понятий (и связанных с ними процедур), как Q-тест, H-тест, K-тестор, тупиковый пакет, тестовое мнение, является продолжающей разработкой (Красавчиков, 1973; Королева, 1974). На практике разграничение это не столь отчетливо, как показано на приведенных примерах, и подразделение разработок вызвано необходимостью некоторой систематизации тестового подхода.

### ТЕСТОВОЕ ИЗУЧЕНИЕ РАЗЛИЧИЙ И СХОДСТВ ОБЪЕКТОВ

Различие и сходство объектов, выявляемое в процессе сравнительного изучения, можно детально рассмотреть с помощью тестовых процедур, причем как в рамках одной таблицы, так и для одновременно исследуемых нескольких таблиц. Приведем перечень тестовых процедур и краткое их истолкование в содержательном смысле.

### Однотабличное изучение различий

Приводимые ниже величины прямо или косвенно вычисляются на базе построения всех тупиковых тестов. Название этих величин производится с учетом содержательной трактовки процедур, с помощью которых эти величины оцениваются.

Пусть  $\{x_1, \dots, x_n, x_{n+1}\}$  — список признаков, организующих признаковое пространство, признак  $x_{n+1}$  фиксирован в качестве целого. Каждый объект таблицы охарактеризован всеми признаками, составляющими признаковое пространство. Каждый признак имеет два значения — 0 или 1, а таблица, составленная из таких признаков, называется бинарной. Бинарная таблица называется допустимой, если все ее строки попарно различны. Набор столбцов  $t^* = t_{i_1}, \dots, t_{i_k}$  допустимой таблицы  $T$  называется тестом, если при удалении из  $T$  всех столбцов, не входящих в  $t^*$ , полученная таблица является допустимой. Тест называется тупиковым, если из него нельзя удалить ни одного столбца без того, чтобы он перестал тестом. Путем построения всех тупиковых тестов для данной допустимой таблицы  $T$  находят числа:  $k$  — число всех тупиковых тестов таблицы  $T$  и  $k_i$  — число всех тупиковых тестов, в которые входит  $i$ -й столбец. На основе этих двух чисел вычисляется различающий вес столбца (признака)

$$P_i = \frac{K_i}{K}. \quad (1)$$

Содержательное истолкование различающего веса выводится из анализа понятий — тупиковый тест и множество тестов, построенных для данной  $T$ . Тупиковый тест таблицы  $T$  (по определению) представляет собой элементарный и несжимаемый акт различения (распознавания) всех строк таблицы  $T$ . Тогда все множество тупиковых тестов является количественной мерой разнообразия  $R_T$  таблицы  $T$

$$R_T = \frac{K}{V}, \quad (2)$$

где  $V = mn$ ;  $m$  — число строк, а  $n$  — число столбцов в  $T$ . Наличие хотя бы одного тупикового теста для таблицы  $T$  свидетельствует о единственном способе несходства всех строк по данному набору признаков. Но, как показала практика, конкретные таблицы обладают высоким потенциалом разнообразия; например, таблица по исследованию редкометального орудения (Модников и др., 1969) размером  $10 \times 27$  содержит более 45 тыс. тупиковых тестов,  $R_T = 166$ . Существует много признаков комбинаций, различающих все строки таблицы  $T$ , и это показывает, что реальные объекты, заданные перечнем характеристических признаков, многократно различимы поднаборами признаков из их общего списка. Наблюдение способности каждого столбца попадать во все тупиковые тесты показало большую неравно-

значность столбцов в процедурах различения строк. Столбцы, наибольшее число раз входящие во все тупиковые тесты, т.е. имеющие высокий различающий вес  $P_i$ , считаются наиболее информативными, поскольку наиболее полно характеризуют разнообразия таблицы (Дмитриев и др., 1966; Васильев и др., 1974<sup>1,2</sup>).

Нередко по содержательным соображениям требуется установить меру выразительности теста по его длине (числу столбцов, составляющих данный тест  $t$ ). Оценим эту выразительность как разность  $L(T) - l_i(t)$ , где  $L(T)$  — средняя длина тупикового теста для  $T$ , а  $l_i(t)$  — длина теста  $t$ . Очевидно, что чем больше разность, тем экономнее тест различает все строки таблицы  $T$  (Дмитриев и др., 1973). Учет длины тестов позволяет более тонко дифференцировать столбцы и строки таблиц, подлежащих исследованию. Так, например, если столбцы  $i_1$  и  $i_2$  имеют близкие значения различающих весов  $P_{i_1}$  и  $P_{i_2}$ , то при необходимости указать лучший различающий столбец следует учитывать среднюю длину тупиковых тестов столбцов. Признаки, попадающие наибольшее число раз в тесты малой длины, называются максимально различающими признаками. Совокупность максимально различающих признаков называется ядром разнообразия<sup>1</sup> таблицы  $T$ . Вычленение этого ядра практически означает обнаружение информационного очага исследуемых объектов. Вне ядра количество информации убывает, так как объекты на его периферии становятся все более тождественными, и внутритабличное исследование разнообразия превращается в исследование однообразия, т.е. черт сходства объектов.

### Однотабличное изучение сходств

При исследовании сходства объектов обращается внимание на такую группу признаков, по которым наибольшее число объектов в классе неразлично. Воспользуемся простой процедурой поиска неразличающих признаков на основе учета различающего веса.

Пусть  $T$  — бинарная таблица, а  $t$  — ее тупиковый тест. Оставшиеся столбцы из  $T$ , в которые не входят столбцы из  $t$ , образуют  $F$ -таблицу. Заметим, что строки из  $F$  более сходны между собой, чем строки из  $T$ , в том смысле, что если  $\rho(j_1, j_2)^{(0)}$  — расстояние Хеминга между  $j_1$  и  $j_2$ -й строками по таблице  $T$ , а  $d(j_1, j_2)^{(0)}$  — расстояние Хеминга между  $j_1$  и  $j_2$ -й строками по таблице  $F$ , то следует, что  $d(j_1, j_2) \ll \rho(j_1, j_2) - 1$  для каждой пары  $(j_1, j_2)$ ,  $1 \leq j_1 < j_2 \leq m$ . Это обстоятельство дает основание назвать отождествителем для  $T$ . Если определить тупиковый отождествитель как отождествитель, который теряет свойство (0) при добавлении любого столбца таблицы, то

<sup>1</sup> Следует указать на содержательное сходство понятия разнообразия в тестовом подходе и понятия разнообразия по Эшби (1959).

получим, что  $F$ -тупиковый отождествитель. Далее непосредственно из определения следует, что и каждому тупиковому тесту  $t$  соответствует такой тупиковый отождествитель  $F$ , что совместно они дополняют  $F$  до  $T$ . Следовательно, число тупиковых отождествителей равно числу тупиковых тестов  $K$ . Число же тупиковых отождествителей, в которые входит  $i$ -й столбец, равно  $K - K_i$ . Следовательно,  $Q = 1 - P_i$ .

Величину  $Q$  можно использовать в процессе их сравнительного изучения.

Задача оценки существенности  $i$ -го столбца из совокупности  $x_1, \dots, x_n$  для объединения строк таблицы  $T$  в сжатые или малоразличимые совокупности объектов решается средствами  $Q$ -тестов. Определение  $Q$ -тестов формируется по аналогии с различающими тестами.

Если  $t_{i_1}^{j_1} = t_{i_2}^{j_2}$ , то строки  $t_{i_1}^{j_1}, t_{i_2}^{j_2}$  называются сходными по столбцу  $t_i$ ; строки, сходные хотя бы по одному столбцу, называются сходными строками.

Таблица  $T$  называется  $Q$ -допустимой, если каждые две строки ее — сходные строки. Набор столбцов  $\langle t_{i_1}, t_{i_2}, \dots, t_{i_1} \rangle$  называется  $Q$ -тестом, если таблица, полученная из  $T$  удалением всех столбцов

$\langle t_{i_1}, t_{i_2}, \dots, t_{i_1} \rangle$ , является  $Q$ -допустимой.  $Q$ -тест называется тупиковым, если удалением любого столбца он преобразуется в не- $Q$ -тест.

Построив полную совокупность тупиковых  $Q$ -тестов таблицы  $T$ ,

можно оценить отождествляющую<sup>1</sup> существенность каждого признака. Мера важности признака  $x_i, i = 1, 2, \dots, n$  берется пропорциональной числу тех тупиковых  $Q$ -тестов, в которых имеется столбец  $t_i$  и равной

$$Q_i = \frac{K_{Q_i}}{K_Q}, \quad (3)$$

где  $K_Q$  — число всех тупиковых  $Q$ -тестов таблицы  $T$ . Содержательно величина  $Q_i$  обозначает похожесть строк в  $i$ -м столбце и является некоторой мерой вырождения разнообразия в единство, т. е. своеобразной оценкой меры убывания информации.

Для оценки того, насколько уместен объект  $l_j, j = 1, 2, \dots, m$ , в таблице  $T$  исследуемого класса объектов, вычисляется величина  $Q(t^j)$  по формуле

$$Q(t^j) = \sum_{i=1}^n \left( t_{i_1}^j P_i + (1 - t_{i_1}^j)(1 - P_i) \right) Q_i, \quad (4)$$

где  $P_i = \frac{1}{m} \sum_{j=1}^m t_{i_1}^j$ . Чем выше значение величины  $Q(t^j)$ , тем уместнее

<sup>1</sup> Следует иметь в виду самостоятельное значение отождествителей; они не симметричны тестовым весам.

объект  $l_j$  в ряду  $l_1, l_2, \dots, l_m$ . Кроме того, оценка  $Q(t^j)$  позволяет устанавливать степень однородности объектов, сведенных в данный класс, а упорядочивание объектов по этой величине вычленяет ядро сходства исследуемого класса. Как правило, признаки ядер сходства и разнообразия объектов для данной таблицы  $T$  не пересекаются. Рассуждения по использованию длин тупиковых  $Q$ -тестов в процедурах изучения во многом сходны с той лишь разницей, что исследуется сходство объектов.

### Многотабличное изучение различий

Необходимость в многотабличном (межтабличном) исследовании возникает, когда число классов объектов, подлежащих исследованию, два, три и более. Наряду с приближенными способами межтабличного изучения используются более детальные и трудоемкие тесторные методы, развившиеся из идей и средств тестового подхода.

Пусть заданы бинарные таблицы, допустимые в тестовом смысле,  $A = \{T_1, \dots, T_j, \dots, T_n\}$  — набор бинарных таблиц с  $n$  столбцами в каждой.

Если одна и та же строка  $S = (a_1, \dots, a_n)$  не может входить одновременно ни в какие две сравниваемые таблицы из  $A$ , то набор  $A$  называется допустимым. Набор признаков  $t = \{x_{i_1}, \dots, x_{i_k}\}$

тестором допустимого набора таблиц  $T_1, \dots, T_M$ , если при удалении из них всех столбцов, отличных от  $i_1, \dots, i_k$ , получим снова допустимый набор таблиц.

Тестор называется тупиковым, если из него нельзя удалить ни одного признака без того, чтобы он перестал быть тестором. Применение тесторного метода для сравнительного изучения двух классов объектов, представленных бинарными таблицами  $T_1$  и  $T_2$  (строки — объекты изучения, столбцы — признаки объектов), производится с помощью построения всех тупиковых тесторов. Задача выбора набора признаков, максимально различающих объекты сравниваемых таблиц  $T_1$  и  $T_2$ , сводится к вычислению тесторных различающих весов  $P^*$ . Оценка тесторных весов признаков производится построением всех тупиковых тесторов (неизбыточных наборов признаков, позволяющих различить каждый объект I класса от каждого объекта II класса)

$$P_i^* = \frac{K_i^*}{K^*}, \quad (5)$$

где  $K_i^*$  — число тупиковых тесторов с участием  $i$ -го признака;  $K^*$  — общее число всех тупиковых тесторов. Совокупность всех признаков с максимальными тесторными весами характеризует межклассовое разнообразие объектов.

В ряде практических случаев необходимо знать вклад тупиковых тесторов (Васильев и др., 1974<sub>1,2</sub>) в оценку существенности каждого

признака. При исследовании двух классов это можно учесть подразделением тесторов по заданному порогу длин

$$P_i^* = \frac{K_i^*(\bar{l})}{K^*(\bar{l})}, \quad (6)$$

где  $K_i^*(\bar{l})$  — число тупиковых тесторов меньше некоторой длины  $\bar{l}$ , с участием  $i$ -го признака (например, естественно принимать  $\bar{l}$  равной средней длине всех тупиковых тесторов данных таблиц);  $K^*(\bar{l})$  — соответствующее число всех тупиковых тесторов.

