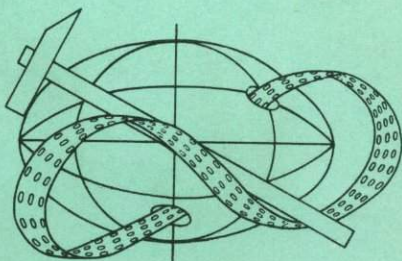


АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ГЕОЛОГИИ И ГЕОФИЗИКИ

**ЛОГИКО-МАТЕМАТИЧЕСКАЯ
ОБРАБОТКА
ГЕОЛОГИЧЕСКОЙ
ИНФОРМАЦИИ**

(ТЕОРИЯ И МАТЕМАТИЧЕСКИЙ АППАРАТ)



НОВОСИБИРСК—1976

АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ГЕОЛОГИИ И ГЕОФИЗИКИ

519.55

ЛОГИКО-МАТЕМАТИЧЕСКАЯ
ОБРАБОТКА
ГЕОЛОГИЧЕСКОЙ
ИНФОРМАЦИИ

(ТЕОРИЯ И МАТЕМАТИЧЕСКИЙ АППАРАТ)

СБОРНИК НАУЧНЫХ ТРУДОВ

НОВОСИБИРСК—1976



1927

А Н Н О Т А Ц И Я

В сборнике освещены вопросы логико-математического исследования описаний геологических объектов на базе ЭВМ. Дана краткая теоретическая и практическая схема решения геологических задач; изложены рекомендации по практической подготовке и кодированию исходной информации; приведены оригинальные математические методы и процедуры, широко апробированные на задачах теоретического и производственного характера. Некоторые из этих задач, в основном, прогнозно-поискового профиля, изложены в данном сборнике (прогнозирование перспективных объектов на железо, медно-никелевые руды, нефть, газ).

Сборник рассчитан на широкий круг научных, инженерно-технических работников, геологов, изучающих вопросы математической обработки геологических данных. Он будет полезен также студентам ВУЗов, аспирантам, специализирующимся в области распознавания образов, и молодым геологам.

Редакционная коллегия

Г.С.Федосеев, В.В.Бабич, В.О.Красавчиков,
Т.И.Штатнова

Ответственный редактор А.Н.Дмитриев

Печатается по решению секции
стратиграфии, тектоники, ли-
тологии и осадочных полезных
ископаемых Ученого совета
Института геологии и геофизики
СО АН СССР

©

Институт геологии и
геофизики СО АН СССР
1976 г.

А. Н. Дмитриев

ВОПРОСЫ ФОРМАЛИЗОВАННЫХ ПОСТАНОВОК ГЕОЛОГИЧЕСКИХ
ЗАДАЧ ПРОГНОЗНО-ПОИСКОВОГО ПРОФИЛЯ

Данная работа основана на обобщении опыта практического применения логико-математических методов при решении задач прогнозно-поискового профиля. Обобщение проводится с небольшими экскурсами в методологию и, где это необходимо, в область конкретных примеров. Исследуемые вопросы постановочного характера и связаны с задачами (например, таксономии, классификации, упорядочения объектов по заданному признаку), которые мы объединяем названием "задачи сравнительного изучения объектов". Логическую схему и содержательную трактовку геологических задач рассмотрим с учетом следующих особенностей исследований:

- а) содержательное фиксирование обстановки возникновения задачи;
- б) вычленение задачи из обстановки ее возникновения;
- в) постановка задачи в соответствии с ее содержательной формулировкой;
- г) формализованная постановка задачи и ее решение в автоматическом режиме.

Принимается во внимание и тот факт, что содержательная специфика описаний геологических объектов и логическая структура наборов характеризующих их признаков требует построения схем диагноза, отличных от схем, возникающих в области технических применений. Да и практические примеры убеждают, что построение специализированных схем исследования совокупности описаний геологических объектов с учетом содержательного характера постановок задач и формулирования целей намного предпочтительнее. Разнообразные математические процедуры, в этом случае, строятся в последовательность, для которой предусмотрен обязательный контроль со стороны специалиста геолога (оценивающего промежуточные и конеч-

ные результаты решения).

Освещение вопросов о процессах возникновения, формирования и постановках геологических задач в прогнозно-поисковом профиле может осуществляться с такими взаимодействующими частями исследования: 1) проведение процедур наблюдения, регистрация и учет фактов геологической природы; 2) построение математических приемов и формализованных схем решения задач; 3) построение логических представлений и заключений, которыми выявляются уровни соответствия между массивами фактического материала и избранными средствами математического подхода.

Источником информации для первого раздела являются собственно геологические исследования (например, данные геологической съемки, поиска, разведки, научных исследований). Во второй части наращивание математических возможностей производится сбором или созданием новых соответствующих алгоритмов и программ решения и в третьей – видовая специфика логических построений задается расширяющейся системой целей исследования.

Конечно же, трудным и дискуссионным является вопрос о том, что считать геологической задачей, подлежащей решению с помощью математических средств. Мы сочли полезным употребить содержательную трактовку некоторых мест, чтобы оттенить возможность профессиональных постановок задач в геологии, с учетом понятийного и терминологического багажа не только формализованных средств, но и прямых возможностей естественного геологического языка.

