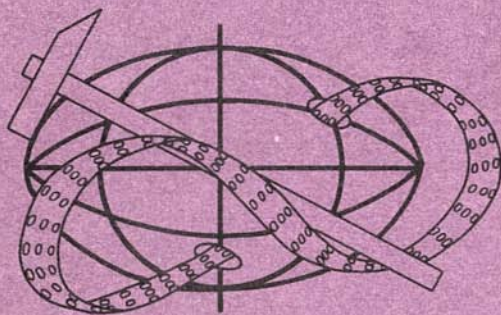


1

ЛОГИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ



НОВОСИБИРСК 1986

612

АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ГЕОЛОГИИ И ГЕОФИЗИКИ

ЛОГИКО- МАТЕМАТИЧЕСКИЕ
МЕТОДЫ
В ГЕОЛОГИЧЕСКИХ
ИССЛЕДОВАНИЯХ
ТЕОРИЯ И ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

СБОРНИК НАУЧНЫХ ТРУДОВ

НОВОСИБИРСК 1986

УДК 519.1:553.062. 98:56+55(09)

Логико-математические методы в геологических исследованиях:
Сб. науч. тр. /АН СССР, Сиб. отд-ние, Ин-т геологии и геофизики;
Редкол.: В.В.Бабич, А.Н.Дмитриев (отв. редакторы) и др. - Новоси-
бирск: ИГиГ, 1986. - 144 с.

Сборник посвящен вопросам развития и применения логико-математических методов. В первой его части представлены работы теоретического характера: вопросы усовершенствования и применения тестового подхода для целей распознавания; результаты исследований по заполнению пропусков в таблицах исходных данных и др. Во второй - приведены примеры решения задач петрологии, прогноза руд, исследования проблем морфологической классификации. В третьей - изложен материал, касающийся истории развития математической геологии.

Расчитан на геологов, интересующихся математическими методами, и специалистов по прикладной математике, имеющих отношение к обработке информации на ЭВМ.

Редакционная коллегия

С.В.Макаров, Г.С.Федосеев, Т.И.Штатнова

Ответственные редакторы

В.В.Бабич, А.Н.Дмитриев

Рецензенты

д-р геол.-мин.наук В.К.Кучай (ИГиГ СО АН СССР),
канд.геол.-мин.наук В.В.Зуенко (НГУ)

© Институт геологии
и геофизики СО АН СССР,
1986 г.

ПРЕДИСЛОВИЕ

Математическая геология как самостоятельное направление научных исследований уже миновала ту стадию развития, когда целесообразность ее существования нередко вызывала сомнение. Но она еще и не достигла уровня широких теоретических обобщений, задающих общую логическую структуру данной области знаний. Именно поэтому математическая геология в настоящее время представляет собой уникальную сферу творческого взаимодействия специалистов разного профиля, в результате которого геология приобретает мощнейший инструмент научного исследования, а математика существенно расширяет область своего практического применения.

Проблемы, возникающие в связи со становлением математической геологии, можно разделить на три группы:

- селекция, модернизация и создание математических методов, которые в принципе могут быть использованы для решения определенных типов геологических задач;

- содержательно-логический анализ (осмысление) математических процедур и методов как необходимое условие осознанного и целенаправленного их выбора при решении конкретной геологической задачи;

- решение конкретных геологических задач математическими методами и содержательная интерпретация результатов.

Указанные группы вопросов отображают путь прохождения математических методов от их зарождения вплоть до практического использования и определяют области и характер взаимодействий геолога и математика в этом процессе. Если первая группа вопросов решается, как правило, математиками (при некотором участии геологов), а последняя - геологами (при некотором участии математиков), то в решении промежуточной группы вопросов в равной степени заинтересованы представители обеих наук, так как именно здесь происходит "передача из рук в руки" методов решения задач или

задач, требующих решения. Очевидно, что от качества "процедуры передачи" во многом зависит надежность и достоверность конечных результатов.

Безусловно, описанная "трехслойная" модель весьма схематично и несколько упрощенно отображает реальное состояние математической геологии, однако она, на наш взгляд, весьма полезна для понимания и объяснения многих явлений (в том числе психологического характера), сопутствующих развитию нового научного направления.

В отличие от многих аналогичных изданий, посвященных проблемам математической геологии, имеющих, как правило, узкотематическую или сугубо методологическую направленность, настоящий сборник затрагивает вопросы всех трех указанных уровней и это не случайно. Авторский коллектив сборника - не формальное объединение специалистов разного профиля, интересующихся вопросами математической геологии, а коллектив единомышленников, связанных многолетней совместной работой, причем (за редким исключением), в стенах одной лаборатории. Поэтому структура сборника в какой-то степени является отражением реализации в Институте геологии и геофизики СО АН СССР одного из подходов к созданию творческих групп математиков и геологов, призванных решать задачи на стыке двух наук. Этим объясняется некоторая кажущаяся пестрота в содержании и стиле изложения материала, что может быть расценено как недостаток, но практически неизбежно при таком всестороннем рассмотрении сложной проблемы.

В первом разделе сборника представлены статьи, посвященные главным образом методическим разработкам. С.В.Макаровым описана модифицированная вычислительная процедура практически не используемого в геологии метода гребневой регрессии и обозначены возможности его применения для решения задач, сводимых к проблеме "восстановления пропусков"^{*)} в таблице данных. В статье А.Н.Дмитриева и С.В.Макарова описаны уже реализованные, а также возможные пути усовершенствования широко известных методов тестового подхода, успешно применяющихся для решения геолого-прогнозных задач. В работе А.А.Бишаева и Г.С.Федосеева описан оригинальный алгоритм целевого преобразования количественных признаков как

^{*)} Правильнее все же говорить "заполнение пропусков". - Примеч. отв. ред.

составной части проблемы кодирования исходной информации. Г.С.Федосеевым предпринята попытка логического обобщения проблемы "Восстановления пропусков" и теоретически рассмотрены возможные пути ее решения.

Во втором разделе объединены статьи, посвященные описанию схем и результатов решения математическими методами разнообразных геологических задач. А.Ф.Белюсовым и А.С.Вакуленко излагаются результаты применения одного из методов кластерного анализа в целях выявления близких по петрохимическим характеристикам совокупностей магматических горных пород. В.Д.Карбышевым с соавторами приведено описание результатов рудопрогнозных построений (на примере Северо-Сибирской никеленосной области) методами распознавания образов, использующих линейные решающие правила. В.О.Красавчиковым и В.Я.Саниным предпринята попытка формальной проверки и уточнения одной из палеонтологических классификаций с привлечением целого спектра математических приемов из кластер-анализа, теории графов, теории распознавания образов.

Заключает сборник статья А.Н.Дмитриева, посвященная истории зарождения математической геологии как организованного научного направления на примере Института геологии и геофизики СО АН СССР. Работа написана ярким, своеобразным языком, подчеркивающим обстановку бурного кипения страстей в начале 60-х годов, сопровождавшего процесс поиска оптимальных путей взаимопроникновения геологии и математики.

Настоящий сборник в целом рассчитан на подготовленного читателя. Он должен вызвать интерес у многих специалистов, интересующихся математической геологией, вопросами прикладной математики и имеющих отношение к обработке информации на ЭВМ.

В.В.Бабич'

А.Н.Дмитриев, С.В.Макаров

ОБ УСОВЕРШЕНСТВОВАНИИ ТЕСТОВОГО ПОДХОДА
В ЗАДАЧАХ ДИАГНОСТИКИ

Значительное количество методов диагностики, появившееся за последние двадцать лет, реализует возможность тестового подхода /5,18/. По мере его теоретического развития и практического применения вскрылись узкие места, которые потребовали значительных модификаций и дальнейших усовершенствований. В данной статье, в её первом разделе, приводятся некоторые рекомендации по усовершенствованию тестового подхода при диагностике. Они связаны некоторыми своими сторонами с вопросами обнаружения скрытых закономерностей, содержащихся в описаниях геологических объектов. Во втором разделе даны соображения о взаимосвязи тестовых методов с линейно-статистическими и намечены некоторые возможности его дальнейшего развития.

I: Усовершенствование диагностики

Тестовый подход к проблеме диагностики рассматривается с учетом обширного опыта решения конкретных геологических задач. Предлагается новый "метод перекладки" как некоторая модификация известной процедуры голосования по тупиковым тестам. Вводится тестовая модель "с крупноблочным покрытием классов". Эта модель базируется на введенных понятиях: "закономерность k -го порядка", "носитель закономерности", "представительность выборки".

I.I. Общие предпосылки

Пусть дано некоторое множество описаний объектов, снабженных замерами (значениями) признаков, причем каждому объекту соответствует один и тот же набор признаков. Имеющаяся информация

(множество замеров признаков) сведена в таблицу "объект-признак" X так, что строки соответствуют объектам, столбцы - признакам и, таким образом, элемент $x_{ik} \in X$ есть значение k -го признака на i -ом объекте. Предполагается, что совокупность объектов таблицы X разбита на два непересекающихся класса A и B так, что каждый объект (строка) принадлежит одному из этих классов.

Множество объектов, составляющих (вместе со значениями признаков) таблицу X , рассматривается как обучающая выборка, взятая из некоторой генеральной совокупности, каждый элемент которой (объект) снабжен значениями тех же признаков.

Распознавание (диагностика) состоит в том, что произвольный объект из генеральной совокупности относят к одному из двух классов A или B на основе информации, содержащейся в обучающей выборке, а также некоторых предположений о генеральной совокупности. Диагностика производится в соответствии с некоторым решающим правилом, эффективность которого зависит от устройства генеральной совокупности и от представительности обучающей выборки (т.е. от их взаимного соответствия). Качество диагностики устанавливается применением процедуры диагностики к некоторому множеству объектов, "классность" которых выявлена процедурой "экзамен".

Пусть каждый элемент x_{ik} может принимать только одно из двух значений: 0 или 1 ("бинарные признаки"). Для диагностики объектов с бинарными признаками применяется один из методов тестового подхода, основанного на применении функций алгебры логики. Описание подхода и результатов его применения можно найти, например, в ряде публикаций /4,5,6,10,14/. Понятия "тест", "тупиковый тест", "голосование по тупиковым тестам" будем считать известными.

Тестовый подход успешно применялся для решения многих задач диагностики, в частности, для диагностики геологических объектов /2,7,9,11/, по мере применения обнаружались также и его недостатки (которые отчасти предвиделись).

Тестовый подход в задачах диагноза состоит из двух основных процедур: 1) составление списка тупиковых тестов; 2) распознающие процедуры, в частности, голосование по тупиковым тестам. Наиболее распространен тестовый подход с голосованием по тупиковым тестам.

Тестовая структура таблицы "объект-признак", включающей два

класса объектов, определяется частотами вхождения признаков в список тупиковых тестов (т.е. "информационными весами" признаков) и распределением тупиковых тестов по длинам (т.е. по количеству входящих в них признаков).

Генеральная совокупность в случае бинарных признаков представляет собой список всех нуль-единичных наборов длины (по числу признаков) с вероятностями появления этих наборов (объектов) (вероятности которых в большинстве случаев являются неизвестными). Генеральная таблица есть множество нуль-единичных наборов, образующих генеральную совокупность, без вероятностей наборов (объектов) и с указанием "классности" для каждого объекта. Ограничимся рассмотрением важного частного случая, когда каждый объект из генеральной таблицы принадлежит одному и только одному из классов (А или В), и "третьего не дано". Тест, различающий объекты разных классов генеральной таблицы, назовем генеральным тестом. Гипотетическому списку генеральных тестов соответствует гипотетическая генеральная тестовая структура. Разбиение массива генеральной таблицы на два непересекающихся класса есть генеральное разбиение. Задача диагностики может быть сформулирована как задача восстановления генерального разбиения по выборочному разбиению, представленному обучающей выборкой с использованием некоторых гипотез о генеральной таблице (генеральном разбиении).

I.2. Недостатки тестового подхода

I.2.1. Алгоритмы составления списков тупиковых тестов очень трудоемки. Основной лимитирующий параметр - количество признаков. При $n = 50$ обработка становится затруднительной (даже с применением современных ЭВМ), а при $n = 100$ - практически нецелесообразной.

I.2.2. Голосование по тупиковым тестам далеко не всегда поддерживает конкуренцию с другими более простыми методами (алгоритмами). Вычислительный эксперимент, описанный в работе /12/, показывает, что распознавание по нетупиковым тестам дает примерно такой же результат (использован один и тот же материал). Отметим, что составление списка нетупиковых тестов, участвующих в распознающей процедуре (голосовании), выполняется, в отличие от

составления списка тупиковых тестов, посредством простого нетрудоемкого алгоритма.

Примечание. В работе /12/ голосование по тупиковым тестам сравнивается с распознаванием по так называемым "к-выборкам" (каждая из которых определяется фиксированным набором "к-признаков") и делается вывод, что применяемый метод "к-выборки" дает несколько лучший результат. Однако из приведенных в этой статье данных следует, что это преимущество кажущееся. Если при голосовании учитывать только такие "к-выборки", которые обладают разделяющим ("тестовым") свойством, то, как нетрудно усмотреть из приведенной в статье сводки результатов, эффективность метода "к-выборки" повышается до эффективности голосования по тупиковым тестам. При этом распознавание по "к-выборкам" превращается в распознавание по нетупиковым тестам.

1.2.3. Процедура голосования по тестам не имеет теоретического обоснования. Более того, нетрудно построить двухклассовую модель, для которой голосование будет давать заведомо неправильный результат. Возьмем, например, генеральное разбиение, в котором классы А и В различаются только одним признаком, т.е. существует единственный тупиковый тест, и его длина равна единице. Далее, возьмем малую выборку объектов А и В таким образом, чтобы тестовая структура таблицы АВ содержала, помимо главного (генерального) теста, задающего генеральное разбиение, несколько других. Назовем их дополнительными или эфемерными, поскольку они исчезают при увеличении объема выборки; так что в предельном случае остается только первичный (генеральный) тест. Если выборка АВ достаточно мала, нетрудно подобрать не вошедший в нее объект, который по главному тесту будет классифицироваться как объект класса А, а по всем остальным – как объект класса В. Такие объекты будут распознаваться посредством голосования по тестам ошибочно. Процедура голосования может быть улучшена, если учитывать тупиковые тесты с разными весами, в зависимости от их длин (коротким тестам следует придавать большее значение). На диагностическую существенность коротких тестов указывалось в работе /3/.

1.2.4. Эффективность тестового подхода существенно зависит от устройства двухклассовой генеральной совокупности, т.е. от генерального разбиения. Между тем в работах по применению под-

хода очень редко встречаются определенно сформулированные гипотезы о генеральном разбиении /4,5,II/. Для некоторых генеральных разбиений (например, описываемых сложными функциями алгебры логики) тестовый подход не может быть эффективным. Пример: генеральное разбиение, задаваемое функцией, известной под названием "счетчик четности" (когда все вершины нечетных ярусов n -мерного куба относятся к одному классу, а все вершины четных ярусов - к другому). Тупиковый подход не будет в среднем эффективным, за исключением вырожденных частных случаев, когда генеральное разбиение задается случайным образом.

1.2.5. Тестовая структура неустойчива по отношению к случайным возмущениям (помехам). Добавление всего лишь одного объекта к обучающей выборке может очень сильно изменить тестовую структуру. В частности, если окажется, что классы А и В пересекаются, список тупиковых тестов становится пустым. Необходимо научиться работать с нечеткими разбиениями и, таким образом, пользоваться нечеткими тестами, различающими не все объекты альтернативных классов.

1.2.6. Выборка А,В (таблица Т) из генеральной таблицы может оказаться непредставительной. В таком случае никакая распознающая процедура не может быть эффективной. Гипотеза о представительности обучающей выборки часто подразумевается, но не формулируется в явном виде.

1.3. Метод перекладки

Пусть дана обучающая выборка $T(A,B)$, разделенная на два класса (А и В), и новый объект С, подлежащий классификации. Присоединив С к классу А, получим таблицу $T_1 = T(AC,B)$; присоединив С к классу В, получим таблицу $T_2 = T(A,BC)$. Каждая из таблиц Т, T_1 , T_2 имеет свою тестовую структуру: TC_0, TC_1, TC_2 , определяемую списками тупиковых тестов ST_0, ST_1, ST_2 и распределениями тестов по длинам. Сравним TC_1 и TC_2 . Выпишем соответствующие распределения тестов по длинам: $r(1), r(2), \dots, r(n)$; $s(1), s(2), \dots, s(n)$; $r(k)$ - количество тестов длины k в TC_1 ; $s(k)$ - количество тестов длины k в TC_2 .

Правило сравнения. Если $r(1) > s(1)$, то будем считать, что

TC_1 проще, чем TC_2 , если $r(I) < s(I)$, то TC_2 проще, чем TC_1 , если $r(I) = s(I)$, переходим к аналогичному сравнению чисел $r(2)$, $s(2)$ и т.д.

Решающее правило. Если TC_1 проще, чем TC_2 , будем считать, что испытуемый (классифицируемый) объект C относится к классу A ; если TC_2 проще, чем TC_1 , будем считать, что C относится к классу B .

Примечание. Из практических соображений (наличие многих "эфемерных" тестов и т.п.) нецелесообразно многократное применение вышеприведенного правила сравнения. Если равенство $r(1) = s(1)$ (причем $r(1) \neq 0$) выполняется два раза подряд, то можно считать, что TC_1 отличается от TC_2 несущественно (TC_1/TC_2); в этом случае принадлежность испытуемого объекта C остается неопределенной.

1.3.1. Другое решающее правило в методе перекладки - это "длина СТ". Возьмем списки тупиковых тестов: ST_0 , ST_1 , ST_2 и обозначим количество их элементов (тестов) через P_0 , P_1 , P_2 соответственно. Возьмем испытуемый элемент C и выполним процедуру голосования по списку ST_0 . Обозначим через T_A , T_B , T_0 количество тестов, "высказавшихся" за принадлежность объекта C к классу A , классу B и воздержавшиеся (соответственно). Список ST_1 отличается от ST_0 отсутствием тестов, голосовавших за принадлежность объекта C к классу A , и наличием дополнительных тестов (N_A штук), появившихся за счет превращения некоторых нетупиковых тестов в тупиковые из-за добавления элемента C . Список ST_2 отличается от ST_0 отсутствием тестов, голосовавших за принадлежность объекта C к классу A , и наличием дополнительных тестов (N_A штук), появившихся за счет превращения некоторых нетупиковых тестов в тупиковые. Поэтому: $P_1 = T_A + T_0 + N_A$; $P_2 = T_B + T_0 + N_B$. Отсюда следует: $P_1 - P_2 = (T_A - T_B) + (N_A - N_B)$.

Разность $(T_A - T_B)$ есть критерий, по которому определяется принадлежность испытуемого объекта с помощью голосования по тупиковым тестам.

1.3.2. Эвристическая гипотеза. Разность $(N_A - N_B)$ "в большинстве случаев" должна быть относительно малой. Эта гипотеза может быть обоснована, если задать подходящую двухклассовую модель генеральной совокупности и формирование обучающей выборки. Таким образом, предполагается, что для большинства практических

задач $P_I - P_2 = T_A - T_B$. Отсюда следует диагностическое правило. Если $P_I - P_2 > 0$, испытуемый объект относится к А; если $P_I - P_2 < 0$, объект относится к В. Распознавание по такому правилу есть модификация распознавания с процедурой голосования по тупиковым тестам. Поэтому подход, включающий вышеизложенный "метод перекладки", является модификацией тестового подхода с голосованием по тупиковым тестам.

Возникает естественный вопрос о сравнении подхода, включающего вышеописанный метод перекладки, с тестовым подходом. Если обрабатывать произвольные данные, не накладывая никаких ограничений на генеральные совокупности, таблицу, разбиение и обучающую выборку, то следует ожидать, что эффективность названных вариантов тупикового подхода будет в среднем примерно одинаковой и не очень высокой. В частности, нетрудно показать, что для генерального разбиения, задаваемого "счетчиком четности", а также для случайных разбиений метод с перекладкой столь же неэффективен, как и обычный по голосованию. Но тестовый подход с методом перекладки обладает, так сказать, большей эвристической силой. Это означает, что для метода перекладки легче выделить класс "выгодных" ситуаций генеральных совокупностей, таблиц, разбиений и обучающих выборок, для которых он будет (в среднем) эффективнее других распространенных методов (в частности, метод перекладки и голосования). Кроме того, ограничиваясь "областью эффективности" метода перекладки, можно наметить некоторые (ранее не использованные) возможности усовершенствования тестовой методики. В качестве примера выгодной ситуации для тестового подхода с методом перекладки можно привести "крупноблочное" генеральное разбиение с представительной выборкой. Для этого необходимо ввести несколько новых понятий.

Пусть n - количество бинарных признаков. Генеральное разбиение можно задавать с помощью характеристических функций так, что

- если $f(x_1, \dots, x_n) = 0$, точка (x_1, \dots, x_n) принадлежит классу А;
- если $f(x_1, \dots, x_n) = 1$, точка (x_1, \dots, x_n) принадлежит классу В.

Если характеристическая функция (принадлежности) зависит только от k -переменных ($k \leq n$), будем говорить, что в генеральной таблице существует закономерность k -го порядка. Имея в виду

практические цели, будем считать, что порядок закономерности (k) значительно меньше размерности куба (n). Наличие закономерности k -го порядка соответствует покрытию классов кубами размерности $n-k$. Переменные, от которых существенно зависит функция принадлежности, определяют k -мерный куб – "носитель закономерности". Обучающая выборка считается представительной, если каждая точка носителя закономерности представлена в ней не менее одного раза или (что то же самое) если каждый из покрывающих классов A , B кубов представлен в обучающей выборке хотя бы одним элементом

I.4. Тестовая модель с крупноблочной структурой классов

Предполагается, что генеральная совокупность генеральной таблицы представляет собой объединение двух непересекающихся классов (A и B), каждый из которых покрывается максимальными интервалами размерности $n-k$. Таким образом, постулируется наличие закономерности k -го порядка. Это значит, что существует k -мерный куб (носитель закономерности), в проекции на который классы A и B не пересекаются. Причем имеется в виду, что среди кубов, обладающих таким свойством, выбирается куб с минимальной размерностью. Переменные (признаки), образующие носитель закономерности составляют основной или определяющий тест. В генеральной таблице этот тест является единственным. С точки зрения приложений желательно, чтобы размерность носителя (порядок закономерности) была существенно меньше n . Иначе методика распознавания по малой обучающей выборке в рамках тестового подхода будет неэффективной.

I.4.I. Образуется малая выборка из генеральной таблицы (вообще говоря, с повторяющимися строками). Эта обучающая выборка должна иметь представителей каждого класса. Желательно, чтобы она была представительной и чтобы в ней содержались все точки носителя закономерности, а это означает, что при проектировании на носитель закономерности выборка должна покрывать его полностью. Допустимо, чтобы обучающая выборка покрывала не все, но большую часть точек носителей закономерности, т.е. чтобы она была близка к представительной. Список тестов обучающей выборки при отсутствии помех содержит основной ("определяющий") тест

генеральной таблицы, а также в большинстве случаев много дополнительных, "эффемерных" тестов, появляющихся случайным образом ввиду малости выборки.

1.4.2. Допускается наличие помехи, которая состоит в том, что в каждом классе обучающей выборки присутствует небольшое количество "чуждых" объектов (взятых из другого класса в генеральной таблице). При наличии хотя бы одного "чуждого" объекта тестовая структура обучающей выборки существенно меняется. Исчезает основной (определяющий) тест генерального разбиения. Вместо него появляется серия дополнительных (эффемерных) тестов, содержащих признаки, входившие в основной (определяющий) тест. Информационные веса частоты вхождения в список тестов основных признаков (т.е. таких, которые входили в основной (определяющий) тест) должны быть относительно большими. Это дает возможность восстановить потерянный за счет помехи определяющий тест на основе анализа информационных весов, вычисляемых по списку тестов.

Задача состоит в том, чтобы при сделанных допущениях найти (или восстановить) основной (или "определяющий") тест и, таким образом, выявить скрытую закономерность, содержащуюся в исследуемых описаниях геологических объектов.

2. Некоторые обобщения и взаимосвязи подхода

Как на необходимость обобщений, так и на существование взаимосвязей тестового подхода с другими методами указывалось в более ранних работах /1,8,13,16/. Здесь мы более подробно остановимся на обобщениях, примыкающих к вопросам попарного сравнения, которые возникают при сравнительном изучении описаний объектов. Рассматриваются также задачи масштабирования признаков и тонкости, связанные с преобразованием исходных признаков в бинарные с помощью приема "ε-различимости" /16,17/. Эти задачи тесно связаны с решением практически важной проблемы о способах преобразования исходных данных, представленных разнотипными признаками. Дано сопоставление тестового подхода с линейными статистическими методами.

2.1. Виды сравнений и отношений

Как уже неоднократно указывалось, в задачах сравнительного изучения фундаментальное значение имеет попарное сравнение объектов, которое производится как по отдельным признакам ("частное сравнение"), так и по всей совокупности учтенных признаков ("общее сравнение").

Частное сравнение двух объектов A_1 и A_2 может привести к одному из трех исходов (по k -му признаку):

- 1) сходство или близость A_1 и A_2 : $S_k(A_1, A_2) = +1$;
- 2) различие или удаленность A_1 и A_2 : $S_k(A_1, A_2) = -1$;
- 3) несравнимость A_1 и A_2 : $S_k(A_1, A_2) = 0$.

Несравнимость фиксируется, например, когда нечего сравнивать из-за отсутствия данных по k -му признаку для одного из объектов. Для бинарных признаков комбинации $(0,0)$, $(1,1)$ означают сходство, а комбинации $(0,1)$, $(1,0)$ - различие.

Общее сравнение $S(A_1, A_2)$ есть некоторая функция частных сравнений. Наделим каждый признак весом P_k , и пусть h^+ (соответственно h^- , h^0) составляет сумму весов P_k , распространенную на признаки x_k , для которых выявлено разнообразие частного сравнения (сходство, различие, несравнимость). Зададим пороги h_0^+ , h_0^- , h_0^0 . Если $h^+ > h_0^+$ ($h^- > h_0^-$, $h^0 > h_0^0$), то утверждается, что имеется общее сходство объектов A_1 и A_2 (различимость, несравнимость). Конечно, пороги следует задавать с расчетом, чтобы обнаружение общей несравнимости исключало различие или сходство, но два последние свойства могут сосуществовать. В зависимости от сравнительной важности этих свойств для работы с конкретной информацией применяются P , Q или N -модели /4/.

Названные модели предполагают наличие измерений по каждому признаку (для тестового подхода предпочтительна бинарная шкала). Парные сравнения не обязательно опираются на детализацию значений признаков (достаточно и меньшей информации), а область применимости тестового подхода можно распространить и на случай признаков, заданных в шкале интервалов.

Суммируем разнообразие задач, подлежащих решению на основе и посредством попарных сравнений, и типизируем задачи следующим образом:

- I. Установление на множестве объектов линейного порядка,

согласованного в некотором смысле с таблицей попарных сравнений (выполненных для всех или для части объектов) и змеющего содержательную интерпретацию (например, соответствие "естественного" порядка объектов задаваемого целевым признаком порядку, обнаруженному алгоритмическим путем). Задачи этого типа возникают при анализе экспертных оценок.

2. Выделение классов (типов) таким образом, чтобы внутри них и между ними выполнялись в соответствующей пропорции заданные соотношения (сходство, различие, несравнимость).

3. Минимизация количества признаков с сохранением заданной системы отношений между опорными* объектами (множество которых может при сильном сходстве объектов совпадать с исходным множеством).

Приведенная типизация задач является узкой в том смысле, что она связана с содержанием не только процедур попарного сравнения, но и с процедурной последовательностью схем тестового подхода. Укажем на наиболее общие варианты отношений между объектами.

В а р и а н т ы		Примечание
внутри классов	между классами	
Несравнимость	Различие	P-модель
Сходство	Различие	
Сходство-различие	Различие	H-модель
Различие-сходство	Несравнимость	$\sim H^+$
Сходство	Различие-несравнимость	$\sim Q$
Несравнимость	Сходство	Q-модель

В некоторых задачах уместно допущение, что отношение несравнимости может переходить в различие или сходство. Эта возможность активно используется при типиковой минимизации с последующим голосованием, при классификации новых объектов.

*) Выделение опорных или "типизирующих" объектов представляет собой трудоемкую самостоятельную задачу, но нередко их можно задать непосредственно.

2.2. Перечень парных сравнений

Проанализируем особенности Q-модели. Естественно считать, что все объекты, связанные отношением сходства, принадлежат одному и тому же классу. Для двух произвольных объектов A, B образуем дизъюнкцию признаков, для которых установлено частное сходство:

$$\omega(A, B) = x_{11} \vee \dots \vee x_{12}.$$

Далее образуем конъюнкцию всех скобок такого вида (для всех пар A, B) и обозначим ее специальным термином — перечень попарных сравнений (ППС):

$$W = \bigotimes_{A, B} \omega(A, B).$$

Пусть h_0^+ , необходимый для констатации общего сходства двух объектов, равен единице.

Определение 1. Полная система представителей (ПСП) для данного перечня парных сравнений есть набор признаков x_{k_1}, \dots, x_{k_2} такой, что любая скобка $\omega(A, B)$, т.е. любой множитель, входящий в ППС W, содержит хотя бы один признак из данного набора.

Определение 2. Тупиковая система представителей (тупиковый тест) для данного перечня парных сравнений есть полная система представителей, переставшая быть таковой при удалении хотя бы одного признака.

Теперь, если ППС представить в развернутом виде раскрытием скобок (умножением) и, полагая, что все признаки являются булевыми переменными, сократить поглощенные термы, то множество оставшихся термов будет взаимно однозначно задавать множество всех тупиковых тестов для данного перечня попарных сравнений. Процедуру получения тупиковых тестов (всех или минимальных) назовем тупиковой минимизацией (полной или частичной). Нередко в соответствии с целями обработки таблиц бывает полезно среди тупиковых тестов выделить

Все ПСП (в нашем случае тупиковые тесты) сохраняют структуру Q-модели. Этот вывод уместен и для P-модели с той лишь разницей, что дизъюнкции $\omega(A, B)$ должны содержать признаки, по которым установлено частное различие. Если же при тупиковой минимизации следует проверить как сходство, так и различие (напри-

мер, в H -модели), то надо построить ППС отдельно по каждому из отношений (сходство и различие), а затем взять их произведение:
 $W = W_1 \cdot W_2$.

Краткий набросок одной из многих схем тупиковой минимизации, который мы привели, не содержит сведений относительно алгоритмических трудностей, связанных с большим перебором. Существует несколько экономных алгоритмов развертывания ППС (раскрытия скобок с учетом поглощений), в которых перебор существенно сокращен (по сравнению с приведенной здесь принципиальной схемой без искусственных приемов). После того, как получен список тупиковых тестов, для определения принадлежности нового объекта к заданному классу применяется процедура голосования, описанная в разных вариантах многими авторами.

2.3. "Эпсилон-различимость"

Для обработки таблиц, в которых признаки объектов заданы последовательностью вещественных чисел (т.е. заданы в шкале интервалов), зачастую прибегают к процедуре преобразования исходных признаков в бинарные. Это преобразование производится в простейшем случае по принципу "больше-меньше" посредством введения пороговых значений для каждого из признаков. Такие процедуры загромождают данные и ведут к значительной потере детальной информации. В целом же в таком загромождении зачастую необходимости нет. Поскольку для парных сравнений достаточно иметь дифференциальные пороги ϵ_k по каждому признаку, то имеет прямой смысл проводить информационное исследование таблиц без потерь информации бинарных вариантов.

Рассмотрим пример R -модели, для которой при тупиковой минимизации следует сохранить систему различий. Условимся, что для регистрации общего различия между какими-нибудь двумя объектами достаточно иметь хотя бы одно частное различие. Далее, и это главное, примем, что произвольно выбранные объекты A, B имеют по признаку x_k частное различие в том случае, когда выполняется неравенство:

$$|x_k(A) - x_k(B)| > \epsilon_k$$

Этих условий достаточно для составления перечня парных

сравнений (ППС) и последующей тестовой минимизации. Назовем этот вид расширения тестового подхода "методом ϵ -различимости" или, принимая во внимание характер определения частного сходства, "методом ϵ -порогов". Длина списка тупиковых тестов зависит от вектора $\tilde{\epsilon}$ -порогов ($\epsilon_1, \dots, \epsilon_n$). Если увеличить синхронно все пороги ϵ_k , то, начиная с нуля, список тестов сначала будет расти, а потом, когда n -мерная точка $\tilde{\epsilon}$ пересечет некоторую границу в пространстве признаков, список тестов начнет сокращаться и "сойдет на нет", когда параметры ϵ_k станут достаточно большими. Таким образом, существует ряд эффективных (для диагностики) векторов $\tilde{\epsilon}$, и подходящий (не обязательно оптимальный) вектор $\tilde{\epsilon}$ может быть установлен с помощью одного из численных методов поиска, применяемых в планировании экспериментов.

2.4. Сопоставление тестового подхода с линейными статистическими методами

Такое сопоставление полезно не только в смысле углубления содержательных трактовок процедур тестовых алгоритмов, но и в смысле поисков процедур, полезных для уточнения результатов статистических решений.

1. Если исходные данные подвержены случайным помехам, и качество тестовой диагностики резко снижается с увеличением массива данных, то возникает много случайных пересечений классов, из-за чего строгое разделение становится невозможным. В подобных ситуациях применение тестового подхода становится затруднительным.

2. Линейные статистические методы (к примеру дискриминантный анализ) могут дать разделение пересекающихся классов, хотя и не строгое. Хорошее разделение достигается только для массивов, имеющих относительно простую структуру, когда, например, точки в многомерном пространстве сгруппированы в небольшое количество сгущений ("кластеров").

Характерно, что для статистической обработки отмечается повышение надежности результатов с увеличением выборки, а количество вычислительных процедур растет медленно (обычно пропорционально первой степени числа объектов). В случае же тестового

подхода отмечается, что с ростом выборки сокращается список тупиковых тестов, а вместе с этим убывает и надежность операций, связанных с пропорциональным количеством тестов (например, голосование по тестам). Кроме того, при тестовой обработке таблиц в ряде алгоритмов объем вычислений растет по экспоненте с возрастанием объема таблиц. Преодолеть эти трудности можно по следующим направлениям:

1. Видоизменить понятие полной системы представителей. Для этого следует потребовать, чтобы не все скобки имели своего представителя в ПСП, а только большая их часть. В связи с этим изменится развертывание перечня попарных сравнений. В результате такого требования достигается "почти-независимость" списка тупиковых тестов от объема выборки. Следует отметить, что задача в такой постановке еще ни кем не решалась.

2. Чередовать применение линейно-статистических и тестовых методов. С помощью статистических (и тестовых в отдельных случаях) можно выделить малое подмножество опорных объектов.

3. В некоторых модификациях задач можно отказаться от процедуры голосования по тестам, заменив ее диагностикой по методу потенциальных функций - например, с евклидовым расстоянием в экспоненте. Параметры же потенциальной функции (масштабные коэффициенты по каждой координате) могут находиться путем анализа перечня попарных сравнений с помощью упрощенного эвристического алгоритма без выписывания тупиковых тестов. Сначала отыскивается признак, имеющий наибольшее число вхождений в ППС, затем вычеркиваются скобки, в которые входит этот признак; оставшаяся часть ППС обрабатывается так же и т.д. - до вычеркивания всех скобок. Максимальные частоты вхождений признаков (на каждом этапе алгоритма) устанавливают своеобразную меру информативности признаков. Эти меры (числа) можно использовать для конкретизации потенциальной функции.

Намеченные пути представляют собой сближение (и взаимное дополнение) тестовых и линейно-статистических методов. Что же связывается с конкретным применением этих разнотипных методов, так сказать, в их "чистом виде"? Согласно накопленному опыту уместно сделать следующее заключение.

Линейно-статистические методы выгоднее применять для таблиц большого объема (особенно по числу объектов). При наличии силь-

ных случайных помех, когда не исключено пересечение представителей классов (даже в обучающей выборке), генеральная совокупность является достаточно простой, такой, что большинство разнотипных объектов все же можно подразделить одной-двумя гиперплоскостями.

Тестовый подход оказывается в ряде случаев намного эффективнее, особенно когда подлежащая исследованию представительная обучающая выборка мала, помехи отсутствуют (или почти отсутствуют), а классы не пересекаются. Однако данные, характеризующие объекты, обычно имеют весьма сложную пространственную структуру /15/, и для разделения классов необходимо строить сложную поверхность. Такие алгоритмы можно применять для обработки таблиц с пропусками, т.е. когда часть замеров отсутствует. Наличие пропусков приводит к потере некоторых дизъюнктивных слагаемых в скобках перечня попарных сравнений, но принципиальная схема работы остается без изменений.

2.5. Масштабирование признаков

При попытке конкретного применения "ε-различимости" в задачах соответствующего профиля возникают серьезные препятствия, связанные с многопараметричностью. Преодоление усматривается нами в построении процедур по соответствующему (точному в смысле требований метода) выбору масштабирования значений характеристических признаков.

Пусть, например, информационные веса признаков (зависящие от масштабирования) между собой равны. Тогда в методе "ε-различимости" можно потребовать условия, чтобы все пороги тоже были одинаковыми: $\epsilon_1 = \dots = \epsilon_n = \epsilon$, а задача таким путем преобразуется в однопараметрическую. Конечно, такое требование имеет сугубо эвристическое происхождение и может в некоторых разновидностях задач привести к результатам, далеко отстоящим от оптимальных. Однако оно, по крайней мере, не хуже других неформальных допущений, применяемых в задачах распознавания, к тому же и сами критерии оптимальности формулируются в весьма условных обстоятельствах.

При наличии нескольких классов или одного нормировка раз-

личная. Для случая с одним классом вполне уместна простейшая нормировка признаков, при которой среднее квадратическое значение каждого делается равным единице. При наличии же нескольких классов такая нормировка может оказаться несостоятельной, поскольку она может усложнить и смазать границу между классами. Возможность выбора масштабов, усиливающих контраст исследуемых объектов, можно рассмотреть на двух примерах.