Содержательно длина тупикового тестора вскрывает концентрацию многотабличного (межтабличного) разнообразия в каждом данном тесторе, который осуществляет различие всех строк двух исследуемых классов объектов. Очевидно, что чем меньше в построении данного тупикового тестора участвует столбцов, тем больше эффективность каждого из них в обнаружении разнообразия классов и тем выше информационная емкость тестора. Отсюда следует, что самые сильные различающие столбцы образуют самые короткие тупиковые тесторы, а совокупность столбцов (признаков), вошедших во все тесторы минимальной длины, образует критерий различия двух исследуемых классов. Признаки, составляющие критерий различия двух классов, минимизируют пространство признаков в исследовании разнообразия классов при их сравнительном изучении.

### Многотабличное изучение сходств

Представляется целесообразным распространить понятия и процедуры, связанные с  $Q$ -тестами на случай многотабличного изучения сходств. Пусть имеется набор бинарных  $Q$ -допустимых таблиц  $A = \{T_1, T_2, \dots, T_M\}$  с  $n$  столбцами.

Набор признаков  $t = \{x_{i_1}, \dots, x_{i_k}\}$ , принадлежащих набору  $Q$ -допустимых таблиц  $T_1, \dots, T_M$ , называется  $Q$ -тестором, если при удалении из каждой таблицы всех столбцов, отличных от  $x_{i_1}, \dots, x_{i_k}$  снова получается набор  $Q$ -допустимых таблиц.  $Q$ -тестор называется тупиковым, если при удалении из него хотя бы одного признака он перестает быть  $Q$ -тестором.

Применение  $Q$ -тесторных процедур для сравнительного изучения двух классов объектов, представленных бинарными таблицами  $T_1$  и  $T_2$ , опирается на построение всех тупиковых  $Q$ -тесторов. Задача выбора набора признаков, максимально связывающих объекты внутри сравниваемых таблиц  $T_1$  и  $T_2$  сводится к вычислению  $Q$ -тесторных связывающих весов  $Q_i^*$ . Оценка  $Q$ -тесторных весов признаков производится с помощью формулы

$$Q_i^* = \frac{U_i^*}{U^*}, \quad (7)$$

где  $U^*$  — число тупиковых  $Q$ -тесторов с участием  $i$ -го признака,  $U_i^*$

общее число всех тупиковых тесторов (неизбыточных наборов признаков, позволяющих связывать объекты внутри I класса и одновременно объекты внутри II класса). Совокупность всех признаков с максимальными  $Q$ -тесторными весами характеризует однообразие объектов внутри классов.

Вклад длин тупиковых  $Q$ -тесторов в оценку существенности признаков при исследовании двух классов объектов учитывается подразделением  $Q$ -тесторов по заданному порогу длин

$$\tilde{Q}_i^* = \frac{U_i^*(\bar{l})}{U^*(\bar{l})}, \quad (8)$$

где  $U_i^*(\bar{l})$  — число тупиковых  $Q$ -тесторов длиной меньше некоторого  $\bar{l}$  с участием  $i$ -го признака, а  $U^*(\bar{l})$  — соответствующее число всех тупиковых  $Q$ -тесторов (в качестве  $\bar{l}$  естественно принимать среднюю длину тупиковых  $Q$ -тесторов).

Содержательная трактовка этого приема усматривается в том, что длина тупикового  $Q$ -тестора вскрывает концентрацию однообразия таблиц в каждом данном  $Q$ -тесторе, посредством которого осуществляется связывание всех строк внутри исследуемых классов объектов. Очевидно, что чем меньше столбцов затрачивается на построение данного тупикового  $Q$ -тестора, тем большая эффективность составляющих его столбцов в обнаружении однообразия классов и тем больше информационная однородность  $Q$ -тестора. Естественно формулировать, что самые сильные связывающие столбцы образуют самые короткие тупиковые  $Q$ -тесторы, а совокупность столбцов, вошедших во все  $Q$ -тесторы минимальной длины, образуют критерий однообразия объектов внутри двух исследуемых классов. Признаки, составляющие критерий однообразия двух классов, минимизируют пространство признаков в исследовании однообразия классов при их сравнительном изучении.

### Совместное изучение различий и сходств

Необходимость поиска согласованных оценок характеристических признаков определяется полезностью знания удельного содержания различия и отождествления для каждого признака из заданного пространства. Кратко рассмотрим процедуры при внутритабличном совместном исследовании сходств и различий. Пусть имеется совокупность объектов, объединенных в один класс, заданный таблицей бинарных символов  $T$ . Предположим, что класс однороден, но индивидуализирован (т.е. каждый объект его может рассматриваться вне связи с другими объектами класса) по значениям целевого признака  $x_{n+1}$ . Охарактеризовать признаки согласованной (равновесной) оценкой как в отношении коллективного взаимодействия объектов (однородность), так и в отношении индивидуальных (независимых) свойств можно следующим путем.

Набор столбцов  $\{t_{i_1}, t_{i_2}, \dots, t_{i_j}\}$  таблицы  $T$  называется  $H$ -тестом, если в таблице  $\{t_{i_1}, t_{i_2}, \dots, t_{i_j}\}$  все строки как попарно сходны, так и

попарно различны.  $H$ -тест называется тупиковым, если удалением любого столбца он переводится в набор, не являющийся  $H$ -тестом.

Если  $K^H$  — числа тупиковых  $H$ -тестов таблицы  $T$ , а  $K_i^H$  — числа тупиковых  $H$ -тестов с участием  $t_i$ ,  $i = 1, 2, \dots, n$ , то необходимость столбца  $t_i$  для отражения совместно однородности и индивидуальности строк из  $T$  равна

$$H_i = \frac{K_i^H}{K^H} \quad (9)$$

Оценка строк осуществляется следующим образом.  $H$ -информационный вес строки  $t^j$  задается формулой

$$H_*(j) = \sum_{i=1}^n t_i^j H_i \quad (10)$$

чем выше  $H_*(j)$ , тем значимее строка  $t^j$  в последовательности  $t^1, t^2, \dots, t^m$ .  $H$ -классификационный вес строки  $t^j$  получаем из выражения

$$H^*(j) = \sum_{i=1}^n \alpha_i^j H_i$$

где  $\alpha_i^j = t_i^j P_i + (1 - t_i^j)(1 - P_i)$ ; чем выше значение  $H^*(j)$ , тем уместнее строка  $t^j$  в последовательности  $t^1, t^2, \dots, t^m$ . И, наконец, оценка строки  $t^j$

$$H(j) = H^*(j) + \sum_{i=1}^k t_i^j P_i H_i \quad (11)$$

выражает согласованность (равновесность) строки  $t^j$  в ряду  $t^1, t^2, \dots, t^m$ .

Совместное изучение сходства — различия для нескольких таблиц (нескольких классов объектов) базируется на расширении понятия тестов в идейном и процедурном отношениях. Рассмотрим случай для двух классов с двузначным целевым признаком  $x_{n+1}$ . Для согласования характеристик объектов  $S_j$  с целевым признаком  $x_{n+1}$  требуется изучить: а) отождествление объектов внутри каждого класса и б) разделение объектов разных классов.

Пусть совокупности классов поставлена в соответствие таблица  $T$  из  $m$  строк  $t^j$ , кодирующих объекты  $S_j$ , и  $n$  столбцов  $t_i$  кодированных признаков  $x_i$ . Строки  $t^1, t^2$  из  $T$  различаются, если существует стол-

бец  $t_i$ ,  $1 \leq i \leq n$ , такой, что  $t_i^1 \neq t_i^2$ ; строки  $t^1, t^2$  отождествляются, если существует столбец  $t_i$ ,  $1 \leq i \leq n$ , такой, что  $t_i^1 = t_i^2$ .

Набор  $(i_1, i_2, \dots, i_l)$  столбцов таблицы  $T$  называется классифицирующим  $K$ -тестом, если при удалении из  $T$  всех столбцов, кроме

$t_{i_1}, t_{i_2}, \dots, t_{i_l}$  в полученной таблице для любой пары строк  $t^{j_1} \in T_{J_1}, t^{j_2} \in T_{J_2}$  выполнены следующие отношения строк  $t^{j_1}, t^{j_2}$ : различаются, если  $J_1 \neq J_2$ ; отождествляются, если  $J_1 = J_2$ .

$K$ -тестор называется тупиковым, если удалением любого столбца он переводится в набор, не являющийся  $K$ -тестором.

Поскольку для  $K$ -тестора условия а, б выполнены, то это понятие можно использовать в процедурах сравнительного изучения, связанных с сортировкой объектов. Построим множество  $\kappa$  всех тупиковых  $K$ -тесторов и найдем числа  $K^l$  всех элементов из  $\kappa$  длины  $l$ ,  $l \geq 1$  и числа  $K_i^l$  всех элементов из  $\kappa$  длины  $l$  с участием столбца  $t_i$ ,  $i = 1, 2, \dots, n$ . Величина  $K_i^l / K^l$  называется классифицирующим весом  $l$ -го порядка столбца  $t_i$ , а величину

$$\tilde{P}_i^k = \sum_{l \geq 1} K_i^l / \sum_{l=1} K^l \quad (12)$$

назовем классифицирующим весом столбца  $t_i$ . Эти величины оценивают существенность каждого признака в способности различать классы и сжимать объекты внутри каждого класса.

### Процедуры распознавания

Сортировка объектов по заданным классам с помощью процедур распознавания может осуществляться методом вычисления расстояний. В этом методе исходная информация представляется числовой матрицей

$$\{t_{ij}\} = T, \quad \begin{cases} i = 1, \dots, n, \\ j = 1, \dots, m, \end{cases}$$

где каждому классу эталонных объектов соответствует фиксированный набор строк  $T_K = S_j^k = \{t_{ij}^k\}$ . Рассмотрим простой случай для двух

классов эталонных объектов

$$T = \begin{cases} T_I = \{S_{jI}\} = \{t_{ijI}\}, & j^I = 1, \dots, r, \\ T_{II} = \{S_{jII}\} = \{t_{ijII}\}, & j^{II} = r+1, r+2, \dots, m, \end{cases}$$

и класса объектов-проб:  $T_n = \{S_{jn}\} = \{t_{ijn}\}$ ,  $j^n = 1, \dots, m^n$ . Из значений

информационных весов  $\{P_i\}$  путем нормировки:  $R_i = P_i / \sum_{i=1}^n P_i$  ( $\sum_{i=1}^n R_i = 1$  —

нормировочное условие) получают нормированные значения информационных весов признаков  $\{R_i\}$ . Для любой пары объектов, один из которых проба, а другой — эталон  $(S_j^n, S_j^k)$ , вычисляется нормированное рас-

стояние (мера различия) —  $d_i^{(j^n, j^k)}$ ,  $0 \leq d_i^{(j^n, j^k)} \leq 1$  по каждому признаку.

Далее, при заданных  $\{R_i\}$  для каждой пары проба — эталон вычисляется нормированное расстояние по всей системе признаков

$$D^{(j^n, j^k)} = \sum_{i=1}^n R_i d_i^{(j^n, j^k)} \quad 0 \leq D^{(j^n, j^k)} \leq 1.$$

Решение об отнесении пробы  $S_j^n$  к тому или иному классу эталонных объектов  $\{S_j^I\}$ ,  $\{S_j^{II}\}$  принимается по величине разности минимальных расстояний до каждого из классов (до ближайших объектов в каждом классе)

$$\Delta_j^n = D_{\min}^{(j^n, j^I)} - D_{\min}^{(j^n, j^{II})} \quad (13)$$

или по разности средних расстояний до классов

$$\bar{\Delta}_j^n = \bar{D}_{II}^{j^n} - \bar{D}_I^{j^n}, \quad \text{где} \quad \begin{cases} \bar{D}_I^{j^n} = \frac{1}{r} \sum_{j^I=1}^r D^{(j^n, j^I)}, \\ \bar{D}_{II}^{j^n} = \frac{1}{r} \sum_{j^{II}=1}^r D^{(j^n, j^{II})}. \end{cases}$$

Если проба ближе к I классу, то величина  $\Delta_j^n$  имеет положительный знак, в другом случае — отрицательный. По абсолютному значению этой величины оценивается достоверность (значимость) распознавания. Случай, когда  $\Delta_j^n = 0$ , означает полную неопределенность в отнесении пробы к тому или иному классу, поэтому обычно задается зона значимого

распознавания:  $|\Delta_j^n| \geq \epsilon$ ,  $|\Delta_j^n| < \epsilon$  — зона неопределенности, где  $\epsilon$  — соответствующий порог. Процедура распознавания, использующая в качестве проб сами эталонные объекты, называется внутренним экзаменом, который проводится для выяснения качества разделения объектов на классы при заданных информационных весах признаков. Очевидно, что качество внутреннего экзамена зависит от значений  $\{R_i\}$  (при одной и той же исходной системе признаков), что позволяет оценить информативную систему признаков, полученную при решении задачи классификации признаков. Следует отметить, что в методе "вычисле-

ние расстояний" обоснованным является использование информационных весов таких признаков, у которых имеется логическая взаимосвязь с процессами разделения (в смысле расстояний) в пространстве признаков объектов разных классов. Под внешним экзаменом понимается процедура распознавания, при которой в качестве проб выступают объекты с известным значением целевого признака, но не включенные в эталонную выборку при поиске информативной системы признаков. Качество внешнего экзамена прежде всего оценивает представительность эталонной выборки, т.е. распознавание ведется с точностью до материала обучения.