Возникновение задач

Учтем следующие сферы, из которых осуществляется "процесс формирования" задач: "сфера целеуказания", "сфера средств", "сфера результатов". Вполне естественно, что этим взаимосвязанным сферам соответствуют различные "части" задачи (подзадачи общей задачи). В "сфере целеуказаний" задачи формулируются в самом общем виде и имеют постановляющий (волевой) характер^{ж)}. Конкрет-

ж) Например, формулировки: "обеспечить черную металлургию высококачественными магнетитовыми рудами", "увеличить добычу нефти на п-милн.тонн" и др.

ное же решение этих задач уходит вглубь соответствующих теоретических разработок и практических реализаций. Как правило, причина возникновения данной "сферы целеуказания" содержится явно (или неявно) в уровнях еще более общих проблем. Задачи, возникающие на самом общем уровне, зачастую формируются и формулируются как отзыв на острую проблемную ситуацию.

"Сфера средств" представляет собой всю совокупность опыта в способах преодоления проблемных ситуаций, то есть, содержит все, что характеризует конкретную обстановку в плане постановок задач и их решения имеющимися на данный момент методами как математического, так и геологического направлений.

"Сфера результатов" возникает непосредственно в "сфере средств" и содержит ответы решения задач, оценка которых выявляет своеобразную меру полноты преодоления проблемной ситуации, сформулированной в "сфере целеуказаний".

Подчеркнем, что это деление на сферы необходимо для понимания процесса формирования на содержательном уровне. "Сфера целеуказаний" очерчивает область и формулировку проблемы, например, постановляет "Разработать систему логико-математических методов для обработки информации по дифференцированным трапповым интрузиям Севера Сибирской платформы, включая Норильский район и на ее основе выдать практические рекомендации с указанием площадей, перспективных на наличие в них полезных ископаемых". Этот уровень содержательного формирования задач называем постановляющим уровнем, ограничивающим исследования сверху. Очевидно, что вся совокупность исследовательских процедур направляется "вниз", к поиску путей реализации цели (из "сферы целеуказаний"), причины возникновения которого лежат за пределами интересов исполнителя. Задачи, возникающие на этом уровне, общи, но не должны выходить за сферу целеуказания "Разработать систему выдать рекомендации".

"Сфера средств", отображающая характер конкретных разработок, содержит структуру и последовательность рабочих (процедур, правил) шагов для достижения цели, сформулированной на постановляющем уровне: "обработкой информации по дифференцированным трапповым интрузиям Севера Сибирской платформы - осуществить определение площадей, перспективных на наличие в них месторождений полезных ископаемых". Этот этап формирования и решения задач, за-

даваемый целеуказанием сверху, наиболее сложен в исследовательском отношении. Обнаруживаемые здесь трудности происходят не только от характера целеуказания сверху, но и от реальной обстановки в области точности и полноты конкретных средств решения задач. Уменьшение разнообразия задач, возникающее от вскрытых трудностей решения, можно производить путем "центрирования" формулировок и процедур практической и исследовательской актуальностью, а также проверенностью средств решения этих задач в сходных обстановках.

"Сфера результатов" обобщает все материалы, связываемые с характером реализации целеуказания. Вопросы эксплуатации методов в сфере средств проходят испытание на эффективность и делаются заключения об их результативной применимости и практической пригодности. Причем, в случае вышеуказанного примера, проверке подлежат как установка "Разработать систему методов ...", так и требование "... выдать рекомендации с указанием перспективных площадей".

В связи с задачами подобного содержания в "сфере средств" для области логико-математического подхода возникают вопросы о создании новых комплексов алгоритмов и реализующих их программ для ЭВМ в соответствии с исходным постановлением (в приведенном выше примере таким постановлением является "Разработать систему логико-математических методов..."). В области же практического внедрения возникают задачи производственного характера, как отзыв на требование "... выдать рекомендации ...".

В последующих параграфах изложены некоторые примеры и детали формирования задач в сферах средств и результатов, задачи же постановляющего уровня, в связи с их спецификой, здесь не рассматриваются.

Объекты

Формирование задач в "сфере средств" содержательно контролируемое заданной проблемой исследования, производится с учетом характера целеуказания, свойств объектов, природы характеристических признаков и специфики математических процедур. Задачи, например, связанные с вопросами систематизации объектов, формиру-

ются в соответствии с реальным разнообразием объектов исследования (рудные районы, месторождения, участки месторождений, рудные тела, рудопоявления и пустые участки), разнообразием целей систематики (вид полезного ископаемого, масштаб запасов, концентрация полезного элемента, генетический тип, локализация во времени, локализация объектов в пространстве) и разнообразием типов характеристик (геологические, геохимические, геофизические, тектонические, стратиграфические и другие комплексы признаков). Целью систематизации объектов является поиск однородных совокупностей объектов в соответствии с целеуказаниями и характеристической обеспеченностью объектов^{*}). Выделенные однородные совокупности объектов принято называть образами или классами, а имеющиеся в распоряжении исследователя конкретные представители этих классов — эталонами. Простейшим случаем систематизации является подразделение объектов по одному признаку (например, подразделение объектов по признаку "промышленное значение месторождения" на две совокупности: "промышленные месторождения" и "непромышленные"). Математические средства исследования позволяют проводить более объективную (в смысле большего учета возможных вариантов и обобщений) систематизацию, основанную на базе всех учтенных признаков. Систематика объектов (как первый и основной результат в организации естественно-научного материала после всестороннего исследования объектов) в случае изучения месторождений зачастую направлена на выявление связи между полезными в промышленном отношении компонентами и комплексом данных о геологических условиях [1,2,3]. Содержательно задачи систематизации завершаются указанием перечня объектов, для которых формулируются дальнейшие требования о детальном сравнительном изучении. Дальнейшая процедура систематизации на базе математических средств сравнительного изучения объектов представляет собой решение задач по до-вольно разнообразным целям.