1. Пусть задана таблица значений характеристических признаков объектов двух классов X и Y так, что x_{ik} (y_{jk}) есть значение k -го признака на i -ом объекте из X (на j -ом объекте из Y). Различие между классами X и Y будем характеризовать функционалом:

$$Q = \frac{1}{n} \sum \log(1 + z_k a_k); \quad a_k^2 = \text{const} \cdot \frac{\sum (x_{ik} - x_{jk})^2}{\sum (x_{ik} - x_{jk})^2}$$

Символ z_{ik} есть общее обозначение для x_{ik} и y_{jk} . В знаменателе выражения для a_k^2 производится суммирование по всем парам объектов, взятых из объединения классов $z = XY$. Величины r_k суть весовые (масштабные) коэффициенты признаков. Будем искать положительные значения r_k при условии $r_k = C$, обращающие Q в максимум.

Применяем обычную технику с множителем Лагранжа:

$$\begin{aligned} \tilde{Q} &= Q + \lambda (\sum z_k - C); \quad \frac{\partial \tilde{Q}}{\partial z_k} = 0, \\ z_k &= \lambda^{-1} - a_k^{-1}; \quad n \lambda^{-1} = C + \sum a_k^{-1}, \\ z_k &= \frac{1}{n} (C + \sum a_k^{-1} - a_k^{-1}); \quad Q_{\max} = \frac{1}{n} \sum \log \left[1 + \frac{a_k}{n} (C + \sum a_s^{-1} - a_k^{-1}) \right] \end{aligned}$$

Отметим, что подбором константы можно обеспечить положительные значения всех коэффициентов r_k .

2. Для классов X , Y задана мера сходства:

$$S = \frac{1}{n} \sum \exp(-z_k a_k); \quad a_k^2 = \text{const} \cdot \frac{\sum (x_{ik} - y_{jk})^2}{\sum (x_{ik} - x_{jk})^2}$$

(как и в предыдущем примере $z_{ik} \in XY$). Следует найти коэффициенты r_k ($r_k > 0$, $\sum r_k = C$), минимизирующие S . Имеем

$$\bar{z} = s + l^{\lambda} (\sum z_k - c); \quad \frac{\partial \bar{z}}{\partial z_s} = 0,$$

$$\dot{z}_s = \lambda a_s^{-1} + a_s^{-1} \log a_s; \quad -\lambda \sum a_k^{-1} + \sum a_k^{-1} \log a_k.$$

Выбором константы c (а также и нормирующего множителя в выражении для a_k^2) можно добиться, чтобы множитель λ был отрицательным, а все масштабные коэффициенты r_s — положительными.

Помимо выявления и организации малого подмножества опорных объектов, имеется и другая, пока неиспользованная возможность улучшения исходного массива данных для диагностических целей — подчеркивание контуров, ограничивающих кластеры (сгущение объектов, выявленное по их описаниям). Для этого можно подвергнуть исходный информационный материал предварительной обработке посредством алгоритмов цифровой голографии. Для этого, конечно, следует модифицировать алгоритм соответственно заданной цели. Известно, что цифровые голограммы Фурье, полученные без рассеивателя, обладают заметным дифференцирующим эффектом, который проявляется в получении контурного изображения вместо силуэтного, причем этот эффект можно получить не только для двумерных, но также и для многомерных изображений. После такой голографической обработки можно устранить объекты, лежащие внутри контуров и, таким образом, снизить объем информации.

Литература

1. Васильев Ю.Л., Дмитриев А.Н. Спектральный подход к сравнению объектов, охарактеризованных набором признаков. — Докл. АН СССР, 1972, т.206, № 6, с.1309-1312.
2. Вышемирский В.С., Дмитриев А.Н., Трофимук А.А. Поискные признаки гигантских нефтяных месторождений. — В кн.: УШ Мировой нефтяной конгресс. М.: Наука, 1971, с.3-16.
3. Дмитриев А.Н. Использование длин тупиковых тестов при обработке таблиц. — В кн.: Дискретный анализ. Новосибирск: Наука, 1970, вып.17, с.18-21.
4. Дмитриев А.Н. Новые тестовые разработки в задачах прогнозирования рудоносности (на примере трапповых интрузий). — В кн.: Математические методы при прогнозе рудоносности. М.: Наука, 1977, с.104-163.

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

6. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. Об одном принципе классификации и прогноза геологических объектов и явлений. - Геол. и геофиз., 1968, № 5, с.50-64.

7. Дмитриев А.Н., Кренделев Ф.П., Бишаев А.А. и др. Организация и обработка геологической информации с помощью ЭВМ на основе построения тупиковых тестов. - В кн.: Логико-информационные решения геологических задач. М.: Наука, 1975, с.83-128.

8. Дмитриев А.Н., Смертин Е.А. Связь тестовых параметров с повторяемостью столбцов. - В кн.: Первая Всесоюзная конференция по проблемам теоретической кибернетики. (Тез. докл.). Новосибирск, 1969. с.104

9. Еремеев А.Н., Модников И.С., Писаревский С.В. и др. Принципы крупномасштабного прогнозирования редкометалльного сруднения в вулканогенных формациях с применением ЭВМ. - В кн.: Математические методы при прогнозе рудоносности. М.: Наука, 1977, с.164-192.

10. Журавлев Ю.И. Об алгоритмическом подходе к решению задач распознавания или классификации. - В кн.: Проблемы кибернетики. М.: Наука, 1978, вып.33, с.4-68.

11. Константинов Р.М. Математические методы количественного прогноза рудоносности. М.: Недра, 1979. 125 с.

12. Константинов Р.М., Королева Э.Е., Кудрявцев Б.О. О комбинаторно-логическом подходе к задачам прогноза рудоносности. - В кн.: Проблемы кибернетики. М.: Наука, 1976, с.5-34.

13. Логико-математическая обработка геологической информации (теория и математический аппарат). Новосибирск: ИГиГ СО АН СССР, 1976. 152 с.

14. Логико-информационные исследования в геологии. Новосибирск: ИГиГ СО АН СССР, 1977. 146 с.

15. Метод согласованных оценок. Новосибирск: ИГиГ СО АН СССР, 1982. 144 с.

16. Макаров С.В. Логико-вероятностный метод типизации многомерных объектов. - В кн.: Математические проблемы геофизики. Новосибирск: Наука, 1971, вып.2, с.330-344.

17. Макаров С.В., Мерекин Ю.В., Шумилов Ю.С. О классификации экспериментальных кривых на ЭВМ. - Там же, с.345-355.

ИВ. Чегис И.А., Яблонский С.В. Логические способы контроля работы электрических схем. — Труды Математического ин-та им. В.А.Стеклова. М., 1958, т.51, с.270—339.

С.В.Макаров

ВОССТАНОВЛЕНИЕ ОДИНОЧНОГО ПРОПУСКА
В ТАБЛИЦЕ "ОБЪЕКТ—ПРИЗНАК"
НА ОСНОВЕ СИНГУЛЯРНОГО РАЗЛОЖЕНИЯ ДАННЫХ

Предлагаемый метод представляет собой вариант метода гребневой регрессии, отличающийся от известных вычислительной схемой и способом регуляризации.

Дана таблица "объект—признак" \hat{X} размера $(m+1) \cdot (n+1)$, в которой i -ая строка соответствует i -му объекту, k -й столбец соответствует k -му признаку и, таким образом, элемент x_{ik} есть значение k -го признака на i -м объекте. Предполагается, что один элемент, расположенный для определенности в правом нижнем углу, неизвестен. Задача состоит в том, чтобы оценить неизвестное значение.

Необходимые математические сведения даются в приложении к статье (см. также /1/).

Обозначения:

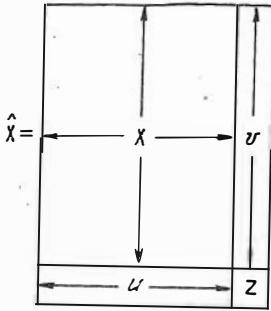
1. x есть таблица, получаемая из \hat{X} удалением $(n+1)$ -го столбца и $(m+1)$ -й строки.
2. Пропущенное значение (пропуск): $z = x_{m+1, n+1}$.
3. Сокращенный столбец, получаемый из последнего столбца таблицы \hat{X} удалением элемента z :

$$v = (x_{1, n+1}, \dots, x_{m, n+1})^T$$

4. Сокращенная строка ("неполный объект"), получаемая из последней строки таблицы \hat{X} удалением элемента z :

$$u = (x_{m+1, 1}, \dots, x_{m+1, n}).$$

Исходная таблица данных разбивается на четыре части: X , u , v , z (см. рисунок).



Возьмем сокращенный вектор-столбец \tilde{v} , соответствующий $(\bar{n}+1)$ -му признаку, и рассмотрим его гребневую оценку

$$\tilde{v} = X [X^T X + \varrho^2 I]^{-1} X^T v. \quad (1)$$

Здесь ϱ^2 - параметр регуляризации, подбираемый экспериментально [2]. При $\varrho = 0$ оценка (1) переходит в регрессионную оценку по методу наименьших квадратов:

$$\tilde{v} = X (X^T X)^{-1} X^T v. \quad (2)$$

По теореме 1 (см. приложение) имеет место приближенное равенство:

$$[X^T X + \varrho^2 I]^{-1} X^T \approx X^+, \quad (3)$$

где X^+ - псевдообратная матрица. Из (1), (3) следует:

$$\tilde{v} \approx X X^+ v. \quad (4)$$

Предполагая, что приближенная линейная зависимость, выражаемая формулой (4), сохраняет силу также для последней (неполной) строки в таблице \tilde{X} , получаем искомую оценку пропуска

$$z = u \cdot \tilde{X}^+ \cdot v. \quad (5)$$

Примечание. Если формулу (4) понимать как точное равенство, она может рассматриваться как новое (видоизмененное) определение гребневой оценки

$$\tilde{v} = X X^+ v. \quad (6)$$

Это определение, приводящее к несколько иным (сравнительно с (1)) численным оценкам, является более удобным для теоретических выкладок.

Рассмотрим вопрос о нахождении матрицы X^+ (или некоторого приближения к ней). По теореме 2 (см. приложение) таблица \tilde{X} может быть представлена в виде:

$$\tilde{X} = PL^{1/2} Q^T. \quad (7)$$

Матрица L - диагональная; на главной диагонали стоят ненулевые собственные числа матрицы $\tilde{X}^T \tilde{X}$, образующие спектр $(\lambda_1, \dots, \lambda_r)$. Будем считать, что собственные числа располагаются в убывающем порядке.

Матрицы L , P , Q связаны соотношениями

$$XX^T P = PL, X^T X Q = QL, P^T P = I, Q^T Q = I. \quad (8)$$

Представление (7) называется сингулярным разложением. Для его реализации (т.е. для нахождения матриц L , P , Q) существует много специальных методов.

Имея L , P , Q , легко найти псевдообратную матрицу X^+ :

$$X^+ = QL^{-1/2}P^T. \quad (9)$$

После этого можно вычислить гребневую оценку (5). Однако ввиду того, что некоторые собственные числа λ_1 могут быть (относительно) очень малыми (сравнительно с главным собственным числом), получаемые оценки могут оказаться весьма неустойчивыми. В таком случае рекомендуется использовать приближенное сингулярное разложение с учетом только "существенно ненулевых" собственных чисел.

Обозначения:

L_k - диагональная матрица k -го порядка, по диагонали которой располагается существенно ненулевая часть спектра ($\lambda_1, \dots, \lambda_k$); ($k \leq r$).

P_k, Q_k - матрицы, образованные первыми k столбцами матриц P, Q соответственно.

Приближенное сингулярное разложение:

$$X \approx P_k L_k^{1/2} Q_k^T. \quad (10)$$

Приближенное псевдообращение:

$$X_k^+ = Q_k L_k^{-1/2} P_k^T. \quad (11)$$

Гребневая оценка вектора \tilde{v} (по аналогии с (6)):

$$\tilde{v}^{(k)} = X_k^+ \tilde{v}. \quad (12)$$

Наконец, распространяя зависимость вида (12) на неполный объект, получаем оценку пропуска

$$\tilde{z}^{(k)} = \mu X_k^+ \tilde{v}. \quad (13)$$

Целочисленный параметр k является эквивалентом непрерывного параметра η^2 в формуле (1) и также подбирается экспериментально. При $k = r$ оценка (13) совпадает с оценкой (5).

Отметим, что гребневые оценки являются смещенными /2/.

При удачном выборе параметра регуляризации достигается повышенная устойчивость при относительно небольшом смещении. Регуляризация посредством параметра k позволяет проследить эффект "сжатия информации", подобно тому, как это делается в методе главных компонент.

Примечание. В работе /3/ рассматривается "метод сдвига" для решения плохо обусловленных систем уравнений. Правая часть системы $A_k x = y$ заменяется выражением:

$$\tilde{y} = A A_k^+ y. \quad (I4)$$

Формула (I4) совпадает с (I2).

П р и л о ж е н и е

МАТЕМАТИЧЕСКАЯ СПРАВКА

(По А.Алберту /I/)

Теорема I. Пусть A - произвольная матрица размера $m \times n$. Существует матрица A^+ определяемая соотношениями

$$A^+ = \lim_{\varepsilon \rightarrow 0} (A^T A + \varepsilon^2 I)^{-1} A^T = \lim_{\varepsilon \rightarrow 0} A^T (A^T A + \varepsilon^2 I)^{-1}, \quad (\varepsilon \rightarrow 0).$$

Матрица A^+ называется псевдообратной (по отношению к A) матрицей Мура-Пенроуза.

Псевдообратная матрица A^+ определяется также свойствами

$$A A^+ A = A, \quad A^+ A A^+ = A^+, \quad A^+ A \text{ и } A A^+ \text{ - симметричны.}$$

Частные случаи. 1) если A - квадратная невырожденная матрица, то $A^+ = A^{-1}$, 2) если строки A линейно независимы, $A^+ = A^T (A A^T)^{-1}$, 3) если столбцы матрицы A линейно независимы, $A^+ = (A^T A)^{-1} A^T$.

Теорема 2. Пусть A - матрица размера $(m \times n)$ и L - диагональная матрица размера $(r \times r)$, диагональ которой состоит из ненулевых собственных чисел матрицы $A A^T$. Существуют матрицы P $(m \times r)$ и Q^T $(r \times n)$ такие, что выполняются равенства

$$A = P L^{1/2} Q^T, \quad A A^T = P L P^T, \quad A A^T = Q L Q^T, \quad P^T P = I, \quad Q^T Q = I.$$

Следствие.

$$A^+ = Q L^{-1/2} P^T$$

Литература

1. Алберт А. Регрессия, псевдорегрессия и рекуррентное оценивание. М.: Наука, 1977. 220 с.

2. Конев В.С., Товаров А.К. Смещенные (гребневые) оценки в построении экономико-математических моделей. - В кн.: Математико-статистические методы в экономическом анализе. Новосибирск: ВЦ СО АН СССР, 1983, с.67-80.

3. Фаддеева В.Н. Сдвиг для систем с плохо обусловленными матрицами. - Журн. вычислит. мат-ки и математ. физики, 1965, т.5, № 5, с.907-911.

Г.С.Федосеев

ВОССТАНОВЛЕНИЕ ПРОПУСКОВ В ТАБЛИЦАХ ДАННЫХ

Пропуски (неизвестные значения признаков) являются обычным атрибутом большеформатных таблиц данных, подготовленных, например, для решения задач распознавания образов. Количество пропусков отражает полноту охарактеризованности объектов в выбранном пространстве признаков, и если число их достаточно велико (> 30-40 %), то пригодность имеющегося материала для достоверных выводов становится сомнительной. В этой связи перед заказчиком (пользователем) обычно выдвигается требование полной охарактеризованности объектов, во всяком случае в той части таблицы, которая представляет материал обучения. Ситуация аналогична и при автоматической классификации (таксономии). Однако выдвигаемое требование не всегда может быть выполнено, особенно в тех случаях, когда речь идет о месторождениях полезных ископаемых (недоразведанность, отсутствие комплексности в исследованиях, частичная или полная отработка месторождений и т.п.). Поэтому при математическом решении конкретной задачи нередко приходится решать и вопрос о пропусках, т.е. о неизвестных или неизмеренных данных /5,6,7 и др./. Существенное значение имеет также и то, что большинство алгоритмов распознавания образов не может работать на материале с пропусками.

Начекается несколько путей решения этой проблемы: от самого радикального - устранение строк и (или) столбцов, содержащих пропуски /1/, до наиболее "лояльного" - не требующего восстановления неизвестных значений /2,3,7/. Сокращение объема таблицы не всегда может быть реализовано, во-первых, ввиду возможного малого числа объектов i , во-вторых, из-за наличия пропусков в каждой строке. Этот прием ведет к нарушению представительности выборки, а, следовательно, к возможному смещению математических оценок.

Второй путь предусматривает восстановление пропусков за пределами таблицы обучения. В наиболее часто встречающемся случае - это определение значений целевого признака, т.е. решение задачи прогнозирования. При этом имеющиеся в исходной таблице пропуски считаются допустимым элементом (в алфавите первоначально бинарного кода). Они вводятся в структуру данных в качестве самостоятельного элемента описания, соответствующего тому или иному объекту.

Промежуточное положение между отмеченными путями занимает способ заполнения пропусков в самой таблице данных, цель которого не устранить проблему, а напротив, найти ее решение. В решении этой задачи можно назвать, по крайней мере, две возможности /7/. В первом случае пропуски восстанавливаются вне зависимости от характера дальнейшей обработки матрицы данных с помощью регрессии переменных, главных компонент, факторного анализа и других методов. Основное внимание при этом уделяется поиску закономерностей внутри таблицы. Во втором случае признаковое пространство расширяется за счет целевого признака. Это означает, что решаемая задача уже известна, и на исходные данные как бы накладывается "целевая ориентация", а восстановление признаков производится с учетом цели. Поэтому каждый пропуск в таблице может принимать различные значения в зависимости от характера цели, (многоцелевая постановка), т.е. структура восстановленных данных адаптируется на цель.

Вообще говоря, без наличия дополнительных предложений относительно зависимостей между признаками (функциональных или статистических) задача становится неопределенной. Даже для аппроксимации пропуска средним значением или при оценке парной и множественной корреляции в явном или неявном виде вводятся предпо-

ложение о представительности выборки. Стремление же к сохранению и использованию всей имеющейся информации вынуждает осуществлять поиск закономерностей в пределах таблицы или ее частей. Применение алгоритмов связано с реализацией правдоподобных гипотез или предположений. В алгоритме ЗЕТ, например, используется два предположения - о "линейной избыточности" (наличие парной связи строк) и "компетентности" (наличие выявляемого сходства) /4/.

Эта задача может решаться в нескольких вариантах: с учетом информации только по признаку, имеющему прочерк, или по двум признакам, один из которых не имеет прочерков и, наконец, учитывать информацию о всей таблице. Для того, чтобы наметить стратегию работы с таблицами, содержащими пропуски, рассмотрим максимально упрощенные случаи сочетаний пропусков у целевого и характеристических признаков (рис.1). В варианте "а" при заполнении пропуска с помощью одного алгоритма имеется единственное решение. Во втором варианте (рис.1, б) их уже несколько, в третьем (рис.1, с) - еще больше. Обобщенный случай представлен четвертым вариантом. Очевидно, что число возможных решений неуклонно возрастает с ростом числа пропусков. Существенную роль при выборе удовлетворительного пути решения задачи будут играть последовательность заполнения пропусков, учет или неучет восстановленных значений и характер сокращения обучающей выборки (только по строкам, только по столбцам или в сочетании).

При наличии нескольких пропусков принципиально важным является выбор способа восстановления - последовательного или одновременного. Первый способ может быть осуществлен в двух вариантах - с учетом восстановленных значений и без них. В процедурном отношении способы могут быть одноактными (неитеративными) или многоактными (итеративными). Из последних предпочтение отдается сходящимся процедурам.

Известно, что статистические закономерности могут быть выявлены при достаточно большом числе объектов, но во многих реальных задачах это условие часто не выполнимо. Далее, если в статистическом варианте принятие решения осуществляется по частоте встречаемости тех или иных значений признака, то восстановление прочерков с учетом характера распределения является, по существу, прогнозом по аналогии. Неопределенность решений

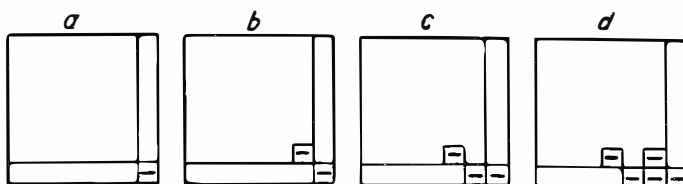


Рис. I. Простейшие варианты сочетания пропусков у целевого и характеристических признаков:

- а - пропуск значения целевого признака у пробы; б - пропуски значений целевого признака у пробы и характеристического - у эталона; в - пропуски значений целевого и характеристического признаков у пробы и характеристического (не совпадающего с пропуском у пробы) у эталона; д - общий случай

таким подходе состоит в том, что в случае независимых признаков восстановленные значения будут одинаковыми для всех испытуемых объектов (проб). Для случая с невыясненным распределением вопрос о восстановлении пропуска остается открытым. Очевидно также, что статистический подход к поиску закономерностей по отдельно рассматриваемым строкам лишен смысла, если не известен характер взаимозависимости признаков. В противном случае речь может идти только о восстановлении пропусков у так называемых выводимых признаков, т.е. получаемых аналитически, путем вычисления функций. Аналогичным примером тривиального восстановления пропуска является строка приведенных значений признаков (приведенных, например, к единице, 100 % и т.п.).

Основанием для создания алгоритмов эвристического подхода является предположение о том, что значения признаков у конкретных (в частности, у геологических) объектов не являются случайными, а обусловлены скрытыми взаимосвязями, не проявленными в очевидной форме. Именно выявление скрытых закономерностей и служит основанием для создания и применения операций по восстановлению пропусков.

Использование того или иного метода в первом приближении определяется соотношением m (число объектов) и n (число признаков): при $m \gg n$ предпочтительнее статистический подход, при

$m \ll n$ - эвристический, при $m \approx n$ трудно отдать предпочтение какому-либо из подходов. Отметим также, что именно таблицы, приближающиеся к квадратным, являются наиболее "удобными" с точки зрения применения более сложных методов, специально направленных на восстановление пропусков. Все перечисленное выше не только ориентирует поиск подходящих алгоритмов, но и определяет порядок заполнения пропусков. Последнее достигается комбинированием различных методов.

Выбор соответствующего алгоритма может быть осуществлен с помощью контрольного примера, решенного в рамках представленной таблицы данных. Эта процедура должна показать способность таблицы и алгоритма обеспечить получение приемлемых по устойчивости результатов. Для ее ориентирующего решения необходимо выбрать схему покрытия подтаблицы (не содержащей пропусков) системой искусственных пропусков. Здесь нет определенных правил и рекомендаций, и, пожалуй, единственным критерием оценки является число аномальных отклонений от известных значений. Не очевидно также и то, насколько полным должно быть покрытие. Предполагается, что приемлемый по надежности результат можно выбрать при полном покрытии исходной таблицы и варьировании числом удаляемых значений признаков - сначала по одному, затем - по два и т.д. Не последней роль должна, очевидно, играть и равномерность покрытия.

Для решения задачи восстановления пропусков обычно применяются итерационные процедуры: сначала используются какие-либо предварительные оценки пропусков, например, коэффициенты регрессии, а затем поочередно улучшаются оценки пропусков и уточняются коэффициенты регрессии и т.д. Но при этом возникает необходимость проверки, в какой степени конечный результат зависит от начальных приближений, а также от очередности заполнения пропусков. Все это требует дополнительных вычислений.

В данной работе использован метод, представляющий собой разновидность гребневой регрессии с модифицированной вычислительной схемой /6/(*). Для элементов таблицы \hat{X} с единственным пропуском согласно методу вводятся буквенные обозначения (рис.2).

Существенную роль в методике играет псевдообращение и сингулярное разложение (SVD), примененные к прямоугольной матрице

*). См. также работу С.В.Макарова в данном сборнике.

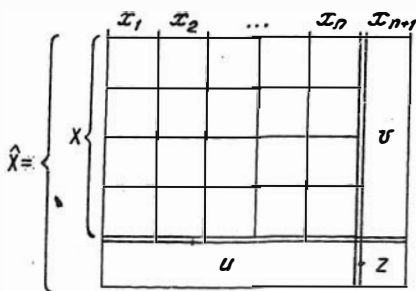


Рис.2. Матрица с одним пропуском:

X - обучающая выборка, v - сокращенный вектор-столбец (неполный целевой признак), u - сокращенный вектор-строка (неполный объект или проба), z - неизвестное значение целевого признака (пропуск)

X размера $m \times n$. Матрица X^+ называется псевдообратной (по отношению к X), если выполняются соотношения

$$XX^+X = X, \quad X^+XX^+ = X^+, \quad X^+X \text{ и } XX^+ - \text{симметричны.}$$

Существенно, что псевдообратная матрица X^+ может быть вычислена по одной из формул:

$$X^+ = \lim (X^T X + \varepsilon^2 I)^{-1} X^T,$$

$$X^+ = \lim X^T (X X^T + \varepsilon^2 I)^{-1}, \quad (\varepsilon \rightarrow 0).$$

ε - параметр регуляризации; X^T - транспонированная матрица; I - единичная матрица.

В частном случае, если столбцы или строки матрицы X линейно независимы, то соответственно $X^+ = (X^T X)^{-1} X^T$ или $X^+ = X^T (X X^T)^{-1}$

В алгоритме используется SVD-разложение таблицы данных

$$X = PL^{1/2}Q^T.$$

При этом квадратные матрицы P (главных компонент) и Q (собственных векторов) и диагональная матрица L (собственных чисел, расположенных в убывающем порядке) находятся из уравнений

$$XX^T P = PL, \quad X^T X Q = QL, \quad P^T P = I, \quad Q^T Q = I.$$

Для решения этих уравнений существует большая серия готовых алгоритмов и программ.

В предложенной С.В.Макаровым модификации гребневой регрессии используется приближенное сингулярное разложение

$$X_K^+ = Q_K L_K^{-1/2} P_K^T,$$

где X_K^+ - приближенная псевдообратная матрица, а P_K, Q_K и L_K

образуются первыми k -столбцами матриц P , Q , L соответственно (k - целочисленный параметр, подбираемый экспериментально).

На материале обучения получают оценки целевого признака

$$\tilde{v} = X X_k^+ v,$$

и предполагается, что эта зависимость сохраняется и для строки u . Значение пропуска находится по формуле

$$z = u X_k^+ v.$$

Отмечается, что при работе по описанной методике вопрос о нормировании или стандартизации не имеет принципиального значения. Однако желательно некоторое выравнивание порядков по таблице в целом (например, один порядок для максимальных значений признаков).

Алгоритм был апробирован на известном примере железорудных месторождений Горного Алтая /8/. В исходную таблицу данных были отобраны только количественные признаки, число объектов определилось материалом обучения. Процедура применения алгоритма носила переборный характер. Во-первых, роль целевого признака последовательно выполнял каждый из характеристических и, во-вторых, искусственным прочерком последовательно заменялись все значения у целевых признаков. В результате получилось полное покрытие таблицы восстановленными значениями. Поскольку первоначальная таблица не содержала пропусков, то, естественно, восстановленные значения в обработку не вовлекались. Эта процедура является своеобразным внутренним экзаменом, поэтому на ее основе можно оценить работоспособность алгоритма. Можно сказать также, какие признаки (или объекты) восстанавливаются с большей надежностью, и для каких (больших, средних или малых) значений это имеет место.

Рассмотрение корреляционных зависимостей между значениями исходных и восстановленных признаков (рис.3,а) показывает*, что целевым признаком с точки зрения восстанавливаемости, а следовательно, и прогнозируемости значений может быть далеко не каждый. Этому условию, согласно величинам коэффициента корреляции, в

* Для унифицированности и компактности графиков введено нормирование по $\{max\}$ по столбцам и строкам (раздельно).

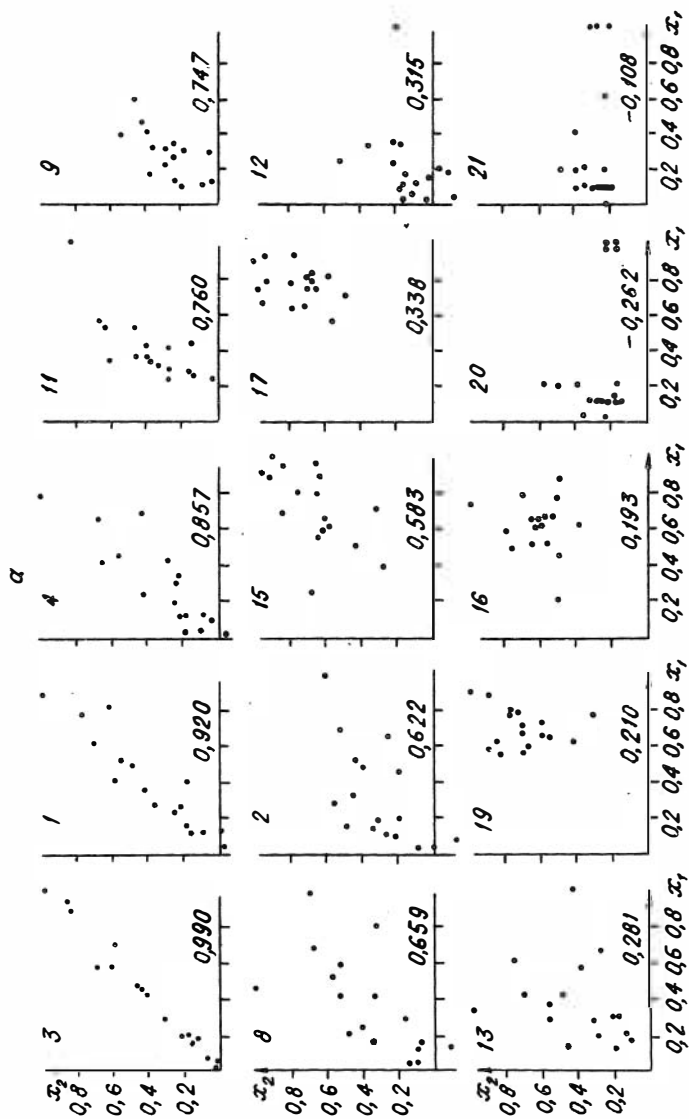
Таблица исходных данных

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	М
I	0,0460	20	12,50	12,50	-1,28	-0,26	105,0	10,40	1525	927	1,64	1,00	130	0,600	0,959	0,590	0,600	0,777	0,825	45	1,0	0,759
	0,0412	6	11,30	11,00	-2,52	-1,12	100,0	15,50	572	333	3,91	2,80	292	0,792	0,842	0,563	0,566	0,550	0,720	20	1,0	0,965
2	0,0480	13	5,90	3,00	-1,70	-1,60	50,0	4,50	970	273	2,90	-2,00	136	0,847	0,842	0,563	0,566	0,590	0,720	20	1,0	0,897
3	0,0300	22	3,00	3,00	-0,30	-0,84	60,0	12,70	852	336	3,40	0,40	470	0,715	0,470	0,563	0,705	0,432	0,605	18	1,7	0,912
4	0,0550	30	7,95	7,73	-0,30	-0,30	10,0	13,22	247	240	1,43	2,70	373	0,610	0,460	0,563	0,640	0,473	0,630	8	1,1	0,930
5	0,0220	146	5,00	3,00	-0,70	-0,05	50,0	8,40	350	170	2,44	2,60	550	1,267	0,184	0,445	0,500	0,275	0,655	15	1,2	0,893
6	0,0502	85	5,90	5,40	-0,45	0,23	50,0	9,40	319	198	1,44	2,60	571	0,266	0,155	0,469	0,507	0,405	0,655	15	1,2	0,893
7	0,0510	37	10,50	7,40	-0,20	-0,21	50,0	23,20	744	815	1,67	4,11	731	0,769	0,305	0,717	0,651	0,715	0,861	16	1,6	0,857
8	0,0338	50	7,00	7,00	-0,20	-0,40	50,0	11,70	480	220	2,18	2,50	410	0,589	0,353	0,472	0,543	0,532	0,735	16	1,4	0,833
9	0,0147	54	3,30	4,50	-0,20	-0,75	40,0	6,00	98	380	1,56	-0,70	212	0,560	0,341	0,563	0,730	0,442	0,770	18	1,1	0,899
10	0,0216	112	3,30	3,30	-0,75	-0,29	50,0	4,00	98	380	1,56	-0,70	212	0,560	0,341	0,563	0,730	0,442	0,770	18	1,1	0,899
I 0	0,0100	124	1,70	1,60	-0,07	-0,10	20,0	2,11	1052	444	2,07	2,00	350	0,500	0,284	0,563	0,668	0,400	0,706	45	2,0	0,893
	0,0118	122	1,69	1,66	-0,07	-0,10	20,0	2,11	1052	444	2,07	2,00	350	0,500	0,284	0,563	0,668	0,400	0,706	45	2,0	0,893
II	0,0090	83	2,32	2,32	-0,40	0	14,8	3,30	600	425	1,45	2,00	474	0,469	0,315	0,506	0,601	0,468	0,610	45	1,8	0,920
	0,0080	89	2,36	2,44	-0,40	0	14,8	3,30	791	514	1,45	2,00	474	0,469	0,315	0,506	0,601	0,468	0,610	45	1,8	0,920
12	0,0061	78	2,20	3,20	-0,35	0,13	19,7	2,20	800	500	1,74	1,12	360	0,647	0,233	0,536	0,625	0,322	0,670	48	2,0	0,942
13	0,0087	35	2,00	3,60	-0,75	-0,05	12,5	5,15	310	240	0,29	1,70	190	0,505	0,432	0,477	0,609	0,580	0,841	8	1,0	0,861
	0,0078	33	1,84	3,60	-0,75	-0,05	12,5	5,15	310	240	0,29	1,70	190	0,505	0,432	0,477	0,609	0,580	0,841	8	1,0	0,861
14	0,0250	5	5,45	6,99	-1,48	-0,80	56,5	18,30	1000	850	1,20	1,80	300	0,650	0,352	0,600	0,825	0,470	0,531	2	1,0	0,788
	0,0114	10	5,45	6,99	-1,48	-0,80	56,5	18,30	1000	850	1,20	1,80	300	0,650	0,352	0,600	0,825	0,470	0,531	2	1,0	0,788
15	0,0156	30	0,32	-0,71	0,08	-0,16	5,4	2,53	1200	620	2,08	1,90	106	0,422	0,320	0,520	0,447	0,365	0,570	18	1,2	0,967
16	0,0168	41	0,70	1,73	-0,35	-0,22	-2,2	3,00	650	450	1,48	0,60	140	0,855	0,510	0,585	0,575	0,430	0,750	7	1,1	0,872
	0,0140	41	0,70	1,73	-0,35	-0,22	-2,2	3,00	650	450	1,48	0,60	140	0,855	0,510	0,585	0,575	0,430	0,750	7	1,1	0,872
17	-0,0034	119	0,17	3,35	-0,10	0,48	-27,0	0,9	0,70	2600	500	0,38	320	0,640	0,463	0,707	0,660	0,366	0,565	16	2,0	0,831
	-0,0004	119	0,17	3,35	-0,10	0,48	-27,0	0,9	0,70	2600	500	0,38	320	0,640	0,463	0,707	0,660	0,366	0,565	16	2,0	0,831
Б	0,0065	41	0,55	2,26	0,31	0,24	24,4	5,24	280	127	0,68	2,53	266	0,185	0,093	0,176	0,132	0,117	0,171	19	1,8	0,842
Б	0,0164	55	4,09	4,06	0,65	0,52	39,1	6,71	580	226	0,94	2,57	228	0,113	0,114	0,153	0,101	0,115	0,105	17	1,8	0,842
М	0,92	0,62	0,99	0,86	0,87	0,90	0,87	0,66	0,75	0,86	0,76	0,31	0,23	0,24	0,58	0,19	0,39	0,37	0,21	-0,26	-0,11	0,842

Примечание. Исходные данные (верхняя строка), восстановленные (нижняя); М -- номера

признаков (в скобках - номера по "Распознавание ..."/8/): I(77) - интенсивность намагничения (ед. СГС), 2(78) - глубина до возмущающего объекта (м), 3(79) - максимальное значение Δ Та (мЭ)†, 4(80) - модальные максимальные значения Δ Та (мЭ), 5(81) - минимальные значения Δ Та (мЭ), 6(82) - модальные минимальные значения Δ Та (мЭ), 7(83) - максимальный горизонтальный градиент ($\gamma^*/\text{м}$), 8(84) - осредненный горизонтальный градиент ($\gamma^*/\text{м}$), 9(87) - длина аномалии L (м), 10(88) - ширина аномалии В (м), 11(89) - вытянутость аномалии (L/V), 12(91) - амплитуда первого уровня (мЭ), 13(92) - радиус первого уровня (м); 14(93) - d_1^1/d_1^1 , 15(94) - d_1^1/d_3^1 , 16(95) - d_2^2/d_3^2 , 17(96) - d_1^2/d_3^2 , 18(97) - d_1^2/d_3^2 , 19(98) - d_2^2/d_3^2 , где d_1^1 - полурадиус аномалии по изолинии 2/3 Δ Та для пологой (1 = 1) и крутой (1 = 2) ветвей, d_2^2 - полурадиус аномалии по изолинии 1/2 Δ Та и d_3^2 - полурадиус аномалии по изолинии 1/3 Δ Та; 20(18) - величина длинной оси интрузива (км), 21(25) - ширина зоны контактового метаморфизма (км). М - номера объектов: 1 - Белоречье, 2 - Инское (центральный участок), 3 - Инское (юго-западный участок), 4 - Холзунское (Перевальное участки АН 7-9), 5 - Холзунское (Перевальный участок АН 6), 6 - Холзунское (Тургусинский участок), 7 - Холзунское (Тургусинские участки АН 1-2), 8 - Холзунское (Перевальный участок АН 5), 9 - Инское (восточный участок), 10 - Геофизическое, 11 - Пономаревское-1, 12 - Пономаревское-П, 13 - Холзунское (северный участок АН 10), 14 - Чесноковское-П, 15 - Новознаменское, 16 - Верхнедровское, 17 - Кузнецовское. б - среднеквадратичное отклонение от действительных, б̄ - среднеквадратичное отклонение от средних по столбцу, $\gamma_{\text{п}}$ - коэффициенты парной корреляции действительных и восстановленных значений для столбцов, $\gamma_{\text{п}}$ - то же для строк.