### Метод пакетов

Рассмотрим процедуры внутриклассового упорядочивания и распознавания объектов по методу пакетов. Пусть исследуемые объекты класса I  $S_1, S_2, \dots, S_m$  представлены  $n$  бинарными характеристическими признаками  $x_1, \dots, x_n$ , а их описания сведены в таблицу  $T(m \times n)$ .

Набор столбцов (признаков)  $x_{i_1}, \dots, x_{i_k}$  называется пакетом, если в таблице, полученной из  $T$  удалением всех столбцов, кроме  $x_{i_1}, \dots, x_{i_k}$  в каждой строке содержится хотя бы одна единица.

Пакет, из которого нельзя удалить ни одного столбца без того, чтобы в полученной таблице не образовалась нулевая строка, называется тупиковым пакетом. Тупиковый пакет — это составной характеристический признак (булева функция)  $x_{i_1} \cup \dots \cup x_{i_k}$ , выполняющий-

ся для всех объектов  $S_1, \dots, S_m$  и избыточный относительно совокупности составляющих его признаков. Эта совокупность признаков составляет часть характеристики исследуемых объектов, а набор всех тупиковых пакетов образует компактную характеристику всего класса объектов, представленных таблицей бинарных символов. Наряду с пакетами таблицы  $T$  рассматриваются пакеты таблицы  $\bar{T}$ , составленной из описаний объектов  $S_1, \dots, S_m$  в пространстве признаков  $\bar{x}_1, \dots, \bar{x}_n$ , где  $\bar{x}_i$  — логическое отрицание признака  $x_i$ . Тупиковый пакет таблицы  $\bar{T}$  называется 0 — пакетом (нуль-пакетом) таблицы  $T$ .

Введем на основании понятия "тупиковый пакет" величины, допускающие процедуры диагностики (сортировку) объектов. Пусть  $I \subset R$ , где  $R$  — более широкий, чем  $I$ , класс объектов. Например,  $R$  — класс оловородных месторождений,  $I$  — класс крупнейших оловородных месторождений. Пусть, далее, объект  $S \in R$ , а  $F(T)$  — общее число тупиковых пакетов таблицы  $T$ . Определим  $F(S, T)$  — число тупиковых пакетов  $T$ , которые выполняются как составные характеристические признаки (булевы функции) для объекта  $S$ . Положим

$$B(S) = \frac{F(S, T)}{F(T)}, \quad \bar{B}(S) = \frac{F(S, \bar{T})}{F(\bar{T})}.$$

Обозначив  $F_i(T)$  числа тушиковых пакетов  $T$  с участием  $i$ -го столбца, получим такие величины

$$A = \frac{F_i(\bar{T})}{F(T)}. \quad (14)$$

Величины  $B(S)$ ,  $\bar{B}(S)$ ,  $A_i$ ,  $\bar{A}_i$  являются диагностическими величинами. Покажем процедуры, предназначенные для сортировки и классификации объектов.

А. Голосование по тушиковым пакетам. Пусть  $T_1$  — эталоны класса  $I$ ,  $T_2$  — эталоны класса  $R/I$ ,  $T_3$  — объекты экзамена класса  $I$ . Объекты, подлежащие сортировке на их принадлежность к классу, представлены таблицей  $T_4$ . Определим  $\rho^+ = \min_{S \in T_3} B(S)$ ,  $\rho^- = \max_{S \in T_3} B(S)$  и потребуем, чтобы  $\rho^+ > \rho^-$ .

Для  $S$  из  $T_4$  вычисляется  $B(S)$ , причем если  $B(S) > \rho^+$ , то  $S$  относится к  $I$ ; если  $B(S) \leq \rho^-$ , то  $S$  относится к  $R/I$ . При  $\rho^+ > B(S) > \rho^-$  сортировка не производится. Аналогично процедуре голосования по тушиковым тестам производится голосование по  $\theta$ -пакетам (нуль-пакетам).

Б. Голосование с проверкой на аномальность пробы. Исходное состояние подобно алгоритму а, с той лишь разницей, что при этом не требуется, чтобы  $\rho^+ > \rho^-$ . Для объектов таблиц  $T_2, T_3, T_4$  вычисляются величины  $B(S)$  и  $\bar{B}(S)$ . Требуется, чтобы ни для какой пары объектов  $S' \in T_2$  и  $S'' \in T_3$  не выполнялось одновременно  $B(S') \geq B(S'')$  и  $\bar{B}(S') \geq \bar{B}(S'')$ . Далее, если для строки  $S$  найдется строка  $S'' \in T_3$ , такая, что  $B(S) \geq B(S'')$  и  $\bar{B}(S) \geq \bar{B}(S'')$ , то  $S$  относится к  $I$ . Если для нее найдется строка  $S' \in T_2$ , то  $B(S) \leq B(S')$  и  $\bar{B}(S) \leq \bar{B}(S')$ , то  $S$  относится к  $R/I$ . При отсутствии таких строк в испытываемой таблице сортировка не производится. Алгоритм б имеет ряд модификаций, позволяющих в некоторых случаях повысить надежность распознавания и упростить систему процедур.

В основном вся совокупность процедур, нацеленных на исследование раздельных и согласованных свойств объектов в логических понятиях сходство и различие, призвана обеспечить математическую сторону сравнительного изучения в широком применении к задачам рудопргноза. Двойственная природа сообщений сходство — различие указывает на способность процедур схватывать изолирующее свойство некоторых признаков по отношению к другим классам объектов и стягивающее свойство этих же признаков для компоновки объектов в данный класс. Эти признаки как бы снимают противоречие пары сходство — различие и, таким образом, превосходят значение отождествляющих и разделяющих признаков.

### Практические применения

Практическое применение тестового подхода, особенно в области исследования различий объектов, интенсивно расширяется (Дмитриев и др., 1968<sub>2,3</sub>; Бугаец и др., 1969; Вышемирский и др., 1971; Гольдман и др., 1974). Как правило, это применение связывается с вопросами

прогнозирования поисков руд. Прогнозируются три типа объектов: рудные районы, месторождения, участки месторождений. Если позволяет детальность и точность информации, возможно прогнозирование не только локализации месторождений, но и вероятных масштабов оруденения (Трофимук и др., 1971; Яблонский и др., 1971; Константинов, 1973).

Кроме чисто практических результатов тестовый подход в соответствии с целеуказанием позволяет получить теоретические результаты, касающиеся классификации объектов, их генезиса (при соответствующей постановке задач), проверки гипотез, формулирования новых гипотез на базе большого объема обработанной и логически переосмысленной информации. В качестве примера практического применения тестового подхода приводится более подробное рассмотрение одной из задач.

## ЗАДАЧА ОЦЕНКИ ПЕРСПЕКТИВ РУДОНОСНОСТИ ТРАПОВЫХ ИНТРУЗИЙ СЕВЕРА СИБИРСКОЙ ПЛАТФОРМЫ

### Постановка задачи

Под постановкой задачи понимается процесс ее формирования по следующим этапам.

1. Формулирование задачи на геологическом уровне.
2. Преобразование геологической задачи по возможности в формализованную постановку.
3. Построение схемы решения и выбор решающих правил.

Постановка геологической задачи производится формулированием цели исследования: "Оценка перспектив рудоносности трапových интрузий севера Сибирской платформы". Эта постановка не содержит, однако, указаний на конкретный предмет изучения и средства оценки; поэтому необходима детализация темы "Логико-математическая обработка геолого-петрологической информации по дифференцированным траповым интрузиям севера Сибирской платформы в связи с оценкой перспектив их рудоносности".

В таком виде формулировка считается удовлетворительной, она содержит необходимые для задач этого профиля элементы: цель (исследование рудоносности), предмет исследования (геолого-петрологическая информация), средства исследования (логико-математическая обработка информации). Информационное обеспечение задачи состоит из 62 объектов по 40 признакам. Формализованная постановка осуществляется после выяснения природы и структуры информации.

### Объекты исследования

Объектами исследования в данной задаче служат дифференцированные траповые интрузии. Интрузии обнаруживают как сходство, так и различие в строении, петрологии, минеральном составе и т.д. Внутреннее строение их, как правило, трехчленное: безоливиновые, оливинсодержа-



Таблица 1

Исходная информация

Интрузия	Признаки и их номера										
	геолого-тектонические										
	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	0	0	1	0	0
2	1	1	1	1	1	0	0	1	1	0	0
3	1	1	1	1	1	1	1	1	0	0	1
4	1	1	1	1	1	1	1	1	0	0	1
5	1	1	1	1	1	1	1	0	0	1	1
6	1	1	1	1	1	1	0	0	1	0	0
7	1	1	0	1	1	1	0	1	0	0	0
8	1	1	0	1	1	1	0	0	1	0	0
9	1	0	0	1	1	1	1	0	0	1	1
10	1	1	0	1	1	1	0	0	1	0	0
11	1	1	1	1	1	1	0	0	1	0	0
12	1	1	0	0	1	0	0	0	1	0	0
13	1	1	0	1	0	1	0	1	0	0	1
14	1	1	1	0	0	1	0	1	0	0	1
15	1	0	0	0	0	0	1	0	0	0	0
16	1	1	0	0	0	1	0	1	0	0	1
17	1	1	1	0	0	1	0	1	0	0	1
18	1	1	0	1	1	1	1	0	0	1	0
19	1	0	0	0	0	0	1	0	0	0	1
20	1	1	1	0	0	0	0	0	1	0	0
21	1	1	0	0	0	1	1	0	0	1	1
22	1	1	1	0	0	0	0	0	1	0	0
23	1	0	0	0	0	0	1	0	0	1	0
24	1	1	0	1	1	1	0	1	0	0	0
25	1	0	0	0	0	0	0	0	1	0	0

Признаки и их номера													
магматические							минералогические						
12	13	14	15	16	17	18	19	20	21	22	23	24	
1	1	1	1	1	1	1	1	20	48	1	96	1	
1	1	1	1	1	1	1	1	20	30	1	90	1	
1	1	1	1	1	1	1	1	22	40	1	90	1	
1	1	1	1	1	1	1	1	23	39	1	95	1	
0	1	1	1	1	1	1	1	17	28	1	90	1	
1	1	1	1	1	1	1	0	22	40	1	86	1	
1	1	1	0	1	1	1	0	16	30	1	88	1	
1	1	1	0	1	1	1	0	17	34	1	90	1	
1	1	1	1	1	0	1	1	19	35	1	71	1	
1	1	1	0	1	1	1	0	25	31	1	80	1	
1	1	1	1	1	1	1	1	16	28	1	79	1	
1	0	1	0	1	1	1	0	-	-	1	-	1	
0	0	1	0	0	1	0	0	33	58	1	70	0	
0	0	1	0	0	1	1	0	39	42	1	71	1	
1	0	0	1	0	0	0	0	41	53	0	74	1	
0	0	1	0	0	1	0	0	33	59	1	70	1	
0	0	1	0	0	1	1	1	45	50	1	50	1	
0	1	0	1	1	1	1	0	10	28	1	90	1	
0	0	0	0	0	1	1	0	36	50	0	62	0	
1	0	1	1	1	0	0	1	26	40	1	78	1	
0	1	1	0	1	0	0	0	31	48	0	67	1	
1	0	1	0	0	0	0	0	41	63	1	70	1	
0	1	1	0	1	1	0	0	30	48	0	65	1	
1	0	0	0	0	1	0	1	28	40	0	69	1	
1	1	1	0	1	0	0	0	32	42	1	65	1	

Таблица 1 (продолжение)