^{*}) Характеристическая обеспеченность объектов, как важное свойство геологических описаний, включает в себя целый список особенностей, таких как: полнота изученности, виды детальности сведений, специфика сведений (геологические, геохимические, геофизические), разнообразие природы данных (логические, количественные и другие особенности).

Систематизация исследуемых объектов по существу своему связана и с проблематикой содержательной классификации, а результаты ее, в первую очередь, зависят от заданной системы признаков (в предположении, что система признаков информативна) и от исходного множества объектов (которое предполагается либо представительным, либо практически исчерпанным, например, в случае систематизации уникальных месторождений).

Примером, иллюстрирующим вышеизложенные представления, может служить такая систематизация: "Все месторождения никеля" → "месторождения Севера Сибирской платформы" → "месторождения Норильского района" → "месторождения Норильского района, охарактеризованные системой петрологических признаков". Здесь очевиден тот факт, что от общего подразделения систематизация идет к детальному, с одновременным изменением и числа объектов, и характера признаков.

Признаки

В соответствии с существующим уровнем исследования, результаты исследования геологических объектов и явлений в большинстве случаев представляются в виде описаний объектов на естественном языке и каждая попытка исследовать информацию о той или иной совокупности геологических объектов производится на основе составления списка характеристических признаков. Зачастую списки бывают очень длинными (до нескольких сотен признаков) и имеют сложную структуру логического соподчинения. Так возникает задача о систематизации и содержательной трактовке самих признаков.

Кроме того, количество регистрируемых признаков, в их общем списке, для данного целеуказания может оказаться либо избыточным, либо недостаточным. Отсюда основной причиной исследования признаков является необходимость выделения групп признаков, согласующихся не только с конкретной формулировкой цели, но и с характером содержательной постановки геологической задачи. Решение вопроса о систематизации признаков завершается, когда составлен "логический граф" признаков и указаны конкретные процедуры и цели для проверки пригодности выделенных групп (или подгрупп) в решении задач. Результатом такой систематизации является выделе -

ние информативной группы признаков для данной цели обработки таблиц (или ее подтаблиц).

Выделение из множества исходных признаков информативной группы (например, поискового комплекса признаков) при реализации цели обработки информации может иметь три практически важных исхода:

Э ф ф е к т и в н ы й и с х о д, когда выделением информативного комплекса признаков (минимизацией признакового пространства согласно цели) цель исследования достигается. Например, из общей совокупности характеристических признаков находится такой комплекс признаков, по которому с помощью системы математических процедур месторождения хорошо упорядочиваются по запасам (коэффициент парной корреляции порядка "по запасам" и порядка по вычисленным "весам" объектов имеет значение $\geq 0,8$), генетическим типам, или по локализации в пространстве и т.д.

М а л о э ф ф е к т и в н ы й и с х о д, когда наилучший из информативных комплексов по заданной цели обработки информации превышает случайный уровень упорядочивания объектов, но не достигает хорошего уровня (парный коэффициент корреляции составляет примерно $0,6-0,8$). Этот исход на практике встречается довольно часто и для каждого такого случая принимается решение в соответствии со степенью практической полезности результата, которая оценивается экспертами геологами [3,4,5,6].

Н е э ф ф е к т и в н ы й и с х о д, когда любой из информативных комплексов по заданной цели обработки информации не превышает случайный уровень упорядочивания объектов. Например, заданная совокупность признаков, подразделенная на информативные комплексы, ни одним из комплексов не упорядочивает месторождения по запасам лучше, чем случайным образом.

Причины эффективного исхода усматриваются в правильной*) организации совокупности исследуемых объектов, их характеристик и системы обработки информации. Причины малоэффективного и неэффективного исходов более разнообразны: неверное целеуказание, плохой

*) Правильность в том смысле, что исследуемые объекты учитываются в строгом соответствии с целью изучения заранее сформулированной и определяющей вид результата (положительный или отрицательный), то есть исключается постановка типа "решать, что получится".

подбор совокупности объектов, изъяны в организации пространства признаков, неподходящие математические процедуры, слабая изученность объектов в их существенной части и т.д. В случае неэффективного исхода следует тщательно пересмотреть целеуказание, содержательную часть геологической и формализованной постановки задачи [1,4,8].

Формализованное подразделение признаков, основанное на вычислении количественных оценок их существенности, может быть проведено по следующим двум направлениям.

"И н д и в и д у а л ь н а я" о ц е н к а существенности каждого признака в достижении поставленной цели при минимизации признакового пространства. Например, выделить совокупность признаков, по отдельности максимально различающую два класса объектов, или оценить способность каждого признака упорядочивать объекты по целевому признаку (запасы, генезис, возраст и другие свойства, взятые в качестве целевых). Простейшим случаем оценки индивидуальной существенности признаков является коэффициент парной корреляции данного признака со значением целевого признака. Более сложные способы индивидуальной оценки признаков, основанные на содержательном геологическом понятии приуроченности тех или иных признаков к объектам, подлежащим поиску, приводятся в [1,4,9,10].