†) I эрстед = 79,5775 ампер/метр.



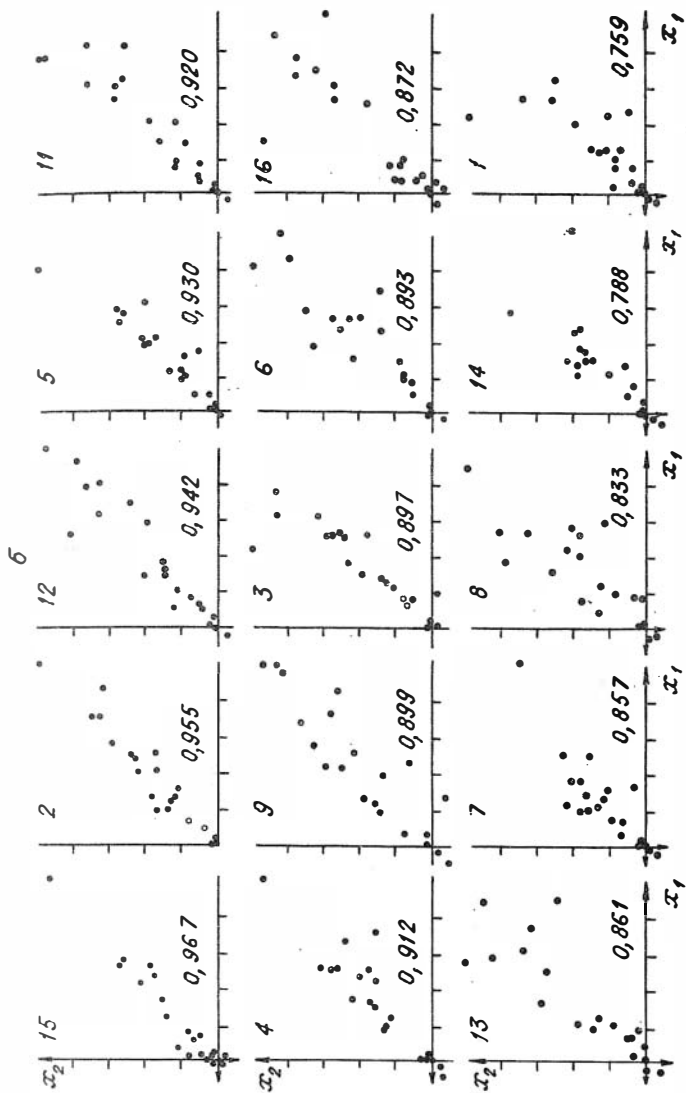


Рис. 3. Зависимости между действительными (x_1) и восстановительными (x_2) значениями:

а — по столбцам, б — по строкам. Цифрами показаны номера строк и столбцов (вверх — ху), коэффициент корреляции (внизу)

наибольшей степени удовлетворяют только два признака — № 3 (максимальные значения ΔT_a) и № 1 (интенсивность намагничивания). Следует отметить, что именно эти признаки в первую очередь характеризуют количественную сторону магнитовозмущающего объекта и коррелируются с запасами. С известной долей условности могут восстанавливаться признаки № 4–6, остальные для целей прогнозирования не пригодны, и в этом отношении выделяются признаки № 20 и 21.

Таким образом, можно утверждать, что в реальных задачах по распознаванию образов целевые признаки обязательно должны проверяться на степень их восстанавливаемости с помощью выбранного алгоритма.

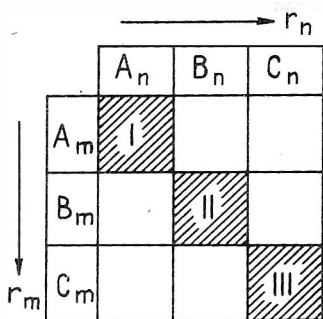


Рис. 4. Выделение "полей" по уровню относительной надежности восстановления пропусков

По величине коэффициентов корреляции $A > B > C$ для r_m и r_n , можно получить таблицу, позволяющую выделить "поля" по уровню надежности восстановления пропусков (рис.4).

Понятно, что пользоваться этой схемой можно только после операции предварительного распознавания проб. При этом в зависимости от положения пропуска становятся возможными выводы о надежности восстановления и в других полях.

Необходимо отметить, что таблицы, предназначенные для восстановления пропусков, должны быть представлены однотипными

Аналогичным образом можно, по-видимому, рассуждать о степени восстанавливаемости пропусков по строкам (рис.3,б), имея в виду парные корреляции между действительными и восстановленными описаниями объектов. Можно заключить, что надежность восстановления пропусков, например, для объекта 5 выше, чем для объекта 8. Более того, упорядочив признаки и объекты по величине коэффициента корреляции и определив 1–2 условные границы (например, по резким перепадам в значениях), можно обозначить структуру рабочей таблицы с точки зрения надежности восстановления пропусков. Поэтому, задавшись укрупненными градациями

признаками, поскольку для каждого типа существуют свои меры сходства и различия. В наибольшей степени они исследованы для признаков, измеренных в сильных шкалах. Базирующиеся на этих мерах алгоритмы не пригодны для восстановления пропусков по таблицам порядковых данных. Восстановление пропусков в бинарных таблицах может быть осуществлено, например, с помощью тесторного метода. Возможности алгоритмов для восстановления порядковых данных ограничиваются небольшим числом допустимых преобразований. Пропуски у точечных признаков, измеряемых в шкале наименований, не должны восстанавливаться в принципе, так как различные объекты могут иметь одно и то же имя и, напротив, разные имена могут быть присвоены одинаковым объектам.

На основании вышеизложенного можно сделать следующие выводы:

1. В решении проблемы восстановления пропусков имеется два основных варианта - "бесцелевой" и "целевой"; в первом случае все признаки считаются равноценными, основой восстановления пропусков являются закономерности в структуре таблицы, во-втором - один из признаков считается главным (целевым) и восстановление осуществляется с учетом целевой ориентации.

2. Высказаны предварительные рекомендации в отношении стратегии выбора алгоритма и схемы обработки конкретных данных в зависимости от вида таблиц и сочетания пропусков у целевого и характеристического признаков.

3. При решении задач распознавания предлагается осуществлять проверочную процедуру по оценке достоверности восстановления значений у целевого признака.

Литература

1. Бабич В.В., Федосеев Г.С. Прогнозная оценка железорудных месторождений Кондомского района Горной Шории. - В кн.: Логико-математическая обработка геологической информации. Новосибирск: ИГиГ СО АН СССР, 1976, с.22-42.

2. Горелик А.Л., Гуревич И.Б., Скрипкин В.А. Современное состояние проблемы распознавания. Некоторые аспекты. М.: Радио и связь, 1985. 160 с.

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

4. Загоруйко Н.Г., Елкина В.Н., Тимеркаев В.С. Алгоритм заполнения пропусков в эмпирических таблицах. - В кн.: Вычислительные системы, № 67. Новосибирск: ИМ СО АН СССР, 1976, с.3-28.

5. Левит В.Е., Переверзев-Орлов В.С. Структура и поле данных при распознавании образов. М.: Наука, 1984. 120 с.

6. Макаров С.В. Обратная схема линейной регрессии (к проблеме восстановления пропусков). - В кн.: Системы и методы анализа данных. Новосибирск: ВЦ СО АН СССР, 1984, с.59-68.

7. Методы анализа данных: подход, основанный на методе динамических сгущений: Пер. с франц. / Кол. авт. под рук. Э.Дидэ; М.: Финансы и статистика, 1985. 357 с. (Математико-статистические методы за рубежом).

8. Распознавание образов в задачах качественного прогноза рудных месторождений / Г.С.Федосеев, В.В.Бабич, В.В.Зайков и др. Новосибирск: Наука, 1980. 208 с. (Тр. ИГиГ, вып.469).

А.А.Бишаев, Г.С.Федосеев

О ЦЕЛЕВОМ ПРЕОБРАЗОВАНИИ ЧИСЛОВЫХ ПРИЗНАКОВ (для случая двух классов объектов)

Предлагается алгоритм, реализующий идею целевого преобразования разнотипных шкал признаков на основе их представления в виде бинарных матриц /2/. Для числовых признаков сущность ее сводится к следующему.

Пусть набор объектов s_1, \dots, s_m охарактеризован упорядоченными по возрастанию числовыми значениями x_1, \dots, x_m признака X . В пределах x_{\min}, x_{\max} задано множество всех допустимых различающихся значений признака X в виде конечного набора упорядоченных градаций $(q_0, q_1, \dots, q_1) = Q_\xi^X$, в котором для всех соседних пар градаций q_j, q_{j-1} ($j = 1, \dots, 1$) известен соответствующий

щий порог их различимости (длина элементарного интервала между двумя соседними градациями):

$$q_j - q_{j-1} = \varepsilon = \text{const}; \quad j = 1, \dots, \ell.$$

Конечное множество Q_ε^x , для элементов которого допустимы преобразования, сохраняющие отношение порядка, мы называем порядковой числовой шкалой признака X . При $\varepsilon = \text{const}$ эта шкала будет равномерной, и если принять $q_0 = x_{\min}$, $q_1 = x_{\max}$, тогда

$$q_z = x_{\min} + \varepsilon \cdot z \quad (z = 0, 1, \dots, \ell).$$

В общем случае вводится представление шкалы Q_ε^x в виде ее полного бинарного отображения в матрицу $D_{l+1, l} = \{d_r^j\}$, в которой каждой градации q_r ($r = 0, 1, \dots, l$) соответствует строка (d_r^1, \dots, d_r^l) , а каждому элементарному интервалу $\{q_j, q_{j-1}\}$ ($j = 1, \dots, l$) — столбец $(d_0^j, \dots, d_{j-1}^j, d_j^j, \dots, d_l^j)$, где $d_r^j \in \{0, 1\}$. Правило заполнения матрицы D следующее: $d_r^j = 0$ для $r = 0, 1, \dots, j-1$ и $d_r^j = 1$ для $r = j, j+1, \dots, l$.

Для столбцов матрицы D задается вектор весовых коэффициентов $\rho^{(0)} = (\rho_1^{(0)}, \dots, \rho_j^{(0)}, \dots, \rho_\ell^{(0)})$, где $\rho_j^{(0)} = \varepsilon$ ($j = 1, \dots, \ell$). Это представление исходной числовой шкалы признака позволяет проводить ее однозначное декодирование, т.е. получать начальные значения градаций признака путем обратного отображения:

$$q_z^{(0)} = \sum_{j=1}^{\ell} d_z^j \cdot \rho_j^{(0)} + x_{\min} \quad (z = 0, 1, \dots, \ell).$$

Таким образом, матрица D вместе с набором весовых коэффициентов $(\rho_1^{(0)}, \dots, \rho_\ell^{(0)})$ имеет взаимно однозначное соответствие с исходной шкалой признака Q_ε^x и является кодом для бинарного отображения $X = (x_1, \dots, x_{2^l})$ в матрицу $X \rho^{(0)}(X, Q_\varepsilon^x)$. Подобное отображение признака X допускает его целевые и сохраняющие порядок преобразования на основе линейных оптимизационных процедур.

Отметим, что в соответствии с [2], данное бинарное представление числовой шкалы признака довольно просто обобщается на случай шкал типа "наименований", допускающих любые упорядочения при их целевом преобразовании. Конкретное же решение задачи целевого преобразования признаков заключается в нахождении новых

(оптимальных в заданном смысле) весовых коэффициентов $\mathcal{P}^* = (\rho_{I_1}^*, \dots, \rho_{I_1}^*)$ относительно явно заданных значений целевого признака $Y = (y_1, \dots, y_m)$.

Целевое преобразование шкалы признака X , имеющее программную реализацию, сводится к следующему. Пусть для объектов a_1, \dots, a_m , известны значения целевого признака $Y = (y_1, \dots, y_m)$. В принятой постановке задачи это означает, что в соответствии с Y , $y_i \in \{0, 1\}$, $i = I_1, \dots, m$, объекты состоят из двух классов: $s_1^I, \dots, s_{m_1}^I$ и $s_1^{II}, \dots, s_{m_2}^{II}$, $m_1 + m_2 = m$. Поэтому в последовательности значений признака X выделяются две упорядоченные подпоследовательности: $X^I = x_1^I, \dots, x_{m_1}^I$ и $X^{II} = x_1^{II}, \dots, x_{m_2}^{II}$. Пусть также задано распределение на шкале Q_ε^X значений $(x_1 \leq x_2 \leq \dots \leq x_m) = X = X^I \cup X^{II}$ и их бинарное представление в виде матрицы \mathcal{Q}^{ρ^0} размерностью $(m \times 1)$. Тогда начальное целевое преобразование шкалы Q_ε^X (относительно заданных двух классов объектов исследования) проводится по следующему правилу.

По всем упорядоченным значениям признака $X_{(m)}$ на его шкале выделяются интервалы $\delta_0, \delta_1, \dots, \delta_L$, в каждый из которых попали значения, связанные отношением соседства и принадлежащие только одному классу. Это означает, что интервалы значений каждого из классов обязательно разделяются значениями другого класса. Отметим, что интервал δ_k ($k = 0, \dots, L$) может содержать как несколько, так и одну градацию шкалы Q_ε^X , поэтому $L \leq 1$. Аналогичные предельные интервалы могут иметь место для двух классов одновременно. В общем случае для каждого интервала вычисляется "идеальная" градация γ_k^c , например, путем осреднения с точностью ε всех значений признака X , попавших в данный интервал:

$$\gamma_k^c = \frac{1}{t_k} \sum_{i=1}^{t_k} x_{ik}, \quad x_{ik} \in \delta_k, \quad \gamma_k^c \in Q_\varepsilon^X.$$

Затем множество всех градаций интервала δ_k "стягивается" в γ_k^c за счет минимизации (зауления) элементарных интервалов внутри δ_k . Что касается совокупностей элементарных интервалов (q_j, q_{j-1}) , попадающих в промежутки между соседними интервалами $\delta_j; \delta_{j-1}$, то каждая из них заменяется одним укрупненным интервалом $(\gamma_j^c, \gamma_{j-1}^c)$ с длиной $\Delta_j = \gamma_j^c - \gamma_{j-1}^c$ ($j = 1, \dots, L$). Таким образом, мы получили преобразованную шкалу Γ_Δ^X и соответ-

ствующую матрицу кодов D_{L+1} , $L(\Gamma_{\Delta}^X)$ для бинарного представления признака: $\mathcal{X}_{\Delta}^y(X, \Gamma_{\Delta}^X)$.

Отметим, что независимо от целевого признака γ аналогичная процедура укрупнения интервалов исходной шкалы признака X , в которые не попадает ни одно из значений (x_1, \dots, x_m) , проводится и на этапе представления исходных значений признака в виде бинарной матрицы $\mathcal{X}_{\rho'}(X, Q_{\Delta}^X)$ путем преобразований: $q_{\xi}^X \rightarrow q_{\Delta'}^X$ и $\mathcal{X}_{\rho}^o(X, Q_{\xi}^X) \rightarrow \mathcal{X}_{\rho'}(X, Q_{\Delta'}^X)$.

Итак, описанный способ начального целевого преобразования числового признака X посредством его представления через шкалу Q_{ξ}^X в виде бинарной матрицы $\mathcal{X}(X, q_{\xi}^X)$ состоит в отображении $q_{\xi}^X \rightarrow \Gamma_{\Delta}^X$, $\mathcal{X}_{\rho}^o(X, q_{\xi}^X) \rightarrow \mathcal{X}_{\rho'}^y(X, \Gamma_{\Delta}^X)$, где $\rho_j = \Delta_j$, $j = 1, \dots, L$. Данное преобразование, наряду с сокращением числа всех интервалов Q_{ξ}^X , увеличивает однородность (сходство) объектов по признаку внутри заданных целевых классов, сохраняя величину средних различий между классами.

Рассмотренные операции являются составной частью обобщенной алгоритмической модели метода целевой итерационной классификации. Поэтому окончательное целевое преобразование числовых признаков многомерного пространства $\{X_1, \dots, X_n\}$ предполагает применение основной в этом методе итерационной процедуры оптимизации $I/$.

Механизм конкретной развертки числового признака в бинарную матрицу с предварительным целевым преобразованием шкалы можно представить в графическом виде (см. рисунок). Исходная конкретная шкала Q_{ξ}^{*X} состоит из всех возможных различающихся числовых градаций $q_0 = 5, \dots, q_{17} = 90$ признака X^* , расстояния между которыми определяются величиной $\xi^* = 5$. Если исходные числовые значения признака X^* получены с более высокой точностью, то они должны быть приравнены значению ближайшей градации на шкале. Таким образом, по заданной шкале значения признака, попадающие в одну градацию, с точностью ξ считаются неразличимыми, причем таковыми могут быть и значения признака у объектов разных классов, например, $x_5^*(q_5^1) = x_6^*(q_6^{\Pi})$. Преобразование $Q_{\xi}^{*X} \rightarrow \Gamma_{\Delta}^X(y)$ заключается в сокращении числа градаций исходной шкалы. Новые (γ^1) градации обуславливаются не всеми исходными градациями (q), а только теми, которым соответствуют реально заданные значения признака X^* . В соответствии с целевым признаком γ^* серии со-

- 1 - - - - -
- 2 - - - - -
- 3 <----->
- 4 <=====>

$S^{(Y^*)}$	$S_1^{(Y)}$	$S_2^{(Y)}$	$S_3^{(Y)}$	$S_4^{(Y)}$	$S_5^{(Y)}$	$S_6^{(Y)}$	$S_7^{(Y)}$	$S_8^{(Y)}$	$S_9^{(Y)}$	$\varepsilon = 5$
X^*	5	15	20	25	30	30	70	75	90	$l = 17$

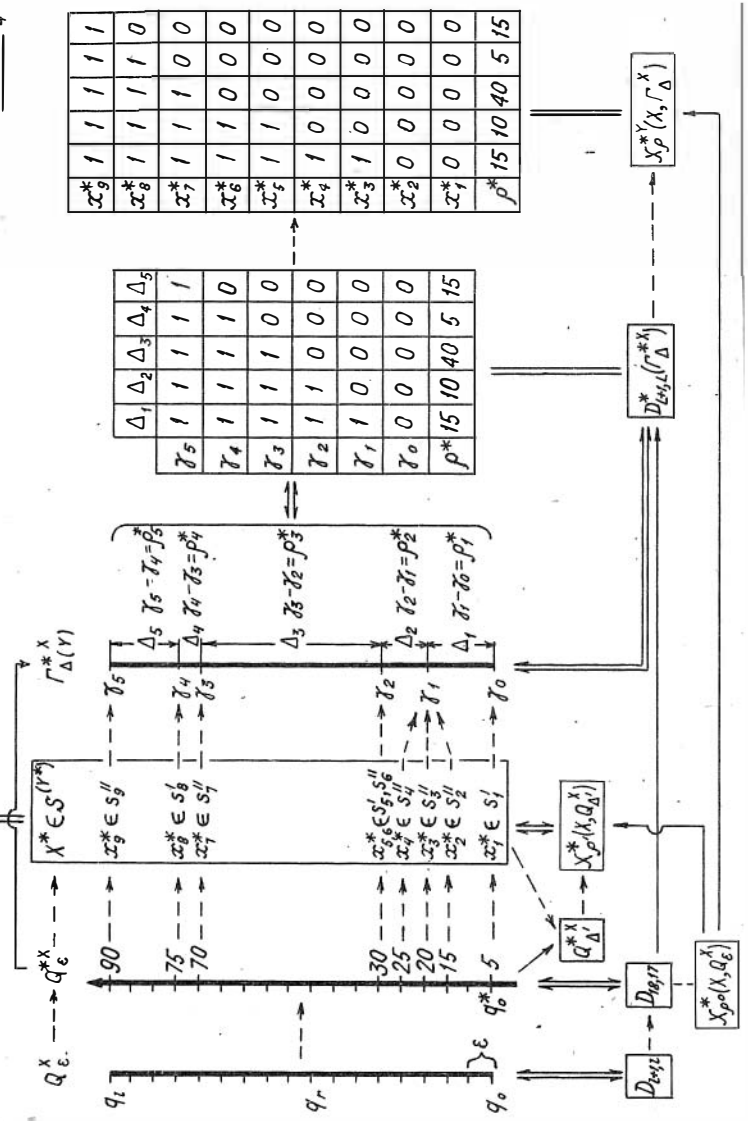


Схема целевого преобразования числового признака $x^{\#}$ в бинарную матрицу. Отношения между элементами схемы: 1 - согласование, 2 - преобразование, 3 - взаимно однозначное соответствие, 4 - равенство. Звездочкой обозначены индексы, относящиеся к конкретным значениям. Остальные обозначения см. по тексту

седних градаций признака, принадлежащих только одному из классов, заменяются усредненной величиной $x_2^{\#} + x_3^{\#} + x_4^{\#} / 3 = \gamma$. Каждому из интервалов между градациями ставятся в соответствие весовые коэффициенты $\rho_1^{\#}, \dots, \rho_5^{\#}$, равные разностям численных значений соседних градаций: $\rho_k^{\#} = \gamma_k - \gamma_{k-1}$; $k = 1, \dots, 5$. Затем шкале $\Gamma_{\Delta}^{\#X}$ ставятся во взаимно однозначное соответствие целевая кодовая матрица, состоящая из шести строк-градаций, пяти столбцов-интервалов и строки весовых коэффициентов ($\rho^{\#}$) столбцов. Заполнение строк целевой матрицы $D^{\#}$ производится по правилу заполнения матрицы $D_{1+1,1}$, причем исходные численные значения $\gamma_0, \dots, \gamma_5$ вычисляются по матрице $D^{\#}$ с помощью уравнения $\gamma_k = q_0^{\#} + \sum_{j=1}^k \rho^{\#j} \cdot d_k^{\#j}$. Значение начальной градации $\gamma_0 = q_0^{\#} = 5$ определяется нулевой строкой. Следующая строка (γ_1) содержит одну единицу, которая соответствует первому интервалу (столбцу), строка γ_2 содержит уже две первые единицы и т.д. Значение последней градации $\gamma_5 = q_{17}^{\#} = 90$ определяется, следовательно, единичной строкой, а вся кодовая матрица является треугольной. И, наконец, каждое значение признака $X^{\#}$ заменяется его бинарным кодом. В результате получается необходимое целевое представление признака в виде бинарной матрицы.

В заключение отметим, что бинарное представление числовых признаков и их начальное целевое преобразование могут применяться как самостоятельно, так и в сочетании с другими методами и алгоритмами.

Авторы выражают благодарность А.С.Вакуленко, составившему программу^{*)} по описанному алгоритму и сделавшему ценные замечания по рукописи работы.

*) Программа GRACOD для ЭВМ ЕС-1052 апробирована при кодировании информации по рудным объектам Енисейского края.

Литература

1. Бишаев А.А. Комплекс программ к методу "Целевая итерационная классификация (ЦИКЛ)". - В кн.: Программные комплексы для целевой обработки информации (оперативно-информационный материал). Новосибирск: ИГиГ СО АН СССР, 1977, с.57-63.
2. Бишаев А.А. Об итеративном целевом преобразовании различных шкал признаков в задачах распознавания. - В кн.: XXI областная научно-техническая конференция, посвященная дню радио. (Тез. докл.). Новосибирск: ИМ СО АН СССР, 1978, с.24-25.

А.Ф.Белоусов, А.С.Вакуленко

АПРОБАЦИЯ МЕТОДА ДИНАМИЧЕСКОГО КЛАСТЕРНОГО АНАЛИЗА
ПРИ ВЫДЕЛЕНИИ ПОПУЛЯЦИИ МАГМАТИЧЕСКИХ ГОРНЫХ ПОРОД

Метод динамического кластерного анализа (МДКА) является не-иерархической классификационной процедурой в многомерном признаковом пространстве. Он разработан Е.Дидеем (Diday) /4/ с целью разбиения исходной совокупности объектов на заданное число групп и выявления наиболее характерных элементов (ядер) этих групп.

Особенностью метода является возможность произвольного задания исходных ядер по числу групп. С формальной точки зрения все ядра равнозначны, поэтому номера групп являются их именами. МДКА избавляет от необходимости полного перебора всех попарных сравнений объектов. Число объектов в ядрах может быть произвольным. Е.Дидей рекомендует включать в одно ядро 5-10 % от суммарного числа объектов. Число ядер может задаваться произвольно, если нет априорной информации о количестве предполагаемых групп объектов.

На первом шаге процедуры выбираются объекты, составляющие ядра, на следующем - вычисляются расстояния между точками и центрами "тяжести" ядер. Рассматриваемые объекты, описанные N вещественными признаками, можно считать точками R^N пространства, а в качестве расстояний между ними выбрать евклидову метрику. Каждой точке присваивается групповой номер, равный номеру ближайшего ядра. В случае равенства расстояний объекту присваивается меньший групповой номер. Таким образом, получается первое разбиение исходной совокупности объектов. На третьем шаге в автоматическом режиме формируются новые ядра из числа объектов тех групп, которые были получены при разбиении на предыдущем шаге. Вычисления продолжают до тех пор, пока решение не перестанет улучшаться по сравнению с предыдущим. В качестве оценки разбиения берется величина, равная сумме расстояний между точками и центрами "тяжести" своих групп. Полученное решение называется локальным, потому что оно зависит от начального множества заданных ядер.

Неоднократное применение МДКА к одному и тому же множеству объектов дает несколько различных локальных разбиений, если начальные множества ядер различны. Вместо поиска лучшего локального разбиения, которое может оказаться истинным разбиением на данное число групп, авторы работы /5/ рекомендуют воспользоваться результатами всех решений. Объекты, остающиеся в одном классе при различных локальных решениях, выделяются в группы, которые называются "сильными" классами. При достаточном числе локальных решений "сильные" классы обеспечивают устойчивое разбиение исходной совокупности объектов. Число "сильных" классов не меньше, а чаще больше заданного числа групп.

Метод использован при обработке петрохимической информации по магматическим породам двух районов. Целью обработки в обоих случаях было выявление горнопородных популяций – совокупностей близкосходных горных пород – в сложных комплексах.

Район края Ветренный пояс, Карелия. Материалом послужила выборка химических анализов (204 анализа, коллекция В.С.Куликова) лав и, возможно, отчасти субвулканитов вулканического комплекса серии Ветреного пояса (нижний протерозой). Химические анализы были пересчитаны на 100 % при исключенных летучих, чтобы приблизить состав пород к первичному. Предварительная обработка с помощью точечных графиков (плоттеров) $MgO-Al_2O_3$ показала наличие трех сгущений (мод) точек анализов. Эти три сгущения петрохимически и с учетом обобщенных петрографических данных /3/ были поставлены в соответствие с тремя квазивидовыми группами пород – базальтоидной, субмеллапикритоидной и меллапикритоидной /1/.

Геологической привязки петрографических описаний для конкретных проанализированных образцов в нашем распоряжении не имелось, поэтому следовало попытаться разбить их на три указанные группы по химическим анализам. Такая постановка задачи оправдывается тем, что квазивидовая группа в нормальном случае должна иметь собственную моду состава, которая и может быть автоматически выявлена при динамическом кластер-анализе в качестве ядра особого класса (кластера). При вычислении евклидовых расстояний содержание каждого окисла нормировалось в интервале 0–100 %; за 100 % принималось максимальное содержание окисла в выборке анализов.

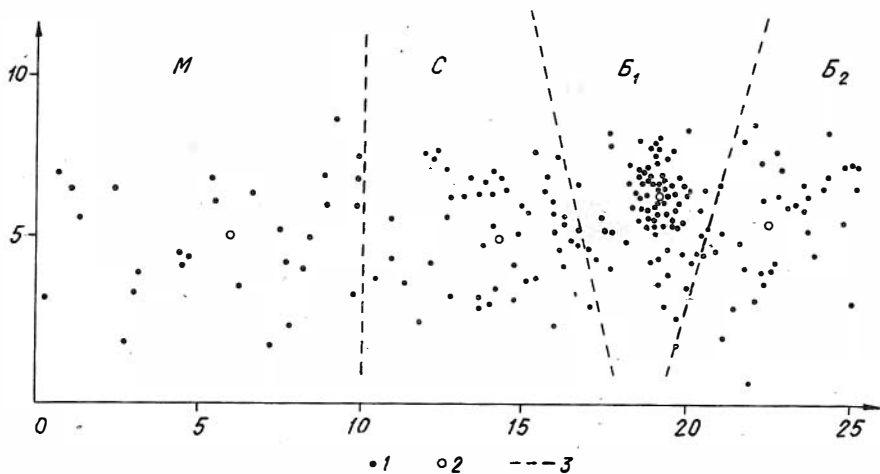
Поскольку при разделении исходной совокупности объектов на три класса выявленные ядра не обнаруживали устойчивой принадлежности к классам и распределялись в основном между первым, четвертым и шестым либо между первым, пятым и шестым ядрами, то было выполнено решение задачи разделения исходной совокупности объектов при четырех ядрах. В этом случае было получено устойчивое разбиение объектов на четыре класса, не зависящее от задания исходных ядер. Значения центров ядер и центров классов для Ветреного пояса приведены в табл. I.

Таблица I

Номер ядра (класса)	Число объек- тов	SiO ₂	TiO ₂	Al ₂ O ₃	MgO	CaO	Na ₂ O	K ₂ O	FeO	Квазивидо- вые группы (популяции)
Центры ядер										
1	10	47,4	0,5	9,0	23,3	7,1	0,8	0,3	11,3}	M
2	10	49,6	0,7	11,4	16,3	9,3	1,4	0,3	10,8}	C
3	10	51,7	0,7	12,8	11,8	9,6	1,7	0,4	10,8}	B ₁
4	15	52,5	0,8	14,0	8,9	10,8	2,1	0,5	10,1}	B ₂
Центры классов										
1	31	47,0	0,4	8,8	23,4	7,2	0,8	0,2	11,6}	M
2	52	50,0	0,7	11,2	16,2	9,1	1,4	0,3	11,0}	C
3	81	51,7	0,7	12,6	11,9	9,7	1,7	0,4	11,0}	B ₁
4	40	52,7	0,8	13,6	9,1	10,5	2,2	0,5	10,1}	B ₂

Класс I представляет популяцию, сопоставимую с мелалакри-тоидной (M) квазивидовой группой, с высокоглиноземистой и повы-шенно-железистой ее частью /I/. Популяция, представленная клас-сом 2, соответствует субмелалакриитоидной (C) группе. Популяции 3 и 4 по петрохимии вполне отвечают базальтоидной группе, при-чем 3 - так называемым базальтовым коматиитам (B₁), а 4 - толеи-товым базальтам (B₂).

Таким образом, МДКА позволяет разбить выборку индивидуаль-ных проб горных пород на классы близкосходных образований - по-пуляций, каждая из которых затем уже в целом может быть сопос-тавлена с той или иной квазивидовой группой горных пород. Это облегчает задачу рассортировки пород на квазивидовые группы, яв-ляющиеся опорными при дальнейшем многоцелевом исследовании комп-лексов (формаций) горных пород.



Проекция точек анализов пород Ветреного пояса:

1 — точки анализов; 2 — центры классов; 3 — границы классов; М — мелапикритоидная группа, С — субмелапикритоидная группа. Базальтоидная группа: B_1 — коматиитовые базальты, B_2 — толеитовые базальты

Проекция исследуемых объектов на плоскость, проходящую через центры классов, показывает их взаимное расположение (см. рисунок). Центры классов находятся приблизительно на одной прямой, образуя последовательность М, С, B_1 , B_2 . Можно видеть, что наиболее четкое ступение точек имеет кластер B_1 . Намечаются минимумы плотности точек между классами С, B_1 и B_2 .

Расстояние между центрами классов уменьшается от 8,3 между М и С, до 5,0 между С и B_1 и до 3,4 между B_1 и B_2 .

Это уменьшение многомерного расстояния между классами отражает возрастающую петрохимическую устойчивость групп пород от пикритоидов к базальтоидам. Такая устойчивость в работе /2/ интерпретирована как котектическая устойчивость, связанная с переходом к плавлению и кристаллизации комбинаций контрастных по составу минералов.

Запад Скандинавского полуострова. Задача состояла в выяснении популяционной структуры пород позднепротерозойских дайковых

поясов, рассекающих гранитоиды западной окраины докембрийской Свелокарельской складчатой области (пояс Блекинг-Даларна) и докембрийский фундамент скандинавских каледонид на предполагаемом северном погружении зоны тех же гранитоидов (пояс Сярв).

Долериты Блекинг-Даларна юга Швеции образуют дугообразный пояс даек, простирающихся субмеридионально вдоль расположенной к западу региональной зоны разломов Протогин, отделяющей докембрийскую складчатую область Юго-Западной Скандинавии от докембрийской Свелокарельской. Дайки рассекают протерозойские "постороженные" граниты западного края свекокарелид. Оценки абсолютного возраста долеритов укладываются в основном в 850-1000 млн лет. Ширина пояса даек достигает 125 км, длина - 650 км. Палеотектоническую обстановку формирования пояса можно рассматривать как материковое плато (молодую платформу). Авторы материалов /6/ интерпретируют ее как материковую рифтовую область.

Долериты Сярв центральных районов Норвегии и Швеции образуют субмеридиональный дайковый пояс шириной до 200 и протяжением свыше 600 км, расположенный на простираии пояса Блекинг-Даларна к северу. Они секут позднепротерозойские аллювиальные и мелководно-морские песчаники Балтоскандинавской миегеосинклинали, обнажающиеся в так называемом покрове Сярв среднего аллохтона скандинавских каледонид. Оценки возраста долеритов - около 650-750 млн лет /6,7/, вмещающих пород - около 650 млн лет. Формирование геосинклинального комплекса с долеритами авторы /7/ связывают с раскрытием палеоокеана Ялетус.

В отмеченных работах устанавливается относительно однородный состав долеритов Блекинг-Даларна, относимых по традиции к толеитам, но петрохимически отвечающих почти без исключения щелочным разновидностям (на диаграмме $SiO_2-(Na_2O+K_2O)$). Долериты Сярв относятся к низкощелочному и в несколько меньшей мере - к щелочному классам.

Нами, во-первых, решалась задача разделения сводной выборки химических анализов долеритов из районов Блекинг-Даларна и Сярв. Химические анализы приведены к 100 % при исключенных летучих. Устойчивое решение получено после нескольких постановок при разном числе ядер только при $n = 2$, причем ядро первого класса сформировано из анализов пояса Блекинг-Даларна, а другое - из пояса Сярв. При расчетах значения содержаний каждого окисла нормированы в интервале 1-100 % по минимуму и максимуму.

Центры ядер для долеритов
запада Скандинавского полуострова

Номер ядра	SiO ₂	TiO ₂	Al ₂ O ₃	MnO	MgO	CaO	Na ₂ O	K ₂ O	P ₂ O ₅	FeO'
1	50,10	3,4	15,7	0,16	4,9	7,0	3,5	1,5	0,9	11,6
2	50,0	1,4	15,4	0,19	6,5	11,0	2,5	0,7	0,17	10,3

В первый класс выделилось 156 анализов, в том числе 146 из пояса Блекинг-Даларна и 10 - из пояса Сярв. Во второй - 170, из них 142 - из пояса Сярв и 28 - из пояса Блекинг-Даларна. В табл. 2 приведены вычисленные центры ядер выделенных классов. Из её данных видно, что классы различаются по FeO', TiO₂, Na₂O, K₂O, P₂O₅ (содержания которых выше в первом классе), CaO и MgO (повышены во втором классе). Неделаящими оказываются SiO₂, Al₂O₃ и MnO.

Таким образом, долеритовые комплексы двух поясов представляют собой две ясно петрохимически выраженные популяции, из которых одна (пояс Сярв) - более меланократовая и менее щелочная. Петрохимические особенности дайкового пояса Блекинг-Даларна, формировавшегося в обстановке материкового плато, свидетельствуют о более глубоком положении фронта магмогенеза, чем при формировании магм долеритов Сярв, ассоциированных с морским геосинклинальным накоплением /2/.

Далее выяснялось возможное присутствие локальных популяций долеритов в каждом из двух поясов.

В поясе Блекинг-Даларна при задании трех ядер выделилось четыре петрохимических популяции. В первую вошли обогащенные кремнеземом и обедненные магнием долериты повышенной щелочности, петрохимически отвечающие трахиандезито-базальтам. Во вторую и третью выделились базальты; вторая от третьей отличается более высокими содержаниями натрия, титана и фосфора и пониженным - магния. Четвертая популяция выделяется благодаря повышенным содержаниям железа и марганца.

Географическая привязка данных для пояса Блекинг-Даларна /6/ позволяет выяснить, не являются ли выделенные по петрохимическим признакам кластеры долеритов территориально обособленными

популяциями. Оказалось, что породы класса 3 (наименее щелочной и наименее кремнеземистой) обособлены в крайнем юго-западном участке пояса. Породы класса I значительно более щелочные и кремнеземистые долериты с повышенными содержаниями титана и фосфора) располагаются северо-восточнее и концентрируются только в южной части пояса. Породы класса 4 (менее щелочные с низкими уровнями титана и фосфора) тяготеют к крайнему юго-восточному участку пояса; однако имеется еще один участок концентрации пород этого класса - в северной части пояса. Дolerиты класса 2 (повышенно-щелочные, наиболее кремнеземистые) тяготеют на большом протяжении к восточной периферии пояса, обращенной в глубь докембрийского палеоконтинента; они ассоциируются с породами I и 4 классов. Таким образом, имеет место как разобщение, так и совмещение максимумов распространенности пород выявленных популяций по площади.