Инту- зия	Признаки и их номера								
	минералогические					петрохимические			
	25	26	27	28	29	30	31	32	33
1	1	1	1	1	0	12,48	31,1	2,28	21,9
2	1	1	1	1	0	10,09	36,9	2,97	25,5
3	1	1	1	1	0	11,30	22,5	2,70	22,5
4	1	1	1	1	0	10,84	20,9	2,93	20,9
5	1	1	1	1	0	10,24	36,9	3,69	18,2
6	1	1	1	0	0	12,60	60,2	3,92	22,4
7	1	1	1	0	0	11,66	32,2	2,33	20,6
8	1	1	1	0	0	9,53	35,2	1,92	25,9
9	1	1	1	0	0	9,69	36,2	2,38	14,0
10	1	1	1	0	0	12,18	38,9	3,70	33,3
11	1	1	1	0	0	11,68	31,1	1,73	16,0
12	1	1	1	0	0	10,36	36,5	1,32	27,8
13	1	1	1	1	0	6,68	49,4	3,16	23,9
14	1	1	1	0	0	7,14	44,1	2,78	17,1
15	1	1	1	0	0	7,21	43,0	2,78	14,6
16	1	1	1	1	0	7,44	47,4	2,72	12,5
17	1	1	1	0	1	4,38	57,4	3,90	14,0
18	1	1	1	0	0	7,80	39,9	2,70	13,0
19	1	1	1	0	0	4,67	56,6	3,58	9,5
20	1	1	1	0	1	9,90	38,9	2,18	12,1
21	1	1	1	1	0	10,81	37,8	2,21	12,1
22	1	1	1	0	0	4,66	60,9	3,55	23,5
23	1	1	1	0	0	9,56	44,9	2,63	15,4
24	1	1	1	0	0	8,60	44,3	2,70	14,6
25	1	1	1	0	0	9,56	41,6	2,75	14,6

Признаки и их номера						
петрохимические						
34	35	36	37	38	39	40
731	937	193,5	24,16	75,6	0,92	27,9
736	923	194,3	23,81	76,5	0,96	26,6
731	917	194,5	21,30	75,00	1,03	1,49
726	921	194,0	20,18	69,5	1,75	17,1
800	909	194,4	19,27	73,8	1,37	17,5
792	905	197,9	21,04	75,4	2,96	14,9
752	938	194,1	22,87	75,1	1,64	20,9
733	949	195,2	16,69	72,6	1,42	30,6
766	936	197,3	22,20	76,1	1,31	25,0
776	910	195,4	20,59	71,7	2,59	20,6
763	946	194,3	21,10	76,9	1,26	13,4
772	967	194,5	19,77	74,3	0,72	26,5
839	919	197,1	-	56,0	-	-
827	932	197,5	9,48	39,6	2,40	-
811	931	196,5	8,30	59,9	2,69	7,3
800	929	196,5	-	35,0	-	-
849	908	199,1	8,38	58,0	2,64	18,3
797	932	196,8	19,30	75,6	1,33	33,7
837	913	198,6	7,49	55,7	2,81	32,0
771	943	195,3	17,42	69,5	1,74	18,3
762	940	194,5	14,22	67,3	1,57	29,8
812	915	197,8	7,97	59,4	2,79	22,0
778	930	190,8	18,87	67,3	1,79	16,9
776	931	195,3	9,37	33,9	2,89	9,0
787	926	195,2	16,57	64,2	2,10	7,0

Таблица 1 (продолжение)

Интрузия	Признаки и их номера										
	геолого-тектонические										
	1	2	3	4	5	6	7	8	9	10	11
26	1	0	0	0	0	0	0	0	1	0	0
27	1	0	0	0	0	1	1	0	0	1	1
28	1	1	1	0	0	1	0	1	0	0	1
29	1	1	0	0	0	0	1	0	0	0	1
30	1	1	0	1	1	1	0	0	1	0	0
31	1	0	0	0	0	0	0	0	1	0	0
32	1	1	0	1	1	1	1	0	0	1	0
33	1	1	1	1	1	1	0	0	1	0	0
34	1	1	0	1	1	0	0	0	1	0	0
35	1	1	0	0	0	0	1	1	0	0	1
36	1	0	0	0	0	0	1	1	0	0	1
37	1	0	0	0	0	0	1	1	0	0	1
38	1	1	0	0	0	1	1	0	0	1	1
39	1	0	0	0	0	0	0	0	1	0	0
40	1	1	1	0	0	0	0	0	1	0	0
41	1	0	0	0	0	0	0	1	0	0	0
42	1	1	0	0	0	0	1	0	0	0	1
43	1	1	0	0	0	0	1	1	0	0	0
44	1	1	1	1	1	1	0	0	1	0	0
45	1	1	0	0	0	0	0	1	1	0	1
46	1	1	0	0	0	0	0	1	0	0	0
47	1	0	0	0	0	1	1	0	0	0	0
48	1	1	1	0	0	0	0	1	0	0	1
49	1	1	1	0	0	0	1	1	0	0	1
50	1	0	0	0	0	0	1	1	0	0	0
51	1	1	1	0	0	0	1	0	0	1	0
52	1	1	0	0	0	0	0	0	1	0	0

Признаки и их номера												
магматические								минералогические				
12	13	14	15	16	17	18	19	20	21	22	23	24
1	0	1	0	1	0	0	0	33	39	1	84	1
0	1	1	0	1	1	0	0	33	59	1	70	1
1	0	0	1	0	1	1	0	32	47	1	81	1
0	1	0	0	1	1	0	0	28	49	1	88	1
1	0	0	0	0	1	0	0	46	-	0	70	0
1	0	0	0	0	1	0	0	41	42	1	72	1
0	0	0	0	1	0	0	0	5	19	1	68	1
1	1	1	0	1	1	1	1	23	33	1	72	1
1	0	0	1	0	0	0	0	33	60	1	67	1
0	0	0	0	0	0	0	0	35	40	0	71	1
0	0	1	1	0	0	0	0	36	50	1	67	1
0	0	1	1	1	0	0	0	31	42	1	67	1
0	0	0	0	0	0	0	0	28	61	0	67	0
1	0	0	0	0	0	0	0	40	61	0	65	0
1	0	0	0	0	1	0	0	25	70	1	87	1
1	1	0	0	0	0	0	0	28	51	0	67	0
1	0	0	0	0	0	1	0	46	49	1	64	1
1	0	0	0	0	1	0	1	21	76	0	76	1
1	0	0	1	1	1	1	1	33	35	0	63	1
1	1	0	0	1	0	1	0	24	30	1	75	1
1	1	0	1	0	0	0	0	46	58	1	68	1
1	0	0	0	0	0	0	1	40	59	1	70	1
0	0	0	0	0	0	0	0	21	80	0	70	1
0	0	0	0	0	0	0	0	32	80	0	72	1
1	0	0	0	0	1	0	0	26	66	0	85	1
0	0	0	0	0	0	1	1	41	-	1	62	1
1	1	0	0	0	0	0	0	42	80	1	64	1

Таблица 1 (продолжение)

Инту- зия	Признаки и их номера								
	минералогические					петрохимические			
	25	26	27	28	29	30	31	32	33
26	1	1	1	0	0	7,67	47,8	3,06	13,3
27	1	1	1	0	1	6,20	50,3	3,60	16,6
28	1	1	1	1	0	7,41	46,0	3,52	19,2
29	1	1	0	0	0	8,77	41,8	2,55	15,8
30	1	1	1	0	0	6,25	53,3	3,33	14,3
31	1	1	1	0	0	7,02	49,9	2,93	13,5
32	1	1	1	0	0	12,71	35,8	1,79	13,4
33	1	1	1	0	1	9,10	38,9	3,70	33,3
34	1	1	1	0	0	7,71	39,4	3,63	9,0
35	1	1	1	0	1	5,41	52,3	4,21	19,0
36	1	1	1	0	0	9,63	36,8	2,41	17,1
37	1	1	1	0	0	8,53	42,4	2,54	15,3
38	1	1	1	0	0	4,96	52,3	2,71	1,1
39	1	0	0	0	0	6,81	46,7	2,35	17,1
40	1	1	1	0	-	6,41	46,3	2,72	20,0
41	1	1	1	0	0	5,21	51,1	3,21	14,9
42	1	1	1	0	1	6,78	49,1	3,03	18,0
43	1	0	0	0	0	4,67	57,8	3,06	10,9
44	1	1	1	0	1	8,37	36,0	2,25	20,3
45	1	1	0	1	0	8,61	42,7	2,63	15,8
46	1	1	1	0	0	3,65	64,8	3,19	20,0
47	1	1	0	0	0	7,45	44,9	2,61	13,2
48	1	0	0	0	1	7,85	42,6	2,55	10,5
49	1	0	0	0	1	5,71	53,0	3,24	18,4
50	1	0	0	0	0	7,23	51,1	2,66	12,5
51	1	0	0	0	1	8,08	44,0	2,76	9,3
52	1	0	0	0	0	4,94	59,1	4,21	21,6

Признаки и их номера						
петрохимические						
34	35	36	37	38	39	40
790	921	196,5	15,56	65,8	2,15	18,9
820	913	197,1	12,20	63,6	2,70	15,9
805	913	196,8	7,96	46,0	3,31	3,8
777	933	195,3	15,70	70,4	1,62	11,2
789	917	196,9	8,44	56,3	1,65	10,2
787	929	196,6	7,02	49,8	2,95	-
735	951	193,4	20,29	64,3	1,79	33,7
776	910	195,4	13,54	71,7	2,59	18,0
758	902	192,8	7,81	59,0	4,99	10,0
792	895	197,8	7,48	51,4	2,69	-
790	964	195,3	11,15	61,0	2,17	3,5
795	962	196,0	9,75	63,5	2,21	9,3
834	950	199,3	5,37	48,7	2,38	4,2
822	931	197,3	7,04	51,1	3,18	-
813	934	198,1	7,73	65,9	1,45	9,9
839	923	197,6	6,93	62,2	2,45	37,0
819	927	197,2	8,07	54,3	2,46	25,6
834	923	198,7	-	-	-	-
774	944	196,2	11,18	67,4	2,01	22,9
800	935	196,1	8,68	39,0	2,26	5,5
817	923	197,4	4,90	54,6	2,50	6,54
802	934	197,0	8,31	51,2	2,43	1,3
784	936	196,1	13,11	64,3	1,93	7,4
796	920	197,5	11,01	73,6	2,82	4,8
782	929	196,4	10,28	60,8	1,20	13,5
779	925	195,1	7,20	59,3	3,11	12,7
813	914	198,7	7,65	55,5	2,60	12,08

Таблица 1 (продолжение)

Интрузия	Признаки и их номера										
	геолого-тектонические										
	1	2	3	4	5	6	7	8	9	10	11
53	1	1	1	0	0	0	1	1	0	0	0
54	1	0	0	1	1	1	0	1	0	0	1
55	1	1	1	1	0	1	0	1	0	0	1
56	1	1	0	0	0	0	0	1	0	0	1
57	1	0	0	0	0	0	0	1	0	0	0
58	1	0	0	0	0	0	0	1	0	0	1
59	1	0	0	0	0	0	0	1	0	0	1
60	1	1	0	0	0	0	0	1	0	0	1
61	1	1	0	1	1	1	0	0	1	0	0
62	1	1	1	1	0	1	1	0	1	0	1

Таблица 1 (окончание)

Интрузия	Признаки и их номера									
	минералогические					петрохимические				
	25	26	27	28	29	30	31	32	33	
53	1	0	0	0	0	6,80	51,1	3,38	11,8	
54	1	1	1	0	0	18,90	25,9	1,64	16,0	
55	1	1	1	0	0	8,57	36,9	3,69	18,2	
56	1	1	0	0	0	5,11	60,2	3,92	22,4	
57	1	1	1	0	0	4,98	54,9	3,14	13,0	
58	1	1	0	0	0	5,62	57,9	2,90	13,9	
59	0	0	0	0	0	6,81	46,8	3,25	7,2	
50	1	0	0	0	0	6,56	51,0	3,26	14,3	
61	1	1	1	0	0	6,97	45,6	3,25	18,8	
62	1	1	1	1	0	11,72	28,1	1,75	12,0	

	Признаки и их номера												
	магматические								минералогические				
	12	13	14	15	16	17	18	19	20	21	22	23	24
	1	0	0	0	0	1	0	1	21	-	0	76	1
	0	1	0	0	1	0	0	0	16	25	1	75	0
	0	1	0	0	1	1	0	1	43	46	1	76	1
	0	0	1	0	0	1	0	0	51	90	1	67	0
	1	0	0	0	0	0	0	0	39	46	0	69	0
	0	0	0	0	0	1	0	0	30	65	1	70	0
	0	0	0	0	0	0	0	0	30	60	1	61	1
	0	0	0	0	0	0	0	0	30	60	1	63	1
	1	0	1	0	1	0	0	1	26	43	1	90	1
	1	1	1	1	1	1	1	1	20	42	1	82	1