"К о л л е к т и в н а я" о ц е н к а признаков в их взаимосвязи. Результатом этой оценки является нахождение наиболее информативной системы признаков, упорядочивающей объекты исследования в соответствии с целевым признаком или исходным подразделением объектов на классы, группы, таксоны и пр. Характерно, что выявление коллективной работы признаков сопровождается нередко построением "сильной" системы признаков из индивидуально "слабых". Вычленение индивидуально "слабых" признаков в такую "сильную" систему при выявлении их "коллективных" особенностей (то есть при числовой оценке коллективной существенности) в ряде случаев приводило к построению более эффективных диагностических комплексов по сравнению с простой суммой индивидуально сильных признаков, взятых без учета их взаимосвязи. Этот интересный момент нуждается в исследовании на обширном перечне практических задач. Имеется ряд соображений, позволяющих предполагать, что логика сравнительного изучения объектов в ряде случаев может по-

требовать совмещения индивидуальной и коллективной оценок (в одном числе или векторе). Некоторые конкретные примеры коллективных оценок и способы их применения в диагностических процедурах описаны в [5, II, I2] .

Общая схема

Накопленный опыт решения конкретных задач показал, что осуществление всего процесса формирования задач зачастую производилось по схеме (см. рисунок), включающей в себя постановку задачи, процесс автоматического режима (или собственно процесс решения) и процесс оценки результатов.

Содержательная постановка прикладной геологической задачи имеет место в том и только том случае, когда исследователь располагает^{ж)}:

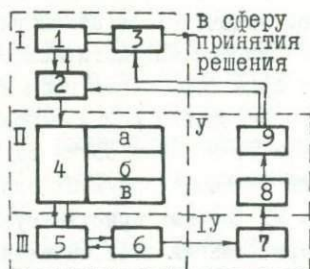
- а) однозначно сформулированным целеуказанием;
- б) отвечающей целеуказанию информацией об исследуемых объектах;
- в) перечнем логических и алгоритмических средств исследования информационного материала;
- г) реальной возможностью применения средств алгоритмического решения вручную или на ЭВМ (ожидаемые затраты машинного времени не превосходят допустимой нормы, имеются рабочие колоды перфокарт и т.п.);
- д) профессиональной заинтересованностью в полученных результатах.

Формализованная постановка задачи выкристаллизовывается таким процессом формирования задачи, в котором выделены все этапы решения от первоначального до завершающего. Процесс автоматического режима, конечно, не исключает, что некоторые формальные процедуры могут производиться "вручную". В разделе оценки результатов выделяются моменты содержательной трактовки. В самом общем виде процесс формирования задачи включает три взаимодействующих блока "Проблема" — "Цель" — "Оператор", в каждом из которых присутствует работа оператора как диспетчера пути формирования цели. Выделение

ж) Несмотря на очевидность и естественность перечисляемых требований, практически задачи, поступающие от "заказчиков", как правило, не удовлетворяют и половине упомянутых требований.

этих блоков связано с необходимостью выявления конкретных особенностей начальных шагов формирования задачи, особенно в части постановок. Интуитивные предположения "оператора" должны охватывать не только возможный способ претворения проблемы в цель исследования и, в конечном итоге, в постановку задачи, но и взвесить ожидаемый результат решения по его роли в разрешении проблемной ситуации. Только после такой предварительной оценки "оператор" может перейти к окончательному формированию задачи и ее конкретной постановке.

Проблемная ситуация возникает всюду, где обостряется необходимость преодоления трудностей нестандартного вида. Непредусмотренность возникшей проблемы часто сочетается с необходимостью ее скорейшего решения, причем это решение зачастую требует построения новых средств. Роль "оператора" в проблемной ситуации, как правило, сводится не только к анализу проблемы на предмет вычленения из нее, по возможности, достижимой цели, но и к тому, чтобы интуитивно уловить характер ожидаемых результатов решения проблемы. Такое "предвосхищение" результатов, базирующееся на профессиональной подготовке и опыте "оператора", играет немаловажную роль в сфере принятия решения, где оно учитывается наряду с результатом применения формальных процедур, как бы контролируя и дополняя его.



Общая схема решения задачи

I - первоначальное формирование задачи: 1-проблемная ситуация; 2-цель; 3-оператор. II - мобилизации информации: 4 - обеспечение информации; а) область мобилизации информации; б) средство мобилизации информации; в) характер информации. III - выбор способа решения: 5 - процедурные средства; 6-блок-схема решения; IV - автоматический режим: 7-ЭВМ. V - оценка результатов: 8-изложение результатов; 9-содержательная трактовка результатов.

{I, II, III} - постановка задачи; {II, III} - процесс окончательного формирования задачи. Встречными стрелками обозначены обратные связи.