В поясе Сярв при задании трех ядер выделилось три петрохимических популяции. Первая из них экстремальна повышенными уровнями щелочей, глинозема, титана и фосфора и пониженным содержанием кремнезема. Вторая и третья популяции - низкощелочные, с повышенным кремнеземом, но вторая против третьей имеет повышенные Al_2O_3 , MgO и CaO и пониженные SiO_2 , TiO_2 , FeO . Первая популяция локализована в средней части пояса в зоне, идущей вкост его простирания (район Эстерсунд и северо-западнее), а породы второй и третьей распространены на всей территории пояса. Территориальное совмещение второй и третьей популяции, возможно, в какой-то мере отражает региональное перемешивание масс в связи с установленным покровно-тектоническим строением пояса.

Таким образом, МДКА позволяет:

1. При наличии априорных данных о присутствии в смешанной совокупности определенного числа популяций выяснить, проявлены ли последние в выбранном конечном многомерном признаковом пространстве.

2. В предположении о небольшой численности присутствующих в смеси популяций найти наиболее устойчивое разбиение, наметив тем самым популяции в выбранном признаковом пространстве. Эти подсоставности удобнее, чем отдельные объекты, использовать при дальнейшем обосновании популяционного разбиения изучаемой совокупности с привлечением дополнительных данных.

3. Использовать это разбиение для частных объектов в качестве способа прогнозного целевого распознавания (например, в отношении возраста, палеотектонического положения, рудоносности горных пород).

Литература

1. Белоусов А.Ф., Кривенко А.П., Полякова З.Г. Вулканические формации. Новосибирск: Наука, 1982. 281 с.

2. Белоусов А.Ф., Кривенко А.П. Магмогенез вулканических формаций. Новосибирск: Наука, 1983. 167 с.

3. Слосарев В.Д., Куликов В.С. Геохимическая эволюция базит-ультрабазитового магматизма протерозоя (юго-восток Балтийского щита). Л.: Наука, 1973. 105 с.

4. Diday E. The dynamic clusters method in nonhierarchical clustering. - Intern. Jour. of Computer and Information Sciences, 1973, N 2, p.61-88.

5. Lefebvre D., David M. Dinamic clustering and strong patterns recognition: new tools in automatic classification. - Canad. Journ. of Earth Sciences, 1977, v.14, N 10, p.2232-2245.

6. Solyom Z., Andereasson P.-G., Johansson I., Hedvall R. Petrochemistry of Late Proterozoic rift volcanism in Scandinavia. In the Blekinge-Dalarna Dolerites (BDD) - volcanism in a failed arm of Iapetus. - Lund publs. Geol., 23, 1984. 56 p.

7. Solyom Z., Andereasson P.-G., Johansson I. Petrochemistry of Late Proterozoic rift volcanism in Scandinavia. II. In the Särvi Dolerites (SD) - volcanism in the constructive arms of Iapetus. - Lund Publs. Geol., 35, 1985. 42 p.

В.Д.Карбышев, Л.Г.Сухов,
А.С.Вакуленко, А.А.Городнянский

ОЦЕНКА ПЕРСПЕКТИВ ПЛОЩАДЕЙ СЕВЕРО-СИБИРСКОЙ
НИКЕЛЕНОСНОЙ ОБЛАСТИ ЛОГИКО-МАТЕМАТИЧЕСКИМИ МЕТОДАМИ
ПО ОТДЕЛЬНЫМ ГРУППАМ ПРИЗНАКОВ*

Территория, для которой проводится рудопрогнозное разделение площадей количественными методами, охватывает северную часть Тунгусской синеклизы и простирается от Енисей-Хатангского прогиба на юг до р.Ниж.Тунгуски, и от Енисея на восток до бассейна р.Котуй. Границы указанной территории определяются ареалом распространения расслоенных гипербазит-базитовых потенциально-никеленосных интрузивов, расположением в ее пределах известных рудопроявлений и месторождений медно-никелевых руд.

Объекты исследования представлены 84 элементарными участками (площадями), за которые были приняты листы крупномасштабной карты. В свою очередь участки объединены в четыре класса по признаку проявления медно-никелевого оруденения. Признаки, которыми охарактеризованы площади, разделены на семь групп: структурно-тектонические, литолого-стратиграфические, магматические, метаморфо-метасоматические, минералогические, геохимические и геофизические. Общее число признаков - 198. Список объектов, признаков и классов приведен в работе Л.Г.Сухова и др. /14/.

Геологическая задача заключается в разделении площадей с известной степенью рудоносности (эталонных площадей) с целью выявления наборов поисковых признаков (критериев) среди признаков, характеризующих эти площади, с последующей оценкой перспективности других участков на основе выявленных критериев. Геологическая задача решалась в трех постановках в соответствии с ныне принятой стадийностью поисковых работ: общие поиски, поисковые и поисково-оценочные работы. Во всех трех постановках задачи разделения площадей и оценки перспективности участков ставились и решались как задачи распознавания образов.

* Работа выполнена в рамках договора о творческом содружестве сотрудников ИГиГ СО АН СССР совместно с геологами Норильской экспедиции ПГО "Севморгеология" с применением логико-математических исследований по оценке перспектив площадей Северо-Сибирской никеленосной области.

0 задачах распознавания образов

В области конкретных применений распознавания существует два основных направления /2 и др./: вероятностно-статистическое и детерминистское (называемое иногда эвристическим). В тех случаях, когда законы распределения для исходных данных являются сложными или неизвестными, а в таблице исходных данных имеются как количественные, так и бинарные признаки, широко используются детерминистские методы распознавания. Их основной принцип заключается в получении такой решающей функции, которая обеспечивала бы безошибочное распознавание образов на этапе обучения, оставаясь по возможности простой /4/. Предварительной оценкой полученного в таком случае решения является эмпирическая ошибка обучения. В конечном же счете целесообразность использования этих алгоритмов обосновывается непосредственной проверкой – подтверждением результатов задач на практике.

Для характеристики задач распознавания введем обозначения: X – матрица исходных данных, состоящая из m столбцов (признаков) и n строк (объектов); Y^0 – столбец значений целевого признака; P – решающая функция, решающее правило – способ вычисления значений Y^0 по известным значениям исходных признаков (x_1, \dots, x_m) . Отличительными чертами этих задач являются следующие: в них обычно предсказываются значения только одного целевого признака Y^0 (признак Y^0 – выражается в шкале наименований).

В любой задаче распознавания можно выделить три основных классификационных элемента: 1) основание классификации объектов – фиксированная совокупность признаков; 2) правило классифицирования – решающее правило (решающая функция), обеспечивающее отнесение объектов к тому или иному классу; 3) классы объектов, где класс – это множество объектов, эквивалентных по каким-либо признакам (объединенных в единое целое) и отличающихся от иных объектов по одному или нескольким другим признакам.

В зависимости от того, какие из этих элементов известны и какие требуется найти, выделяются три основных типа задач распознавания /5/: 1) поиск решающего правила (функции) P , когда зафиксированы классификация объектов (Y^0) и совокупность признаков (m) – (матрица X); 2) выбор информативного набора (системы) признаков, когда зафиксированы классификация объектов (Y^0)

и класс решающих функций; 3) построение классификации объектов (таксономия), когда зафиксированы совокупность признаков (m) и класс решающих функций.

В настоящем случае геологические задачи рассматривались и решались как первый и второй типы задач распознавания: получение решающего правила из класса линейных решающих функций и выбор информативной системы признаков. Для разделения известных (эталонных) площадей по каждой группе признаков в отдельности использованы логико-математические методы: линейные алгоритмы распознавания "Каскад-П" и "Цикл-2" /12/.

Краткая содержательная характеристика используемых алгоритмов

В основу алгоритмов "Цикл-2" и "Каскад-П" положены процедуры поиска линейной решающей функции вида:

$$Y_i = \sum_{j=1}^{m'} P_j x_{ij} \quad , \quad i = 1, \dots, n \quad , \quad m' \leq m \quad ,$$

где Y_i - нагрузка (значение) i -й строки в исходной таблице, x_{ij} - значения признаков в матрице X , m - общее число признаков, n - число эталонных объектов - объектов обучения, P_j - весовые коэффициенты (информационные веса) признаков, m' - оставшееся в результате решения число признаков.

В каждом из алгоритмов можно выделить три основные последовательные стадии их работы: подготовка исходной информации к счету, выбор информативной системы признаков, классифицирование объектов в соответствии с решающим правилом.

Стадия подготовки информации к счету включает в себя следующие операции:

1. Признаки, замеренные в разных шкалах, приводятся к сопоставимому виду, т.е. нормируются. При этом используется нормировка типа "ограничение вариаций" /6/, после которой значения признаков x_j изменяются в пределах от 0 до 1.

2. Для объектов вычисляются по каждому из признаков, включая целевой, меры различия и сходства. Причем объекты здесь сравниваются не друг с другом, а с некоторым "идеальным" объектом. На "идеальном" объекте все признаки условно принимают минимальные значения.

3. Нормируются исходные информационные веса признаков так, что

$$\sum_{j=1}^n P_j = 1.$$

4. Если по какому-либо признаку в исходной матрице X значение x_{1j} для тех или иных объектов неизвестно ("прочерк"), то при вычислении мер различия и сходства в алгоритмах эти меры принимаются равными нулю. Такая же операция проводится для объектов-проб, представленных для классифицирования. Таким образом, в алгоритмах предусмотрена возможность обработки данных с "прочерком". Стметим, что наличие "прочерка" у признака и объекта уменьшает информативность признака (P_j) и строчную нагрузку объекта (Y_1).

Стадия выбора информативной системы признаков. Цель её установить между признаками взаимную соподчиненность и найти минимизированный набор признаков, дающий удовлетворительное решение. Информативная система признаков отыскивается без полного перебора, путем последовательных приближений с помощью итерационной процедуры /9/. В рассматриваемых случаях информативность отдельного признака оценивается с учётом других признаков, т.е. дается совместная индивидуально-коллективная оценка.

Следует отметить, что большое число алгоритмов, особенно статистических, осуществляет выбор информативных наборов признаков на основе индивидуальной оценки информативности каждого признака, что ведет к потере информации, либо для учета коллективной составляющей требует дополнительных операций, например, преобразования исходной матрицы методом главных компонент. При решении некоторых задач, как показано, например, в работе /3/, такой подход отрицательно сказывается на результатах решения.

Отметим также, что основное отличие рассматриваемых алгоритмов друг от друга заключается в том, что в алгоритме "Каскад-II" максимально учитывается коллективная составляющая в совместной оценке информативности каждого признака.

Особенность алгоритмов заключается в возможности отбраковки признаков, неинформативных относительно цели (Y^0). В ряде работ /10,13/ даны обоснования и некоторые рекомендации по сокращению признакового пространства. В рассматриваемых алгоритмах

при достижении информационным весом какого-либо признака минимально допустимого значения происходит удаление этого признака из дальнейших расчетов. Минимизация, как правило, приводит к значительному сокращению исходного числа признаков с одновременным улучшением результатов обучения.

Стадия классифицирования объектов включает в себя две последовательные операции. Сначала вводится оценка качества надежности обучения, т.е. оценка информативности полученной системы признаков для цели разделения классов и в связи с этим строится решающее правило. Затем на базе решающего правила классифицируются объекты-пробы. Последние представляют собой объекты, не участвующие в обучении, т.е. те, для которых требуется решить вопрос о принадлежности их к одному из классов.

Оценка информативности полученной системы признаков производится на эмпирическом уровне по результатам "внутреннего" экзамена - процедуры распознавания, основанной на сравнении самих объектов обучения друг с другом. Для этого вычисляется ошибка - точечная оценка надежности обучения, равная отношению числа неправильно диагностируемых эталонных объектов к их общему числу. Если ошибка достаточно мала (равна нулю или меньше 5 %), то обучение считается надежным, а полученная система признаков информативной.

Геологическая задача в ее трех постановках ранее была решена с использованием статистико-игровых методов распознавания /14/. В этой работе были рассмотрены особенности подготовки геологических данных для рассматриваемой территории и их обработки в условиях информационных массивов большой размерности; установлены - индивидуальная информативность признаков и относительная информативность групп признаков, выделены рудоконтролирующие факторы и для главных из них сделана попытка содержательной интерпретации; дана оценка перспектив северо-запада Сибирской платформы с охватом для этой цели практически всех признаков.

В настоящей работе приведены результаты решения геологической задачи логико-математическими методами, и дана краткая характеристика методов. Оба этих подхода отличаются друг от друга, и в каждом из них имеются свои возможности и ограничения. Не ставя перед собой цели подробного сравнительного анализа обоих

подходов, отметим следующие возможности логико-математических методов, в частности, алгоритмов "Цикл-2" и Каскад-II":

1. Задачи разделения классов в рассматриваемом случае решаются на исходном признаковом пространстве без преобразования признаков, а при статистическом подходе /I4/ - на наборе новых, преобразованных с помощью метода главных компонент. Причем процедура преобразования является в общем-то самостоятельной и независимой по отношению к процедуре поиска решающего правила.

2. Сокращение исходного набора признаков происходит в процессе работы алгоритма на стадии выбора информативной системы, при этом для каждого признака рассчитывается совместная индивидуально-коллективная оценка информативности /I, II/. В статистическом подходе для сокращения числа признаков использованы способы оценки индивидуальной информативности признаков по Кульбаку и преобразование исходных данных методом главных компонент. Причем процедуры сокращения являются, в общем, самостоятельными и независимыми по отношению к процедуре поиска решающего правила.

Результаты разделения эталонных объектов (площадей) и информативность признаков

Первая постановка - стадия поисково-оценочных работ. Производилось разделение отдельно по каждой группе признаков 10 объектов первого класса (A1 - участки, содержащие месторождения) и 56 объектов второго класса (B1, B11 и B111 - участки, не содержащие рудопроявления или содержащие неперспективные рудопроявления). По всем группам признаков, кроме геофизических, получены удовлетворительные решения.

Структурно-тектоническим признакам алгоритмом "Каскад-II" получены две информационные системы* (табл. I). Расчет их проводился последовательно. Сначала была выявлена первая система признаков, включающая четыре качественных признака и дающая наиболее контрастное (табл. I, 2)

* В таблицах № I, 3, 4, 6, 8, 10 нумерация признаков использована из работы /I4/. Признаки, отсутствующие в этих таблицах, являются неинформативными.

Таблица I

Информационные веса структурно-тектонических признаков

Номер признака	Стадия работ			Номер признака	Стадия работ		
	поисково-оценочные		поисковые		поисково-оценочные		поисковые
	P_j^I	P_j^2			P_j^I (ИР)	P_j^I	
2	0	0,02	0	I2	0	0,11	0,03
3	0	0,05	0,03	I3	0,24	0	0
4	0	0	-0,02	I4	0	0	-0,06
5	0,25	0	0,03	I7	0	0	-0,02
7	0	0	0,044	I8	0,28	0	0,103
8	0	0,14	0,06	20(195)	0	0,13	0
9	0	0,38	0,29	21(196)	0	0,04	0,074
10	0	-0,05	-0,095	22(197)	0	0,04	0,04
II	0	0,03	0,072	23(198)	0,23	0	-0,03

Таблица 2

Результаты разделения классов по структурно-тектоническим признакам

Стадия работ						
поисково-оценочные				поисковые		
Класс	Номер объекта	Y_1^1	Y_2^1	Класс	Номер объекта	Y_1^1 (ИР)
I	2	3	4	5	6	7
	I	0,77	0,94		I	0,95
	2	0,77	0,61		2	0,68
	3	1,00	0,82		3	0,87
	4	0,77	0,59		4	0,67
	5	1,00	0,56		5	0,70
AI	6	1,00	0,56	AI	6	0,70
	7	0,72	0,43		7	0,50
	8	0,76	0,54		8	0,73
	9	0,76	0,53		9	0,72
	10	0,77	0,74		10	0,83

Продолжение табл.2

I	2	3	4	5	6	7
	29	0	0,25		II	0,50
	30	0,25	0,29		I2	0,44
	3I	0	0,2I		I3	0,44
	32	0,24	0,16		I4	0,44
	33	0,24	0,20	API	I5	0,44
	34	0	0,07		I6	0,44
	35	0,24	0,3I		I7	0,57
	36	0,24	0,13		I8	0,49
	37	0	0,18		I9	0,66
	38	0,25	0,3I		20	0,64
	39	0,25	0,3I		29	0,35
	40	0,23	0,09		30	0,34
	4I	0,23	0,25		3I	0,4I
	42	0	0,19		32	0,39
	43	0	0,17		33	0,4I
	44	0,24	0,17		34	0,23
	45	0	0,27		35	0,42
	46	0	0,2I		36	0,30
	47	0,25	0,30		37	0,40
БИ	48	0,25	0,30		38	0,38
	49	0	0,12		39	0,38
	50	0	0,27		40	0,26
	5I	0,24	0,24		4I	0,29
	52	0	0,23		42	0,23
	53	0	0,14		43	0,33
	54	0,24	0,14	БИ	44	0,42
	55	0,24	0,20		45	0,36
	56	0,24	0,25		46	0,28
	57	0	0,06		47	0,42
	58	0	0,13		48	0,42
	59	0	0,13		49	0,24
	60	0	0,19		50	0,39
	6I	0,24	0,22		5I	0,32
	62	0,24	0,27		52	0,37

Окончание табл.2

I	2	3	4	5	6	7
	63	0,24	0,27		53	0,31
	64	0	0,25		54	0,37
	65	0	0,27		55	0,25
	66	0	0,27		56	0,34
	67	0,24	0,23		57	0,23
	68	0,24	0,23		58	0,35
	69	0	0,17		59	0,35
	70	0,24	0,20		60	0,41
	71	0	0,08		61	0,42
БИ	72	0,24	0,25		62	0,34
	73	0,23	0,24		63	0,34
	74	0,23	0,24		64	0,40
	75	0	0,29		65	0,42
	76	0,24	0,30		66	0,42
БП	77	0	0,24	БИ	67	0,39
	78	0	0,21		68	0,39
	79	0	0,21		69	0,35
	80	0	0,23		70	0,37
	81	0	0,23		71	0,28
	82	0	0,23		72	0,35
БП	83	0	0,23		73	0,42
	84	0,24	0,27		74	0,42

разделение классов ($y_{\min}^I = 0,72$ - минимальная строчная нагрузка для эталона первого класса, $y_{\max}^{II} = 0,25$ - максимальная строчная нагрузка для эталона второго класса). Данная система признаков позволяет сформулировать простое решающее правило: наличие на площади хотя бы любых трех из этих признаков говорит в пользу её перспективности. Вторая система признаков была получена при исключении первых четырех. В неё вошли 10 признаков (см. табл.1,2), дающих менее контрастное разделение классов ($y_{\min}^I = 0,43$; $y_{\max}^{II} = 0,31$).

Наиболее информативными из структурно-тектонических признаков являются: в первой системе - признаки № 5, 13, 18, 198, соот-

ветственно названия – расположение в пределах областей замыкания отрицательных структур 4-го порядка, наличие систем линейных разрывов, инъекций вещества осадочного чехла и трубок взрыва, эруптивных аппаратов; во второй системе – признаки № 8, 9, 12, 195, соответственно – наличие флексур, густота разрывных нарушений, наличие пологих и прямолинейно параллельных разрывов.

Отметим, что наблюдается хорошее сходство результатов с полученными ранее статистическими оценками наибольшей информативности – признаки № 5, 8, 9, 13, 18, 195, 197 /14/.

Среди литолого-стратиграфических признаков алгоритмом "Каскад-II" (табл.3) выявлена только одна информативная система признаков, которая дает малоконтрастное разделение классов (диапазон нагрузок строк в первом классе $0,64 \leq Y^I \leq 0,72$, а во втором $0,23 \leq Y^{II} \leq 0,51$).

В систему вошло достаточно большое количество признаков (тринадцать), имеющих близкие между собой информационные веса. Наиболее информативные признаки – № 23, 29, 31, 34, 35, 36, 46. Половина признаков имеет отрицательные веса^{*}, т.е. отсутствие этих признаков на объектах первого класса играет положительную роль в оценке перспективности участков. Здесь наблюдается меньшая сходимость со статистическими оценками информативности, где положительно оценены признаки № 20, 23, 25, 32, 35, 38, 41, 44, 46. Обοими методами выделились: наличие отложений курейско-разведочнинского комплекса, полнота разреза стложений фюкинского комплекса, фациальный состав нижнего карбона.

По магматическим признакам при разделении получено пять информативных систем (табл.4), включающих в свой состав только половину признаков (54 %). Первые две системы, последовательно одна за другой выявленные алгоритмом "Каскад-II", включают в свой состав каждая по четыре признака и позволяют сформулировать достаточно простые решающие правила (см. табл.12). Первая из них дает более контрастное разделение классов ($Y_{\min}^I = 0,72$, $Y_{\max}^{II} = 0,37$), чем вторая ($Y_{\min}^I = 0,37$, $Y_{\max}^{II} = 0,28$). Следующие три системы получены алгоритмом "Цикл-2"

^{*}) Отрицательный информационный вес получает признак, имеющий отрицательную корреляцию с целевым признаком; другими словами, величины значений признака в I-ом классе меньше, чем во 2-м.

Результаты разделения классов
по литолого-стратиграфическим признакам
и информационные веса признаков

Стадия поисково-оценочных работ							
Класс	Номер объекта	γ_1^2	Класс	Номер объекта	γ_1^2	Номер признака	P_j^1
AI	I	0,65	BI	53	0,37	20	0
	2	0,72		54	0,47	21	0
	3	0,64		55	0,48	22	0
	4	0,64		56	0,40	23	- 0,08
	5	0,71		57	0,37	24	0
	6	0,71		58	0,49	25	0
	7	0,71		59	0,49	26	- 0,05
	8	0,64		60	0,40	27	0
	9	0,71		61	0,47	28	0
	10	0,64		62	0,49	29	0,12
BI	29	0,47	63	0,49	30	- 0,04	
	30	0,47	64	0,40	31	- 0,14	
	31	0,31	65	0,37	32	0,045	
	32	0,51	66	0,37	33	0	
	33	0,42	67	0,29	34	0,13	
	34	0,51	68	0,29	35	- 0,09	
	35	0,43	69	0,51	36	- 0,07	
	36	0,43	70	0,39	37	0,054	
	37	0,49	71	0,45	38	0,06	
	38	0,50	72	0,51	39	0	
	39	0,50	73	0,50	40	0	
	40	0,49	74	0,50	41	0	
	41	0,49	75	0,45	42	0	
	42	0,37	76	0,50	43	0	
	43	0,42	77	0,46	44	0	
	44	0,44	78	0,39	45	0	
	45	0,43	79	0,39	46	- 0,09	
	46	0,36	80	0,42	47	0	
	47	0,32	81	0,42	48	0	
	48	0,32	82	0,44	49	0,03	
49	0,26	83	0,44	50	0		
50	0,31	84	0,49				
51	0,23						
52	0,47						

Таблица 4

Информационные веса магматических признаков

Номер признака	С т а д и я р а б о т							
	поисково-оценочные					поисковые		
	P_j^1	P_j^2	P_j^3	P_j^4	P_j^5	P_j^1 (ПР)	P_j^2 (ПР)	P_j^3 (ПР)
53	0	0	0	0	0,22	0	0	0
54	0	0	0	0	0,09	0	0	0
55	0	0	0	0	0,02	0	0	0
59	0	0	0	0	0	0	0,02	0
60	0,17	0	0,09	0	0	0	0	0
61	0	0	0	0,23	0	0	0	0
64	0	0	0	0	0	0	0,14	0
65	0	0	0	0	0,04	0	0,05	0
66	0	0,09	0	0	0	0	0	0
67	0	0,65	0,14	0	0	0	0	0
68	0	0	0	0	0,04	0	0	0
69	0	0	0	0	0	0	0,12	0
72	0	0	0	0	0	0	0,15	0
74	0	0	0	0	0	0	0,02	0,04
75	0	0,08	0	0,05	0	0	0,03	0
76	0,20	0	0,04	0	0	0,25	0	0,35
77	0	0,18	0	0,15	0	0	0	0
78	0,29	0	0	0,11	0	0	0,07	0
79	0	0	0	0	0,26	0	0,11	0
80	0	0	0	0	0,09	0	0,08	0
81	0	0	0	0,07	0	0,31	0	0,11
83	0	0	0	0	0,12	0,21	0	0
84	0	0	0,08	0	0	0	0	0
87	0	0	0	0,11	0	0	0	0
88	0	0	0	0	0,12	0,23	0	0,25
89	0	0	0,23	0	0	0	0,02	0,03
91	0	0	0,07	0	0	0	0,10	0,04
93	0,34	0	0,35	0	0	0	0	0,08
95	0	0	0	0,04	0	0	0,09	0,07
96	0	0	0	0,24	0	0	0	0

независимо от предыдущего алгоритма и включает соответственно 7, 8 и 9 признаков с последовательным уменьшением контрастности разделения классов (табл.4,5).

Наиболее информативными из магматических признаков являются в первой системе - признаки № 60, 76, 78 и 93, соответственно названия - наличие на площади оливинофировых базальтов, выдержанного горизонта ультрабазитов в интрузивных телах, максимальной мощности в интрузивах горизонтов пород повышенной кислотности, наличие такситовых долеритов; во второй системе - признаки № 67 и 77 - мощность пикритовых базальтов и максимальная мощность горизонтов ультрабазитов в интрузивах. В третьей, четвертой и пятой системах наиболее информативными признаками, помимо вышеперечисленных, являются признаки № 53, 61, 79, 89 и 96 - наличие на площади андезитовых и пикритовых базальтов, относительная мощность горизонта ультрабазитов в интрузивных телах, наличие среди интрузивных пород "лейкократовых" габбро и "пикритовых" габбро-норитов.

Для магматических признаков сходство по информативности с результатами статистической обработки, где выделены признаки № 53, 54, 60, 61, 67, 68, 76, 77, 79, 84, 88, 92, 94, 95, наблюдается в половине случаев.

По метаморфо-метасоматическим признакам получено пять информативных систем, включающих в себя почти все признаки (табл.6). Первые три выявлены алгоритмом "Каскад-П" с использованием разных режимов его работы, две из них (первая и третья) дают достаточно контрастное разделение классов (табл.7). Наиболее простое решающее правило с ошибкой обучения 3 % дает вторая система, состоящая из двух признаков (№ 98, I03) - наличие кремнекисло-щелочных новообразований и серпентинизации или бруситизации. Выполнение этих двух признаков на площади говорит в пользу её перспективности с ошибкой в 3 %. Последние две системы признаков рассчитаны алгоритмом "Цикл-2" последовательно, одна за другой. В целом наиболее информативными из числа метаморфо-метасоматических признаков, помимо названных двух, являются признаки № 99, I00, I01, I05, I10 и III, соответственно - наличие магнезиального и известкового скарнирования, метасоматических преобразований с привнесом калия (в том числе - биотитизации), хлоритизации, расположение рассматриваемых преоб-

Результаты разделения классов по магматическим признакам

С т а д и я р а б о т											
поисково-оценочные						поисковые					
Класс	Номер объекта	Y_1^3	Y_2^3	Y_3^3	Y_4^3	Y_5^3	Класс	Номер объекта	$Y_1^3(ПР)$	$Y_2^3(ПР)$	$Y_3^3(ПР)$
I	2	3	4	5	6	7	8	9	10	11	12
AI	1	0,84	0,76	0,96	0,65	0,60	AI	1	0,48	0,48	0,89
	2	0,93	0,57	0,95	0,82	0,68		2	0,48	0,51	0,89
	3	0,95	0,69	0,97	0,78	0,63		3	0,48	0,58	0,89
	4	0,90	0,60	0,97	0,75	0,58		4	0,48	0,42	0,89
	5	0,77	0,41	0,70	0,74	0,58		5	0,79	0,35	0,97
	6	0,75	0,48	0,93	0,71	0,55		6	0,48	0,35	0,89
	7	0,72	0,37	0,83	0,57	0,58		7	0,48	0,35	0,89
	8	1,00	0,54	0,94	0,84	0,58		8	0,48	0,45	0,89
	9	0,77	0,60	0,87	0,74	0,54		9	0,48	0,37	0,89
	10	0,81	0,89	0,77	0,71	0,57		10	0,46	0,39	0,62
AII	29	0,00	0,13	0,01	0,24	0,37	AII	11	0,48	0,36	0,71
	30	0,00	0,10	0,01	0,25	0,27		12	0,48	0,35	0,71
	31	0,00	0,13	0,02	0,28	0,25		13	0,48	0,41	0,89
	32	0,00	0,05	0,00	0,00	0,26		14	0,48	0,41	0,89
	33	0,08	0,09	0,00	0,04	0,02		15	0,69	0,39	0,81
	34	0,02	0,03	0,07	0,02	0,36		16	0,48	0,35	0,71
	35	0,20	0,10	0,11	0,07	0,36		17	0,48	0,35	0,71
	36	0,00	0,09	0,00	0,02	0,33		18	0,48	0,44	0,74
	37	0,00	0,01	0,00	0,00	0,23		19	0,54	0,35	0,44
	38	0,08	0,05	0,00	0,04	0,36		20	0,48	0,36	0,67
	39	0,02	0,07	0,14	0,14	0,45		29	0,00	0,03	0,04
	40	0,27	0,28	0,09	0,39	0,30		30	0,00	0,01	0,04
	41	0,12	0,05	0,04	0,04	0,17		31	0,00	0,13	0,10
	42	0,20	0,11	0,00	0,31	0,07		32	0,00	0,03	0,00
	43	0,00	0,07	0,00	0,01	0,03		33	0,00	0,07	0,00
	44	0,04	0,04	0,04	0,02	0,26		34	0,23	0,12	0,29
	45	0,20	0,25	0,00	0,22	0,23		35	0,25	0,26	0,47
	46	0,00	0,03	0,00	0,00	0,00		36	0,00	0,04	0,00
	47	0,0	0,04	0,07	0,01	0,09		37	0,00	0,01	0,00
	48	0,16	0,05	0,00	0,31	0,25		38	0,00	0,18	0,04
49	0,05	0,04	0,00	0,19	0,15	39	0,00	0,14	0,29		
50	0,06	0,08	0,00	0,15	0,18	40	0,00	0,16	0,08		
51	0,00	0,06	0,00	0,00	0,00	41	0,23	0,13	0,29		
52	0,00	0,01	0,00	0,12	0,00	42	0,25	0,07	0,35		
53	0,00	0,00	0,00	0,00	0,00	43	0,00	0,04	0,00		
54	0,00	0,00	0,00	0,00	0,00	44	0,23	0,12	0,29		
55	0,00	0,00	0,00	0,00	0,00	45	0,25	0,26	0,47		
56	0,00	0,03	0,00	0,00	0,03	46	0,00	0,04	0,04		
57	0,00	0,00	0,00	0,00	0,00	47	0,00	0,07	0,04		
58	0,00	0,02	0,00	0,00	0,03	48	0,23	0,25	0,33		
59	0,00	0,02	0,00	0,00	0,03	49	0,23	0,13	0,32		

I	2	3	4	5	6	7	8	9	10	II	I2
	60	0,00	0,05	0,00	0,01	0,02		50	0,23	0,09	0,29
	61	0,08	0,04	0,00	0,05	0,15		51	0,00	0,02	0,04
	62	0,10	0,04	0,00	0,04	0,16		52	0,00	0,08	0,04
	63	0,00	0,04	0,00	0,01	0,10		53	0,00	0,05	0,04
	64	0,00	0,03	0,00	0,00	0,02		54	0,00	0,02	0,04
	65	0,00	0,03	0,00	0,00	0,02		55	0,00	0,15	0,00
	66	0,00	0,04	0,00	0,01	0,02		56	0,00	0,01	0,00
	67	0,00	0,01	0,00	0,00	0,02		57	0,00	0,02	0,00
	68	0,00	0,04	0,00	0,02	0,02		58	0,00	0,09	0,00
	69	0,03	0,02	0,00	0,14	0,04		59	0,00	0,10	0,00
BI	70	0,01	0,00	0,00	0,01	0,15		60	0,00	0,07	0,00
	71	0,00	0,01	0,00	0,00	0,02		61	0,23	0,07	0,25
	72	0,00	0,01	0,00	0,00	0,04		62	0,00	0,25	0,04
	73	0,00	0,02	0,00	0,00	0,02		63	0,00	0,23	0,00
	74	0,00	0,05	0,00	0,02	0,02		64	0,00	0,10	0,04
	75	0,28	0,07	0,04	0,05	0,29		65	0,00	0,09	0,04
	76	0,28	0,03	0,04	0,27	0,21	BI	66	0,00	0,12	0,00
	77	0,20	0,04	0,04	0,01	0,19		67	0,00	0,07	0,04
BII	78	0,00	0,03	0,00	0,02	0,00		68	0,00	0,07	0,00
	79	0,24	0,14	0,04	0,40	0,40		69	0,00	0,23	0,04
	80	0,36	0,28	0,39	0,53	0,54		70	0,23	0,03	0,25
	81	0,00	0,26	0,04	0,24	0,36		71	0,00	0,19	0,04
	82	0,17	0,03	0,09	0,01	0,06		72	0,00	0,25	0,04
BIII	83	0,37	0,06	0,28	0,19	0,38		73	0,00	0,19	0,04
	84	0,19	0,08	0,07	0,14	0,38		74	0,00	0,18	0,00

разований пород вне контактов с картируемыми интрузивами, мощность (в метрах) зон преобразования.

Сходимость со статистическими оценками, где информативными являются признаки № 97, 98, 99, 100, 102, 103, 105, III, - в 60 % случаев.

Минералогические признаки дали четыре информативные системы, в которые вошли 70 % минералогических признаков (табл.8). Из них первые две получены последовательно алгоритмом "Каскад-II", другие - алгоритмом "Цикл-2". Наибольший информативный вес имеют признаки № II3, II8, I24, I25 и I26, соответственно - максимальная магнезиальность оливинов в интрузивных породах, наличие шпинели (хромшпинелидов), наличие в интрузивных породах борнита, халькозина и самородной меди. Четвертое решающее правило, в которое вошло три признака, можно сформулировать

Таблица 6

Информационные веса
метаморфо-метасоматических признаков

Номер признака	С т а д и я р а б о т					
	поисково-оценочные					поисковые
	P_j^1	P_j^2	P_j^3	P_j^4	P_j^5	P_j^1 (ИР)
97	0,05	0	0	0	0	0
98	0,10	0,50	0,16	0,33	0	0,33
99	0,10	0	0	0	0,55	0,06
100	0,10	0	0,15	0,10	0	0,05
101	0,11	0	0,11	0,17	0	0
103	0,11	0,50	0,18	0,36	0	0
104	0,08	0	0	0	0,09	0,25
105	0,09	0	0	0	0,15	0
106	0,07	0	0	0	0	0
107	0,05	0	0	0	0,09	0,18
108	-0,04	0	0	0	0	-0,09
109	0,04	0	0	0	0	0
110	0	0	0	0	0,12	0
111	0,04	0	0,40	0,04	0	0,04

так: если в интрузивных породах имеется шпинель (хромшпинеллиды) и халькозин, а в коренных породах скопления магнетита, то эта площадь перспективна на обнаружение месторождения медно-никелевых руд. Отметим, что все четыре системы дают достаточно контрастное разделение классов (табл.9).

Со статистическими оценками информативности признаков наблюдается сходимость в 60 % случаев.