	Признаки и их номера						
	петрохимические						
	34	35	36	37	38	39	40
	794	915	196,8	8,04	77,0	1,89	-
	752	950	191,7	25,8	75,4	1,14	24,7
	800	909	194,4	9,51	45,0	2,74	0,01
	792	905	197,9	6,30	47,2	3,40	33,0
	834	924	198,8	5,05	47,2	2,91	7,45
	815	927	198,0	5,09	36,7	2,96	10,6
	791	914	196,2	2,91	24,8	4,04	0,6
	800	915	196,8	9,00	57,1	3,20	8,0
	771	919	198,0	9,45	54,4	3,76	8,13
	769	956	194,5	20,25	75,8	1,36	4,7

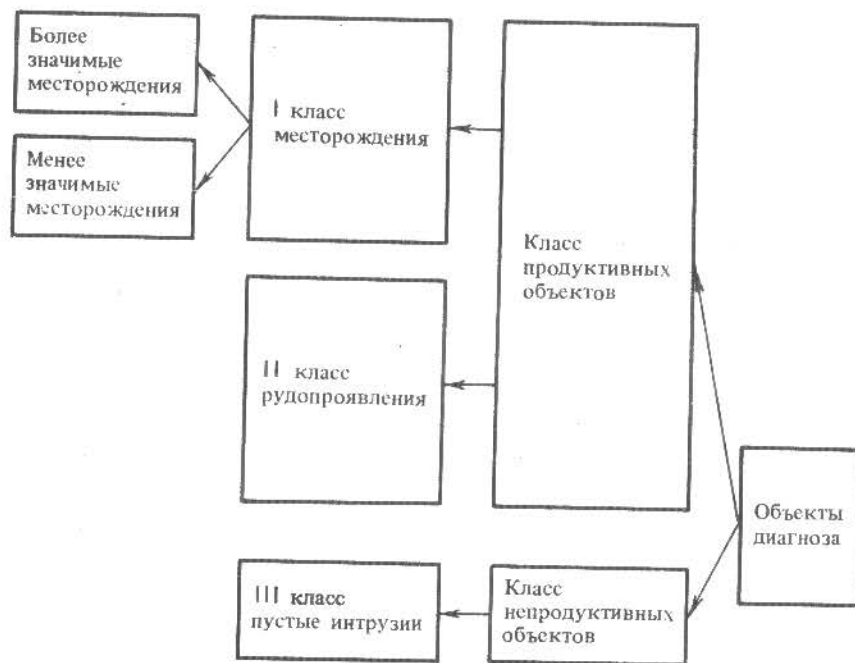


Рис. 1. Общая схема распознавания

щие и обогащенные оливином габбро-долериты. Изученность интрузий неравномерна. Эта неравномерность изученности сказывается на особенностях общего списка характеристических признаков объектов, подлежащих исследованию.

Профессиональный учет качеств объектов позволил выделить среди них три группы объектов, из которых были организованы три основных класса обучения: месторождения, рудопроявления и пустые интрузии. По сравнению с предыдущими исследованиями (Дмитриев и др., 1968<sub>2</sub>) количество рассматриваемых объектов увеличено с 27 до 62 за счет массивов северного и восточного бортов Тунгусской синеклизы, в то время как основная масса изученных интрузий сосредоточена в северо-западной части Сибирской платформы. Состав классов был определен по следующей схеме (рис. 1). Класс "месторождения" составлен из 10 объектов, класс "рудопроявления" – из 18, набор "пустые" интрузии состоит из 10 объектов. Класс "пробы", объединяющий объекты с неустановленной принадлежностью к первым трем классам, насчитывает 34 объекта.

В этот класс (пробы) на этапах обучения вводились также объекты, заведомо соответствующие одному из трех классов, экзамен для отработки системы распознавания. Распознавание производилось для объектов, нуждающихся в уточнении принадлежности к классу, и для объектов с неустановленной принадлежностью к основным классам обучения (табл. 1).

## Формирование пространства признаков

Организация пространства признаков начинается с регистрации эмпирически установленных фактов. Так, давно вскрыта связь сульфидного оруденения с дифференцированными трапповыми интрузиями, обладающими специфическими признаками: высокой степенью дифференцированности, присутствием пикритовых и такситовых габбро-долеритов и т.д.

Но эти же признаки могут быть характерны и для нерудных объектов. Поэтому за основной поисковый критерий принимался лишь признак – наличие сульфидного оруденения. Естественной была попытка установить комплекс косвенных поисковых признаков с целью прогнозировать скрытое оруденение, не обеспеченное регистрацией прямых признаков.

Однако такие комплексы, сформированные на хорошо изученных объектах, оказались непригодными на малоисследованных интрузиях: возникли своеобразные "ножницы", отсекавшие именно те объекты, которые подлежали первоочередной оценке. Во избежание подобного положения оказалось необходимым предусмотреть такой список признаков, который был бы общим для всех изученных интрузий.

Это требование общности ограничило первоначальное число признаков, но по мере продвижения решения задачи появлялась возможность для отдельных групп объектов расширять пространство признаков. На первых же шагах работы по сравнительному изучению интрузий (Дмитриев и др., 1974<sub>1,2</sub>) полученный список поисковых признаков косвенного характера имел практическое значение для ориентации поисков.

Окончательная организация пространства признаков (табл. 2), характеризующего исследуемые объекты, производилась с учетом общего уровня изученности для всех объектов. Оказалось целесообразным увеличить число признаков за счет большей детализации петрохимической и минералогической информации. В общем списке признаки подразделены следующим образом:

- I – геолого-тектоническая обстановка,
- II – признаки строения интрузий,
- III – минералогические,
- IV – петрохимические.

Количественно и качественно состав признакового пространства контролировался результатами предыдущих работ данного профиля, профессиональной интуицией, логическими посылками о взаимосвязи признаков, однозначностью формулировок признаков и их фактической основой.

## Формализованная постановка задачи

Формализация постановки задачи после исследования предоставленной информации и геологического целеуказания состоит в построении схемы конкретного решения и выбора математических процедур на каждом шаге решения. Особенностью предъявленной к обработке информации является короткий список признаков, а специфика

Таблица 2

Характеристические признаки траптовых интрузий севера Сибирской платформы

№ п/п	Признак	Информационные веса		
		1-й шаг	2-й шаг	3-й шаг
1	2	3	4	5
Геолого-тектонические				
1	Связь интрузий с глубинными разломами	-	-	-
2	Наличие метасоматитов в экзо- и эндоконтактных зонах	0,50	-	-
3	Наличие скарнов	0,38		Целевой
4	Присутствие дифференцированных лав во вмещающих толщах	0,26	0,38	Отождествляющий
5	Наличие пикритовых базальтов во вмещающих вулканогенных образованиях	0,26	0,38	То же
6	Наличие субшелочных траппов во вмещающих вулканогенных образованиях	-	0,03	"
7	Туфогенно-лавовая толща и интрузивные породы базальтоидного состава в нижнем экзоконтакте	-	0,24	"
8	Конгломераты, песчаники, сланцы в нижнем экзоконтакте	0,54	0,16	0,05
9	Карбонатно-глинистые породы (ангидрит, гипс) в нижнем экзоконтакте	0,27	-	0,18
10	Туфогенно-лавовая толща и интрузивные породы базальтоидного состава в верхнем экзоконтакте	-	0,19	-
11	Конгломераты, песчаники, сланцы в верхнем экзоконтакте	0,28	0,46	0,21
12	Карбонатно-глинистые породы (ангидрит, гипс) в верхнем экзоконтакте	-	0,30	0,28
	Средние значения весов	0,36	0,27	0,18

Таблица 2 (продолжение)

1	2	3	4	5
Магматические				
13	Характер дифференцированности (трех- или двучленное строение интрузии)	0,15	0,30	Отождествляющий
14	Наличие нижних такситовых габбро-долеритов	0,29	0,22	То же
15	Наличие верхних такситовых габбро-долеритов	-	0,44	0,90
16	Наличие пикритовых габбро-долеритов	0,20	-	Отождествляющий
17	Присутствие кислых гибридных пород в кровле интрузий	-	0,52	То же
18	Форма интрузии (хонолит или другая)	-	0,93	"
19	Наличие эруптивных брекчий в периферических частях интрузий	0,04	0,71	0,90
	Средние значения весов	0,18	0,52	0,90
Минералогические				
20	Железистость оливинов нижних горизонтов	0,20	0,93	0,44
21	Железистость оливинов верхних горизонтов	0,76	0,68	0,37
22	Присутствие ортопироксена в нижних дифференциатах интрузий	-	-	Отождествляющий
23	Основность плагиоклаза нижних горизонтов	-	0,76	1,27
24	Наличие биотита	0,57		Отождествляющий
25	Присутствие неправильных выделений титаномagnetита	-		То же
26	Наличие сульфидной вкрапленности в нижних горизонтах	0,46		"
27	Наличие сульфидной вкрапленности в верхних горизонтах	0,98		"
28	Проявление жильного оруденения	-	-	Целевой
29	Наличие скаполита	0,27	-	Отождествляющий
	Средние значения весов	0,46	0,79	0,70

Таблица 2 (продолжение)

1	2	3	4	5
Петрохимические				
30	Магнезиальность предполагаемого исходного расплава (по средневзвешенным или средним величинам, %)	0,18	0,22	-
31	Железистость предполагаемого исходного расплава $\frac{(\text{FeO} + \text{Fe}_2\text{O}_3)}{\text{FeO} + \text{Fe}_2\text{O}_3 + \text{MgO}} \times 100, \text{ мол. \%}$	0,30	-	0,37
32	Содержание щелочей в предполагаемом исходном расплаве, вес. %	0,81	-	0,65
33	Калиевость предполагаемого исходного расплава $\frac{\text{K}_2\text{O}}{\text{K}_2\text{O} + \text{Na}_2\text{O}} \times 100, \text{ мол. \%}$	0,30	0,55	0,23
34	Кремнекислотность предполагаемого исходного расплава (в атомных количествах $\text{SiO}_2$ )	0,31	0,38	0,46
35	Щелочность - кислотность предполагаемого исходного расплава $\frac{\text{SiO}_2 + \text{Al}_2\text{O}_3}{\text{Na}_2\text{O} + \text{K}_2\text{O} + \text{SiO}_2 + \text{Al}_2\text{O}_3} \times 100, \text{ ат. \%}$	0,50	-	0,65
36	Условный ионный потенциал предполагаемого исходного расплава	0,34	0,35	1,30
37	Магнезиальность наиболее богатых оливином горизонтов, вес. %	0,20	-	
38	Степень магнезиальности наиболее богатых оливином горизонтов $\frac{\text{MgO}}{\text{FeO} + \text{Fe}_2\text{O}_3 + \text{MgO}} \times 100, \text{ мол. \%}$	0,39	Целевой	0,37

Таблица 2 (окончание)

1	2	3	4	5
39	Содержание щелочей в наиболее богатых оливином горизонтах, вес. %	0,27	0,85	0,60
40	Степень дифференцированности интрузий К.ф. = $\left[ \frac{\text{FeO} + \text{Fe}_2\text{O}_3}{\text{FeO} + \text{Fe}_2\text{O}_3 + \text{MgO}} \times 100 \right]$ Верхний дифференциат $\left[ \frac{\text{FeO} + \text{Fe}_2\text{O}_3}{\text{FeO} + \text{Fe}_2\text{O}_3 + \text{MgO}} \times 100 \right]$ Нижний дифференциат	-	-	0,58
Средние значения весов		0,40	0,46	0,54

целеуказания состоит в многоцелевых мотивах обработки информации. Поэтому оказалось целесообразным в общей постановке задачи выделить три последовательных шага решения.

1-й шаг состоял в решении задачи наиболее общего вида: поиск информативной системы признаков для сортировки объектов-эталонов на продуктивные и пустые; на выбранной системе признаков необходимо было произвести распознавание объектов-проб на их принадлежность к одному из этих двух классов.

2-й шаг состоял в решении задачи поиска информативной системы признаков для сортировки объектов-эталонов класса "продуктивные" на классы месторождения и рудопроявления и в распознавании соответствующей принадлежности объектов-проб.