Цель или полное определение — целеуказание в вопросах практических постановок задач — играет роль причины ограничения всевозможных рассмотрений материала, подлежащего исследованию. Таким образом, целеуказание выводит исследование из области объективного существования (где объекты исследуются в собственной природе многообразия) в сферу субъективного исследования (где объекты исследуются в собственной природе единообразия, фиксированного нашим интересом к данному свойству объектов, например, "промышленные месторождения"). Основным фактором, определяющим конкретный вид формулировки, является острота и природа самой необходимости исследования, она-то и производит "редакцию" целеуказания. Естественно, конечно, чтобы формулировка целеуказаний производилась людьми, осведомленными как в области представлений об исследуемых объектах, так и в области средств достижения цели.

В профиле прогнозно-поисковых задач целеуказание обычно формулируется с учетом содержания целевого признака. В нашем опыте значения целевого признака исчерпывались следующими вариантами.

Для случая с одним классом объектов значения целевого признака задавались двумя способами [13,14] :

1. Порядковыми значениями, когда известен номер очередности каждого объекта в данном классе (например, объект № 10 в порядке значимости объектов занимает второе место, а объект № 25 занимает седьмое и т.д.).

2. Количественными конкретными значениями, когда для каждого объекта в классе известно числовое значение целевого признака (например, объект № 10 имеет запасы 200 млн.тонн руды, а объект № 25 — 100 млн.тонн и т.д.).

Для случая двух классов помимо указанных значений встречаются также:

1. Бинарные значения, когда целевой признак для объектов первого класса принимается равным единице, а для объектов второго класса равным нулю.

2. Комбинированные значения, когда внутри данного класса для всех объектов известны конкретные значения, а для объектов другого класса известно только, что они месторождения, без сведений о масштабах.

Для объектов, подлежащих диагнозу (проб) целевой признак, как правило, либо неизвестен, либо нуждается в серьезном уточнении [15,19] .

Приведем некоторые примеры наиболее часто формулируемых целей в задачах прогнозно-поискового профиля [16,17,18]. На практике зачастую приходится иметь дело со следующими разновидностями объектов: месторождения, рудопоявления (нефтепоявления), пустые (непродуктивные) и объекты прогноза, как отдельный класс объектов, подлежащих распознаванию на классах эталонов. Месторождения, в свою очередь, подразделяются на базисные группы: уникальные, гигантские, крупные, средние и мелкие. В соответствии с разновидностями способов задания целевого признака и характером классов исследуемых объектов производится выбор цели обработки информации, основанный на формулировке конкретных целеуказаний: определить принадлежность испытываемых объектов к одному из заданных классов; установить место данной пробы в определенной последовательности объектов и т.д.

Процесс окончательного формирования задачи, помимо операций сбора и хранения информации, в основном заключается в организации решения задач уровня разработок и построения формальной блок-схемы решения, иллюстрирующей целесообразное применение выбранных процедурных средств. На начальных ступенях постановки задачи, когда еще присутствуют элементы содержательного формирования задачи, сосуществует двойная формулировка. В целом можно представить себе постановку задачи по следующей схеме (см. табл.): геологическая задача, возникающая из данной проблемной ситуации, формируется, согласно целеуказанию, процессом мобилизации исходной информации. В процессе преобразования содержательной формулировки геологической задачи (приемами первичной обработки информации) в эту формулировку привносятся элементы формализации такие, как операция сбора, автоматического хранения, использования и систематизации информации. Так возникает как бы двойственность "геолого-формализованной постановки задачи", которая преобразуется в формализованную постановку задачи при переходе к автоматическому режиму решения, что не прослеживается во многих задачах, включая и технические [1,4,7,8].

Не следует считать, что подобная схема понимания постановки задачи является единственной или исчерпывающей. Но эта схема, в той или иной мере, неизбежно встречается во всех практических случаях решения геологических задач. Наблюдались также и различ-

Геологическая постановка задачи	Геолого-формализованная постановка задачи	Формализованная постановка задачи
Процесс первоначального формирования задания	Процесс первоначального формирования и мобилизации информации	Процесс выбора средств решения

ные модификации данной схемы, зависящие не только от геологической специфики и характера учтенной информации, но и от математических средств решения.

Вопросы постановки и средств решения

Конкретный процесс решения задач удобно рассматривать с этапа геологической постановки задачи, поскольку именно эта постановка возникает в сложных условиях проблемной ситуации. На следующем этапе реализации "погружения" конкретного содержания геологической формулировки в изображающие средства формализации можно проследить преобразование геологической постановки задачи на глубину формализованной постановки.

Снова подчеркнем, что геологическая постановка задачи возникает как целеуказание, направленное к решению проблемной ситуации: например, проблема "необходимость увеличения прироста запасов нефти и газа (с минимальными экономическими затратами) на территории Западной Сибири в заданные сроки^{ж)}".

Интуитивные предположения "оператора", происходящие из анализа формулировки проблемы, могут, например, заключаться в том, что возрастание эффективности поиска и разведки месторождений нефти

ж) Характер формулировки проблемы обусловлен информационной обеспеченностью и обнаруженной актуальностью в решении проблемы на постановляющем уровне. Экономические мотивы обусловили возникновение этой задачи как геологической, поскольку на постановляющем уровне отраслевое подразделение лишь зарождается.