Среди геохимических признаков выявлены четыре информативные системы, охватывающие половину признаков (табл.10). Первые три получены алгоритмом "Каскад-П", последняя - "Цикл-2". Наиболее информативные признаки - показатель контрастности дифференциации (№ 133) - это разница показателя

Результаты разделения классов
по метаморфо-метасоматическим признакам

С т а д и я р а б о т									
поисково-оценочные							поисковые		
Класс	Номер объек- та	Y_1^4	Y_2^4	Y_3^4	Y_4^4	Y_5^4	Класс	Номер объек- та	Y_1^4 (ПР)
I	2	3	4	5	6	7	8	9	10
AI	1	0,94	1,0	1,00	1,00	0,79	AI	1	0,82
	2	0,97	1,0	0,76	0,97	0,88		2	0,97
	3	0,99	1,0	0,97	0,99	0,88		3	0,98
	4	0,97	1,0	0,80	0,98	0,88		4	0,95
	5	0,77	0,5	0,61	0,64	0,33		5	0,58
	6	0,96	1,0	0,65	0,96	0,88		6	0,96
	7	0,83	1,0	0,61	0,96	0,79		7	0,78
	8	0,96	1,0	0,65	0,96	0,88		8	0,96
	9	0,96	1,0	0,66	0,96	0,88		9	0,95
	10	0,96	1,0	0,73	0,97	0,88		10	0,93
BI	29	0,12	0,0	0,00	0,00	0,00	BI	II	0,57
	30	0,04	0,0	0,00	0,00	0,00		12	0,50
	31	0,13	0,0	0,00	0,00	0,00		13	0,93
	32	0,03	0,0	0,00	0,00	0,00		14	0,93
	33	0,21	0,0	0,00	0,00	0,09		15	0,93
	34	0,13	0,0	0,08	0,01	0,00		16	0,33
	35	0,13	0,0	0,08	0,01	0,24		17	0,53
	36	0,27	0,0	0,02	0,00	0,00		18	0,53
	37	0,41	0,0	0,20	0,18	0,15		19	0,56
	38	0,23	0,0	0,16	0,10	0,12		20	0,58
	39	0,23	0,0	0,16	0,10	0,12		29	0,06
	40	0,26	0,0	0,11	0,17	0,09		30	0,07
	41	0,03	0,0	0,00	0,00	0,00		31	0,08
	42	0,03	0,0	0,00	0,00	0,00		32	0,05
	43	0,42	0,0	0,00	0,00	0,32		33	0,33
	44	0,42	0,0	0,00	0,00	0,32		34	0,07
45	0,38	0,0	0,18	0,10	0,09	35	0,07		

Продолжение табл.7

I	2	3	4	5	6	7	8	9	10
	46	0,04	0,0	0,00	0,00	0,00		36	0,31
	47	0,47	0,0	0,13	0,17	0,24		37	0,10
	48	0,31	0,0	0,04	0,00	0,24		38	0,12
	49	0,28	0,0	0,03	0,00	0,09		39	0,12
	50	0,28	0,0	0,03	0,00	0,09		40	0,30
	51	0,01	0,0	0,00	0,00	0,00		41	0,05
	52	0,02	0,0	0,00	0,00	0,00		42	0,05
	53	0,02	0,0	0,00	0,00	0,00		43	0,50
	54	0,02	0,0	0,00	0,00	0,00		44	0,50
	55	0,02	0,0	0,00	0,00	0,00		45	0,37
	56	0,03	0,0	0,00	0,00	0,00		46	0,07
	57	0,39	0,0	0,00	0,00	0,32		47	0,32
	58	0,02	0,0	0,00	0,00	0,00		48	0,32
	59	0,03	0,0	0,00	0,00	0,00		49	0,32
	60	0,26	0,0	0,00	0,00	0,09		50	0,32
	61	0,19	0,0	0,00	0,00	0,09	БИ	51	0,00
	62	0,18	0,0	0,01	0,00	0,09		52	0,02
	63	0,17	0,0	0,01	0,00	0,09		53	0,03
	64	0,12	0,0	0,01	0,00	0,00		54	0,03
	65	0,03	0,0	0,00	0,00	0,00		55	0,02
БИ	66	0,03	0,0	0,00	0,00	0,00		56	0,04
	67	0,03	0,0	0,00	0,00	0,00		57	0,46
	68	0,20	0,0	0,07	0,01	0,24		58	0,03
	69	0,29	0,0	0,02	0,00	0,24		59	0,04
	70	0,29	0,0	0,02	0,00	0,24		60	0,26
	71	0,03	0,0	0,00	0,00	0,00		61	0,26
	72	0,31	0,0	0,00	0,00	0,24		62	0,23
	73	0,26	0,0	0,00	0,00	0,24		63	0,21
	74	0,12	0,0	0,02	0,00	0,00		64	0,05
	75	0,30	0,0	0,02	0,00	0,18		65	0,04
	76	0,25	0,0	0,01	0,00	0,18		66	0,05
	77	0,30	0,0	0,00	0,00	0,18		67	0,05
БИ	78	0,13	0,0	0,00	0,00	0,00		68	0,19
	79	0,35	0,5	0,18	0,36	0,55		69	0,30

I	2	3	4	5	6	7	8	9	10
БП	80	0,45	0,0	0,18	0,10	0,32		70	0,30
	81	0,45	0,0	0,18	0,10	0,32		71	0,04
БШ	82	0,19	0,0	0,00	0,00	0,09		72	0,20
	83	0,19	0,0	0,00	0,00	0,09		73	0,20
	84	0,28	0,0	0,18	0,10	0,09		74	0,06

Таблица 8

Информационные веса минералогических признаков

Номер признака	С т а д и я р а б о т					
	поисково-оценочные				поисковые	
	P_j^1	P_j^2	P_j^3	P_j^4	P_j^1 (ПР)	P_j^2 (ПР)
II2	0	0,11	0	0	-0,13	0
II3	0	0,33	0	0	0,28	0,22
II5	0	0,13	0	0	0,24	0
II6	0,14	0	0,175	0	0,06	0,08
II7	0	0	0	0	0,05	0,12
II8	0,18	0	0	0,47	0	0
II9	0	0	0,10	0	0	0
I20	0	0,07	0,09	0	0	0
I21	0,14	0	0	0	0	0,11
I23	0,16	0	0,18	0	0,10	0,38
I24	0,21	0	0,455	0	0	0
I25	0	0,15	0	0,37	0,08	0
I26	0	-0,21	0	0	0	0
I27	0	0	0	0	0,06	0,09
I29	0,17	0	0	0,16	0	0

$$\frac{MgO + FeO + Fe_2O_3}{Na_2O + CaO + Al_2O_3} \times 100$$

для наиболее основной и наиболее кислой разновидностей интрузивных пород; магнезиальность и доля калия в сумме щелочей в наиболее основных интрузивных породах (№ I38, I44); гидрохими-

Таблица 9

Результаты разделения классов по минералогическим признакам

Стадия работ										
поисково-оценочные					поисковые					
Класс	Номер объек- та	γ_1^5	γ_2^5	γ_3^5	γ_4^5	Класс	Номер объек- та	$\gamma_1^5(\text{ПР})$	$\gamma_2^5(\text{ПР})$	
I	2	3	4	5	6	7	8	9	10	
AI	I	I,0	0,90	I,0	I,0	AI	I	0,76	I,0	
	2	I,0	0,95	I,0	I,0		2	0,89	I,0	
	3	I,0	0,95	I,0	I,0		3	0,89	I,0	
	4	I,0	0,95	I,0	I,0		4	0,89	I,0	
	5	I,0	0,8I	0,9I	I,0		5	0,77	0,95	
	6	I,0	0,86	0,9I	I,0		6	0,82	0,97	
	7	I,0	0,72	0,8I	0,63		7	0,74	0,97	
	8	I,0	0,78	0,9I	I,0		8	0,77	0,93	
	9	I,0	0,73	0,9I	I,0		9	0,72	0,9I	
	10	I,0	0,78	0,9I	I,0		10	0,7I	0,96	
BI	29	0,0	0,48	0,0	0,0	AII	II	0,72	0,87	
	30	0,0	0,50	0,0	0,0		12	0,53	0,39	
	3I	0,0	0,47	0,0	0,0		13	0,74	0,90	
	32	0,0	0,38	0,0	0,0		AII	14	0,77	0,9I
	33	0,0	0,48	0,0	0,0			15	0,59	0,58
	34	0,I7	0,47	0,0	0,I6			16	0,5I	0,50
	35	0,I7	0,5I	0,0	0,I6			17	0,68	0,69
	36	0,3I	0,38	0,I7	0,I6			18	0,55	0,66
	37	0,0	0,3I	0,0	0,0			19	0,53	0,53
	38	0,30	0,59	0,09	0,I6			20	0,5I	0,45
	39	0,30	0,35	0,09	0,I6	29		0,25	0,08	
	40	0,0	0,3I	0,0	0,0	30		0,29	0,I0	
	4I	0,0	0,39	0,0	0,0	3I.		0,33	0,09	
	42	0,0	0,46	0,0	0,0	32	0,27	0,05		
	43	0,30	0,42	0,0	0,I6	BI	33	0,3I	0,I7	
	44	0,30	0,59	0,0	0,I6		34	0,39	0,28	
	45	0,32	0,37	0,I7	0,47		35	0,42	0,32	
	46	0,0	0,32	0,0	0,0		36	0,3I	0,I2	
	47	0,I7	0,47	0,09	0,I6		37	0,35	0,II	
	48	0,0	0,56	0,09	0,0		38	0,33	0,43	
49	0,30	0,55	0,0	0,I6	39		0,3I	0,32		
50	0,30	0,37	0,0	0,I6	40		0,II	0,00		

I	2	3	4	5	6	7	8	9	10
	51	0,0	0,37	0,0	0,0		41	0,23	0,04
	52	0,0	0,55	0,0	0,0		42	0,31	0,09
	53	0,0	0,44	0,0	0,0		43	0,31	0,39
	54	0,0	0,43	0,0	0,0		44	0,48	0,46
	55	0,0	0,13	0,0	0,0		45	0,26	0,24
	56	0,18	0,37	0,0	0,47		46	0,21	0,16
	57	0,0	0,44	0,0	0,37		47	0,48	0,28
	58	0,0	0,29	0,0	0,0		48	0,40	0,18
	59	0,18	0,46	0,0	0,47		49	0,39	0,34
	60	0,0	0,52	0,0	0,0		50	0,29	0,25
	61	0,0	0,54	0,0	0,0		51	0,26	0,04
	62	0,0	0,54	0,0	0,0		52	0,46	0,13
	63	0,0	0,57	0,0	0,0		53	0,31	0,06
БИ	64	0,0	0,46	0,0	0,0		54	0,29	0,06
	65	0,0	0,43	0,0	0,0		55	0,20	0,04
	66	0,17	0,28	0,0	0,16		56	0,26	0,05
	67	0,0	0,28	0,0	0,0		57	0,39	0,24
	68	0,31	0,44	0,17	0,16		58	0,31	0,24
	69	0,30	0,53	0,0	0,16		59	0,32	0,07
	70	0,30	0,46	0,0	0,16	БИ	60	0,33	0,13
	71	0,0	0,46	0,0	0,0		61	0,46	0,15
	72	0,31	0,40	0,17	0,16		62	0,36	0,22
	73	0,14	0,15	0,17	0,0		63	0,32	0,15
	74	0,0	0,50	0,0	0,0		64	0,28	0,09
	75	0,30	0,59	0,18	0,37		65	0,34	0,09
	76	0,34	0,59	0,45	0,37		66	0,21	0,10
БII	77	0,0	0,20	0,0	0,0		67	0,15	0,01
	78	0,35	0,39	0,28	0,47		68	0,29	0,15
	79	0,35	0,60	0,28	0,47		69	0,37	0,31
	80	0,30	0,59	0,0	0,53		70	0,26	0,29
	81	0,30	0,41	0,0	0,53		71	0,30	0,10
БIII	82	0,14	0,40	0,0	0,0		72	0,26	0,13
	83	0,14	0,60	0,0	0,0		73	0,21	0,11
	84	0,14	0,50	0,0	0,0		74	0,37	0,30

Информационные веса геохимических признаков

Номер приз- нака	С т а д и я р а б о т					Номер приз- нака	С т а д и я р а б о т				
	ПОИСКОВО-ОЦЕНОЧНЫЕ						ПОИСКОВО-ОЦЕНОЧНЫЕ				
	P_j^1	P_j^2	P_j^3	P_j^4	$P_j^1(ПР)$		P_j^1	P_j^2	P_j^3	P_j^4	$P_j^1(ПР)$
I31	0	0,14	0	0	0	I58	0	0	0,15	0	0,27
I32	0	0	0	0	-0,18	I60	0	0,10	0	0	0
I33	0,43	0	0	0,13	0	I65	0	0,05	0	0	0
I34	0	-0,05	0	0	0	I67	0	-0,04	0	0	0
I35	0	0	-0,13	0	0	I70	0	0,16	0	0	0
I38	0	0,06	0	0,46	0,18	I71	0,135	0	0	0,08	0
I41	0	0	-0,04	0	0	I72	0	0	0	0,09	0
I42	0	0	0	0	0,14	I73	0	0	0,15	0	0
I43	0	0,07	0	0	0	I75	0	0,18	0	0,04	0
I44	0	0,15	0	0	0	I76	0	0	0,12	0	0
I45	0	0	0,05	0	0	I77	0	0	0,03	0	0
I52	0,26	0	0	0	0	I78	0	0	0,14	0	0
I53	0	0	0,08	0	0	I82	0,175	0	0	0,20	0
I54	0	0	-0,11	0	0	I83	0	0	0	0	0,23

ческие параметры - концентрация хлор-иона (в г/л), а также сульфат- и натрий-иона в экв. % (№ I70, I73, I75); геохимические коэффициенты - величина отношений $\frac{Cu}{V}$ в оливиновых долеритах и $\frac{Cr}{Fe}$ в наиболее основной разновидности интрузивных пород (№ I52, I58); наличие рудных валунов с кондиционным содержанием никеля или (и) меди (№ I83). Разделение классов достаточно контрастное (табл. II).

Сходимость со статистическими оценками информативности имеется в менее половины случаев.

Таким образом, на стадии поисково-оценочных работ получено более двух десятков (21) информативных систем признаков, включающих в свой состав 52 % признаков от общего их числа. Если оценить информативность по классам признаков, то она наиболее высокая у метаморфо-метасоматических, магматических, минералогических и геохимических. Далее идут структурно-тектонические и литолого-стратиграфические. Геофизические признаки отнесены к числу неинформативных. Сопоставление оценок информативности с таковыми, полученными ранее статистическими методами /I4/, показывают сходимость результатов в более половины случаев.

На основе полученных систем признаков составлены уравнения и построены решающие правила (табл. I2), по которым было проведено классифицирование (распознавание) "объектов-проб".

Вторая постановка - стадия поисковых работ. По каждому классу признаков проводилось разделение 20 объектов первого класса (A1, AП - участки, содержащие месторождения и перспективные рудопроявления) и 46 объектов второго класса (B1 - участки, не содержащие рудопоявлений). Удовлетворительное решение получено только в пространстве магматических и геохимических признаков.

Среди магматических признаков (см. табл. 4) получено три информативные системы, включающие в себя 40 % признаков, первые две получены последовательно с использованием алгоритма "Каскад-П", третья - "Цикл-2". Наиболее информативными оказались признаки № 64, 72, 76, 81, 83, 88, соответственно названия - доля туфов в объеме вулканических продуктов, средняя площадь интрузивных тел, наличие выдержанного горизонта пород наиболее основного состава, гранит-порфиров или (и) грано-диоритов, или (и)

Таблица II

Результаты разделения классов по геохимическим признакам

С т а д и я р а б о т								
поисково-оценочные						поисковые		
Класс	Номер объек- та	γ_1^6	γ_2^6	γ_3^6	γ_4^6	Класс	Номер объек- та	γ_1^6 (ПР)
I	2	3	4	5	6	7	8	9
AI	I	0,66	0,54	0,48	0,82	AI	I	0,4I
	2	0,48	0,44	0,48	0,70		2	0,34
	3	0,44	0,45	0,48	0,64		3	0,39
	4	0,47	0,57	0,48	0,55		4	0,34
	5	0,36	0,45	0,48	0,58		5	0,40
	6	0,59	0,69	0,62	0,78		6	0,43
	7	0,47	0,45	0,48	0,62		7	0,50
	8	0,42	0,44	0,58	0,53		8	0,42
	9	0,44	0,44	0,49	0,63		9	0,50
	10	0,4I	0,44	0,47	0,55		10	0,30
BI	29	0,04	0,19	0,20	0,08	BI	II	0,30
	30	0,00	0,10	0,27	0,03		I2	0,36
	3I	0,02	0,12	0,23	0,06		I3	0,46
	32	0,0I	0,18	0,3I	0,07		I4	0,48
	33	0,02	0,07	0,22	0,04		I5	0,30
	34	0,02	0,07	0,25	0,07		I6	0,30
	35	0,03	0,16	0,24	0,07		I7	0,32
	36	0,2I	0,16	0,36	0,27		I8	0,39
	37	0,02	0,08	0,25	0,07		I9	0,32
	38	0,00	0,07	0,27	0,03		20	0,30
	39	0,0I	0,06	0,27	0,03		29	0,04
	40	0,02	0,10	0,27	0,07		30	0,07
	4I	0,02	0,10	0,25	0,06		3I	0,04
	42	0,0I	0,10	0,29	0,06		32	0,08
	43	0,02	0,2I	0,3I	0,10		33	0,04
	44	0,00	0,09	0,24	0,05		34	0,09
	45	0,00	0,10	0,24	0,05		35	0,07
	46	0,00	0,19	0,2I	0,04		36	0,07
	47	0,02	0,09	0,29	0,06		37	0,09
	48	0,02	0,08	0,26	0,05		38	0,03

I	2	3	4	5	6	7	8	9
	49	0,01	0,13	0,18	0,06		39	0,10
	50	0,00	0,12	0,21	0,05		40	0,19
	51	0,00	0,13	0,19	0,06		41	0,09
	52	0,06	0,09	0,22	0,05		42	0,09
	53	0,06	0,15	0,19	0,03		43	0,06
	54	0,03	0,16	0,27	0,07		44	0,05
	55	0,03	0,08	0,22	0,04		45	0,05
	56	0,01	0,12	0,32	0,06		46	0,03
	57	0,04	0,10	0,21	0,04		47	0,21
БИ	58	0,04	0,13	0,26	0,08		48	0,19
	59	0,04	0,13	0,28	0,13		49	0,03
	60	0,02	0,12	0,36	0,05		50	0,06
	61	0,01	0,11	0,26	0,06		51	0,04
	62	0,02	0,12	0,31	0,03		52	0,04
	63	0,02	0,12	0,27	0,05		53	0,10
	64	0,03	0,11	0,29	0,09		54	0,21
	65	0,03	0,17	0,35	0,08		55	0,04
	66	0,06	0,17	0,37	0,17		56	0,05
	67	0,02	0,13	0,27	0,07		57	0,14
	68	0,03	0,12	0,28	0,07	БИ	58	0,18
	69	0,02	0,30	0,37	0,15		59	0,11
	70	0,06	0,20	0,23	0,09		60	0,04
	71	0,08	0,30	0,19	0,11		61	0,04
	72	0,21	0,09	0,25	0,26		62	0,10
	73	0,20	0,16	0,37	0,28		63	0,17
	74	0,21	0,15	0,35	0,27		64	0,17
	75	0,13	0,21	0,37	0,29		65	0,08
	76	0,13	0,20	0,33	0,26		66	0,14
	77	0,12	0,23	0,31	0,19		67	0,04
БИ	78	0,00	0,09	0,32	0,05		68	0,10
	79	0,26	0,30	0,36	0,49		69	0,21
	80	0,16	0,27	0,37	0,32		70	0,06
	81	0,01	0,14	0,12	0,05		71	0,03
	82	0,01	0,15	0,31	0,07		72	0,15
БШ	83	0,12	0,18	0,28	0,33		73	0,19
	84	0,14	0,22	0,28	0,15		74	0,15

Линейные решающие правила (уравнения)	Эмпирическая ошибка обучения, %		Условия отнесения к I классу и коэффициенты принадлежности (П)					
	$y \geq$	$y \leq$	3	4	5	6	7	8
I								
СТАДИЯ ПОИСКОВО-ОЦЕНОЧНЫХ РАБОТ								
1. Структурно-тектонические признаки (y^1).								
$y_1^1 = 0,25x_5 + 0,24x_3 + 0,28x_{18} + 0,23x_{198}$	0	0,72	1	0,5	0,5	0,5	0,25	0
$y_2^1 = 0,02x_2 + 0,05x_3 + 0,14x_8 + 1,36x_9 - 0,006x_{10} + 0,03x_{11} + 0,11x_{12} + 0,13x_{195} + 0,04x_{196} + 0,04x_{197} + 0,05x_{199}$	0	0,43	1	-	-	-	0,31	0
2. Литолого-стратиграфические признаки (y^2).								
$y_1^2 = -0,08x_{23} - 0,05x_{26} + 0,12x_{29} - 0,005x_{30} - 0,14x_{31} + 0,045x_{32} + 0,15x_{34} - 0,09x_{35} - 0,07x_{36} + 0,054x_{37} + 0,06x_{38} - 0,09x_{46} + 0,03x_{49} + 0,56$	0	0,64	1	0,58	0,5	0,5	0,51	0
3. Магматические признаки (y^3).								
$y_1^3 = 0,17x_{60} + 0,2x_{76} + 0,0019x_{78} + 0,34x_{93}$	0	0,72	1	0,54	0,5	0,5	0,37	0
$y_2^3 = 0,00005x_{66} + 0,0008x_{67} + 0,00018x_{75} + 0,0015x_{77}$	0	0,48	1	-	-	-	0,28	0
$y_3^3 = 0,09x_{60} + 0,00175x_{67} + 0,04x_{76} + 0,08x_{84} + 0,23x_{89} + 0,07x_{91} + 0,35x_{93}$	0	0,7	1	0,55	0,5	0,5	0,39	0

	1	2	3	4	5	6	7	8
$Y_4^3 = 0,23x_{61} + 0,0001x_{75} + 0,0013x_{77} + 0,0007x_{78} + 0,07x_{81} + 0,11x_{87} + 0,04x_{95} + 0,24x_{96}$		3	0,65	1	0,57	0,5	0,40	0
$Y_5^3 = 0,22x_{53} + 0,09x_{54} + 0,02x_{55} + 0,0011x_{65} + 0,0002x_{68} + 0,0046x_{79} + 0,0014x_{80} + 0,12x_{83} + 0,12x_{88}$		4,5	0,57	1	-	-	0,45	0
4. Метаморфо-метасоматические признаки (Y^4).								
$Y_1^4 = 0,05x_{97} + 0,1x_{98} + 0,1x_{99} + 0,1x_{100} + 0,11x_{101} + 0,11x_{103} + 0,08x_{104} + 0,09x_{105} + 0,07x_{106} + 0,05x_{107} - 0,001x_{108} + 0,04x_{109} + 0,00017x_{111} + 0,043$		0	0,77	1	0,62	0,5	0,47	0
$Y_2^4 = 0,5x_{98} + 0,5x_{103}$		3	1	1	0,5	0,5	0	0
$Y_3^4 = 0,16x_{98} + 0,15x_{100} + 0,11x_{101} + 0,18x_{103} + 0,0017x_{111}$		0	0,61	1	0,45	0,5	0,2	0
$Y_4^4 = 0,33x_{98} + 0,1x_{100} + 0,17x_{101} + 0,36x_{103} + 0,00017x_{111}$		0	0,64	1	0,5	0,5	0,36	0
$Y_5^4 = 0,55x_{99} + 0,09x_{104} + 0,15x_{105} + 0,09x_{107} + 0,12x_{14}$		3	0,79	1	-	-	0,32	0
5. Минералогические признаки (Y^5).								
$Y_1^5 = 0,14x_{116} + 0,18x_{118} + 0,14x_{121} + 0,16x_{123} + 0,21x_{124} + 0,17x_{129}$		0	1	1	-	-	0,5	0
$Y_2^5 = 0,0027x_{112} + 0,0057x_{113} + 0,004x_{115} + 0,07x_{120} + 0,15x_{125} - 0,21x_{126} - 0,28$		0	0,72	1	-	-	0,6	0

	1	2	3	4	5	6	7	8
$Y_3^5 = 0,175x_{116} + 0,1x_{119} + 0,09x_{120} + 0,18x_{123} + 0,455x_{126}$		0	0,81	1	-	-	0,45	0
$Y_4^5 = 0,47x_{118} + 0,37x_{125} + 0,16x_{129}$		0	0,63	1	-	-	0,53	0
6. Геохимические признаки (Y^6).								
$Y_1^6 = 0,00087x_{133} + 0,033x_{152} + 0,0018x_{171} + 0,175x_{182} - 0,01$		0	0,36	1	0,34	0,8	0,26	0
$Y_2^6 = 0,000275x_{131} - 0,0016x_{134} + 0,0025x_{138} + 0,002x_{143} + 0,0023x_{144} + 0,052x_{160} + 0,403x_{165} - 0,039x_{167} + 0,011x_{170} + 0,0024x_{175} + 0,08$		0	0,44	1	0,38	0,5	0,3	0
$Y_3^6 = -0,0031x_{135} - 0,0009x_{141} + 0,0012x_{145} + 0,63x_{153} - 0,00025x_{154} + 0,075x_{158} + 0,0018x_{173} + 2,5x_{176} + 0,86x_{177} + 1,66x_{178} + 0,34$		0	0,47	1	-	-	0,37	0
$Y_4^6 = 0,00026x_{133} + 0,019x_{138} + 0,0011x_{171} + 0,0003x_{172} + 0,0005x_{175} + 0,2x_{182} - 0,095$		3	0,55	1	-	-	0,33	0

СТАДИЯ ПОИСКОВЫХ РАБОТ

I. Структурно-тектонические признаки.

$$Y_1^1 (\text{ПР}) = 0,03x_3 - 0,02x_4 + 0,03x_5 + 0,044x_7 + 0,06x_8 + 1,036x_9 - 0,001x_{10} + 0,072x_{11} +$$

	I	2	3	4	5	6	7	8
	$0,03x_{12}^{-0},06x_{14}^{-0},0003x_{17}^{+0},103x_{18}^{+0},074x_{196}^{+0},04x_{197}^{-0},03x_{198}^{+0},225$	20	0,49	1	0,44	0,5	0,4	0
3. Магматические признаки.								
$Y_1^3(ПР) = 0,25x_{76}^{+0},31x_{81}^{+0},21x_{83}^{+0},23x_{88}$		0	0,48	1	0,45	0,8	0,25	0
$Y_2^3(ПР) = 0,02x_{59}^{+0},0018x_{64}^{+0},0014x_{65}^{+2}x_{69}^{+0},0024x_{72}^{+0},02x_{74}^{+0},000007x_{75}^{+0},0005x_{78}^{+0},0031x_{79}^{+0},0014x_{80}^{+0},02x_{89}^{+0},1x_{91}^{+0},09x_{95}$		0	0,35	1	0,33	0,8	0,26	0
39 $Y_3^3(ПР) = 0,03x_{61}^{+0},04x_{74}^{+0},35x_{76}^{+0},11x_{81}^{+0},0,25x_{88}^{+0},03x_{89}^{+0},04x_{91}^{+0},07x_{95}$	4,5	0,62	1	0,58	0,8	0,35	0	0
4. Метаморфо-метасоматические признаки.								
$Y_1^4(ПР) = 0,33x_{98}^{+0},06x_{99}^{+0},05x_{100}^{+0},25x_{104}^{+0},18x_{107}^{-0},0031x_{108}^{+0},00017x_{111}^{+0},1$	11	0,53	1	-	-	0,32	0	0
5. Минералогические признаки.								
$Y_1^5(ПР) = -0,0037x_{112}^{+0},0048x_{113}^{+0},01x_{115}^{+0},06x_{116}^{+0},05x_{117}^{+0},1x_{123}^{+0},08x_{125}^{+0},06x_{127}^{-0},14$	6	0,51	1	-	-	0,42	0	0

	2	3	4	5	6	7	8
I							
$\chi^2_{(IP)} = 0,0038x_{113} + 0,08x_{116} + 0,12x_{117} + 0,11x_{121} + 0,38x_{123} + 0,09x_{127} - 0,15$	8	0,5	1	-	-	0,34	0
6. Геохимические признаки.							
$\chi^6_{(IP)} = -0,0029x_{132} + 0,0075x_{138} + 0,0031x_{142} + 0,135x_{158} + 0,0043x_{154} + 0,11$	0	0,3	1	-	-	0,21	0

габбро-сиенитов, пегматоидных или (и) гранофировых диоритов (гранодиоритов), кварцевых или кварцосодержащих долеритов.

Здесь наблюдается незначительная сходимость со статистическими оценками информативности, где информативные признаки - № 75, 76, 77, 78, 79, 80, 88, 89, 91, 94. Общими являются только наличие основных дифференциатов и кварцевых долеритов.

По геохимическим характеристикам получена одна информативная система (см. табл. I0), включающая пять признаков, близких между собой по информативности, № I32, I42, I38, I58, I83, соответственно: основность

$$\frac{MgO + FeO + Fe_2O_3}{Na_2O + CaO + Al_2O_3} \times 100$$

и щелочность

$$\frac{Na_2O + K_2O}{Na_2O + K_2O + MgO + FeO + Fe_2O_3} \times 100$$

наиболее кислой разновидности интрузивных пород, магнизиальность и величина отношения $\frac{Cr}{Ti}$ в наиболее основной разновидности интрузивных пород, индекс дифференциации - разницы железистости

$$\frac{FeO + Fe_2O_3}{MgO + FeO + Fe_2O_3} \times 100$$

наиболее кислой и наиболее основной разновидностей. В целом магматические и геохимические признаки дают

полное, но малоконтрастное разделение (см. табл. 5, II).

Для трех других групп признаков получены приближенные решения по разделению эталонных объектов. По структурно-тектоническим признакам рассчитана система, дающая разбиение, которое малоконтрастно для 20 % объектов (см. табл. I, 2, I2). Для метаморфо-метасоматических признаков полученная система дает II % ошибок на стадии обучения (см. табл. 6, 7, I2). Среди минералогических признаков выявлено две системы, дающие соответственно 8 и 6 % ошибок (см. табл. 8, 9, I2).

По литолого-стратиграфическим и геофизическим признакам решения неудовлетворительные.

Таким образом, для стадии поисковых работ выявлены информативные системы только среди магматических и геохимических признаков. На основе полученных систем признаков составлены уравнения и построены решающие правила (см. табл. I2), по которым проведено классифицирование участков ("объектов-проб"). Для данной постановки задачи в связи с использованием приближенных решений оценка перспективности участков считается предварительной, поскольку выявление объединенных информативных систем по признакам разных групп может улучшить распознавание.

Третья постановка - стадия общих поисков. По каждой группе признаков проводилось разделение 18 объектов первого класса (АI - участки, содержащие перспективные рудопроявления, АII - модели информации об участках, содержащих месторождения, на период, предшествовавший открытию последних) и 46 объектов второго класса (БI - участки, не содержащие рудопроявлений). Ни по одной из групп признаков не получено информативных систем, безошибочно разделяющих классы эталонов. В связи с этим в данной постановке задачи предварительная оценка перспективности участков ("объектов-проб") не проводилась. Целесообразно в дальнейшем выявить обобщенные информативные системы по признакам разных групп, потому что подобное объединение при статистико-игровом подходе /I4/ дало более удовлетворительные результаты.

Отметим, что при переходе от первой постановки ко второй и третьей все сложнее найти линейные разделяющие правила. И это логично, так как в целевом отношении участки, содержащие месторождения, отличаются более существенно даже от участков, содержащих рудопроявления, чем последние от "пустых" участков. Таким

образом, признаковое пространство в целом находится в соответствии с ранжировкой эталонных площадей по масштабу оруденения в геологической задаче прогноза.

Оценка перспективности участков на разных стадиях поисков

Для выделения участков, требующих постановки поисково-оценочных и поисковых работ, проведено классифицирование 165 площадей на основе полученных линейных решающих правил. В это число участков вошли площади, выделенные ранее статистическими методами /I4/ для поочередной постановки общих поисков (56 участков), поисковых (49 участков) и поисково-оценочных (10 участков) работ, а также 51 участок, признанных неперспективными.

Распознавание в данном случае проводилось по следующей схеме. Для 165 "объектов-проб" рассчитывались строчечные нагрузки Y^n по всем решающим правилам, приведенным в табл. I2. Если значение строчечной нагрузки для объекта-проба больше или равно минимальному значению строчечной нагрузки для эталонного объекта первого класса ($Y^n \geq Y_{\min}^I$), то "объект-проба" относится к первому классу и её коэффициент принадлежности (мера сходства) к этому классу равен единице ($\Pi = 1$). Если $Y^n \leq Y_{\max}^II$, то $\Pi = 0$. В тех случаях, когда разрыв между классами достаточно велик, вводилась промежуточная оценка (см. табл. I2), при выполнении которой $\Pi = 0,5$. На каждой стадии поиска рассчитывалось среднее значение $\bar{\Pi}$ (средняя мера сходства с I классом), а для поисково-оценочных работ дополнительно - среднее $\bar{\Pi}$ по каждому классу признаков. Рассмотрим результаты классифицирования по стадиям поиска (табл. I3).

Стадия поисково-оценочных работ. Наибольшие средние коэффициенты принадлежности ($\bar{\Pi} \geq 0,2$, см. табл. I2) по 2I решающему правилу получили 10 участков, которые ранее /I4/ с использованием статистических методов были рекомендованы для поочередной постановки поисково-оценочных работ. Отметим, что девять объектов (кроме № 258) получили большие средние коэффициенты принадлежности также и на стадии поисковых работ, т.е. они последовательно проходят по стадии поисковых ра-

Результаты классифицирования участков ("объектов-проб")
на стадиях поисково-оценочных и поисковых работ

Таблица 13

	Номер			Коэффициенты принадлежности (меры сходства)								
	п/п	квadrата	объекта	Б	$\bar{\Pi}_1$	$\bar{\Pi}_2$	$\bar{\Pi}_3$	$\bar{\Pi}_4$	$\bar{\Pi}_5$	$\bar{\Pi}_6$	$\bar{\Pi}$	$\bar{\Pi}(\text{ПР})$
ПОИСКОВО-ОЦЕНОЧНЫЕ РАБОТЫ	1	0725	240	0,74	0,25	0,5	0,6	1,0	1,0	0,3	0,67	1,0
	2	1323	70	0,71	0,0	0,0	0,6	1,0	0,75	0,7	0,66	0,9
	3	0723	274	0,63	0,25	0,5	0,5	0,4	0,0	0,0	0,26	0,75
	4	0823	275	0,50	0,50	1,0	0,3	0,7	0,0	0,0	0,33	0,9
	5	0727	71	0,47	0,0	0,0	0,2	1,0	0,0	0,0	0,28	0,9
	6	0625	239	0,45	0,0	1,0	0,9	0,3	0,0	0,6	0,45	0,9
	7	0827	205	0,43	0,0	1,0	0,6	1,0	0,25	0,0	0,48	0,9
	8	0822	292	0,41	0,50	0,50	0,5	0,5	0,0	0,0	0,31	0,7
	9	1130	127	0,39	0,0	0,0	0,3	0,9	0,0	0,7	0,43	0,75
	10	0624	258	0,38	0,75	1,0	0,6	0,0	0,0	0,0	0,26	0,25
	Номер			Меры сходства		Номер			Меры сходства			
	п/п	квadrата	объекта	Б	$\bar{\Pi}(\text{ПР})$	п/п	квadrата	объекта	Б	$\bar{\Pi}(\text{ПР})$		
ПОИСКОВЫЕ РАБОТЫ	1	3431	123	0,87	0,75	15	3031	119	0,62	0,60		
	2	0320	324	0,84	0,55	16	1703	661	0,62	0,60		
	3	0221	2	0,81	0,75	17	1009	564	0,62	0,50		
	4	1706	624	0,76	0,75	18	1603	660	0,62	0,75		
	5	0521	307	0,70	0,65	19	1707	605	0,60	0,60		
	6	0810	548	0,70	0,60	20	0708	577	0,59	0,60		
	8	2105	646	0,69	0,50	21	0914	483	0,59	0,50		
	9	0814	482	0,68	0,70	22	0910	549	0,59	0,50		
	10	1906	626	0,66	0,50	23	1214	486	0,55	0,70		
	11	1208	582	0,66	0,60	24	3331	122	0,55	0,60		
	12	2005	645	0,64	0,60	25	5818	1388	0,53	0,50		
	13	0710	547	0,64	0,50	26	0909	563	0,53	0,50		
	14	0319	351	0,64	0,70	27	1308	583	0,52	0,60		
						28	1108	581	0,51	0,60		
		29	1407	602	0,50	0,60	41	0722	291	0,50	0,10	
		30	2106	6	0,90	0,50	42	1907	607	0,50	0,30	
		31	1314	7	0,75	0,50	43	3231	121	<0,5	0,50	
		32	3032	9	0,72	0,50	44	4231	1076	<0,5	0,50	
		33	1212	75	0,70	0,75	45	3531	124	<0,5	0,60	
		34	1531	105	0,63	0,50	46	4829	1138	<0,5	0,50	
	35	4131	1075	0,62	0,70	47	2202	808	<0,5	0,50		
	36	3731	8	<0,50	0,50							
	37	3131	120	0,50	0,60							
	38	3329	175	0,50	0,30							
	39	0926	222	0,50	0,00							
	40	1026	223	0,50	0,00							

Примечание. Б - мера сходства с положительными эталонами, полученная статистическими методами /14/; $\bar{\Pi}_1, \bar{\Pi}_2, \bar{\Pi}_3, \bar{\Pi}_4, \bar{\Pi}_5$ и $\bar{\Pi}_6$ - коэффициенты принадлежности проб к первому классу отдельно по структурно-тектоническим, литолого-стратиграфическим, магматическим, метаморфо-метасоматическим, минералогическим и геохимическим признакам; $\bar{\Pi}$ - коэффициент принадлежности проб к первому классу по всем признакам (средняя мера сходства) на стадии поисково-оценочных работ; $\bar{\Pi}(\text{ПР})$ - коэффициент принадлежности проб к первому классу по всем признакам на стадии поисковых работ; номера квадратов и объектов соответствуют таковым в работе /14/.

бот в качестве перспективных. Это очень важный момент. В рассматриваемых задачах прогноза, связанных между собой определенной последовательностью, объект распознавания ("объект-проба") сначала оценивается на стадии поисковых работ и, в случае перспективности, допускается к следующей поисково-оценочной стадии.

В качестве первоочередных ($\bar{\Pi} > 0,6$) выделены два участка (объекты № 240, 70, см. табл.13), которые ранее статистическими методами получили наибольшие меры сходства (Б) с I классом. К объектам второй очереди ($\bar{\Pi} > 0,4$) отнесены № 205, 239, 127. Остальные пять объектов на данном этапе решения задачи прогноза по отдельным классам признаков можно отнести к разряду малоперспективных.

К участкам, рекомендуемым для поисково-оценочных работ первой и второй очереди, относятся два участка к северо-северо-западу от Норильска (№ 240, 239), в пределах которых распространена хорошо расслоенная Пясинско-Вологочанская интрузия. Для участка № 240 установлено благоприятное сочетание ($\Pi_2, \Pi_3, \Pi_4, \Pi_5 \geq 0,5$) литолого-стратиграфических, магматических, метасоматических и минералогических признаков; неблагоприятное сочетание ($\bar{\Pi}_1, \bar{\Pi}_6 < 0,5$) по структурно-тектоническим и геохимическим характеристикам. Для участка № 239 малоинформативную роль сыграли структурно-тектонические, метаморфо-метасоматические и минералогические признаки.

В число рекомендуемых для поисково-оценочных работ первой очереди вошел также участок (№ 70), известный под названием Имангдинского, на котором имеется ряд расслоенных интрузивных тел. Для него установлено благоприятное сочетание по четырём классам признаков, кроме структурно-тектонических и литолого-стратиграфических. Отметим, что интрузии Имангдинская и Пясинско-Вологочанская (особенно Вологочанская ветвь) отнесены к перспективным при сравнительном математическом изучении дифференцированных трапловых интрузий /7/ и петрохимических особенностей "шкритовых" габбро-долеритов /8/.

К участкам второй очереди отнесены объекты № 205, 127. Первый из них расположен к северо-северо-западу от Талнахского месторождения и в его пределах известны так называемые Тулаек-Тасаские дифференцированные интрузии. Благоприятным сочетанием для участка № 205 являются литолого-стратиграфические, магматичес-

кие и метаморфо-метасоматические признаки. Для объекта № I27 (к нему приурочена Тальминская дифференцированная интрузия) информативными оказались только две группы признаков: метаморфо-метасоматические и геохимические. По-видимому, данный участок может быть рекомендован для дальнейшего изучения по другим четырем группам признаков с учетом того, что большая его часть закрыта четвертичными отложениями.