3-й шаг, соответствующий наиболее детальному уровню решения задачи, состоял в поиске информативной системы признаков для сортировки объектов-эталонов класса "месторождения" на классы более значимых и менее значимых и в распознавании объектов-проб на выделенных классах. Выбранные средства решения представляют систему тесторных процедур, опыт практического применения которой по существу еще мало известен. Каждый шаг решения задачи включает пять видов процедур (рис. 2):

а) подготовка исходной информации, ее предварительная обработка и кодирование (этот вид процедур в основном ручной и лишь процедуры кодирования имеют программное обеспечение);

б) первичная минимизация признакового пространства (в данной задаче необходимая лишь на первом шаге);

в) вычисление тесторных информационных весов (Дмитриев и др., 1966; Смертин, Дмитриев, 1970) и минимизация признакового простран-



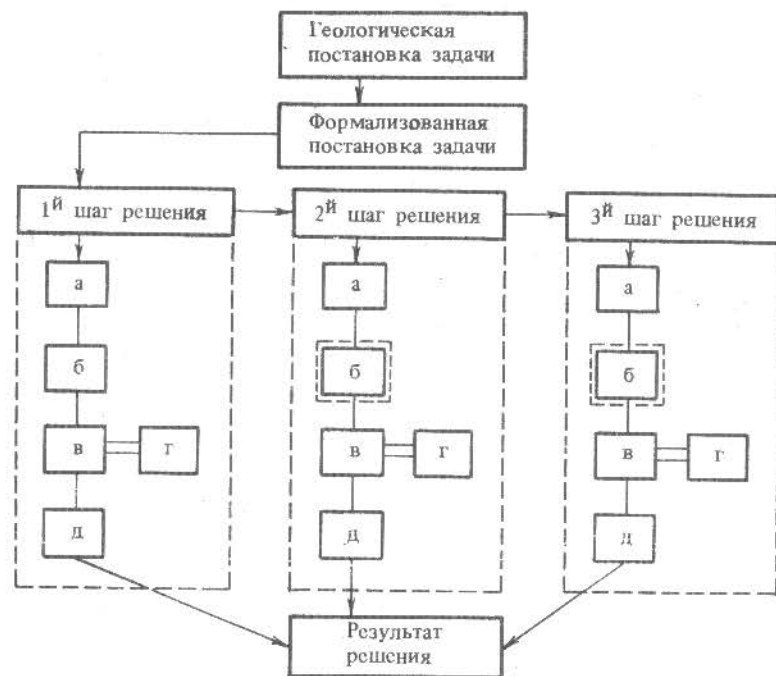


Рис. 2. Схема формализованной постановки задачи

ства с учетом длин тесторов (Дмитриев, 1970) – формирование информативной системы;

г, д) сортировка проб с помощью процедур распознавания на базе метода вычисления расстояний (Дмитриев, Смертин, 1970; Дмитриев и др., 1973).

### Первый шаг решения

На первом шаге решения задачи было отобрано для обучения 20 объектов из класса "продуктивные" и 10 объектов из класса "пустые", остальные 32 объекта были взяты в качестве проб. Из последних 4 представляют собой объекты с неизвестной принадлежностью к классу, а 28 принадлежат к продуктивным и взяты для внешнего экзамена (табл.3).

При подразделении эталонов на продуктивные и пустые для количественных признаков был использован метод целевого кодирования, т.е. произведен выбор разделяющей границы значений: значения, находящиеся по одну ее сторону, принимались за 1, а по другую – за 0. Целевое кодирование основано на принципе получения максимального числа различий для объектов разных классов и является исходной процедурой для подразделения классов (Бишаев, 1973; Бабич и др., 1974). Для качественных признаков выполнение обозначено 1, а невыполнение (отсутствие признака) – 0.

Целеуказание на данном шаге решения сводится к умению различать эталоны продуктивного класса от эталонов класса пустых объектов. Значения целевого признака бинарные, т.е. известно, какие объекты

из класса продуктивные, а какие пустые. Поиск информативной системы признаков представляет определенные трудности, поскольку целевой признак не несет информации о степени однородности объектов – эталонов внутри своего класса. Кроме того, целевой признак не формирует об интенсивности разграничения (расстоянии) между классами. Поэтому использовалось предположение о неоднородности объектов внутри каждого класса и о незначительном различии между классами. Поэтому же вопрос о хорошем разделении объектов между классами, вообще говоря, недостаточен точен, что в свою очередь влечет приближенность решения. Дальнейшее уточнение решения производится при переходе к последующим шагам решения поставленной геологической задачи.

На базе процедур учета длин тесторов при многотабличном исследовании произведено сокращение (минимизация) списка признаков, были изъяты признаки 6, 7, 10, 12, 15, 17, 18, 22, 23, 25, 28, 40 как слабые, мало влияющие на различение классов. На оставшихся признаках была выделена наиболее информативная система (процедура *в*, см. рис. 2), которая дает наилучший результат при внутреннем экзамене (процедуры *г*, см. рис. 2). Вслед за этим проводилось распознавание проб (процедуры *д*, рис. 2) и контрольных объектов известных классов. Объекты проб 12, 13, 14, 16, 30, 31, 53 имеют признаки с неизвестными значениями (прочерки), для них распознавание производилось дважды; в первом случае прочерки обозначались как 1, во втором – 0.

Результаты первого шага решения геологической задачи в вопросах минимизации признаков и сортировки объектов сводятся к следующему. На рис.3 приведено распределение признаков по информативности; признаки, которые были предварительно сокращены, имеют на этом рисунке нулевые значения информационных весов. Из остальных признаков составлена информативная система, в которой наиболее существенным из них соответствуют максимальные значения информационных весов. Значительная роль ряда признаков в распознавании очевидна с петрологической точки зрения (признаки 2, 3, 21, 24, 26, 27, 32, 39, табл.2).

На внутреннем экзамене по информативной системе получено полное разделение объектов обучения; область между ближайшими из них, принадлежащими к разным классам, является зоной неопределенности решения. В случае попадания объекта на нулевую линию этой зоны решение о принадлежности к одному из классов обучения принято быть не может. В результате диагноза на классах обучения объекты экзамена



Рис. 3. Тесторные веса признаков по первому решению  
Веса отложены по вертикали

Таблица 3

Кодированная информация по первому шагу решения

Объекты	Различающие признаки																												
	21	30	31	16	27	36	14	37	13	6	5	4	34	39	40	38	15	26	18	33									
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	0	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
10	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
21	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
25	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1
28	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1	1	1	1	1
32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
34	0	1	1	0	1	1	0	0	0	1	1	1	1	0	1	1	0	1	0	0	1	1	0	1	1	0	0	0	0
36	1	1	1	0	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
37	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1

ПРОДУКТИВНЫЕ

Таблица 3 (продолжение)

Объекты	Различающие признаки																												
	12	8	23	20	11	17	32	7	22	35	3	9	24	2	19	10	29	28	25										
1	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	0	1	0	0	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
5	0	0	1	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	0	1	1	0	1	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	1	0	0	1	1	0	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	0	1	0	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	0	1	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	0	1	1	0	1	1	0	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	1	0	1	1	0	1	0	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1
25	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	1	1	1	0	1	1	0	0	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	1	0	0	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
34	1	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	1	0	0	1	0	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	1	0	0	1	0	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ПРОДУКТИВНЫЕ

Таблица 3 (продолжение)

Объекты	Различающие признаки																			
	21	30	31	16	27	36	14	37	13	6	5	4	34	39	40	38	15	26	18	33
23	1	1	0	1	1	1	1	1	0	0	0	0	1	1	1	1	0	1	0	1
35	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
38	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
40	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1
48	0	1	0	0	1	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0
49	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1
56	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1
58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
59	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
43	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	0
12	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	0	1	1	1
12'	0	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1	0	1	1	1
13	1	0	0	0	1	0	1	0	1	0	1	0	1	1	1	0	0	1	0	1
13'	1	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	1	0	1
14	1	1	0	0	1	0	1	1	0	1	0	0	0	0	1	0	0	1	1	1
14'	1	1	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	1	1	1

Пустые

Плоды

Таблица 3 (продолжение)

Объекты	Различающие признаки																		
	12	8	23	20	11	17	32	7	22	35	3	9	24	2	19	10	29	28	25
23	0	0	0	0	1	1	1	1	0	1	0	0	1	0	0	1	0	0	1
35	0	1	0	0	1	0	0	1	0	0	0	0	1	1	0	0	1	0	1
38	0	0	0	1	1	0	1	1	0	1	0	0	0	1	0	1	0	0	1
40	1	0	1	1	0	1	1	0	1	1	1	1	1	1	0	0	1	0	1
48	0	1	0	1	1	0	1	0	0	1	1	0	1	1	0	0	1	0	1
49	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	0	1	0	1
56	0	1	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0	0	1
58	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1
59	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0
60	0	1	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	1
39	1	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	1
43	1	1	0	1	0	1	0	1	0	0	0	0	1	1	1	0	0	0	1
12	1	0	1	1	0	1	1	0	1	1	0	1	1	1	0	0	0	1	1
12'	1	0	0	0	1	1	0	1	1	0	1	1	1	1	0	0	0	0	1
13	0	1	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0	1	1
13'	0	1	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0	1	1
14	0	1	0	0	1	1	0	0	1	1	1	0	1	1	0	0	0	0	1
14'	0	1	0	0	1	1	0	0	1	1	1	0	1	1	1	0	0	0	1

Пустые

Плоды

Таблица 3 (продолжение)

Объекты	Различающие признаки																			
	21	30	31	16	27	36	14	37	13	6	5	4	34	39	40	38	15	26	18	33
15	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1
16	0	1	0	0	1	1	1	0	1	0	0	0	0	1	1	0	0	1	0	0
16'	0	1	0	0	1	1	1	0	1	0	0	0	0	0	1	0	0	1	0	0
17	1	0	0	0	1	0	1	0	1	0	0	0	0	0	1	0	0	1	1	0
18	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	0
19	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0
20	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	0	0
22	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1
24	1	1	0	0	1	1	0	1	0	1	1	1	1	0	0	0	0	1	0	1
26	1	1	0	1	1	1	1	0	0	0	0	1	1	1	1	1	0	1	0	0
27	0	0	0	1	1	0	1	1	1	1	0	0	0	0	1	1	0	1	0	1
29	1	1	1	1	0	1	0	1	1	0	0	0	1	1	0	1	0	1	0	1
30	1	0	0	0	1	0	0	0	1	1	1	1	1	1	0	0	0	1	0	0
30'	0	0	0	0	1	0	0	0	1	1	1	1	1	1	0	0	0	1	0	0
31	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0
31'	1	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
33	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1
41	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	1
42	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1

Пробы

Таблица 3 (продолжение)

Объекты	Различающие признаки																		
	12	8	23	20	11	17	32	7	22	35	3	9	24	2	19	10	29	28	25
15	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
16	0	1	0	0	1	1	0	1	1	0	0	1	1	0	0	0	0	1	1
16'	0	1	0	0	1	1	1	0	1	1	0	0	1	1	0	0	0	1	1
17	0	1	0	0	1	1	0	0	1	0	1	0	1	1	1	0	1	0	1
18	0	0	1	1	0	1	1	1	1	1	0	0	1	1	0	1	0	0	1
19	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1
20	1	0	1	0	0	0	1	0	1	1	1	1	1	1	1	0	1	0	1
22	1	0	0	0	0	0	0	0	1	0	1	1	1	1	1	0	0	0	1
24	1	1	0	1	0	1	1	0	0	1	0	0	1	1	0	0	0	0	1
26	1	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1
27	0	0	0	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0	1
29	0	0	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	1
30	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1
30'	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1
31	1	0	0	0	0	1	0	0	1	1	0	1	1	0	0	0	0	0	1
31'	1	0	0	0	0	1	0	0	1	1	0	1	1	0	0	0	0	0	1
33	1	0	0	1	0	1	0	0	1	0	1	0	1	1	0	0	0	0	1
41	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
42	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	1

Пробы

Таблица 3 (окончание)

Объекты	Различающие признаки																			
	21	30	31	16	27	36	14	37	13	6	5	4	34	39	40	38	15	26	18	33
44	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1
45	1	1	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	1	1	1
46	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	0	1
47	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
50	0	1	0	0	0	1	0	1	0	0	0	1	1	1	1	1	0	0	0	0
51	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0
52	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
53	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
53'	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
54	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1
55	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	1	0	1
57	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
61	1	1	0	1	1	0	1	1	0	1	1	1	1	0	0	0	0	1	0	1
62	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0

Пробы

Таблица 3 (окончание)

Объекты	Различающие признаки																		
	12	8	23	20	11	17	32	7	22	35	3	9	24	2	19	10	29	28	25
44	1	0	0	0	0	1	1	0	0	1	1	1	1	1	1	0	1	0	1
45	1	1	0	1	1	0	1	0	1	1	0	1	1	1	1	0	0	1	1
46	1	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1
47	1	0	0	0	0	0	1	1	1	1	0	0	1	0	1	0	0	0	1
50	1	1	0	1	0	1	1	1	0	1	0	0	1	0	0	0	0	0	1
51	0	0	0	0	0	0	1	1	1	0	1	0	1	1	1	1	1	0	1
52	1	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	1
53	1	1	0	1	0	1	0	1	0	0	1	0	1	1	1	0	0	0	1
53'	0	1	0	1	1	0	1	0	1	1	0	0	1	1	1	0	0	0	1
55	0	1	0	0	1	1	0	0	1	0	1	0	1	1	1	0	0	0	1
57	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
61	1	0	1	1	0	0	0	0	1	0	0	1	1	1	1	0	0	0	1
62	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1

Пробы

разделились на продуктивные и пустые; из 25 заведомо продуктивных объектов на внешнем экзамене 11 уверенно относятся в свой класс, 11 попало в зону приближенного решения, для трех объектов (41, 22, 45), находящихся вблизи нулевой линии, решение об их принадлежности к классу продуктивных не принимается. Три объекта: 16, 50, 52, попавшие в зону приближенного решения, должны быть отнесены к классу пустых; это противоречие объясняется неоднородностью объектов в классе.