и газа может происходить не только путем увеличения объема буровых работ, но и путем решения задачи как информационной проблемы. То есть проводится утверждение о необходимости мобилизации информации и ее точной обработки для тех же целей, для которых проводятся трудоемкие процессы инструментальных поисков и разведки. Для этого предлагается долю математического участия в задаче обработки и обобщения разнообразных геологических данных увеличить до необходимости решения специальных задач целевой обработки информации с помощью автоматических средств. Формулировка проблемы и предположения оператора позволяют однозначно указать цель, которая по существу является общей геологической постановкой задачи^{ж)}.

Таким образом, выявление проблемной ситуации, целеуказание, учет предположений оператора представляет собой первый этап формирования задачи в виде общей геологической постановки. В последующем осуществляется уточнение и детализация геологической задачи и преобразование ее в формализованную. Причем эта детализация и уточнение приводят к более широкому истолкованию геологической задачи в терминах и объеме мобилизованной информации.

Нередко регистрация характеристических признаков и отбор объектов исследования производят дальнейшее сужение геологической задачи. Задача "сдвигается" в сторону формализованной постановки. Например, ранее приведенная формулировка конкретной задачи преобразовалась в следующую: "Сортировка продуктивности локальных поднятий по заданной совокупности продуктивных и "пустых" структур по алгоритмам и реализующим их программам распознавания". Данная постановка задачи приводит исследователя к чисто математическим средствам, в частности, требуется найти подходящую для данного случая систему процедур и наметить схему распознавания. В этом месте исследований требуется тщательная последовательность каждого шага, чтобы геологическая задача, перемещенная в область математических идей и средств, не "уточнилась" до своего исчезновения, то есть не претворилась в "чисто" математическую.

Процесс выбора математических средств осуществляется в условиях, которые выявляют формальную сторону геологической задачи. Формализация постановки задачи после содержательного исследования
ж) См. статью В.А.Каштанова и А.Д.Соколова в настоящем сборнике.

ния и отбора исходной информации, а также геологического целеуказания, состоит в построении блок-схемы конкретного решения и выбора математических процедур на каждом шаге решения, о чем изложено в работах [4,13,19]. Особенностью, предъявленной к обработке информации, является довольно длинный список признаков при небольшом числе исследуемых объектов. Эта особенность подсказывает наиболее приемлемую схему распознавания и членения общего решения задачи на отдельные шаги, которые могут состоять, например, в следующем:

1. Поиск информативной системы признаков для сортировки объектов-эталонов на продуктивные и пустые.
2. Сортировка объектов-проб на выбранной системе признаков на их принадлежность к одному из упомянутых эталонных классов.
3. Поиск информативной системы признаков для сортировки объектов-эталонов на нефтяные и газовые.
4. Поиск информативного комплекса признаков, которые позволили бы упорядочить газовые и нефтяные месторождения в соответствии с их запасами.
5. Определение по комплексу признаков (полученному на 4-м шаге) возможного масштаба запасов для проб, попавших в классы нефтяных и газовых месторождений.

1927
4061

Можно указать ту или иную группу алгоритмов, пригодных для решения задачи на каждом шаге. Так, тестовый подход (включая и метод пакетов [9] для задач небольшого объема) пригоден для любого из указанных шагов решения задачи. Зачастую на практике тестовый подход привлекался для получения наиболее подробных результатов на 4-м и 5-м шагах решения [5]. На первом, втором и третьем шагах решения устойчивые результаты получаются по алгоритмам методов согласованных оценок ("Качели") и суммарного учета мер приуроченности и согласования ("Каскад") [5]. При достаточно хорошо организованной информации полные и подробные результаты решения можно получать для шагов 1-4 методом целевой итерационной классификации ("Цикл")*).

Естественно, что приведенный пример не исчерпывает разнообразие геологических целеуказаний, и последовательность шагов не

*) См. статью А.А.Бишаева в настоящем сборнике.



единственна. Нередки случаи более сложных постановок и блок-схем решения, где применяются комплексы алгоритмов и программ. Процесс выбора способа решения заканчивается переводом задачи на ручной или автоматический режим (выполнения множества решающих процедур). Решение задачи в автоматическом режиме осуществляется с помощью электронных вычислительных машин, для чего имеющийся перечень математических процедур программируется на определенные типы машин (в нашем случае М-222, БЭСМ-6). Результаты машинного счета выводятся в явную форму (численные величины, оценивающие строки (объекты) и столбцы (признаки)), и поступают на последующий этап решения задачи, именуемый оценкой результатов.

Оценка результатов

Вопросы оценки результатов и их интерпретации являются ответственным связующим звеном между процессами формирования задачи и процессами формализованной обработки информации с одной стороны, и принятием решений на основании полученных результатов - с другой [1,4].

Сложность оценки и интерпретации обусловлены тем фактом, что необходимо дать окончательную трактовку самым разнообразным данным. Здесь снова объединяются все сообщения, которые были подразделены в процессе решения. Кроме того, требуется установить, по возможности, исчерпывающим образом степень совместимости специфических особенностей изучаемых предметов и явлений с полученными результатами [2,3].

В связи с вышеизложенным возникает необходимость сформулировать некоторые содержательные требования к проведению процесса оценки результатов. Поскольку представление результатов в значительной степени определяет ход интерпретации и тем самым может повлиять на ее итог, начнем с требований к изложению.