Стадия поисковых работ. Наибольшие средние коэффициенты принадлежности \bar{P} (ПР) $> 0,5$, (см. табл. I3) по восьми решающим правилам получили 42 участка. Из них только первые 29 объектов из табл. I3 (60 %) совпали с рекомендованными ранее по результатам статистической обработки, 10 вошли из числа рекомендованных ранее для общих поисков и 3 из числа неперспективных (№ I24, II38, 808).

В качестве первоочередных на стадии поисковых работ рекомендуются девять участков, для которых \bar{P} (ПР) $> 0,7$ (см. табл. I3; № I23, 2, 624, 482, 35I, 660, 486, 75, IO75). Отметим, что кроме двух последних, все они входили в число перспективных и ранее /I4/, и на этой стадии.

Для данной постановки задачи прогноза оценка перспективности участков, тем не менее, считается предварительной, так как она получена с использованием приближенных решений (см. табл. I2).

Для стадии поисковых работ проводилась оценка пятидесяти участков, отнесенных ранее по результатам статистической обработки к неперспективным. По использованным решающим правилам только три из них (6 %) получили удовлетворительные средние коэффициенты принадлежности (см. табл. I3). Один из них (№ II38 - Ханарский участок) получил $\bar{P} = 0,5$. Ханарская интрузия была отнесена к перспективным и при сравнительном математическом изучении дифференцированных интрузий /7/. В связи с этим участок заслуживает дальнейшего рассмотрения при совместном анализе признаков разных групп.

В заключение отметим, что в оценке перспективности участков в самом надежном варианте - на стадии поисково-оценочных работ, а также по отношению к 50 неперспективным участкам, статистический и логико-математический подходы дали достаточно хорошую сходимость результатов. Это ещё раз подтверждает правомочность использования обоих подходов к решению рассмотренных задач по ме-

таллогеническому районированию и оценке перспектив рудоносности территории Северо-Сибирской никеленосной области.

Литература

I. Бишаев А.А. Метод нахождения целевой информативности признаков. - В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск: ИГиГ СО АН СССР, 1973, с.55-56.

2. Бугаец А.Н., Дуденко Л.Н. Математические методы при прогнозировании полезных ископаемых. Л.: Недра, 1976. 270 с.

3. Геология и математика / Ю.А.Воронин, Н.А.Ионина, Г.Н.Каратаева и др. Новосибирск: Наука, 1970. 224 с.

4. Загоруйко Н.Г. Структура проблемы распознавания слуховых образов и методы её решения. - В кн.: Распознавание слуховых образов. Новосибирск: ИГиГ СО АН СССР, 1970, с.3-54.

5. Загоруйко Н.Г. Эмпирическое предсказание. Новосибирск: Наука, 1979. 123 с.

6. Йереског К.Г., Клован Д.И., Реймент Р.А. Геологический факторный анализ. Л.: Недра, 1980. 223 с.

7. Карбышев В.Д., Бишаев А.А., Золотухин В.В. и др. Перспективы рудоносности дифференцированных интрузий севера Сибирской платформы (на основе логико-математических методов обработки геологической информации). - В кн.: Петрология и перспективы рудоносности траппов севера Сибирской платформы. Новосибирск: Наука, 1978, с.216-257.

8. Карбышев В.Д. Использование процентных весовых величин в задачах классификации и прогноза (на примере габброидов и пикритов северо-запада Сибирской платформы). - В кн.: Петрохимия. Аспекты петрологии и металлогении. Новосибирск: ИГиГ СО АН СССР, 1982, с.129-146.

9. Коллатц Л. Функциональный анализ и вычислительная математика, М.: Мир, 1969. 447 с.

10. Лбов Г.С. Методы обработки разнотипных экспериментальных данных. Новосибирск: Наука, 1981. 158 с.

11. Логико-математическая обработка геологической информации (теория и математический аппарат). Новосибирск: ИГиГ СО АН СССР, 1976. 152 с.

12. Программные комплексы для целевой обработки информации. - Новосибирск: ИГиГ СО АН СССР, 1977. 164 с.

13. Раудис Ш.Ю. Ограниченность выборки в задачах классификации. - В кн.: Статистические проблемы управления. Вильнюс, 1976. 15 с.

14. Сухов Л.Г., Дуденко Л.Н., Наторхин И.А. Количественные методы прогнозирования эндогенных рудных месторождений. Л.: Недра, 1981. 139 с.

В.О.Красавчиков, В.Я.Санин

ИССЛЕДОВАНИЕ ЛОГИКО-МАТЕМАТИЧЕСКИМИ МЕТОДАМИ
ПРОБЛЕМЫ МОРФОЛОГИЧЕСКОЙ КЛАССИФИКАЦИИ
(на примере класса *Bivalvia*)

Критерии выделения вида на протяжении многих десятилетий обсуждаются на страницах специальной литературы /15/. Информация о видовой принадлежности индивидов имеет определяющее значение и при выделении таксонов более высокого ранга. Исходя из этого, авторы попытались рассмотреть следующие вопросы: что такое вид с точки зрения общности свойств описаний составляющих его индивидов, и нельзя ли построить математическую модель для выделения таксонов более высокого, чем вид, ранга на основе анализа описаний индивидов с известной видовой принадлежностью. Это позволило бы наметить путь к проверке и уточнению классификации для класса *Bivalvia*, т.е. составить перечень характеристических признаков и, исходя из общности свойств описаний, рассортировать индивиды по сериям, относящимся к одному и тому же виду; затем, используя результаты сортировки и математическую модель для выделения таксона более высокого ранга, дополнительно обосновать выделение родов, семейств и т.д.

На данном этапе исследований в число характеристических признаков включались только морфологические. Их отбор осуществлялся исходя из двух требований: а) максимально полного отражения разнообразия индивидов и б) однозначности определения признака любым квалифицированным специалистом.

Принципиальная применимость математических методов при биологической классификации (в частности, в палеонтологии) несомненна. С другой стороны, общеизвестна определённая осторожность, с которой опытные систематики относятся к машинной таксономии /4,8/. Не в последнюю очередь подобное отношение обусловлено тем, что стандартные методы обработки палеонтологического материала не извлекают всей информации, необходимой для достаточно адекватных моделей таксонов. Но проблема автоматической классификации палеонтологических данных остаётся актуальной. При этом "перспективы развития и применения математических методов в таксономии сводятся к моделированию традиционных методов работы палеонтолога" /4, с.2/. Настоящая работа отличается от имеющихся публикаций более детальным учётом специфики трёх групп признаков, традиционно указываемых в описаниях индивидов (первичные мерические /Ю, с.33/, соотносительные и качественные (структурные) признаки). Делается попытка установить по каждой из трёх групп признаков, содержит ли она (с точки зрения объективных критериев) полезную информацию для классификации индивидов внутри таксона, и какими математическими методами эта информация может быть извлечена.

I. Исходный материал

Объектом исследования является группа видов надсемейства *Arcasea* (подкласс *Pteriomorpha*). Из четырех семейств, входящих в состав надсемейства (*Arcidae*, *Cucullaeidae*, *Noetidae* и *Grammatodontidae*), в верхнеюрских-нижнемеловых отложениях на севере СССР встречаются лишь представители семейств *Cucullaeidae* и *Grammatodontidae*.

Родовой и видовой состав мезозойских аркаей на севере СССР относительно беден. Проведенная ревизия прежних работ по мезозойской фауне севера СССР установила 10 видов аркаей, относящихся к родам *Grammatodon*, *Cosmetodon* (семейство *Grammatodontidae*), три вида рода *Lopatania* и один вид рода *Paralopatania* (семейство *Cucullaeidae*) /1,3,11,12,18,19,20,22,23,24/.

В первом специальном исследовании мезозойских аркаей севера СССР впервые описано девять видов: *Dicranodonta yatriensis*,

Lopatinia woodsi, *L.taimyrensis*, *L.uralica*, *L.gigantae* (семейство *Cucullaeidae*) и *Grammatodon lopsiensis*, *G.gracialis*, *Cosmetodon bojarkensis*, *C.chetaensis* (семейство *Grammatodontidae*) /18/.

К материалам предъявлялись следующие требования: достаточная представительность выборок (исключались виды, представленные единичными экземплярами), хорошая сохранность (продиктовано необходимостью полных замеров) и достаточная детальность описаний (в ряде предыдущих работ описания ограничивались несколькими определяющими фразами и 2-3 замерами).

Указанным требованиям полностью отвечают описания видов аркачей: *Dicranodonta yatriensis*, *Lopatinia taimyrensis*, *L.arctica*, *L.uralica*, *Grammatodon lopsiensis*, *Cosmetodon bojarkensis* и *C. borissiakii*. При анализе качественных признаков использовался также вид *L.woodsi* (5 экз.), так как качественные признаки постоянны в пределах вида и объем выборки в данном случае особой роли не играет.

Для проверки некоторых предположений, возникших в результате обработки представителей надсемейства *Arcacea*, привлекались данные по семейству *Buchiidae* (надсемейство *Pterioidea*) - виды *Buchia keyserlingi* и *B.tolmatschowi* и семейства *Astartidae* (подкласс *Heterodonta*) - *Astarte lyapunensis* и *A.supraextensa* /9,10/.

Признаковое пространство для представителей надсемейства *Arcacea* (подкласс *Pteriomorphia*) включает из первичных мерических признаков длину раковины (Д), высоту (В), толщину створки (Т), длину передней части (ДПЧ), высоту ареи (u_a), высоту макушки (к), ширину (s_a) и длину (l_a) ареи, апикальный (макушечный) угол ($\angle\alpha$) и угол схождения связочных борозд ($\angle\gamma$) (линейные размеры в мм - пример - табл.3).

Соотносительные признаки включают некоторые традиционные отношения: удлиненность (В/Д), выпуклость (Т/В), равносторонность (ДПЧ/Д). Кроме того, привлекается ряд следующих отношений: u_a/k , $l_a/Д$, s_a/l_a , Т/Д, s_a/u_a , u_a/l_a , $s_a/Т$, В/ l_a , Т/ l_a , ДПЧ/ l_a . Указанные признаки традиционным путём разбиты на градации, наиболее удобные для работы /9,10,17,18/. Каждой градации присвоен код (табл.2). Качественные признаки перечислены в табл.1.

Таблица 1

В и д ы	Номер признака												
	I6	I7	I8	I9	20	2I	22	23	24	25	26	27	28
<i>Dicranodonta yatriensis</i>	I	0	4	2	2	I	I	0	0	0	0	0	I
<i>Lopatinia arctica</i>	0	0	3	2	2	I	I	0	0	0	0	I	2
<i>Lopatinia uralica</i>	I	0	3	I	2	I	I	0	0	0	0	I	2
<i>Lopatinia taimyrensis</i>	0	0	2	2	2	I	I	0	0	0	0	I	2
<i>Grammatodon lopsiensis</i>	I	I	3	0	2	2	I	0	0	I	I	2	0
<i>Cosmetodon bojarkensis</i>	I	0	2	2	2	2	I	2	I	I	I	2	0
<i>Cosmetodon borissiaki</i>	I	0	3	2	2	2	I	2	I	I	I	2	0

Примечание. I6 - киль (да - I, нет - 0), I7 - ушко (да - I, нет - 0), I8 - рёбра (нет - 0, концентрические - I, радиальные - 2, смешанные - 3, радиальные двух порядков - 4), I9 - линии нарастания (0 - нет, I - тонкие, 2 - грубые), 20 - арка (0 - нет, I - равносторонняя, 2 - неравносторонняя), 2I - лигамент (I - амфидетный, 2 - опистодетный), 22 - прозогирная макушка (I - да, 0 - нет), 23 - зияние (0 - нет, I - переднее, 2 - брюшное), 24 - биссусный вырез (I - да, 0 - нет), 25 - замок (0 - симметричный, I - асимметричный), 26 - боковые зубы (0 - задние по длине равны передним, I - задние длиннее передних), 27 - форма боковых зубов (0 - прямые, I - свисающие, 2 - передние скошенные), 28 - центральные зубчики (0 - не обособлены, I - одинаковые, 2 - двух видов).

Таблица 3

№ п/п	Д	В	Т	ДПЧ	u_a	к	s_a	l_a	$\angle \alpha$	$\angle \gamma$
I	29,8	19,0	9,6	II,5	I,3	3,5	2,0	24,0	I00	I45
2	28,6	18,5	8,0	II,0	2,0	4,0	2,2	25,4	I00	I42
3	27,7	19,7	8,8	I2,3	2,5	4,1	2,1	2I,6	I02	I50
4	27,6	18,5	8,5	IO,6	I,5	3,5	2,1	22,2	I04	I50
5	25,8	I7,3	8,2	IO,5	I,5	4,1	I,4	20,5	I02	I60
6	25,5	I7,7	6,8	9,3	I,1	3,0	I,3	20,7	I02	I55
7	22,4	I4,4	5,8	9,2	0,9	2,7	I,1	I8,7	I05	I55
8	22,1	I3,6	6,3	7,9	0,7	2,3	0,8	I8,5	I05	I60
9	2I,8	I3,2	5,8	8,6	0,5	2,0	I,4	I8,0	I02	I60
IO	2I,4	I2,7	6,7	8,3	I,2	3,5	I,5	I9,5	I03	I55
II	2I,3	I2,6	5,7	7,7	0,4	I,7	I,0	I6,5	I02	I55
I2	20,6	I2,5	5,8	7,5	I,0	2,3	I,0	I8,1	I02	I55

Таблица 2

1	2	3	4	5
В/Д	Т/В	ДПЧ/Д	Т/Д	$\angle \alpha$
.95-.85	1	.60-.50	0	.50-.40
.86-.75	2	.51-.40	1	.41-.30
.76-.65	3	.41-.30	2	.31-.20
.66-.55	4	.31-.20	3	.21-.10
.56-.45	5			
				80-85
				86-90
				91-95
				96-100
				101-105
				106-110
				111-115
				116-120

6	7	8	9	10
u_a/k	s_a/l_a	s_a/u_a	$\angle \gamma^*$	u_a/l_a
.30-.40	1	.00-.05	0	.50-.70
.41-.50	2	.06-.10	1	.61-.90
.51-.60	3	.11-.15	2	.91-1.10
.61-.70	4	.16-.20	3	1.11-1.30
		.21-.30	4	1.31-1.50
			5	1.51-1.70
			6	1.71-1.90
			7	1.91-2.10
			8	2.11-2.30
			9	2.31-2.50
			10	

11	12	13	14	15
l_a/d	s_a/T	В/ l_a	Т/ l_a	ДПЧ/ l_a
.20-.30	0	.05-.10	1	.50-.70
.31-.40	1	.11-.15	2	.71-.90
.41-.50	2	.16-.20	3	.91-1.10
.51-.60	3	.21-.25	4	1.11-1.30
.61-.70	4	.26-.30	5	1.31-1.50
.71-.80	5	.31-.35	6	1.51-1.70
.81-.90	6	.36-.40	7	1.71-1.90
			8	1.91-2.10
			9	2.11-2.30
			10	2.31-2.50
			11	2.51-3.00

2. Анализ данных по качественным, соотносительным и первичным мерическим признакам

Данные по каждому из трех типов признаков вначале проанализированы отдельно с учетом их специфики. Так, для первичных мерических признаков различия их абсолютных величин, как правило, не имеют диагностического значения в связи с возрастными различиями индивидов (фактор роста) и проблематичностью зафиксированных пределов изменчивости этих признаков внутри таксонов. С другой стороны, наличие четко выраженных корреляционных полей, указывающих на наличие линейной зависимости между этими признаками, в принципе позволяет использовать их при проверке гипотезы о принадлежности совокупности индивидов к одному и тому же виду на основе использования регрессионных методов.

Соотносительные признаки в известной мере свободны от влияния фактора роста, так что сходство-различие соотносительных признаков может нести полезную информацию о принадлежности индивидов к таксону.

Качественные признаки постоянны в пределах вида, что делает их наиболее удобными для машинной обработки и проверки классификационных построений.

2.1. Анализ данных по качественным признакам

Для каждой пары видов Q_p , Q_r (из восьми отобранных) подсчитаем число попарных различий по качественным признакам (табл.4).

Согласно данным табл.4, наблюдается определенная связь между степенью родства видов и числом попарных различий. Так, для видов, относящихся к различным семействам, имеется не менее семи различий (7-9 или 54-69 %); для видов, относящихся к различным родам одного семейства - 4-5 различий (31-38 %), для видов одного рода - не более трех (0-23 %). Исходя из вышесказанного, рассмотрим табл.4 как матрицу данных о близости видов и на основе этих данных опишем свойства таксонов (родов, семейств) традиционной классификации с целью нахождения (применительно к анализируемому материалу) математической модели таксона более высокого чем вид ранга:

Таблица 4

В и д ы	I	2	3	4	5	6	7	8
1. <i>Dicranodonta yatriensis</i>	4	4	4	4	8	8	8	
2. <i>Lopatinia arctica</i>	4	2	1	2	8	9	8	
3. <i>L.uralica</i>	4	2	3	0	7	9	8	
4. <i>L.taimyrensis</i>	4	1	3	3	9	8	9	
5. <i>L.woodsii</i>	4	2	0	3	7	9	8	
6. <i>Grammatodon lopsiensis</i>	8	8	7	9	7	5	4	
7. <i>Cosmetodon bojarkensis</i>	8	9	9	8	9	5	1	
8. <i>C.borissiakii</i>	8	8	8	9	8	4	1	

а) для видов, относящихся к одному семейству, число различий не превосходит пяти; относящихся к разным семействам, оно не менее семи;

б) для видов, относящихся к одному роду, число различий не превосходит трёх, к разным родам одного семейства – не менее четырёх.

Приходим к следующей математической модели таксона более высокого чем вид ранга: пусть для каждой пары видов Q_p, Q_r из анализируемой совокупности Q_1, \dots, Q_N задана характеристика близости α_{pr} . Величины α_{pr} – элементы упорядоченного множества A , причем если $\alpha_{pr} \leq \alpha_{ij}$, то это означает, что близость между видами Q_p и Q_r не меньше, чем между Q_i и Q_j . Совокупность видов $F = \{Q_j, \dots, Q_k\}$ образует кластер, если найдётся элемент h из A такой, что $A) \alpha_{pr} \leq h$ для любых Q_p, Q_r из F ;

Б) если $Q_p \in F, Q_r \notin F$, то $\alpha_{pr} > h$.

Если данные о близости берутся из табл.4, то все анализируемые роды и семейства рассматриваемой группы являются кластерами в указанном выше смысле. Однако чтобы проверить адекватность этой модели, необходимо также выяснить, есть ли еще совокупности, являющиеся кластерами, и не приводят ли дополнительные кластеры к классификации, существенно отличной от традиционной.

Возникает, таким образом, алгоритмическая проблема перечисления всех кластеров. Немаловажны и вопросы наглядного представления результатов кластеризации. Для решения этих проблем воспользуемся математическим аппаратом, описанным в работе Д.В. Матулы /16/, поскольку модель А, Б укладывается в рамки развиваемого в /16/ подхода к кластер-анализу, а применение теории графов для рассматриваемой модели приводит к нетрудоёмким алгоритмам и наглядному представлению результатов обработки информации^{*)}.

Граф $G = (V, E)$ состоит из непустого множества вершин $V = \{Q_1, \dots, Q_N\}$ и множества рёбер $E = \{e_1, \dots, e_m\}$, $m = N(N-1)/2$, где каждое ребро e_k есть отдельная пара $Q_i, Q_j \in V$, обозначаемая как $e_k = Q_i Q_j$. Если $m = N(N-1)/2$, то G является полным графом, а если на множестве рёбер задано отношение предпорядка $e_k \leq e_j$, то это упорядоченный граф^{**)}.

В моделях кластер-анализа, основанных на теории графов, вершины (элементы множества V) принято рассматривать как объекты, а рёбра (элементы множества E) — как связи. В частности, граф близости $P = (V, E)$ — это упорядоченный граф, у которого $V = \{Q_1, \dots, Q_N\}$ — множество объектов (в нашем случае — видов), подвергаемых кластеризации, а $E = \{e_1, \dots, e_M\}$ — множество пар объектов, называемых связями. Отношение предпорядка на связях определяется данными о близости объектов. Точнее говоря, если $e_i = Q_p Q_r$, $e_j = Q_s Q_t$, то выражение $e_i \leq e_j$ означает, что объекты Q_p и Q_r не менее сходны, чем объекты Q_s и Q_t . Ситуация, когда $e_i < e_j$ и $e_j \leq e_i$ означает, что различные в принципе связи e_i и e_j имеют тот же порядок, что и в равенстве $e_i = e_j$; в этом случае будем писать $e_i \equiv e_j$. Таким образом, полный граф близости представляет в сжатом виде порядковое отношение близости.

Если рёбра e_i и e_j таковы, что $e_i \leq e_j$ и $e_i \neq e_j$, то будем писать $e_i < e_j$. Запишем рёбра таким образом, что если $e_i < e_j$, то $i < j$. Уровнями расщепления графа близости назовем уровни $s=0, s=M$, а также все s , $1 \leq s \leq M-1$, для которых $e_s < e_{s+1}$.

*) Математический аппарат для анализа моделей типа А, Б методами теории графов неоднократно рассматривался в литературе (см. например, работу /30/).

**) Отношение aRb называется отношением предпорядка на множестве E , если: 1) aRa , 2) aRb и bRc влечёт aRc , 3) aRb или bRa для любых $a, b, c \in E$.

Для каждого уровня расщепления рассмотрим граф $H_s = (V, E_s)$, где $E_s = \{e_1, \dots, e_s\}$ не предполагается упорядоченным. Граф H_s называется пороговым подграфом s -го порядка для P ; наряду с H_s будем использовать также обозначение $H_{i,j}$, где $e_s = Q_i Q_j$. На рис.1 изображена диаграмма полного графа близости для восьми рассматриваемых видов, у которых отношение порядка для 28 связей определяется данными о близости из табл.4:

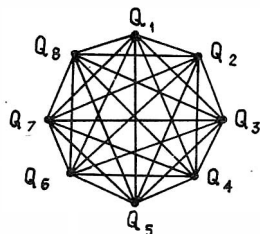


Рис.1. Полный граф близости для восьми видов

$Q_3 Q_5 < Q_2 Q_4 = Q_7 Q_8 < Q_2 Q_3 = Q_2 Q_5 < Q_3 Q_4 = Q_4 Q_5 < Q_1 Q_2 =$
 $= Q_1 Q_3 = Q_1 Q_4 = Q_1 Q_5 = Q_6 Q_8 < Q_6 Q_7 < Q_3 Q_6 = Q_5 Q_6 < Q_1 Q_6 = Q_1 Q_7 = Q_1 Q_8 = Q_2 Q_6 = Q_2 Q_8 =$
 $Q_3 Q_8 = Q_4 Q_7 = Q_5 Q_8 < Q_2 Q_7 = Q_3 Q_7 = Q_4 Q_6 = Q_4 Q_8 = Q_5 Q_7$.

Уровни расщепления: $s = 0, 1, 3, 5, 7, 12, 13, 15, 23$ (т.е. номера связей в приведенном выше упорядочении, для которых $e_s < e_{s+1}$). Пороговые подграфы H_s изображены на рис.2.

Сформулируем вышеописанную модель таксона в терминах теории графов. Напомним, что путь в произвольном графе $G = (V, E)$ — это последовательность попарно несовпадающих вершин Q_{j_1}, \dots, Q_{j_k} , $k \geq 1$, для которой любые соседние вершины $Q_{j_s}, Q_{j_{s+1}}$, $s = 1, \dots, k-1$ соединены ребром. Подмножество V вершин графа G называется компонентой связности, если: 1) V связно, т.е. для любых вершин Q, Q' из V существует путь Q_{j_1}, \dots, Q_{j_k} (где k — произвольное натуральное число) такой, что $Q_{j_1} = Q, Q_{j_k} = Q'$ и $Q_{j_s} \in V, s = 1, \dots, k$; 2) любое множество вершин V' такое, что $V \subset V', V \neq V'$ не является связным. Подмножество вершин $C \subset V$ является кликой, если любые две вершины из C связаны ребром. Клика максимальна, если любое множество вершин C' такое, что $C \subset C', C \neq C'$ не является кликой. Очевидно, что описанная ранее модель таксона в терминах теории графов получает следующее определение: совокупность видов $T = \{Q_{j_1}, \dots, Q_{j_k}\}$ образует кластер, если для некоторого порогового подграфа H_s совокупность T является компонентой связности и максимальной кликой.

Из рис.2 видно, что все таксоны традиционной классификации — кластеры, и, кроме того, кластерами являются совокупности $\{Q_3, Q_5\}, \{Q_2, Q_4\}$. В частности, таксон $\{Q_3, Q_5\}$ — компонента связности и максимальная клика графа H_1 ; $\{Q_2, Q_4\}$ — графа H_3 , как

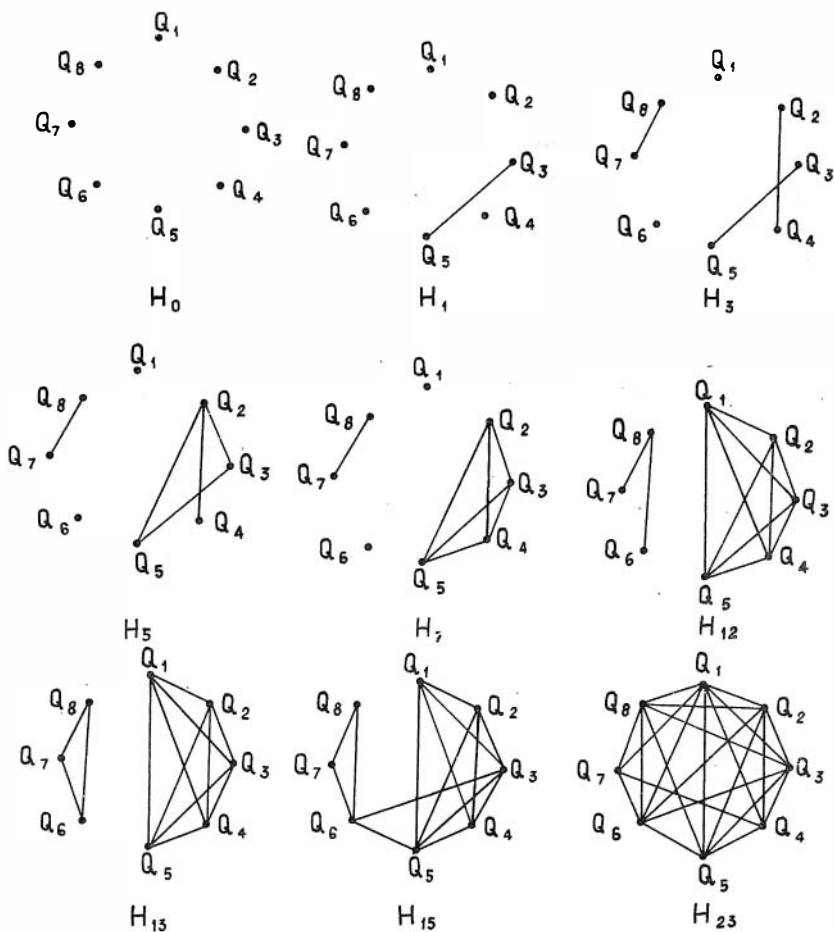


Рис.2. Пороговые подграфы для данных о близости по качественным признакам

и таксон $\{Q_7, Q_8\}$; род *Loratinia* - графа H_7 и т.д. Формирование таксонов происходит на уровнях I, 3, 7, 12, 13. На рис.3 представлена результирующая классификация. Выделение дополнительных, по сравнению с традиционной классификацией, кластеров свидетельствует, по-видимому, о целесообразности выделения на базе подро-

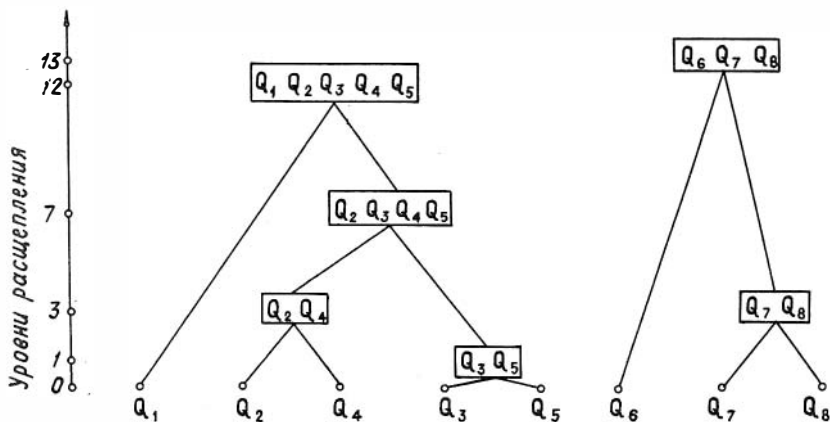


Рис.3. Результирующая классификация по качественным признакам

да *Lopatinia s.str.* двух новых самостоятельных подродов, объединяющих виды, относящиеся к вышеуказанным кластерам; таким образом, дополнительные кластеры приводят к детализации сложившейся систематики, не нарушая её.

2.2. Анализ данных по соотносительным признакам

Чтобы проверить, содержат ли соотносительные характеристики полезную информацию с точки зрения выделения таксонов более высокого, чем вид, ранга и подтвердить адекватность модели таксона, покажем, что эти характеристики позволяют построить матрицу данных о близости видов, для которой вышеуказанная модель дает классификацию, совпадающую с традиционной.

Предполагается, что для каждого индивида S исследуемой группы видов известны значения соотносительных признаков $X_1(S)$, ..., $X_{n-1}(S)$ и целевого признака $X_n(S)$. Целевой признак $X_n(S)$ для объекта S j -го вида принимает значение j . Соотносительные признаки замерены в шкале наименования /2I/, т.е. используются не сами количественные значения, а их градации, причём для сходства-различия индивидов существенно только совпадение или

несовпадение градаций. В табл.2 представлены градации, по которым велось кодирование признаков. Градации кодировались натуральными числами. Анализ данных проводился в четыре этапа:

1) для каждой пары видов Q_p, Q_r , исходя из анализа закодированных описаний составляющих их индивидов, выделяются сочетания целевого и характеристических (в данном случае соотносительных) признаков, взаимосвязь между которыми является значимой в некотором, оговорённом ниже, смысле;

2) для каждой пары Q_p, Q_r из множества значимых взаимосвязей отбираются те, которые позволяют однозначно предсказать для любого объекта обучения из Q_p или Q_r значение целевого признака по значениям характеристических. Для каждого сочетания признаков, взаимосвязь между которыми является значимой, фиксируется длина, т.е. число характеристических признаков, входящих в это сочетание;

3) для каждой пары Q_p, Q_r вычисляются последовательности $S_1^{pr} = (a_1, \dots, a_l)$, где a_j - число отобранных сочетаний длины j , $j = 1, \dots, l$ и $l = 1, 2, \dots$. S_1^{pr} называется спектром глубины l пары Q_p, Q_r . Спектр рассматривается как носитель информации о близости видов Q_p, Q_r по подобию числа попарных различий между качественными признаками: матрица с N строками и N столбцами, на пересечении i -й строки и j -го столбца которой расположен спектр S_1^{pr} , принимается в качестве матрицы данных о близости;

4) на множестве спектров глубины l задается линейный порядок (так называемое лексикографическое упорядочение) и принимается та же модель таксона, что и для качественных признаков. В результате при $l = 4$ получаем классификацию, совпадающую с традиционной.

Для отыскания взаимосвязей, упомянутых выше, используется один из алгоритмов метода табличных свойств /I3, I4/. Этот метод является модификацией тестового подхода к классификации предметов и явлений /7/ и предназначен для изучения совместного поведения признаков на множестве объектов.

Пусть виды Q_p, Q_r представлены индивидами S_1, \dots, S_m , где $m = m(p, r)$. Данные об объектах S_1, \dots, S_m , подлежащие обработке, сведены в таблицу (матрицу) "объект-признак" $T = (t_{ij})_{m \times n}$, в которой строки соответствуют объектам, а столбцы - признакам

так, что на пересечении i -й строки и j -го столбца стоит значение признака X_j объекта S_i .

Для формулировки определений используется аппарат функций алгебры логики /5/. Пусть M_1, \dots, M_n - множества значений признаков X_1, \dots, X_n соответственно. Каждому X_j сопоставим переменную x_j , пробегающую множество M_j , $j = 1, \dots, n$. Для $\sigma \in M_j$ определим на M_j функцию

$$x_j = \begin{cases} 1 & \text{при } x_j = \sigma, \\ 0 & \text{при } x_j \neq \sigma. \end{cases}$$

Пусть $\Omega \subseteq \{1, 2, \dots, n\}$ - произвольная непустая совокупность номеров столбцов таблицы T .

О П Р Е Д Е Л Е Н И Е 1. Функция Z_Ω вида

$$\bigvee_{i=1}^m \left(\bigwedge_{j \in \Omega} x_j^{t_{ij}} \right)$$

называется табличным свойством матрицы T , если $Z_\Omega \neq 1$.

Здесь $\bigwedge_{j \in \Omega} x_j^{t_{ij}} = x_p^{t_{ip}} \wedge x_q^{t_{iq}} \wedge \dots \wedge x_2^{t_{i2}}$ - конъюнкция $x_j^{t_{ij}}$

по всем $j \in \Omega$ для фиксированного i , а $\bigvee_{i=1}^m$ означает дизъюнкцию всех таких конъюнкций. Табличное свойство $(TC) Z_\Omega$ естественно интерпретировать как высказывание, характеризующее совместное поведение признаков с номерами из Ω на множестве объектов S_1, \dots, S_m . Условие $Z_\Omega \neq 1$ означает, что имеется, по крайней мере, одна комбинация значений этих признаков, не встречающаяся у объектов S_1, \dots, S_m .

О П Р Е Д Е Л Е Н И Е 2. $TC Z_\Omega$ называется тупиковым, если

$$\left\{ \Omega' \mid \Omega' \subsetneq \Omega, \quad \exists \Omega' \neq 1 \right\} = \emptyset, \quad (I)$$

$$\left(\left(\bigwedge_{\Omega' \subset \Omega} \exists \Omega' \right) \rightarrow \exists \Omega \right) \neq 1.$$

Здесь " \rightarrow " означает булеву импликацию, и конъюнкция берётся по всем $TC Z_{\Omega'}$, где Ω' - собственное подмножество Ω . Можно показать, что (I) в данном случае равносильно соотношению

$$\bigwedge_{\Omega' \subset \Omega} Z_{\Omega'} \neq Z_\Omega \quad (2)$$

Условие (I) означает, что совместное поведение признаков, чьи

номера составляют множество Ω , невозможно однозначно восстановить по их совместному поведению в поднаборах данного набора признаков. Взаимосвязь между целевым признаком X_n и характеристическими признаками считается значимой, если этому сочетанию признаков соответствует тупиковое ТС Z_{Ω} .

О П Р Е Д Е Л Е Н И Е 3. Тупиковое ТС Z_{Ω} называется тестовым, если $n \in \Omega$ и для любой пары объектов S_k, S_r (где $1 \leq k, r \leq m$) таких, что $X_n(S_k) \neq X_n(S_r)$, найдется отличный от n номер $j \in \Omega$ такой, что $X_j(S_k) \neq X_j(S_r)$. Тестовые табличные свойства Z_{Ω} соответствуют взаимосвязям, упомянутым в п.2 схемы анализа данных. Важнейшей характеристикой Z_{Ω} является его длина, т.е. число членов множества Ω , не считая целевого признака^{*)}.

На третьем этапе анализа данных для каждой пары Q_p, Q_r , $p \neq r$ найдём все тестовые тупиковые Z_{Ω} и для $l = 1, 2, \dots$ составим спектры $S_l^{pr} = (a_1, \dots, a_l)$, где $a_1 \geq 0$ - число тестовых тупиковых ТС длины один, $a_2 \geq 0$ - длины два и т.д. Так, на таблицах 5-8 представлены спектры для 15 пар шести исследованных видов при $l = 1, 2, 3, 4$.

Из табл.7 видим, что объекты, относящиеся к разным видам, имеют несовпадающие описания, причем для различения видов, относящихся к разным семействам, достаточно одиночных признаков; для различения видов внутри одного семейства всегда существуют сочетания из двух-четырех признаков (см. табл.5-8). При фиксированном l на множестве всех последовательностей натуральных чисел вида $\tilde{\beta} = (b_1, \dots, b_l)$ введем отношение порядка: $(a_1, \dots, a_l) < (b_1, \dots, b_l)$, если первая отличная от нуля разность $b_1 - a_1, b_2 - a_2, \dots, b_l - a_l$ положительна. Так, для спектров из табл.8 $(0, 0, 1, 1) < (0, 0, 1, 4) < (0, 0, 12, 20) < (0, 6, 3, 0) < (1, 25, 2, 0) < (2, 15, 7, 0) < (4, 4, 17, 2) < (6, 0, 7, 0) < (6, 3, 0, 0) < (7, 9, 3, 0) < (7, 10, 1, 0) < (7, 12, 0, 0) < (7, 14, 0, 0) < (8, 4, 1, 0) < (8, 7, 1, 0)$ или, что то же самое, $c^{2,3} < c^{1,3} < c^{1,2} < c^{5,6} < c^{4,5} < c^{4,6} < c^{1,4} < c^{2,4} < c^{3,4} < c^{1,6} < c^{1,5} < c^{3,5} < c^{2,5} < c^{2,6} < c^{3,6}$.

(Нижний индекс 4 для удобства записи опущен). Если рассматривать

*) На целесообразность учёта длин тестов при автоматической классификации впервые указал А.Н.Дмитриев /6/. Им же сформулированы основные принципы анализа распределения тестов по длинам и, в частности, положение о том, что чем меньше длина теста, тем он информативнее.