По-видимому, указанные объекты являются рудопроявлениями со специфическими чертами, отличающими их от интрузий обычного типа. Три объекта из класса пустых (47, 51, 53) на внешнем экзамене уверенно относятся в свой класс. Таким образом, из 28 объектов внешнего экзамена три распознаны неправильно, о трех решение не принималось, остальные 22 были распознаны верно.

Результаты распознавания неизвестных объектов-проб следующие: 3 отнесены к продуктивным (54, 42, 55) (Нижне-Фокинская, Нижне-Томбинская, Средне-Чангадинская), принадлежность одной (57) остается неопределенной.

### Второй шаг решения

Исходя из результатов первого шага решения для обучения было взято 8 объектов класса "месторождений" и 8 - из класса "рудопроявления" (см. рис. 1). 32 объекта представляют класс проб. Из них 26 объектов входят как неизвестные (в их число включены продуктивные объекты и 4 объекта, фигурирующие на предыдущем шаге как неизвестные - 42, 54, 55, 57) и 6 объектов оставлены для внешнего экзамена (5 месторождений и 1 рудопоявление), среди них два объекта (4, 37) не имеют близких аналогов в обучении (табл. 4).

Целеуказание на втором шаге - разделять объекты-эталоны класса месторождения от объектов-эталонов класса рудопоявления. Значения целевого признака бинарные, т.е. известна принадлежность объектов-эталонов к классам, но информация о взаимоотношении их внутри каждого класса отсутствует, хотя известно, что класс месторождения неоднороден (более значимые и менее значимые месторождения).

Внутри класса рудопоявления можно предполагать как однородность, так и неоднородность.

Ход решения задачи в основном такой же, как и на первом шаге. Для объектов-проб, содержащих прочерки, распознавание производилось, как и на первом шаге, в двух вариантах, равно как и для тех объектов-проб, для которых в кодировании некоторых количественных признаков возникла неопределенность из-за совпадения значения количественного признака с выбранной разделяющей границей (объекты 24, 29, 41, 55).

На данном шаге решения задачи из пространства признаков были выделены: совокупность отождествляющих признаков (имеющих одинаковые значения на эталонных объектах разных классов) и совокупность разделяющих признаков; последняя (34 признака) использовалась для поиска информативной системы признаков. Это позволило применить метод длины тесторов без предварительной процедуры сокращения признаков.

Основные результаты этого шага решения показаны на рис. 4. Признаки, имеющие нулевые значения тесторных весов, были минимизированы в процессе поиска информативной системы. Наиболее существенными оказались признаки 11, 15, 17, 18, 19, 20, 21, 23, 33, 39; большинство из них находят петрологическое объяснение. Это признак наличия в составе интрузий нижних и верхних такситовых габбро-долеритов (14, 15), что подтверждает давно известную связь проявления основных пегматоидов и сульфидного оруденения и свидетельствует в пользу их генетической связи; характер железистости оливинов

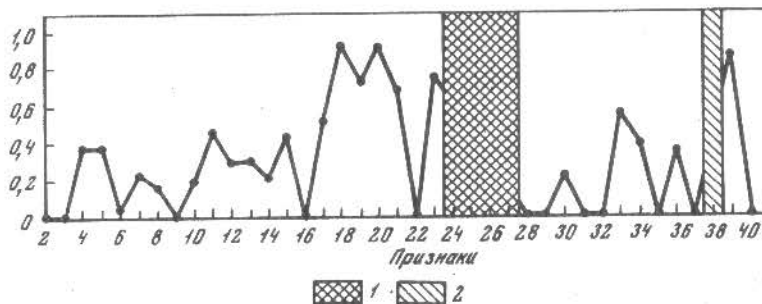


Рис. 4. Тесторные веса признаков по второму шагу решения  
Веса отложены по вертикали; заштрихованы области нулевых значений весов отождествляющих (1) и целевого (2) признаков

нижних и верхних дифференциатов интрузий (20, 21); основность плагиоклаза нижних горизонтов (23); содержание щелочей в обогащенных оливином горизонтах (которые фиксируются широким распространением биотита и вторичной альбитизации в рудоносных интрузиях - 39). Однако в ряде случаев признаки, несомненно играющие большую петрологическую роль, оказываются малоинформативными. Так, важный признак наличия в интрузиях пикритовых горизонтов (16) не работает в принятой постановке, так как многие разделяемые объекты содержат этот горизонт.

На внутреннем экзамене для 8 эталонов месторождений и 8 эталонов рудопоявлений получено полное распознавание с зоной неопределенности в пределах значений от  $(-)$ 0,1 до  $(+)$ 0,2. Шесть объектов внешнего экзамена были уверенно отнесены к соответствующим классам. Из 26 проб (неизвестных объектов) уверенно отнесены в класс месторождений и интрузии 62 и 18, а в класс рудопоявлений - интрузия 21. В зону неопределенности попали интрузии 20, 44 и сюда же относятся интрузии 24, 30, 29, а после получения недостающей информации по отдельным признакам из группы месторождений может попасть 12-я. Из-за недостатка информации несколько неопределенно также положение объектов 13, 41, 55 из класса рудопоявлений. Объекты, отнесенные в класс месторождений и в зону неопределенности (20, 44, 61), на следующем шаге использовались как неизвестные объекты в классе проб.

## Кодированная информация по второму шагу решения

Объ- ект	Признаки и их номера																																		
	отождествляющие														целевой																				
	25	26	27	24	38	37	20	18	5	4	6	40	13	16	14	22	19	28	30	38	37	20	18	5	4	6	40	13	16	14	22	19	28	30	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
8	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
23	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
25	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
28	1	1	1	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Таблица 4 (продолжение)

Объ- ект	Признаки и их номера																																			
	отождествляющие														целевой																					
	23	17	12	2	39	31	15	3	36	34	33	21	35	32	9	8	11	8	10	29	38	37	20	18	5	4	6	40	13	16	14	22	19	28	30	
1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
5	1	1	0	1	1	1	1	1	1	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
9	0	0	1	0	1	1	1	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
21	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
25	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
28	1	1	1	1	1	0	0	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
32	0	0	0	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
34	0	0	1	1	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1







Таблица 4 (окончание)

Объ- ект	Признаки и их номера															прочие			
	стожествляющие														целевой				
	25	26	27	24	38	37	20	18	5	4	6	40	13	16			14	22	19
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	0	0	1	0	0	1	1	1	1	0	0	1	1	0	1	0
13'	1	1	1	0	0	0	0	0	1	1	1	0	0	0	1	1	0	1	0
19	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
29	1	1	0	1	1	0	1	0	0	0	0	1	1	1	0	1	0	0	0
29'	1	1	0	1	1	0	0	0	0	0	0	1	1	1	0	1	0	0	0
30	1	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
30'	1	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
41	1	1	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0
41'	1	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
45	1	1	0	1	0	0	1	1	0	0	0	1	1	0	1	0	1	0	1
50	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
52	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0
54	1	1	1	0	1	0	1	0	1	1	1	1	1	1	0	1	0	1	1

Таблица 4 (окончание)

Объ- ект	Признаки и их номера															прочие				
	стожествляющие														целевой					
	23	17	12	2	39	31	15	3	36	34	33	21	35	32			9	8	11	8
6	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	0	0	0	0	0
11	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0
13	0	1	0	1	1	0	0	0	0	0	1	0	1	1	0	1	1	0	0	1
13'	0	1	0	1	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	1
19	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0
29	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
29'	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
30	0	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0
30'	0	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
41	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0
41'	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0
45	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0
50	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0
52	0	0	1	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0
54	0	0	0	0	0	1	0	0	1	0	1	1	1	1	0	1	0	1	0	0

Таблица 5

Кодированная информация по третьему шагу решения

Объекты	Признаки и их номера								
	отождествляющие								
	4	5	6	13	14	16	18	22	

## Месторождения

Более значи- мые	1	1	1	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1
	3	1	1	1	1	1	1	1	1
	4	1	1	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1
Менее значи- мые	7	1	1	1	1	1	1	1	1
	8	1	1	1	1	1	1	1	1
	9	1	1	1	1	1	1	1	1
	10	1	1	1	1	1	1	1	1
Пробы	6	1	1	1	1	1	1	1	1
	11	1	1	1	1	1	1	1	1
	33	1	1	1	1	1	1	1	1
	62	1	0	1	1	1	1	1	1
	18	1	1	1	1	0	1	1	1
	61	1	1	1	0	1	1	0	1
	12	0	1	0	0	1	1	1	1
	12'	0	1	0	0	1	1	1	1
	44	1	1	1	0	0	1	1	0
	54	1	1	1	1	0	1	0	1
20	0	0	0	0	1	1	0	1	

Признаки и их номера											
отождествляющие					целевые			различающие			
24	25	26	27	29	28	3	2	7	8	9	10

## Месторождения

1	1	1	1	0	1	0	1	0	0	1	1
1	1	1	1	0	1	1	1	0	1	1	1
1	1	1	1	0	1	1	1	1	1	0	0
1	1	1	1	0	1	1	1	1	1	0	0
1	1	1	1	0	1	1	1	1	0	0	1
1	1	1	1	0	0	0	1	0	1	0	0
1	1	1	1	0	0	0	1	0	0	1	0
1	1	1	1	0	0	0	0	1	0	0	1
1	1	1	1	0	0	0	1	0	0	1	0
1	1	1	1	0	0	0	1	0	0	1	0
1	1	1	1	0	0	0	1	0	0	1	0
1	1	1	1	1	0	1	1	0	0	1	0
0	1	1	1	0	0	0	0	0	1	0	0
1	1	1	1	1	0	1	1	0	0	1	0

Таблица 5 (окончание)

Объекты	Признаки и их номера							
	различающие							
	11	12	15	17	19	20	21	23
	Месторождения							
Более значи- мые	1	1	1	1	1	1	1	1
	2	1	1	1	1	1	0	1
	3	1	1	1	1	1	1	1
	4	1	1	1	1	1	1	1
	5	1	0	1	1	1	0	0
Менее значи- мые	7	0	1	0	1	0	0	0
	8	0	1	0	1	0	0	1
	9	1	1	1	0	1	0	0
	10	0	1	0	1	0	1	0
Пробы	6	0	1	1	1	0	1	1
	11	0	1	1	1	1	0	0
	33	0	1	0	1	1	1	0
	62	1	1	1	1	1	1	0
	18	0	0	1	1	0	0	0
	61	0	1	0	0	1	1	1
	12	0	1	0	1	0	1	1
	12	0	1	0	1	0	0	0
	44	0	1	1	1	1	1	0
	54	1	0	0	0	0	0	0
	20	0	1	1	0	1	1	1

Признаки и их номера										
различающие										
30	31	32	33	34	35	36	37	38	39	40
Месторождения										
1	1	0	1	1	0	1	0	1	1	0
1	0	1	0	1	1	1	0	1	1	0
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	0	0	1
1	0	1	1	0	1	1	1	1	0	1
0	0	0	1	0	0	1	0	1	0	0
0	0	0	0	1	0	0	1	0	0	0
0	0	0	1	0	0	0	0	1	0	0
1	0	1	0	0	1	0	1	0	0	0
1	0	1	1	0	1	0	1	1	0	1
1	1	0	1	0	0	1	1	1	0	1
0	0	1	0	0	1	0	1	0	0	0
1	1	0	1	0	0	1	1	1	0	1
0	0	1	1	0	0	0	1	1	0	0
0	0	1	1	0	1	0	1	0	0	1
1	0	0	0	0	0	1	1	1	1	0
1	0	0	0	0	0	1	1	1	1	0
0	0	0	1	0	0	0	1	0	0	0
1	1	0	1	0	0	1	0	1	1	0
0	0	0	1	0	0	0	1	0	0	0

### Третий шаг решения

В соответствии с возросшей детальностью подразделения исходных интрузий на классы для обучения были взяты пять объектов – более значимые месторождения и четыре – менее значимые. В связи с небольшим общим количеством эталонных объектов все они участвовали в обучении и внешний экзамен не проводился. 11 объектов были взяты в качестве проб. При организации признакового пространства выделялась совокупность отождествляющих признаков, в дальнейшем

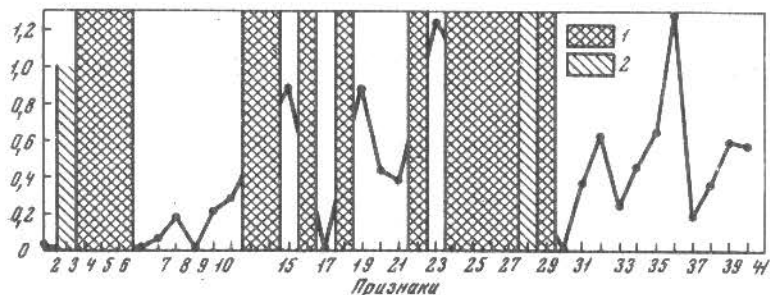


Рис. 5. К-тесторные веса признаков по третьему шагу решения. Веса отложены по вертикали; заштрихованы области нулевых значений весов отождествляющих (1) и целевых (2) признаков

не участвовавшая в решении. В поиске информативной системы признаков использовалось 24 оставшихся разделяющих признака (табл. 5).