К о н к р е т н о с т ь, простота и наглядность изложения. В ряде случаев интерпретаторы используют общие фразы изложения результатов решения. Эта форма изложения результатов, допустимая для устных сообщений, не должна быть нормой при представлении результатов решения в полном виде. Следует четко задать количественный материал, на базе которого строятся выводы и обобщения. Причем явная форма представления результатов не ограничивается

только табличной формой, но могут строиться соответствующие графики, сравнительные кривые и т.п.

П о л н о т а и последовательность использования результатов решения. Данное требование нацелено против попытки "кусочной" интерпретации "нужного" места задачи. При такой интерпретации используется лишь некоторая, возможно, не самая существенная часть полученных результатов, не производится предварительного выявления главных, определяющих и второстепенных фактов. Функция и форма участия геолога в процессе оценки результатов должны определяться после выяснения общей картины полученных результатов. В этом случае можно более четко и удобно представить данные, полезные для геолога, за счет удаления второстепенных деталей. Кроме того, если понадобится подробно рассмотреть задачу в иных геологических предположениях, то просмотр приведенного в удобный вид материала не составит большого труда.

Н е з а в и с и м о с т ь от целеуказания. Это требование говорит о том, что при интерпретации не следует пытаться путем натяжек совместить полученные результаты с результатами, ожидаемыми согласно целеуказанию. Эта независимость может помочь обнаружить факты, не предусмотренные первоначальной постановкой задачи. Кроме того, бывает полезным иметь представление о ценности полученных результатов относительно видоизмененной формулировки целеуказания.

Н е з а в и с и м о с т ь от "внешней" информации. Данное требование родственно предыдущему, но является особенно важным. Его важность связана с тем, что зачастую интерпретатор (геолог) начинает бессознательно (а иногда и сознательно) расширять совокупность используемых признаков, причем делать это раньше того момента, когда будет готово обобщение материала, основанное на исходном табличном перечне признаков. Эта ситуация является основным источником обесценивания результатов решения, поскольку они попадают в сферу произвола, обусловленного уровнем осведомленности и профессиональной подготовки интерпретатора. Таким образом, привнесение "дополнительных" соображений и "новых данных" до момента принятия решения является предметом опасным, а потому нежелательным.

Если возникает необходимость задействовать новые важные сообщения, то их следует учесть на новом шаге решения, с учетом пре-

дыдущего. При этом возможны два случая:

- а) новые сообщения подтверждают полученные результаты;
- б) новые сообщения опровергают полученные результаты.

Во втором случае надо пересмотреть не только пространство при — знаков, но и постановку задачи. Конечно, совершенно недопустимо отбрасывание данных, не согласующихся с представлениями заказчика. Что же касается оценки практической пригодности результатов, то она устанавливается общепринятым образом — через соотношение теоретических и практических данных (например, корреляция запасов месторождений с информационными весами).

ж ж ж

Взаимопроникновение математики и геологии осуществляется во встречном движении. Результатом этого движения должны быть новые разделы прикладной математики и превращение геологической отрасли знания в точную науку с новыми целями и средствами исследования. Этот трудоемкий и длительный процесс имеет сложную структуру, применим в наиболее развитых отраслях геологии и имеет отношение к теоретическим и практическим задачам. Разделы геологии, связанные с учением о полезных ископаемых, являются вполне приемлемой областью применения математических средств. Причем это обусловлено не только наибольшей готовностью и необходимостью ускорить исследования прогнозно-поискового характера, но и созрелыми формами математических приемов в задачах подобного содержания. Изложенные схемы организации геологического материала в области формализованных понятий и процедур по мере накопления практического опыта будут несомненно расширяться и уточняться.

Л и т е р а т у р а

1. КОНСТАНТИНОВ Р.М. Математические методы в металлогенетических исследованиях рудных районов. — В кн.: Металлогенетический анализ рудоконтролирующих факторов в рудных районах. М., "Недра", 1972, с.1-114.
2. РОДИОНОВ Д.А. Статистические методы разграничения геологических объектов по комплексу признаков. М., "Недра", 1968. 153с.
3. Формальные методы разграничения геологических объектов и их применение при решении некоторых задач минералогии и металло-

гении. - В кн.: Применение математических методов в геологии. Алма-Ата, "Наука", 1968, с. 5-27.-Авт.: Бугаец А.Н., Гражданцев Н.Г., Дорофеев А.А. и др.

4. Организация и обработка геологической информации с помощью ЭВМ на основе построения тупиковых тестов. - В кн.: Логико-информационные решения геологических задач. М., "Наука", 1975, с.83-148. - Авт.: Дмитриев А.Н., Кренделев Ф.П., Бишаев А.А. и др.

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

6. ЖУРАВЛЕВ Ю.И., НИКИФОРОВ В.В. Алгоритмы распознавания, основанные на вычислении оценок.-"Кибернетика", 1971, № 3, с. I-II.

7. ЗАГОРУЙКО К.Г. Методы распознавания и их применение. М., "Советское радио". 1972. 192 с.

8. Алгоритмы и программы решения геологических задач на ЭВМ "Минск-2" и "БЭСМ-3М". Алма-Ата, Изд. Казахского НИИМС, вып.3, 1970. 133с. -Авт.:БУГАЕЦ А.Н., ДВОРНИЧЕНКО Г.К., МАЦАК А.П.и др.

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

10. ТРОФИМУК А.А., ВЫШЕМИРСКИЙ В.С., ДМИТРИЕВ А.Н. Геохимические основы поисков нефти и газа в Сибири. Новосибирск, 1971.81 с.