1-1

Таблица 5

В и д ы	I	2	3	4	5	6
1. <i>Dicranodonta yatriensis</i>		0	0	4	7	7
2. <i>Lopatinia uralica</i>	0		0	6	7	8
3. <i>Lopatinia taimyrensis</i>	0	0		6	7	8
4. <i>Grammatodon lopsiensis</i>	4	6	6		I	2
5. <i>Cosmetodon bojarkensis</i>	7	7	7	I		0
6. <i>Cosmetodon borissiaki</i>	7	8	8	2	0	

1-2

Таблица 6

В и д ы	I	2	3	4	5	6
I		(0,0)	(0,0)	(4,4)	(7,10)	(7,9)
2	(0,0)		(0,0)	(6,0)	(7,14)	(8,4)
3	(0,0)	(0,0)		(6,3)	(7,12)	(8,7)
4	(4,4)	(6,0)	(6,3)		(1,25)	(2,15)
5	(7,10)	(7,14)	(7,12)	(1,25)		(0,6)
6	(7,9)	(8,4)	(8,7)	(2,15)	(0,6)	

1-3

Таблица 7

В и д ы	I	2	3	4	5	6
I		(0,0,I2)	(0,0,I)	(4,4,I7)	(7,10,I)	(7,9,3)
2	(0,0,I2)		(0,0,I)	(6,0,7)	(7,14,0)	(8,4,I)
3	(0,0,I)	(0,0,I)		(6,3,0)	(7,12,0)	(8,7,I)
4	(4,4,I7)	(6,0,7)	(6,3,0)		(1,25,2)	(2,15,7)
5	(7,10,I)	(7,14,0)	(7,12,0)	(1,25,2)		(0,6,3)
6	(7,9,3)	(8,4,I)	(8,7,I)	(2,15,7)	(0,6,3)	

Виды	1	2	3	4	5	6
1		(0,0,12,20)	(0,0,1,4)	(4,4,17,2)	(7,10,1,0)	(7,9,3,0)
2	(0,0,12,20)		(0,0,1,1)	(6,0,7,0)	(7,14,0,0)	(8,4,1,0)
3	(0,0,1,4)	(0,0,1,1)		(6,3,0,0)	(7,12,0,0)	(8,7,1,0)
4	(4,4,17,2)	(6,0,7,0)	(6,3,0,0)		(1,25,2,0)	(2,15,7,0)
5	(7,10,1,0)	(7,14,0,0)	(7,12,0,0)	(1,25,2,0)		(0,6,3,0)
6	(7,9,3,0)	(8,4,1,0)	(8,7,1,0)	(2,15,7,0)	(0,6,3,0)	

спектр c_1^{pr} как данные о близости видов Q_p, Q_r , то при $l=4$ для анализируемых таксонов (родов, семейств) каждый представляет собой, как и в случае качественных признаков, максимальную клику и компоненту связности для некоторого порогового подграфа $N_v = N_{pr}$. На рис.5 представлены пороговые подграфы исходного графа близости для $l = 4$. Кластерами, как видно из рисунка, являются все традиционно выделенные роды и семейства, других кластеров нет. Таким образом, с помощью одной и той же модели таксона традиционная классификация может быть получена исходя как из качественных (структурных) признаков, так и из соотносительных характеристик.

Из табл.5 видно, что при $l = 1$ лексикографическое упорядочение спектров имеет вид $c^{1,2} \equiv c^{1,3} \equiv c^{2,3} \equiv c^{5,6} < c^{4,5} < c^{4,6} < c^{1,4} < c^{2,4} < c^{3,4} < c^{1,5} < c^{3,5} < c^{1,6} < c^{2,5} < c^{2,6} < c^{3,6}$. Уровни расщепления: $s = 0, 4, 5, 6, 7, 9, 13, 15$. На рис.4 представлены пороговые подграфы этих уровней.

Как видно из рис.4, кластерами, помимо некоторых исходных видов, оказываются совокупности $\{5,6\}, \{1,2,3\}, \{4,5,6\}$, т.е. род *Sometodon* и оба семейства. При $l = 2, 3$ кластеры те же самые. Таким образом, с ростом глубины спектра изменение классификации либо не происходит, либо сводится к дроблению таксонов; при $l = 4$ классификация совпадает с традиционной.

Следует подчеркнуть, что в настоящей работе модель таксона (род, семейство и т.д.) принята без учета его ранга. Опознание ранга таксона по описаниям составляющих его индивидов представляет собой самостоятельную задачу и, видимо, более сложную, чем выделение таксона. Некоторые подходы к её решению обсуждаются далее применительно к понятию вида с привлечением данных первичных измерений (мерических признаков) индивидов.

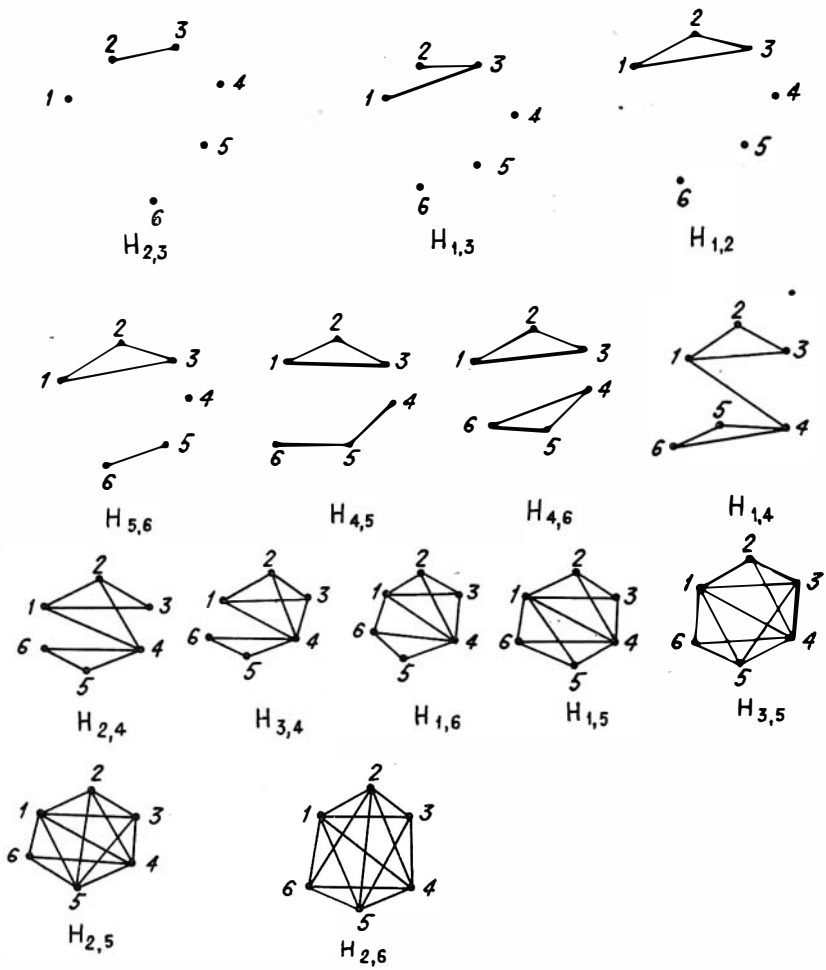


Рис.5. Пороговые подграфы для данных о близости по спектрам глубины 4

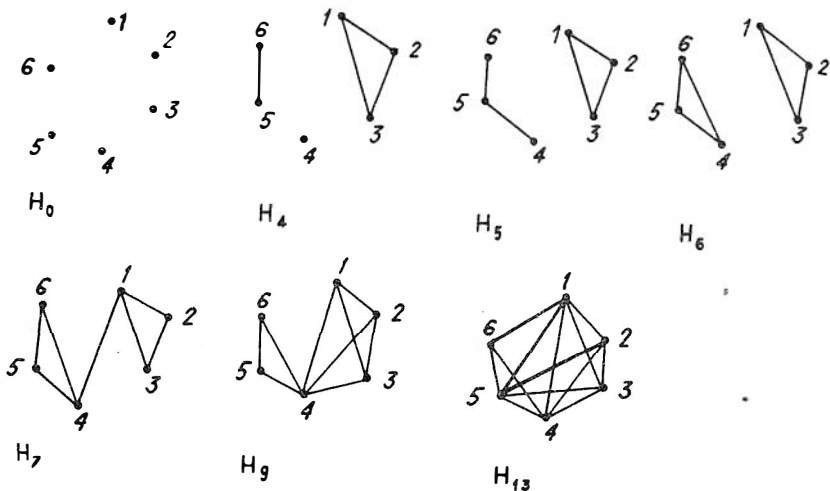


Рис.4. Пороговые подграфы для данных о близости по спектрам глубины I

2.3. Анализ данных по первичным мерическим признакам

Стандартный прием для описания зависимости между двумя признаками заключается в выборе коэффициентов a и b линейной функции $y = ax + b$, исходя из минимума среднеквадратического отклонения /28/

$$\frac{1}{m} \cdot \sum_{i=1}^m (ax^{(i)} + b - y^{(i)})^2, \quad (I)$$

где $(x^{(i)}, y^{(i)})$, $i=1, \dots, m$ - значения признаков X, Y у объектов S_i , $i=1, \dots, m$. Однако такой выбор коэффициентов не учитывает фактор роста организма, вследствие которого, в частности, малые отклонения от прямой $y=ax+b$ у особей малых размеров могут иметь более существенное значение для классификационных построений и практической диагностики, чем гораздо большие (по абсолютной величине) отклонения у особей больших размеров. Учитывая это, нами внесены некоторые изменения в стандартную регрессионную модель. Эти изменения, помимо вышесказанного, имеют своей целью придать линейной зависимости между признаками роль

характеристики, жёстко выделяющей фиксированный палеонтологический таксон на основе общности свойства (соответствия в определенных рамках линейной зависимости) всех составляющих данный таксон индивидов. При этом оценку соответствия объекта этой зависимости (соответствие всегда будет, естественно, приближенным) мы не будем связывать с теоретико-вероятностными предположениями в связи с малыми объемами выборок, типичными для ископаемого материала.

Отмеченных недостатков модели (I) можно избежать, используя для выбора коэффициентов a и b минимизацию по среднеквадратичному относительному отклонению (в предположении, что $y^{(1)}, \dots, y^{(m)} > 0$)

$$\frac{1}{m} \sum_{i=1}^m \left(\frac{ax^{(i)} + b - y^{(i)}}{ax^{(i)} + b} \right)^2, \quad (2)$$

где a и b выбираются таким образом, чтобы для любого $i=1, \dots, m$ значение $ax^{(i)} + b$ всегда было больше нуля. Однако и в этом случае возникает проблема: как, не прибегая к предположениям вероятностного характера, указать обусловленные данным способом выбора a и b дискретные рамки, определяющие, какой индивид удовлетворяет (опять-таки в приближенном смысле) зависимости $y=ax+b$, а какой - нет. Наличие подобных рамок позволило бы однозначно решать вопрос о невключении индивида в таксон на основе анализа информации по признакам X и Y .

Более приемлемой в этом смысле представляется модель, в которой требуется найти a и b такие, что $ax^{(i)} + b$ положительны для всех $i=1, \dots, m$ и минимизируют максимальное относительное отклонение на обучающей выборке

$$\min_{i=1, \dots, m} \max \left| \frac{ax^{(i)} + b - y^{(i)}}{ax^{(i)} + b} \right| \quad (3)$$

в предположении, что все $y^{(1)}, \dots, y^{(m)} > 0$. Как и в стандартном регрессионном анализе [28], задача (3) обобщается на случай, когда Y зависит линейно от n признаков X_1, \dots, X_n , и эта связь приближенно описывается соотношением $y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$. Задача (3) в данном случае заключается в следующем. При заданных положительных $y^{(1)}, \dots, y^{(m)}$ и неотрицательных $x_j^{(k)}$; $k=1, \dots, m$, $j=1, \dots, n$ найти числа a_1, \dots, a_n, b , минимизирующие максимальное относительное отклонение на обучающей выборке

$$\min_{i=1, \dots, m} \max_{x_1, \dots, x_n} \frac{|a_1 x_1^{(i)} + a_2 x_2^{(i)} + \dots + a_n x_n^{(i)} + b - y^{(i)}|}{a_1 x_1^{(i)} + a_2 x_2^{(i)} + \dots + a_n x_n^{(i)} + b} \quad (4)$$

при условии, что $a_1 x_1^{(i)} + \dots + a_n x_n^{(i)} + b$ положительно для всех $i=1, \dots, m$. Задача (4) может быть записана как задача нелинейного математического программирования [27]: найти a_1, \dots, a_n, b, z , при которых достигается минимум величины z

$$\min z \quad (5)$$

при ограничениях на переменные a_1, \dots, a_n, b, z

$$\begin{cases} \frac{|a_1 x_1^{(i)} + a_2 x_2^{(i)} + \dots + a_n x_n^{(i)} + b - y^{(i)}|}{a_1 x_1^{(i)} + a_2 x_2^{(i)} + \dots + a_n x_n^{(i)} + b} \leq z, \quad i=1, \dots, m \\ 0 \leq z \leq z_0 \\ a_1 x_1^{(i)} + a_2 x_2^{(i)} + \dots + a_n x_n^{(i)} > 0 \end{cases} \quad (5')$$

Здесь $z_0 > 0$ - порог значимости относительного отклонения, задаваемый с целью уменьшения трудоёмкости вычислений.

Пусть задача (5), (5') для выборки из некоторого таксона решена. Рассмотрим, какие очевидные практические приложения даёт её решение для палеонтолога.

I. Пусть для индивида s , относящегося к данному таксону, известны признаки x_1, \dots, x_n , а признак y неизвестен (например, из-за плохой сохранности створки). Значения известных признаков обозначим $\hat{x}_1, \dots, \hat{x}_n$, неизвестное значение y обозначим y . Пусть z^* оптимальное значение z , и a_1^*, \dots, a_n^*, b^* соответствующие z^* оптимальные значения остальных неизвестных. Тогда по принципу аналогии естественно принять, что для y , как и для объектов обучения,

$$\frac{|a_1^* \hat{x}_1 + a_2^* \hat{x}_2 + \dots + a_n^* \hat{x}_n + b^* - y|}{a_1^* \hat{x}_1 + a_2^* \hat{x}_2 + \dots + a_n^* \hat{x}_n + b^*} \leq z^*$$

Обозначив выражение в знаменателе через \hat{y} , при $\hat{y} > 0$, получаем

$$(1 - z^*) \hat{y} \leq y \leq (1 + z^*) \hat{y} \quad (6)$$

Чем меньше z^* , тем более точную оценку для y мы можем полу-

чить (опуская вопрос о представительности обучающей выборки). Важно отметить, что если для данных Y и X_1, \dots, X_n у всех видов исследуемого таксона максимальное относительное отклонение измеренного значения Y от расчётного не превосходит постоянной величины $\gamma \ll 1$, то, используя вместо (6) соответствующую оценку

$$(1 - \gamma) \hat{y} \leq y \leq (1 + \gamma) \hat{y}, \quad (6')$$

мы получаем границы изменения для y менее зависящие от степени представительности выборки данного вида.

2. Пусть для индивида S с известными значениями x_1, \dots, x_n и y характеристических признаков X_1, \dots, X_n , Y рассматривается вопрос о его принадлежности к таксону. Если значения признаков таковы, что неравенство (6) не выполняется, то следует ожидать, что S не относится к данному таксону. В случае же выполнения (6) при дальнейшем диагнозе следует проанализировать зависимости между другими первичными мерическими признаками.

Для численного решения задачи (5), (5') рассмотрим тесно связанную с ней, но более удобную для применения численных методов решения задачу:

$$\min z \quad (7)$$

при ограничениях

$$\left| \sum_{j=1}^n a_j x_j^{(i)} + b - y^{(i)} \right| \leq z \left(\sum_{j=1}^n a_j x_j^{(i)} + b \right), \quad i=1, \dots, m \quad (7')$$

$$0 \leq z \leq z_0.$$

Нетрудно убедиться, что при сделанных предположениях о положительности величин $x_j^{(i)}$, $y_j^{(i)}$ для $i=1, \dots, m$, $j=1, \dots, n$ разрешимость (5), (5') равносильна разрешимости (7), (7'), причем всякое решение задачи (7), (7') является решением задачи (5), (5') и наоборот - всякое решение задачи (5), (5') является решением задачи (7), (7'). Поэтому в дальнейшем вместо задачи (5), (5') будем рассматривать (7), (7'). "Удобство" задачи (7), (7') заключается в том, что среди её ограничений нет строгих неравенств, и отсутствуют дробные выражения в ограничениях. Используемый нами численный метод решения этой задачи основан на идее Чяни (Cheny) и Лоеба (Loeb) /26/, которая заключается в сведении решения нелинейной задачи к решению последовательности линейных задач и позволяет получать приближенные значения для

Таблица 9

Признак и аргу- мент- функция	В и д ы					
	Dicranodon- ta yatrien- sis	Lopatinia uralica	Lopatinia taimyrensis	Grammatodon lopsiensis	Cosmetodon bojarkensis	Cosmetodon borisiaki
I	2	3	4	5	6	7
I-2	.I0I7	.0542	.0320	.06I5	.0792	.I00I
2-I	.0870	.0595	.03I2	.0509	.0902	.II20
I-3	.I382	.I509	.I029	.0970	.I009	.II48
3-I	.I02I	.I755	.I034	.0820	.I066	.I025
I-4	.I477	.III6	.I202	.0845	.II24	$z^{\bar{x}} > .20$
4-I	.II48	.I673	.II53	.0623	.I403	.I862
I-5	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$
5-I	.I206	.I550	.0937	.I534	.20	.I27I
I-6	$z^{\bar{x}} > .20$.I6I2	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$
6-I	.II98	.I329	.I034	.I64I	$z^{\bar{x}} > .20$.I546
I-7	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$
7-I	.I304	.I698	.II65	.II07	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$
I-8	.I575	.I9II	.I596	.08I2	.0558	.I083
8-I	.I00I	.0943	.0853	.0790	.0509	.0927
2-3	.I427	.I608	.II33	.I395	.I00I	.I00I
3-2	.0972	.I559	.0972	.I563	.I00I	.0820
2-4	.I329	.0960	.I2I4	.09II	.II99	$z^{\bar{x}} > .20$
4-2	.I542	.II65	.I050	.I0I8	.I396	.I395
2-5	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$
5-2	.I509	.I034	.0845	.I82I	$z^{\bar{x}} > .20$.I083
2-6	$z^{\bar{x}} > .20$.II24	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$.I977	$z^{\bar{x}} > .20$
6-2	.I349	.0943	.I025	.I989	.I788	.I222
2-7	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$	$z^{\bar{x}} > .20$
7-2	.I706	.I2I4	.I066	.I780	$z^{\bar{x}} > .20$.I608
2-8	.I89I	.I626	.I366	.I0I7	.II48	.I59I
8-2	.I485	.0632	.0673	.I436	.0970	.I403
3-4	.I64I	.I386	.0824	.I034	.I509	$z^{\bar{x}} > .20$
4-3	$z^{\bar{x}} > .20$.I542	.I378	.I296	.I928	.I954

1	2	3	4	5	6	7
3-5	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
5-3	.I96I	.I50I	.I756	$z^{\#}>.20$	$z^{\#}>.20$.0968
3-6	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
6-3	.I920	.I542	.I493	.I780	$z^{\#}>.20$.I446
3-7	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
7-3	$z^{\#}>.20$.I4II	.I354	.I7I4	.I429	$z^{\#}>.20$
3-8	.I477	$z^{\#}>.20$.I993	.II48	.II85	.I8I3
8-3	.I427	.I477	.I444	.I32I	.0984	.I66I
4-5	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
5-4	.I862	.I087	.I296	.I626	$z^{\#}>.20$	$z^{\#}>.20$
4-6	$z^{\#}>.20$.I527	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
6-4	.I743	.I099	.I263	.I624	.I772	$z^{\#}>.20$
4-7	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
7-4	.I846	.II36	.I233	.I296	.I772	$z^{\#}>.20$
4-8	.I903	$z^{\#}>.20$	$z^{\#}>.20$.II50	.I4II	$z^{\#}>.20$
8-4	.I649	.II32	.I280	.I460	.0976	.20
5-6	.2002	.I477	.I0I7	.I7I4	.I628	.I600
6-5	$z^{\#}>.20$	$z^{\#}>.20$.I362	$z^{\#}>.20$.I809	.I82I
5-7	$z^{\#}>.20$.I673	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
7-5	$z^{\#}>.20$.I854	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
5-8	.I543	$z^{\#}>.20$.I66I	.I247	$z^{\#}>.20$.I940
8-5	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
6-7	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
7-6	$z^{\#}>.20$.I725	.I723	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
6-8	$z^{\#}>.20$	$z^{\#}>.20$.I993	.III6	$z^{\#}>.20$	$z^{\#}>.20$
8-6	$z^{\#}>.20$.I329	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$
7-8	.I444	$z^{\#}>.20$	$z^{\#}>.20$.0878	$z^{\#}>.20$	$z^{\#}>.20$
8-7	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$	$z^{\#}>.20$

Примечание. Принята следующая нумерация признаков: I - Д, 2 - В, 3 - Т, 4 - ДПЧ, 5 - u_a , 6 - к, 7 - s_a , 8 - l_a .

$a_1^{\#}, \dots, a_n^{\#}, b^{\#}, z^{\#}$ с требуемой точностью. Использование подхода Чинни и Лоеба, например, в задаче дробно-линейной аппроксимации,

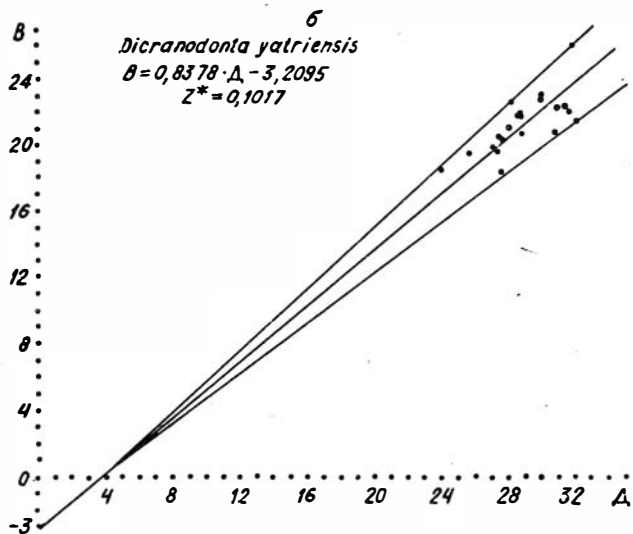
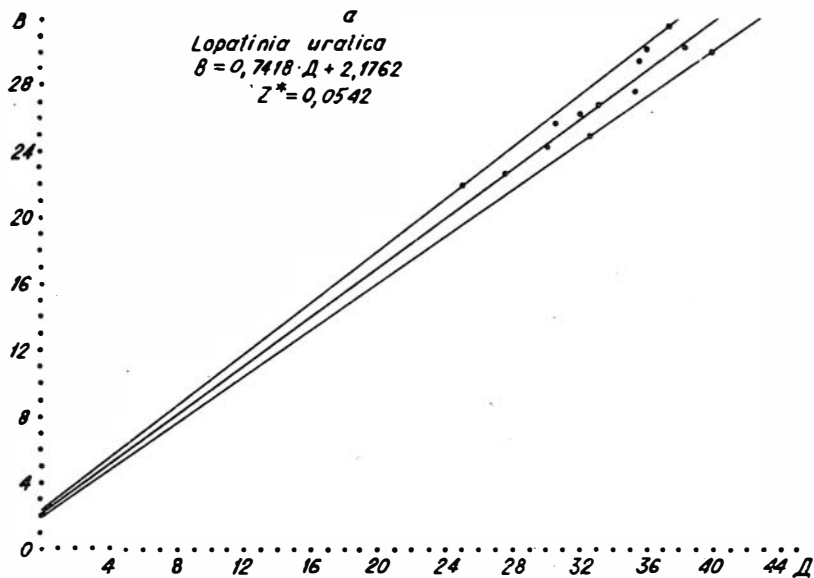


Рис.6. Графики линейной зависимости высоты створки раковины от ее длины при оптимальном относительном отклонении:

(а) - для вида *Lopatinia uralica*, (б) - для вида *Dicranodonta yatriensis*

имеющей с точки зрения вычислительных трудностей сходную с (7), (7') специфику, показало эффективность такого подхода /25/.

В данной работе задача (7), (7') решалась в частном случае для постановки (3) и для всех пар X, Y первичных мерических признаков. Так как этих признаков в данном случае восемь, имеем 28 пар.

Поскольку каждый из признаков может выступать как в качестве аргумента, так и в качестве функции, для каждого из обследованных видов решалось 56 задач типа (7), (7'). На табл.9 представлены значения максимального относительного отклонения z^{\max} для решенных задач. Наиболее четко линейная зависимость у аркачей проявляется для пар высота-длина (аргумент D, функция B) (рис.6, а, б) и длина - длина ареи (аргумент l_a , функция D). В обоих случаях z^{\max} находится в пределах 10 % (учитывая погрешность измерений и тот факт, что 0,0017 по относительному отклонению в абсолютном выражении в данном случае не превосходит 0,05 мм)^{*)}. Сказанное позволяет дополнить модель для вида аркачей двумя условиями: а) B линейно зависит от D с максимальным относительным отклонением в пределах 10 %; б) D линейно зависит от l_a с максимальным относительным отклонением в пределах 10 %.

Проверено, что в рамках использованного материала (см. раздел I) условие (а) соблюдается также для бухид и астартид.

* *
*
*
*

Используя описанный в 2.1-2.3 аппарат, наметим рациональную последовательность действий палеонтолога по выделению таксонов (видов, родов, семейств) при условии, что совокупность признаков заранее задана. Прежде всего, индивиды объединяются в группировки с полным совпадением всех заданных качественных признаков. Затем для каждой такой группировки с помощью регрессионного метода, описанного в 2.3, проверяется гипотеза о принадлежности индивидов данной группировки к одному виду. Если максимальное

^{*)} И в этом смысле десятипроцентный барьер позволяет говорить о линейных зависимостях между признаками как о своего рода "интегралах движения" /29/ для вида.

относительное отклонение по зависимости В от Д и (для аркачей) Д от l_a - в пределах 10 %, то принимается гипотеза, что исследуемая группировка индивидов с одинаковыми значениями качественных признаков состоит из особей одного вида. Если хотя бы одно из вышеуказанных отклонений выходит за рамки 10 %, то, как видно из изложенных результатов, в выборке скорее всего представлена смесь близкородственных видов, как например *Lopatinia uralica* и *Lopatinia woodsii*. Выделение видов из такой группировки авторами не рассматривалось в связи с отсутствием представительного материала. Поэтому ограничимся рассмотрением случая, когда для каждой совокупности индивидов с одинаковыми качественными признаками принимается гипотеза о принадлежности индивидов к одному виду. Тогда с помощью модели таксона по качественным и соотносительным признакам выделяются кластеры. Наибольшего внимания, по-видимому, заслуживают кластеры, выделяющиеся как по качественным, так и по соотносительным характеристикам. Вопрос об определении ранга кластера (таксона) более высокого таксономического положения чем вид, требует специального исследования, выходящего за рамки настоящей работы. Отметим только, что определенную информацию о ранге даёт максимальное число попарных различий между качественными признаками для входящих в данный таксон видов.

В заключение авторы выражают свою признательность А.Ф.Белосову, В.Г.Бердышеву, Е.А.Куширенко и Э.Ю.Куширенко за полезные советы и обсуждение результатов.

Литература

1. Бодылевский В.И. Новая раннемеловая *Cucullaea* центральной части Северной Сибири. - В кн.: Новые виды древних растений и беспозвоночных СССР. Вып. II, ч. I. М.: Наука, 1968, с. 179-180.
2. Бодылевский В.И., Шульгина Н.И. Юрские и меловые фауны низовьев Енисея. М.: Госгеолтехиздат, 1958. 196 с. (Тр. НИИГА, т. 39).
3. Борисьяк А.А. *Pelecypoda* юрских отложений Европейской России. Вып. II, Спб., 1905. 63 с. (Тр. Геол. ком., нов. сер., вып. 19).

4. Ванчуров И.А. Математические модели в палеонтологии. Автореф. докт. дис...., М.: МГУ, 1979. 24 с.
5. Дискретная математика и математические вопросы кибернетики. Т.1 М.: Наука, 1974. 311 с.
6. Дмитриев А.Н. Использование длин тупиковых тестов при обработке таблиц. - В кн.: Дискретный анализ. Новосибирск: ИМ СО АН СССР, 1970, вып.14, с.18-21.
7. Дмитриев А.Н., Журавлёв Ю.И., Кренделев Ф.П. О математических принципах классификации предметов и явлений. - В кн.: Дискретный анализ, Вып.7, Новосибирск: ИМ СО АН СССР, 1966, с.3-15.
8. Ёлкина В.Н., Ёлкин Е.А., Загоруйко Н.Г. О возможности применения методов распознавания образов в палеонтологии. - Геол. и геофиз., 1967, № 9, с.75-79.
9. Захаров В.А. Позднеюрские и раннемеловые двустворчатые моллюски севера Сибири и условия их существования. Ч.2. Семейство *Astartidae*. М.: Наука, 1970. 143 с. (Тр. ИГиГ, вып.113).
10. Захаров В.А. Бухиды и биостратиграфия бореальной верхней юры и неокома. М.: Наука, 1981. 270 с. (Тр. ИГиГ, вып.458).
11. Захаров В.А., Месежников М.С. Волжский ярус Приполярного Урала. Новосибирск: Наука, 1974. 214 с.
12. Захаров В.А., Щурьгин Б.Н. Биогеография, фации и стратиграфия средней юры Советской Арктики. Новосибирск: Наука, 1978. 205 с. (Тр. ИГиГ, вып.352).
13. Красавчиков В.О. Модификация тестового подхода к анализу таблиц описаний на основе понятия тупикового табличного свойства. - Автореф. канд. дис...., Саратов, 1982. 16 с.
14. Красавчиков В.О., Кушниренко Э.Ю., Милицын Б.Л. О возможности связи между физически измеримыми параметрами растений и их ботаническими и химическими характеристиками. - В кн.: Методы физических измерений параметров экологических систем и математическая обработка получаемых данных. Новосибирск: ВЦ СО АН СССР, 1982, с.27-41.
15. Майр Э. Принципы зоологической систематики. М.: Мир, 1971. 454 с.
16. Матула Д.В. Методы теории графов в алгоритмах кластер-анализа. - В кн.: Классификация и кластер. М.: Мир, 1980, с.83-111.

17. Санин В.Я. Раннемеловые ктенодонтиды севера Сибири. Новосибирск: Наука, 1976. 70 с. (Тр. ИГиГ, вып.310).

18. Санин В.Я. Позднеюрские и раннемеловые *Argasaea* (*Bivalvia*) севера СССР. Новосибирск: Наука, 1984. 126 с. (Тр. ИГиГ, вып.585).

19. Соколов Д.Н. Сообщение об окаменелостях с острова Преображения и с острова Бегичев. Спб., 1910, с.41-45. (Тр.Геол.музея, т.4, вып.3).

20. Соколов Д.Н. Мезозойские окаменелости из Большеземельской тундры и Кампура. - Тр. Геол.музея АН СССР, 1928, т.3, с.15-62.

21. Супес П., Эйнс Д. Основы теории измерений. - В кн.: Психологические измерения. М.: Наука, 1967, с.9-110.

22. Orbigny A.d'. *Système jurassique (Etage oxfordien) Mollusques.*- In *Géologie de la Russie d'Europe et des montagnes de l'Oural* by R.Z.Murchison E.de Verneuil, a.A.de Keyserling. v.2, London & Paris, 1845, p.419-488.

23. Rouillier Ch. Explication de la coupe géologique des environs du Moscou. Bull.Soc.Natur.Moscou. 1846, v.1, N 4, p.182-270.

24. Sokolov D., Bodylevsky W. Jura- und Kreidefaunen von Spitzbergen. Oslo, 1931, 157 S. (Skr. Svalbard og Ishavet, N 37).

25. Петрак Л.В. Приближенные функции одного переменного рациональными дробями. - В кн.: Программы оптимизации (приближение функций). Вып.6. Свердловск, 1975, с.110-129.

26. Cheney E., Loeb H. Two New Algorithms for Rational Approximation.- *Numer. Math.*, 1961, v.3, p.72-75.

27. Карманов В.Г. Математическое программирование. М.: Наука, 1980. 256 с.

28. Дрейпер Н., Смит Г. Прикладной регрессионный анализ.М.: Мир, 1973. 391 с.

29. Кушниренко Е.А. Основные параметры для описания состояния наземной экосистемы. - В кн.: Методы физических измерений параметров экологических систем и математическая обработка получаемых данных. Новосибирск: ВЦ СО АН СССР, 1982, с.5-17.

30. Соловьев Н.А. Разбиение множества объектов на однородные группы. - В кн.: Методы дискретного анализа в решении экстремальных задач. Новосибирск: ИМ СО АН СССР, 1979, вып.33, с.90-98.

А.Н.Дмитриев

ПЕРВЫЙ ЭТАП РАЗВИТИЯ МАТЕМАТИЧЕСКОЙ ГЕОЛОГИИ
В ИНСТИТУТЕ ГЕОЛОГИИ И ГЕОФИЗИКИ СО АН СССР

Развитие математической геологии является неотъемлемой частью становления и работы Института геологии и геофизики (ИГиГ). С первых лет его организации для многих ведущих сотрудников проблема "математизации геологии" представлялась самой неотложной. Она определяла специфику научных семинаров и необходимость в самообразовании. Именно первое десятилетие существования института характеризуется "бурей и натиском математизаторов", и именно математики дали повод острословам считать ИГиГ "математической Меккой в геологии"^{*)}.

Развитие математической геологии в институте можно разделить на два периода: 1961-1965 и 1966-1970 гг. Первому присущ поиск перспективных направлений математизации геологии. Второй характеризуется высокой методической активностью математиков, которая, как казалось, должна была убедить геологов в необходимости принятия "новых концептуальных направлений".

I. Период поисков

Применение готовых методов математической статистики создавало доброжелательное отношение к математическим методам, явно служебная роль которой устраивала "традиционную геологию". Но обстановка начала резко меняться с появлением на семинарах и в печати многообещающих утверждений со стороны математиков, острой критики в адрес геологической отрасли знания, попыток традиционные проблемы геологии решать математическим путем.

Уже в самом начале математизация геологии пошла в двух направлениях - построения теоретической геологии на формализованных

^{*)} Мы хотим передать читателям терминологическую среду того времени и просим быть снисходительными к жаргонизмам.

принципах и создания новых математических подходов и методов для решения разнообразных геологических задач теоретического и прикладного характера.

Оба направления интенсивно обсуждались сначала в отделе геофизики, затем на общеинститутских семинарах. Эти обсуждения вызвали тревогу геологов за судьбу устоявшихся форм работы, которая переросла в скептическое отношение к недостаточно проверенным предположениям и схемам. Математики, несмотря на явное неприятие математических методов большинством сотрудников, продолжали интенсивную работу. В короткий срок было организовано изучение проблемы классификаций в геологии, результаты которого вскоре стали дублироваться /5,6/. Существующие геологические классификации подверглись довольно резкой, временами необоснованной, критике. В качестве альтернативы, в частности, выдвигалась малопрактичная для геологов классификация перечисления. Вышедший в 1964г. сборник под редакцией Э.Э.Фотиади /40/ информировал о законченной серии работ по математической геологии. Именно этот сборник вызвал первую открытую дискуссию противников и последователей формализации геологии.

Сборник, посвященный применению математических представлений при построении и анализе геологических классификаций, вызвал интерес, но многие усмотрели в нем "попытку подорвать основы геологической науки". Материалы стали темой обсуждения на геолого-математических семинарах. Слухи о "концептуальных спорах" в ИГиГ довольно быстро заинтересовали методологов Ленинграда, которые включились в дискуссию. После этого спор о "дальнейших парадигмах геологии" становится общесоюзным.

Было проведено Первое Сибирское совещание по применению математических методов в геологии (декабрь 1965 г.). На нем обсуждались результаты работ, была принята программа дальнейших исследований. Характер обсуждаемых материалов, прогностический характер совещания дали повод для дальнейших острых дискуссий. Большую роль это совещание сыграло для широкого поиска новых подходов к старым проблемам геологии: классификации, основным определениям, терминологии и пр. /3/. Именно в этот период в институте произошло размежевание сторонников и скептиков относительно развития математических методов. Поляризация подходов стимулировала общеметодологические поиски.

К концу первого периода (1965 г.) математическая геология выдвинула широкий список задач, большинство которых позже оказались неразрешимыми. Слишком трудны были задачи формализации, очень неопределенны многие геологические постановки. Просматривая этот список, приходится удивляться и наивной самонадеянности, и прозорливости энтузиастов. В нем было всё, с чем связаны последующие достижения и неудачи. Приводим некоторые из этих задач:

1. Необходимость перевода геологической науки из описательной в вычислительную, из естественной в точную.

2. Необходимость замены традиционного хода производственной деятельности геологов на логически непротиворечивые, а потому более эффективные технологические цепочки, состоящие из операций в ключе "единого ответственного".

3. Требования коренного преобразования геологической отчетности как полевой, так и камеральной.

4. Рекомендации избегать принципиально непроверяемых положений, классификаций и гипотез.

5. Создание возможностей автоматизации обработки информационных потоков по списку "разумных целей".

6. Необходимость построения математических моделей основных процессов и явлений в геологии.

7. Обеспечение решения проблем прогноза, поиска и разведки полезных ископаемых с помощью ЭВМ.

2. Первые результаты

Большая настойчивость Ю.А.Воронина обеспечила "беспокойную жизнь" геологам, в круг проблем математизации было вовлечено большое число сотрудников. Поляризация мнений математиков и геологов становится естественной рабочей атмосферой в институте. Следующее пятилетие 1966-1970 гг. прошло с завидной результативностью. Этому способствовало появление второго поколения вычислительных машин и получение ряда существенных практических результатов.

К 1966 г. после напряженных поисков в развитии математической геологии в институте наметились три основных направления исследований:

1. Методология и теория формализованных построений в геологии.

2. Прогнозно-поисковые и разведочные задачи.