Согласно целеуказанию на третьем шаге необходимо одновременно подразделить объекты-эталон на классы более значимых и менее значимых месторождений и отождествлять их внутри каждого из этих классов. Предполагается, что различия внутри классов незначительны по сравнению с различиями между классами. Здесь при выборе информативной системы признаков впервые был применен метод длины *K*-тесторов (Бугаец и др., 1969). Среди признаков наиболее существенными оказались: 15, 19, 23, 32, 35, 36, 39, 40 (рис. 5). На внутреннем экзамене достигнуто полное распознавание с зоной неопределенности в пределах значений от  $(-)$ 0,170 до  $(+)$ 0,272. При распознавании проб пять объектов попадают в эту зону, при этом объекты 62 и 61 тяготеют к классу более значимых месторождений, а объекты 6 и 18 – к классу менее значимых, интрузия 11 занимает неопределенное положение. Четыре объекта (20, 44, 33 и 12) уверенно относятся к классу менее значимых месторождений, хотя интрузия 12 по отмеченным выше причинам может значительно приблизиться к межклассовой границе.

#### Общие результаты решения

Обобщая изложенное выше, кратко отметим наиболее важные моменты в решении задачи прогнозирования продуктивности дифференцированных трапповых интрузий севера Сибирской платформы.

Полный список признаков, характеризующих исследуемые объекты, подразделен на четыре группы: 1) геолого-тектоническая обстановка,

2) строение интрузий, 3) минералогические особенности, 4) петрохимические особенности.

В решении задачи выделены три последовательных шага решения. Роль групп признаков на каждом из шагов несколько меняется. На первом шаге наибольший информативный вес имеют признаки минералогической и петрографической групп, на втором и третьем шагах наряду с ними приобретает значимость группа признаков по строению интрузий. Для каждой пары классов объектов, подлежащих сравнительному изучению, с помощью тесторных процедур найдены комплексы максимально различающих признаков. Установлена специфичность этих комплексов для каждого шага решения, т.е. существенность одних и тех же характеристических признаков для различия объектов разных классов неодинакова. Так, количественные минералогические и петрохимические признаки наиболее существенно различают месторождения от рудопоявлений.

Необходимо отметить, что решение данной задачи вызвало развитие новых методов тестового подхода: были созданы и апробированы вычислительные процедуры, учитывающие длины тупиковых тесторов. Этот учет позволил более качественно провести минимизацию характеристического пространства признаков и оценить каждый признак на его различающую способность при сравнительном изучении объектов двух классов. По мере необходимости одновременной оценки тождества – различия исследуемых объектов двух классов построена серия процедур, дающих эти оценки на базе понятия *K*-тестора. Процедура *K*-тесторы позволяет сортировать объекты (пробы) на выделенных заранее классах (эталонах) с одновременным учитыванием сходств и различий диагностируемых объектов.

#### ЗАКЛЮЧЕНИЕ

Тестовый подход, некоторые методы которого были изложены в настоящей работе, позволяет широко и результативно углубляться в решение все более сложных задач рудных разделов геологии. По мере этого углубления возникают новые проблемы как в области математических разработок, так и в области геологической отрасли знания. В математическом отношении изложенные методы требуют дальнейшей детализации процедур и более высокого уровня разработок. В области геологии требуются большая маневренность в отношении целей исследования и выход на новые направления исследования.

#### ЛИТЕРАТУРА

- Бабич В.В., Федосеев Г.С., Красавчиков В.О. Комплексная оценка информативности признаков при решении прогнозных задач. – В кн: Состояние и направление исследований по металлогении траппов. Красноярск, 1974.
- Бишаев А.А. Метод нахождения целевой информативности признаков. – В кн: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск, 1973.
- Бугаец А.Н., Гражданцев Н.Г., Дорофеюк А.А., Мацак А.П., Нарсеев В.А. Формальные методы разграничения геологических объектов и их применение при решении некоторых

задач минералогии и металлогении. — В кн.: Применение математических методов в геологии. Алма-Ата, "Наука", 1968.

*Бугаев А.Н., Дворниченко Г.К., Мацак А.П., Серова Т.Т.* Алгоритмы и программы решения геологических задач на ЭВМ "Минск-2" и "БЭСМ-3М", вып.2. Алма-Ата, 1969.

*Васильев Ю.Р., Виленский А.М., Дмитриев А.Н., Золотухин В.В., Карбышев В.Д., Рябов В.В., Щедрин Н.Ф.* Новый этап исследований в применении дискретных математических методов в оценке перспектив рудоносности дифференцированных трапповых интрузий Севера Сибирской платформы. — В кн.: Состояние и направление исследований по металлогении траппов. Красноярск, 1974<sub>1</sub>.

*Васильев Ю.Р., Дмитриев А.Н., Золотухин В.В.* Оценка существенности основных признаков дифференцированных трапповых интрузий с медно-никелевым оруденением логико-математическими средствами анализа для поисковых целей. — В кн.: Состояние и направление исследований по металлогении траппов. Красноярск, 1974<sub>2</sub>.

*Вышемирский В.С., Дмитриев А.Н., Трофимук А.А.* Поисковые признаки гигантских нефтяных месторождений. — Материалы VIII мирового нефтяного конгресса, СД-8. М., 1971.

Геология и математика. Новосибирск, "Наука", 1967.

*Гольдман Р.С., Москаленко Ю.С., Николаев В.В.* Использование логических методов для диагноза объектов естествознания на примере прогноза месторождений меди в Приморье. — В кн.: Управление и информация, вып.14. Владивосток, 1974.

*Дмитриев А.Н.* Некоторые табличные числа. — В кн.: Дискретный анализ, вып.12. Новосибирск, "Наука", 1968.

*Дмитриев А.Н.* Использование длин тупиковых тестов при обработке таблиц. — В кн.: Дискретный анализ, вып.17. Новосибирск, "Наука", 1970.

*Дмитриев А.Н.* Анализ геологических задач, сводимых к проблеме распознавания образов. — В кн.: Ма-

тематика в геологии. Новосибирск, 1973.

*Дмитриев А.Н., Журавлев Ю.И., Кределев Ф.П.* О математических принципах классификации предметов и явлений. — В кн.: Дискретный анализ, вып.7. Новосибирск, "Наука", 1966.

*Дмитриев А.Н., Журавлев Ю.И., Кределев Ф.П.* Об одном принципе классификации и прогноза геологических объектов и явлений. — "Геология и геофизика", 1968<sub>1</sub>, № 5

*Дмитриев А.Н., Золотухин В.В., Васильев Ю.Р.* Опыт применения дискретной математической обработки информации по дифференцированным трапповым интрузиям северо-запада Сибирской платформы. — "Сов. геология", 1968<sub>2</sub>, № 12.

*Дмитриев А.П., Васильев Ю.Р., Золотухин В.В.* Логико-математическая обработка информации при выявлении перспективности сульфидного оруденения в некоторых трапповых интрузиях Севера Сибирской платформы. — "Геология и геофизика", 1968<sub>3</sub>, № 7.

*Дмитриев А.Н., Смертин Е.А.* Алгоритмы вычисления тестовых параметров бинарных таблиц в задачах распознавания. — В кн.: Алгоритмы и программы решения геологических задач, вып.3. Алма-Ата, 1970.

*Дмитриев А.Н., Бишаев А.А., Красавчиков В.О., Смертин Е.А., Штаннова Т.И.* Распознавание на базе построения всех тупиковых тестов. — В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск, "Наука", 1973.

*Константинов Р.М.* Основы формационного анализа гидротермальных рудных месторождений. М., "Наука", 1973.

*Константинов Р.М., Дмитриев А.Н.* Использование математических методов для анализа геологических факторов, влияющих на масштабы оруденения (на примере месторождений касситерит-сульфидной формации). — "Геол. рудных месторожд.", 1970, № 2.

*Константинов Р.М., Кристалый Б.В.* и др. Математические методы при изучении рудных месторождений. — В кн.:

Итоги науки, серия геол. Рудные месторождения. М., изд. ВИНТИ, 1970.

*Константинов Р.М., Дмитриев А.Н.* Простые методы определения формационной принадлежности рудного месторождения по геологическим данным. — "Изв. АН СССР, серия геол.", 1971, № 2.

*Константинов Р.М., Белякова Ю.А., Сиротинская С.В., Хетчиков Л.Н.* Логическое планирование экспериментов по выращиванию окрашенного кварца. — "Докл. АН СССР", 1974, 218, № 1.

*Королева З.Е.* О некоторых характеристиках тестовых алгоритмов распознавания. — Тезисы III Всес. конф. по проблемам теоретической кибернетики. Новосибирск, 1974.

*Красавчиков В.О.* Об одном алгоритме построения всех тупиковых тестов бинарной таблицы. — В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск, 1973.

*Кределев Ф.П., Дмитриев А.Н., Журавлев Ю.И.* Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики. — "Докл. АН СССР", 1967, 173, № 5.

*Кределев 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. Новосибирск, "Наука", 1969.

*Модников И.С., Еремеев А.Н., Писаревский В.И., Россман Г.И., Дмитриев А.Н., Кределев Ф.П., Резник Т.Л.* Оценка масштаба редкометального оруденения, локализованного в вулканических аппаратах (с помощью ЭВМ). — "Сов. геология", 1969, № 11.

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

*Нестеренко Г.В., Дмитриев А.Н., Кределев Ф.П., Осинцев С.Р., Штан-*

*нова Т.И.* Сравнительное изучение россыпей с помощью логико-дискретного анализа (на примере Восточного Забайкалья). — Тезисы II Всес. совещ. по россыпям золота, Магадан, 1969.

*Николаев В.В.* О получении тупиковых тестов при изменении диагностических таблиц. — В кн.: Управление и информация, вып.14. Владивосток, 1974.

*Петровская Н.В., Константинов Р.М., Сиротинская С.В.* О применении математических методов для систематики золотых месторождений по минеральному составу. — "Сов. геология", 1971, № 2.

*Родионов Д.А.* Статистические методы разграничения геологических объектов по комплексу признаков. М., "Недра", 1968.

*Родионов Д.А., Серых В.И.* Проблема выбора поисковых признаков и ее статистическое решение. — В кн.: Математические методы в геологии. М., "Наука", 1968.

*Смертин Е.А.* Q-тесты в задачах тестового распознавания. — В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск, 1973.

*Смертин Е.А., Дмитриев А.Н.* Дополнение к алгоритму распознавания "голосованием" по тестам и тесторам. — В кн.: Алгоритмы и программы, вып.3. Алма-Ата, 1970.

*Трофимук А.А., Вышемирский В.С., Дмитриев А.Н., Дробот Д.И., Карагодин Б.Н., Сулимов И.Н.* Распознавание образов гигантских нефтяных месторождений. — В кн.: Проблемы нефтегазоносности Сибири. Новосибирск, "Наука", 1971.

*Чегис И.А., Яблонский С.В.* Логические способы контроля работы электрических схем. — "Труды Матем. ин-та им. В.А.Стеклова", 1958, 51.

*Эшби Р.У.* Введение в кибернетику. М., ИЛ, 1959.

*Яблонский С.В., Дежидова Н.Г., Константинов Р.М., Королева З.В., Кудрявцев В.Б., Сиротинская С.В.* Тестовый подход к количественной оценке геолого-структурных факторов и масштабов оруденения. — "Геол. рудных месторожд.", 1971, № 2.