3. Задачи статистической обработки больших массивов количественных данных.

Каждое из этих направлений имело своих лидеров. Налаживались межинститутские связи, росло число геологических задач, подлежащих решению математическими методами. Но возрастало и психологическое сопротивление тех, кто в математической геологии видел "угрозу классике". Резко ощущалась необходимость организационных мероприятий. Активная позиция руководства Сибирского отделения АН СССР по отношению к работам прогнозно-поискового профиля /20/ заинтересовала Министерство геологии СССР, и уже в 1967г. в докладе комиссии под председательством М.А.Садовского на коллегии ГКНТ при Совете Министров СССР отмечалась весомая заслуга Сибирского отделения в постановке применения прикладной математики, в решении задачи общегосударственного значения. Привлечение средств математики и вычислительной техники признавалось необходимым условием для развития минерально-сырьевой базы СССР.

Последовавшие вслед за работой "Комиссии по применению математических методов и вычислительных средств в геологии" постановления закрепили основную ориентацию математической геологии с применением ЭВМ на решение сырьевых проблем. Это упрочило возможности разработок прогнозно-поискового профиля. Сохранялись перспективы развития теоретических разработок и формализованных построений в плане математизации геологии вообще. Набравшие темп и силу три вышеозначенных направления исследований дали нетривиальные результаты.

2.1. Методология и теория формализованных построений

Данное направление имело огромное значение для развития исследований на стыке геологии и математики. В содержательные формулировки геологических задач внедрялись приемы формализации. Ревизия и переосмысливание ряда геологических принципов и определений были подчас болезненно бескомпромиссными. Ведущие "фор-

мализаторы" - Ю.А.Воронин, С.В.Гольдин, Э.А.Еганов, В.А.Соловьев - сосредоточили внимание на проблемах классификаций и построения формальных предпосылок для введения геологических понятий и определений с тем, чтобы навести "хотя бы элементарный порядок" в общих вопросах теории геологических формаций, фациальных представлений и др. /12,13,14/. Они попытались рассмотреть спорные геологические понятия "фация" и "формация" в рамках последовательной и логически непротиворечивой схемы /13,15/.

Работавшие в институте творческие геолого-математические семинары заинтересовали некоторых ведущих геологов. В этот процесс активно включились тектонисты, и уже к 1966 г. на классические построения геологов было поведено глобальное (представления о геологическом пространстве) и тотальное (геологический язык и терминологическая ревизия) "наступление". Эрудиция и авторитет академика Ю.А.Косыгина, его участие как организатора, воплотились в серии статей установочного характера, которая обозначила реальный фронт работ на стыке математики и создаваемой теоретической геологии. Изложенная Ю.А.Косыгиным и Ю.А.Ворониным формализация геологического пространства /31/ и геологических процессов осадконакопления /32/ подготовила почву для решения ряда вопросов усовершенствования геологического языка. В.А.Соловьев связывает перспективу применения математики в геологии с формализацией понятий и терминологической определенностью /33, 43/.

В это же время предпринималась попытка согласования позиций математиков в вопросах методологических, теоретических и организационных. Под редакцией Э.Э.Фотиади выходит новый сборник "Геология и математика" /17/. Однако жесткий отбор "настоящих формализаторов", по-видимому, превзошел оптимальные пределы. Авторский коллектив оказался недостаточно представительным, а математика сборника - отстраненной от конкретного геологического знания. Часть математиков и геологов восприняли содержание сборника как сугубо абстрактное.

И всё-таки сборник отражал результаты по всем основным содержательным направлениям и формализованным разработкам. Авторский коллектив (включая сотрудников Института математики и СНИИТМСа) изложил свои соображения в пяти разделах сборника. Первый - подводил итоги методологических разработок, освещал су-

существующее положение математических методов в геологии и намечал эффективные (с точки зрения авторов) пути дальнейшего развития математических методов и применения ЭВМ в геологии.

Во втором разделе излагались взгляды на основы теории классификаций с попыткой анализа классификационных построений геологов с позиций формального задания множества объектов. Стремясь к созданию целостного подхода, авторы увязывают построение α -классификаций-перечислений с построением диагностических процедур и вероятностной трактовкой геологических классификаций. В этой модификации формальных построений видна "уступка естественным классификациям".

Двенадцать параграфов третьего раздела, в принятой авторами последовательности, посвящены формализации основных представлений геологии. Вводится понятие "геологического пространства". Здесь же приведена формализованная трактовка геологических пространств, тел, границ и их элементарных единиц.

Четвертый раздел содержит математическое изложение вопросов о геологическом времени, генезисе, картировании. Организационные вопросы рассмотрены в пятом разделе с достаточной для того времени широтой (от внедрения ЭВМ в геологию до подготовки кадров). Сборник явился своеобразной кульминацией усилий формализаторов.

С учетом постановлений о состоянии дел по применению ЭВМ в геологии был предложен перечень новых задач, связанных с прогнозированием. Очередной сборник "Математические методы в геологии и геофизике" /39/ имел уже существенно прикладную направленность. Все же около трети его объема посвящено вопросам аналитического описания сложных геологических тел (Ю.А.Воронин, Э.А.Еганов); процедурного сопоставления сложных геологических тел по их аналитическим описаниям (Ю.А.Воронин и др.); классифицирования плоских геологических тел (Е.Н.Эшштейн). В целом в работах этой части сборника суммируются результаты формализации и построения классификаций, способных войти в логическое сцепление с эмпирическими классификациями.

Активная пропаганда формализованных построений в 1963-1967 гг. приобщила к ним ряд инициативных исследователей-геологов института. Состав "формализаторов" к середине 70-х годов стал более геологическим.

2.2. Прогнозно-поисковые и разведочные задачи

Формализаторское направление, преодолевавшее основное сопротивление сторонников традиционных взглядов на геологическую науку, облегчило развитие второго направления математизации геологии. Риск математиков, внедрившихся в основной раздел геологии — учение о полезных ископаемых в начале 60-х годов, оправдывался крайне важными задачами обнаружения полезных ископаемых. О необходимости математизации процедур анализа геолого-геофизических данных и применения ЭВМ говорили создавшиеся "информационные зазоры", с одной стороны, и "недостача информации", с другой. Быстрый рост возможностей прикладной математики и вычислительной техники удешевлял математическую обработку информации. Кроме того, результаты, получаемые математиками с помощью ЭВМ, могли либо приниматься во внимание геологами на этапе принятия решения, либо рассматриваться как экспериментальные, что устраивало обе стороны. Лишь через два десятилетия поиски и разведка будут "вестись с магнитной ленты". Такова вкратце методолого-информационная обстановка, в которой зародились и развивались работы по математизации прогнозно-поисковых задач.

Задачи прикладного геолого-математического характера начали рассматриваться в институте одновременно с вопросами формализации геологии. В это время систематически проблему автоматической обработки данных развивали геофизики, что было естественным для решаемых ими задач. По мере увеличения разнообразия алгоритмов и реализующих программ, область применения математических методов быстро расширялась. Этому своеобразно способствовала конкуренция двух взглядов на роль математики в геологии: с одной стороны — повышенная математическая требовательность к геологическим задачам, вплоть до формирования "математического языка в геологии" и с другой — умеренная требовательность к языку геологии и модификация математических методов "в сторону содержательно интерпретируемых математических операций".

Выразителем первого направления оставался Ю.А.Воронин, который требовал математической строгости и необходимой модификации геологических задач "до нормальных". Автор данной статьи настаивал на сохранении естественности языка геологов, а необходимую модификацию адресовал математическим методам ("если готовых

методов к данной геологической задаче не имеется, то такой метод следует создать"). С течением времени эти два взгляда существенно сблизились, что оказалось взаимно полезным. Но в середине 60-х годов размежевание было довольно острым, как противопоставление приоритетов математики или геологии.

2.2.1. Приоритет математических требований предполагал поиск "хороших задач" и создание алгоритмического арсенала на заранее построенной математической основе. Соответствующая методологическая ориентация была намечена в начале 60-х годов, а ее некоторые положения были опубликованы более подробно в 1965 г. /45/. Первый альбом алгоритмов и программ для обработки геологических данных содержал примеры решения конкретных задач и ознакомил читателей не только с увеличением возможностей ЭВМ, но и очертил круг задач, подлежащих решению /4/. В этот же период выходит ряд работ Ю.А.Воронина /8-II/, в которых освещаются подходы к построению принципиальных схем подсчета запасов и анализируются различные процедуры диагноза в задачах, сводимых к распознаванию образов. Одновременно создаются, обеспечиваются программами и апробируются алгоритмы "Голотип" для решения задач распознавания (в нефтяной геологии и палеонтологии) /16,46/.

Итогом работ по этому направлению был упоминавшийся уже сборник /39/. В основу разработки широкого набора прикладных задач по нефтяной геологии легла "задача века" о разделении водоносных и нефтеносных пластов по данным промысловой геофизики. Следующий сборник "Геология и математика (задачи диагноза и распознавания в геологии, геохимии и геофизике)" /19/ объединил специалистов-"распознавальщиков" из разных институтов и, по существу, отобразил общесоюзное состояние решения задач распознавания, достигнутое к концу 60-х годов. Сборник строился по принципу математического приоритета. Материал в нем изложен в четырех разделах, первый из которых вводил читателя в проблемы распознавания (от общих положений и требований к подходу до конкретных прикладных вопросов распознавания в решении задач геологии, геохимии, геофизики).

Второй, наиболее обширный, показывал алгоритмическое разнообразие, достигнутое к 70-м годам, и апробацию вычислительных программ. Впервые была решена трудоемкая задача сравнения алгоритмов на их диагностическую способность (по одной и той же за-

даче анализа данных). Хотя, с точки зрения авторов, результаты не были блестящими, но, как показало время, был первый образец применения различных алгоритмов к решению одной и той же задачи.

Третий раздел имел проблемный характер и содержал материалы, касающиеся дальнейшего развертывания работ. Рекомендации даны на основе оценки результатов распознавания, осуществленной по заданному списку программ.

Четвертый – по существу альбом программ, часть которых так и не обрела жизнь в сфере геологических задач. "Раскрученная пружина алгоритмопостроения и программпроизводства" нашла, однако, в существенной своей части применение в возникающих задачах геологии, многие из которых оказались "хорошими" с точки зрения разработанных алгоритмов. Накал "формализаторства" по-прежнему вызывал настороженность геологов, а вторжение математиков в чисто геологические задачи – выжидательную реакцию.

2.2.2. Приоритет геологических требований получил поддержку в 1964 г. на семинарах по теоретической кибернетике (в отделении кибернетики Института математики СО АН СССР). Высокая научная активность А.А.Ляпунова среди естественников коснулась и геологии. Автор имел частые встречи с А.А.Ляпуновым, который убеждал "ориентировать геологическое мышление на исследование организменной модели Земли". Именно для организационных моделей считалось уместным применять идеи и средства теоретической кибернетики. Серьезным шагом в применении ЭВМ при обработке геологической информации было предложение Ю.И.Журавлева об использовании в задачах естествознания тестового подхода (комбинаторно-логическое направление). Автор этой статьи принял это предложение и участвовал в поиске "подходящей задачи".

Такая задача (по обнаружению металлоносных древних конгломератов) была предложена Ф.П.Кренделевым. Несколько месяцев, потраченных на мобилизацию информации, завершились постановкой задачи, и к лету 1965 г. были получены первые результаты.

Основа тестовых построений была заложена С.В.Яблонским по разделу дискретной математики еще в 50-х годах. В начале 60-х логико-комбинаторные методы нашли применение в контроле режима работы электрических схем. Применение тестового подхода в решении геологических задач потребовало существенного видоизменения приемов извлечения информации из получаемых списков тупиковых

тестов. Оно выделялось в самостоятельный раздел исследований и впоследствии вошло в число способов распознавания. С тестовым подходом связано широкое использование логических переменных, которыми богата исходная геологическая информация. Именно этот подход, поставивший логические высказывания в один ряд с количественными данными, вывел математиков и геологов к широкому классу задач прогнозно-поискового профиля.

Первые публикации направления отставившего приоритет геологических требований были посвящены классификации предметов и явлений на примере прогнозных задач /20,23/. Новое направление, имевшее немного сторонников, старалось закрепиться решением конкретных задач, на которые ориентировали постановляющие документы о применении ЭВМ в геологии.

Приемлемый для геологов вид требуемой информации с включением большого числа признаков логического содержания способствовал закреплению и расширению заинтересованности геологов. Уже после первой публикации от них поступил ряд предложений, часть которых была принята. Трудоемкую мобилизацию исходных данных стало возможным проводить совместно. Встречные усилия математиков и геологов оказались крайне полезными в плане выявления проблем доалгоритмической обработки данных. Обнаружились сложности стыковки геологических и формализованных постановок задач. Складывался язык общения, формализовались итгитивные правила по подготовке исходной информации.

Работа "на половине геологов" подстегивалась острыми нападками "чистых формализаторов", которые обвиняли "тестовиков" в "агрессивной форме математической безграмотности" и получали ответ, что "лучше посредственно решенная конкретная задача, чем абстрактное совершенство". Эти споры принесли немалую пользу — обе стороны обнаруживали свои слабые места и старались работать над усилением методов /22,23,26/.

Итак, к началу 70-х годов логико-математическая группа в институте вступила в тесный контакт с геологами. Начались систематические разработки по углублению и расширению алгоритмического и программного обеспечения и по уточнению сфер применимости подхода. Расширялись межинститутские связи с математическими и геологическими коллективами. Постепенно устанавливалось необходимое равновесие между "формализаторами" и "прикладниками".

По мере накопления опыта решения задач они выступали более согласованно.

Ориентация на практику ставила "прикладников" в раги исполнителей, что было чревато сужением теоретического поиска. Именно поэтому прикладники выбирали сложные, по мнению формализаторов, "умопомрачающие задачи". Постепенно решение задач с помощью тестового подхода получило признание. Основные результаты в данном направлении были получены по прогнозу рудоносности золотороссых конгломератов путем сравнительного изучения небольшого числа месторождений /34,35/, никеленосности трапховых интрузий северо-запада Сибирской платформы в сравнении с Норильскими /21, 24/, редкометалльного оруденения вулканических аппаратов /25/; по сравнительному изучению крупных зон нефтенакопления в Западной и Восточной Сибири /43,44/.

2.3. Применение статистических методов

Применение математико-статистических методов в ИГиГ началось с момента его организации, что объяснялось известностью статистического подхода в сфере геологии и резким нарастанием количественных данных.

Применение этих методов в институте велось независимо разными авторами. Каждый исполнитель, сообразуясь со своим материалом и уровнем осведомленности, обрабатывал массивы количественных данных по тому или иному критерию, ссылаясь на зарубежный (В.А.Кутюлин) или отечественный (Н.Л.Добрецов) опыт.

Вклад математической статистики в "бурное десятилетие" становления математических методов в ИГиГ заключается в том, что она стимулировала интерес к математическим методам вообще. Масштаб применения статистических методов постепенно возрастал как по количеству применяемых методов, так и по числу пользователей. Были разработаны некоторые оригинальные алгоритмы. В описываемый период основные направления применения статистических методов в институте задавались В.А.Кутюлиным и Н.Л.Добрецовым.

Было достаточно широким по геологическому набору задач и по коллективу авторов петрохимическое направление: определение формационной принадлежности гипербазитов по петрохимическим данным

на достаточно большой выборке /16,36/; многомерный статистический анализ петрохимических данных по траппам Сибирской платформы /37/; выявление петрохимических критериев формационной принадлежности для базальтов и долеритов /38/, а также изучение химизма базальтов; применение многомерного статистического анализа петрохимических свойств для разделения альпинотипных гипербазитов по возрастным группам /4/.

Применение статистических методов в задачах прогнозно-поискового профиля было начато Н.Л.Добрецовым и расширилось в следующем десятилетии. Многомерный статистический анализ был применен для прогноза алмазности кимберлитов /27/.

Было поставлено преподавание вероятностно-статистических методов в геологии в университете /28/. Предпринят опыт применения методов распознавания (на эвристических алгоритмах) в задачах классификации и диагноза палеонтологических объектов /29/. Дискриминантные функции были применены для классификации хлоритов /30/. Все задачи этого профиля показывали геологам нарастающие возможности математической геологии в целом.

В этот же период было проведено статистическое изучение состава разнообразных вулканических комплексов (лаборатория академика Ю.А.Кузнецова). Выявление неодномодальности распределения петрохимических признаков позволило А.Ф.Белюсову /1/ сформулировать положение о породных группах как элементарных естественных породных совокупностях. Каждому типу породных групп в связи с его унимодальностью был поставлен в соответствие некий специфический генетический механизм. Такое представление о породных группах вполне сопоставимо с представлением о популяциях, разработанным в те же годы У.Крамбеином в статистических моделях геологических объектов. Отличие от известного математико-статистического понятия генеральной совокупности состоит в том, что понятие популяции (и породной группы) делает возможным конкретное-научное истолкование данных об однородности-неоднородности геологических совокупностей, позволяет смоделировать смеси и т.д. Именно в этом направлении была выполнена трудоемкая работа для вулканических комплексов рифея и нижнего палеозоя Алтая /2/. Но пройдет еще почти два десятка лет, прежде чем вскрытая статистическим подходом общая закономерность горнопородного видообразования и воспроизводства начнет обретать права "общенаучного гражданства".

В целом первое десятилетие работы с использованием аппарата математической статистики в институте обобщало опыт применения математических методов в геологических постановках разнообразных задач и уже предчувствовался "монографический урожай" в очередном десятилетии.

О некоторых итогах

Итоги первого десятилетия развития и применения математической геологии в институте довольно разнообразны и интересны. Вот некоторые из них.

1. К 70-м годам в Институте геологии и геофизики положительные результаты по созданию математической геологии стали очевидны, да и математическая геология в целом начала обретать черты нового раздела знаний наук о Земле. Этот раздел был проработан от его методолого-философских положений до практического внедрения результатов анализа геолого-геофизических данных. Была построена "технологическая цепочка" по разработке проблем от постановок задач до принятия решений с помощью ЭВМ.

2. Усилия формализаторов показали возможную альтернативу основному направлению исследований традиционной геологии. Появился новый стимул развития теории геологии, обновления ее концептуального багажа и построения новых общих схем и моделей.

3. Усиливающийся поток новой информации получил соответствующие терминологические и логико-математические средства для переработки.

4. Полученные результаты решения конкретных задач математическими методами уже на первых опытах показали большие возможности ЭВМ при решении минерально-сырьевых проблем.

5. Связи математиков с производственными учреждениями сложились по многим направлениям, причем основные усилия были направлены на помощь геологам-производственникам в поисках и разведке рудных и нерудных полезных ископаемых, месторождений нефти и газа.

Таким образом, к 70-м годам был предложен новый набор исследовательских средств для работы как в плане теоретического развития геологии, так и для решения широкого круга производст-

венных задач, связанных с анализом больших массивов геолого-геофизических данных. Эти средства требовали от геологов не только модификации привычного образа мышления и деятельности, но и дополнительных знаний для "понимания задач нового поколения". Эти требования остаются препятствием для совместной работы. Надежды математиков на то, что "первые конкретные результаты повернут геологов" (к началу 70-х годов), оказались в значительной мере тщетными. По обе стороны барьера зрели новые планы и выдвигались очередные задачи. Готовились условия к организации в институте структурных геолого-математических ячеек.

Литература

1. Белоусов А.Ф. Неоднородность распределения составов в ассоциациях горных пород и представление о породных группах. - Геол. и геофиз., 1967, № 5, с.26-34.
2. Белоусов А.Ф., Кочкин Ю.Н., Полякова Э.Г. Вулканические комплексы рифея и нижнего палеозоя Горного Алтая, Горной Шории и Салаирского кряжа. М.: Наука, 1969. 283 с.
3. Борукаев Ч.Б., Каратаева Г.Н. О работе секции формальных построений на первом Сибирском совещании по применению математических методов и ЭВМ в геологии. - Геол. и геофиз., 1966. № 7, с.147-149.
4. Великий В.Б., Пинус Г.В., Каратаева Г.Н. Применение методов многомерного статистического анализа для разделения разновозрастных альвинотипных гипербазитов по их петрохимическим свойствам. - Геол. и геофиз., 1968, № 1, с.83-88.
5. Воронин Ю.А., Гольдин С.В. Вопросы теории конечных геологических классификаций. - Геол. и геофиз., 1964, № 8, с.90-100.
6. Воронин Ю.А., Гольдина Н.А. О математико-логическом анализе геологических классификаций. - В кн.: Геология нефтегазоносных районов Сибири. Новосибирск, 1964, с.94-110. (Тр. СНИИГТИМС, вып.32).
7. Воронин Ю.А. К итогам Первого Сибирского совещания по применению математических методов и ЭВМ в геологии. (Новосибирск, дек. 1965 г.). - Геол. и геофиз., 1966, № 7, с.145-147.
8. Воронин Ю.А. О построении принципиальной схемы подсчета

запасов полезных ископаемых. - Геол. и геофиз., 1966, № 2, с.69-76.

9. Вороник Ю.А. Об исследованиях, связанных с применением математических методов и ЭВМ в геологии. - Геол. и геофиз., 1967, № 10, с.75-81.

10. Вороник Ю.А. О построении меры неправильности одномерной сети (в задачах геофизики и геологии). - Геол. и геофиз., 1969, № 7, с.63-66.

11. Вороник Ю.А., Каратаева Г.Н. Об одном возможном способе определения гололита и его использование для решения задач диагноза (распознавания). - Геол. и геофиз., 1967, № 4, с.70-75.

12. Вороник Ю.А., Еганов Э.А. Вопросы теории формационного анализа. Новосибирск: ИГиГ СО АН СССР, 1968. 59 с.

13. Вороник Ю.А., Еганов Э.А. К построению формальных основ учения о формациях. - В кн.: Геологические формации (Материалы к совещанию, 21-24 мая 1968 г.). Л.: Наука, 1968, с.38-41.

14. Вороник Ю.А., Еганов Э.А. К теории фациального анализа. Новосибирск, 1969. 89 с. (Деп. ВИНТИ № 689-69).

15. Вороник Ю.А., Еганов Э.А. Логические основы теории фациального и формационного анализов. - Бюлл. МОИП. Отд. геол., 1969, т.44, вып.3, с.148-149.

16. Вороник Ю.А., Каратаева Г.Н., Эшштейн Е.Н. и др. Программа "Гололит" для решения задач распознавания образов / Отв. ред. Э.Э.Фотиади. Алма-Ата: КазИМС, 1968. 100 с.

17. Геология и математика. Методологические, теоретические и организационные вопросы геологии, связанные с применением математических методов в ЭВМ / Отв.ред. Э.Э.Фотиади. Новосибирск: Наука, 1967. 251 с.

18. Гольдия С.В., Кутюлия В.А. К петрохимии трапсов катангского и кузьмовского комплексов западной окраины Сибирской платформы. - Сов. геол., 1964, № 12, с.133-140.

19. Геология и математика. Задачи диагноза и распознавания в геологии и геофизике / Отв. ред. Э.Э.Фотиади. Новосибирск: Наука, 1970. 224 с.

20. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. О математических принципах классификации предметов и явлений. - В кн.: Дискретный анализ. Новосибирск: Наука, 1966, с.3-15. (Тр. Ин-та математики СО АН СССР, вып.7).

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

22. Дмитриев А.Н. Некоторые табличные числа. - В кн.: Дискретный анализ. Новосибирск: Наука, 1968, с.54-58. (Тр. Ин-та математики СО АН СССР, вып.12).

23. Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П. Об одном принципе классификации и прогноза геологических объектов и явлений. - Геол. и геофиз., 1968, № 5, с.50-64.

24. Дмитриев А.Н., Золотухин В.В., Васильев Ю.Р. Опыт применения дискретной математической обработки информации по дифференцированным рудоносным трапловым интрузиям северо-запада Сибирской платформы. - Сов. геол., 1968, № 12, с.98-108.

25. Дмитриев А.Н., Кренделев Ф.П., Модников И.К. и др. Оценка масштаба редкометалльного оруденения, локализованного в вулканических аппаратах (с помощью ЭВМ). - Сов. геол., 1969, № II, с.100-109.

26. Дмитриев А.Н. Использование длин тупиковых тестов при обработке таблиц. - В кн.: Дискретный анализ. Новосибирск: Наука, 1970, с.18-21. (Тр. Ин-та математики СО АН СССР, вып.14).

27. Добрецов Н.Л., Харьков А.Д., Шемякин М.Л. Применение многомерного статистического анализа для решения прогнозных задач на примере алмазности кимберлитов. - Геол. и геофиз., 1966, № 8, с.15-22.

28. Добрецов Н.Л., Харьков А.Д., Шемякин М.Л. Применение вероятностно-статистических методов в геологии. Курс лекций для студентов НГУ. Новосибирск: Изд. НГУ, 1967. 123 с.

29. Елкин Е.Н., Елкина Е.А. О возможности применения методов распознавания образов в палеонтологии. - Геол. и геофиз., 1967, № 9, с.75-79.

30. Кепежинская К.Б., Кузнецова И.К. Применение дискриминантных функций для классификации хлоритов. - Геол. и геофиз., 1966, № 10, с.108-112.

31. Косыгин Ю.А., Воронин Ю.А. Развитие представлений о геологическом пространстве, как возможная основа применения математики в геологии. - Изв. АН СССР. Сер.геол., 1966, № 12, с.17-25.

32. Косыгин Ю.А., Воронин Ю.А. О формализации явлений геологии в связи с проблемами применения математических методов и ЭВМ. – Геол. журн., 1967, т.27, № 2, с.48–53.

33. Косыгин Ю.А., Соловьев В.А. Проблема усовершенствования геологического языка и "математизация" геологии. – Изв. АН СССР. Сер.геол., 1967, № II, с.157–163.

34. Кренделев Ф.П., Дмитриев А.Н., Журавлев Ю.И. Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики. – Докл. АН СССР, 1967, т.173, № 5, с.1149–1152.

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

36. Кутолин В.А., Волохов И.М., Каратаева Г.Н. К оценке возможности определения формационной принадлежности гипербазитов по петрохимическим данным. – Геол. и геофиз., 1966, № 5, с.87–93.

37. Кутолин В.А., Каратаева Г.Н. Применение методов многомерного статистического анализа для трапшов Сибирской платформы по их петрохимическим особенностям. – Тр. Ин-та геологии Арктики, т.151. Л., 1967, с.116–123.

38. Кутолин В.А. Статистические петрохимические критерии формационной принадлежности для базальтов и долеритов. – Докл. АН СССР, 1968, т.178, № 2, с.434–437.

39. Математические методы в геологии и геофизике. Новосибирск: 1968. 173 с. (Тр. СНИИГТИМС, вып.79).

40. Опыт анализа и построения геологических классификаций на основе представлений конечной математики / Отв.ред. Э.Э.Фотиади. Новосибирск: Наука, 1964. 119 с.

41. Применение математических методов в геологии. Алма-Ата: Наука, 1968. 192 с.

42. Соловьев В.А. Формализация понятий – необходимое условие применения математики в геологии. – Сов. геол., 1968, № 8, с.152–156.

43. Трофимук А.А., Вышемирский В.С., Дмитриев А.Н. и др. О сравнительном изучении гигантских месторождений нефти с ис-

пользованием логико-дискретного анализа. - Геол. нефти и газа, 1969, № 6, с.17-19.

44. Трофимук А.А., Вышемирский В.С., Карбышев В.Д. и др. Опыт информационно-энтропийного исследования признаков гигантских нефтяных месторождений. - Геол. и геофиз., 1969, № 7, с.116-118.

45. Фотиади Э.Э., Воронин Ю.А. Методологические вопросы внедрения математических методов и электронно-вычислительных машин (ЭВМ) в практику геологических исследований. - Геол. и геофиз., 1965, № 12, с.15-20.

46. Эшштейн Е.Н. Применение ЭВМ для прогноза качества нефтей. - Докл. АН СССР, 1968, т.180, № 2, с.425-427.

РЕФЕРАТЫ

УДК 519.1:553.3/9

ОБ УСОВЕРШЕНСТВОВАНИИ ТЕСТОВОГО ПОДХОДА В ЗАДАЧАХ ДИАГНОСТИКИ. А.Н.Дмитриев, С.В.Макаров. – В кн.: Логико-математические методы в геологических исследованиях. Новосибирск: ИГиГ СО АН СССР, 1986, с.6–25.

В первом разделе дан критический анализ диагностических возможностей тестового метода (комбинаторно-логического направления) в решении прогнозных задач геологии. Указаны недостатки метода "голосования по тупиковым тестам" и предложен новый "метод перекладки". Вводится тестовая модель с "крупноблочным покрытием классов" (для задач с двумя классами объектов), базирующаяся на понятиях: "закономерность k -го порядка", "носитель закономерности", "представительность выборки". На их основе намечаются пути к обнаружению скрытых закономерностей, содержащихся в описаниях геологических объектов. Во втором – приведены результаты обобщающего характера; выявлены и исследованы взаимосвязи, сближающие тестовый подход с линейно-статистическими методами; подняты вопросы дальнейшего развития теории и практических применений.

Библ. 18 назв.

УДК 519.25

ВОССТАНОВЛЕНИЕ ОДИНОЧНОГО ПРОПУСКА В ТАБЛИЦЕ "ОБЪЕКТ-ПРИЗНАК" НА ОСНОВЕ СИНГУЛЯРНОГО РАЗЛОЖЕНИЯ ДАННЫХ. С.В.Макаров. – В кн.: Логико-математические методы в геологических исследованиях. Новосибирск: ИГиГ СО АН СССР, 1986, с.25–29.

Рассматривается таблица "объект-признак", в которой один элемент неизвестен. Требуется оценить неизвестное значение (пропуск). Используется теорема о сингулярном разложении, на основе которой выводится формула оценки. Предложенный метод является модификацией метода гребневой регрессии.

Библ. 3 назв.

УДК 519.28:553.3

ВОССТАНОВЛЕНИЕ ПРОПУСКОВ В ТАБЛИЦАХ ДАННЫХ. Г.С.Федосеев. - В кн.: Логико-математические методы в геологических исследованиях. Новосибирск: ИГиГ СО АН СССР, 1986, с. 39-42.

Восстановление пропусков решается как задача распознавания образов в постановке обнаружения скрытых закономерностей. Связи элементов таблицы данных (структура которой обуславливает выбор математического метода) устанавливаются с помощью алгоритма регрессионного характера. Процедуре распознавания предшествует исследование таблицы по восстановлению пропусков для фиксированных целевых признаков. Конкретные расчеты проводились на табличных данных железорудных месторождений Горного Алтая. Проведено последовательное восстановление искусственно введенных пропусков по всей таблице. Выявлены корреляционные зависимости по столбцам (признакам) и строкам (месторождениям).

Ил.4, табл.1, библи. 8 назв.

УДК 519.26

О ЦЕЛЕВОМ ПРЕОБРАЗОВАНИИ ЧИСЛОВЫХ ПРИЗНАКОВ (Для случая двух классов объектов). А.А.Бишаев, Г.С.Федосеев. - В кн.: Логико-математические методы в геологических исследованиях. Новосибирск: ИГиГ СО АН СССР, 1986, с. 42-49.

В работе описан алгоритм, реализующий идею представления числовых признаков в виде бинарных матриц и проведения кусочно-линейных (допустимых в шкале порядка) целевых преобразований признаков с помощью методов линейной оптимизации. Основная область применения этого алгоритма связана с проблемой оптимального целевого кодирования информации и построением нелинейных решающих функций в задачах классификации и распознавания.

Ил.1, библи. 2 назв.

УДК 519.25:552.3

АПРОБАЦИЯ МЕТОДА ДИНАМИЧЕСКОГО КЛАСТЕРНОГО АНАЛИЗА ПРИ ВЫДЕЛЕНИИ ПОПУЛЯЦИЙ МАГМАТИЧЕСКИХ ГОРНЫХ ПОРОД. А.Ф.Белуосов, А.С.Вакуленко. - В кн.: Логико-математические методы в геологических исследованиях. Новосибирск: ИГиГ СО АН СССР, 1986, с.49-57.

Реализованный на ЕС ЭВМ алгоритм динамического кластерного анализа использован с целью выделения по петрохимическим данным горнопородных популяций на примерах протерозойских вулканитов серии Ветреного пояса (Карелия) и долеритов запада Скандинавского полуострова. В гетерогенной ассоциации Ветреного пояса выделены три квазивидовые группы пород - мела-, субмелалапикритоидная и базальтоидная, а в базальтоидной группе - еще две популяции, отвечающие так называемым толеитовым и коматиитовым базальтам. В относительно однородной ассоциации долеритов Скандинавии выделяются две разные популяции, отвечающие соответственно геосинклинальному поясу и области материкового палеоплато, и более мелкие локальные популяции.

Ил.1, табл.2, библи. 7 назв.

УДК 553.062/067+519.95

ОЦЕНКА ПЕРСПЕКТИВ ПЛОЩАДЕЙ СЕВЕРО-СИБИРСКОЙ НИКЕЛЕНОСНОЙ ОБЛАСТИ ЛОГИКО-МАТЕМАТИЧЕСКИМИ МЕТОДАМИ ПО ОТДЕЛЬНЫМ ГРУППАМ ПРИЗНАКОВ. В.Д.Карбышев, Л.Г.Сухов, А.С.Вакуленко, А.А.Городнянский. - В кн.: Логико-математические методы в геологических исследованиях. Новосибирск: ИГиГ СО АН СССР, 1986, с. 57-93.

Задача прогноза в варианте распознавания образов, в соответствии со стадиями поисковых работ, решается в три этапа. Исходные данные представлены таблицей 84 x 198 (84 площади и 198 признаков). Обработкой данных выявлено 25 информативных систем (52 % от общего числа признаков). Максимально информативными (в диагностическом смысле) оказались метаморфо-метасоматические, магматические, минералогические и геохимические системы признаков. Оценена сопоставимость решения статистическими и логико-математическими методами. По системам признаков проведена коли-

чественная оценка прогнозируемых участков на рудоносность. Первоочередными названы два объекта, содержащих ряд расслоенных интрузивных тел. На стадии детальных поисков по магматической и геохимической группам признаков выявлено 13 участков в качестве первоочередных. Результаты решения задачи использованы для металлогенического районирования северо-запада Сибирской платформы.

Таб. I3, библ. I4 назв.

УДК 519.25:56

ИССЛЕДОВАНИЕ ЛОГИКО-МАТЕМАТИЧЕСКИМИ МЕТОДАМИ ПРОБЛЕМЫ МОРФОЛОГИЧЕСКОЙ КЛАССИФИКАЦИИ (на примере класса *Bivalvia*). В.О.Красавчиков, В.Я.Сантин. - В кн.: Логико-математические методы в геологических исследованиях. Новосибирск: ИГиГ СО АН СССР, 1986, с. 93-121.

Рассматривается проблема выделения вида (и на этой основе таксонов более высокого ранга) по морфологическим данным с помощью математических методов. Обсуждаются известные модели кластер-анализа и предлагаются новые методы для более обоснованного определения границ видов двухстворчатых моллюсков.

Ил.6, табл.9, библ. 30 назв.

УДК 51(09)+55(09)

ПЕРВЫЙ ЭТАП РАЗВИТИЯ МАТЕМАТИЧЕСКОЙ ГЕОЛОГИИ В ИНСТИТУТЕ ГЕОЛОГИИ И ГЕОФИЗИКИ СО АН СССР. А.Н.Дмитриев. - В кн.: Логико-математические методы в геологических исследованиях. Новосибирск: ИГиГ СО АН СССР, 1986, с. 121-139.

Охарактеризовано зарождение и развитие математической геологии в Институте геологии и геофизики с момента его организации и до 1970 г. Отмечено два этапа становления математических методов по трем направлениям: формализованные построения в геологии, разработка математических методов для решения задач прогнозно-поискового профиля, применение статистических методов. Кратко

анализируется каждое из направлений и приводятся сведения по основным задачам того периода: геологический прогноз; диагностика в палеонтологии; задачи рудной и нефтяной геологии; формализация геологических терминов; разработка классификаций перечисления и др. Отражен напряженный и острый процесс становления этого научного направления в институте.

Библи. 46 назв.

СОДЕРЖАНИЕ

Предисловие	3
-------------------	---

ТЕОРЕТИЧЕСКИЕ ВОПРОСЫ

Дмитриев А.Н., Макаров С.В. Об усовершенствовании тестового подхода в задачах диагностики	6
Макаров С.В. Восстановление одиночного пропуски в таблице "объект-признак" на основе сингулярного разложения данных	25
Федосеев Г.С. Восстановление пропусков в таблицах данных	29
Бишаев А.А., Федосеев Г.С. О целевом преобразовании числовых признаков (для случая двух классов объектов).	42

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

Белюсов А.Ф., Вакуленко А.С. Апробация метода динамического кластерного анализа при выделении популяций магматических горных пород	49
Карбышев В.Д., Сухов Л.Г., Вакуленко А.С., Городнянский А.А. Оценка перспектив площадей Северо-Сибирской никеленосной области логико-математическими методами по отдельным группам признаков	57

Красавчиков В.О., Санин В.Я. Исследование логико-математическими методами проблемы морфологической классификации (на примере класса *Bivalvia*) 93

К ИСТОРИИ МАТЕМАТИЧЕСКОЙ ГЕОЛОГИИ

Дмитриев А.Н. Первый этап развития математической геологии в Институте геологии и геофизики СО АН СССР I2I
Р е ф е р а т ы I39

План выпуска изданий СО АН СССР
на 1986 г., поз.8I

ЛОГИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Сборник научных трудов

Ответственные редакторы

Алексей Николаевич Дмитриев, Валерий Васильевич Бабич

Утверждено к печати
Институтом геологии и геофизики СО АН СССР

Редактор Р.Н. Ильина
Младший редактор М.И. Чиркова
Технический редактор Н.Н. Александрова

Подписано к печати 12.09.86. МН 12235.
Бумага 60x84/16. Печ.л. 9.0. Уч.-изд.л. 8.6.
Тираж 400. Заказ 263. Цена 60 коп.

Институт геологии и геофизики СО АН СССР
Новосибирск, 90. Ротапринт.