11. МАЗУР М. Качественная теория информации. М., "Мир", 1974. 232 с.

12. КОНСТАНТИНОВ Р.М., ДМИТРИЕВ А.Н. Методы обработки качественной геологической информации для определения формационного типа месторождений.-"Изв. АН СССР, сер. геол.", 1971, № 8, с.3-15.

13. БАБИЧ В.В., ФЕДОСЕЕВ Г.С. Типизация прогнозно-поисковых задач и некоторые подходы к их решению. - В кн.: Состояние и направление исследований по металлогении траппов. Тезисы докладов к третьему Всесоюзному совещанию 23-25 октября 1974 г., Красноярск, 1974, с.118-119.

14. ДМИТРИЕВ А.Н., ЖУРАВЛЕВ Ю.И., КРЕНДЕЛЕВ Ф.П. Об одном принципе классификации и прогноза геологических объектов и явлений.-

"Геология и геофизика", 1968, № 5, с.50-64.

15. ВАСИЛЬЕВ Ю.Л., ДМИТРИЕВ А.Н. Спектральный подход к сравнению объектов, охарактеризованных набором признаков.—"ДАН СССР", 1972, т.206, № 6, с.1309-1312.

16. РОДИОНОВ Д.А., СЕРЫХ В.И. Проблема выбора поисковых признаков и ее статистическое решение. — В кн.: Математические методы в геологии. М., "Наука", 1968, с.83-104.

17. Распознавание образов гигантских нефтяных месторождений. — В кн.: Проблемы нефтегазоносности Сибири, Новосибирск, "Наука", 1971, с.34-50. — Авт.: Трофимук А.А., Вышемирский В.С., Дмитриев А.Н. и др.

18. ДМИТРИЕВ А.Н., ЗОЛУХИН В.В., ВАСИЛЬЕВ Ю.Р. Опыт применения дискретной математической обработки информации по дифференцированным трапповым интрузиям Северо-Запада Сибирской платформы.—"Советская геология", 1968, с.98-108.

19. ВАСИЛЬЕВ Ю.Р., ДМИТРИЕВ А.Н., ЗОЛУХИН В.В. Распознавание и оценка никеленосных дифференцированных трапповых интрузий Севера Сибирской платформы.—"Геология и геофизика", 1973, № 1, с.95-101.

Г.С.Федосеев

ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ В ЛОГИКО-МАТЕМАТИЧЕСКИХ ИССЛЕДОВАНИЯХ

Процедура распознавания является традиционным методом всех естественных наук, в том числе и геологии. Все разнообразие задач по диагностике и идентификации геологических объектов, прямо или косвенно затрагивающих проблему распознавания образов, сводится в конечном итоге к двум ситуациям. Во-первых, распознавание по прямым признакам, т.е. признакам, присущим только самим объектам и набираемым в процессе обучения и описания последних. При наличии необходимого и достаточного количества таких признаков решаются задачи классификации (таксономия и построение классификаций). Вторая ситуация возникает при распознавании геологических объектов по косвенным признакам, которые не являются обя-

зательными атрибутами самих объектов и не связаны требованиями адекватности при описании свойств последних. Выполнение косвенного признака само по себе не может быть основанием для однозначного, суждения о наличии объекта. Это связано с тем, что большинство косвенных признаков относится к разряду конвергентных: одним и тем же признаком могут характеризоваться различные объекты, процессы или явления. Таким образом, даже большое количество косвенных признаков не может устранить не только ограниченности характеристики объекта, но подчас даже в явной форме не проливает свет на факт наличия или отсутствия самого объекта. Тем не менее, именно косвенные признаки являются главной основой для прогнозирования в геологических науках.

Математическая обработка косвенных признаков начала привлекать внимание исследователей в последнее время в связи с тем, что были найдены логические процедуры, имеющие прямой выход в практику научного прогнозирования. Действительно, многие задачи геологического прогнозирования (определение перспективных площадей, участков и рудопроявлений, оценка перспективности геомагнитных аномалий, идентификация сейсмических волн, определение генетической принадлежности минералов и т.д.) сводятся к проблеме распознавания образов [1,2,3,4]. Описательный характер значительной части информации в геологии вынуждает прибегать к специальным математическим методам, в частности, к дискретному анализу, весьма перспективному, но, к сожалению, еще недостаточно разработанному в плане решения прикладных геологических задач. В последних диагностика и классификация объектов обычно сводятся к разбиению многомерного признакового пространства на самостоятельные или частично пересекающиеся области, каждая из которых соответствует одному образу. Качество распознавания зависит, естественно, от степени взаимного пересечения классов в заданном признаковом пространстве. В этой связи становится понятным, что этап сбора и предварительной обработки геологических сообщений заслуживает не только самостоятельного, но и первоочередного рассмотрения. Изложению некоторых приемов, применяемых для предварительной обработки исходной информации, и посвящена данная статья.

Исходное признаковое пространство

Под признаковым пространством понимается рациональная с профессиональной и формальной точек зрения система признаков, полный перечень которых составлен с обязательным учетом целевого задания (прогноз запасов, предсказание рудоносных районов и перспективных площадей и т.д.). Качество работы по организации списка признаков, зависящее от уровня геологической культуры, определяет в конечном итоге достоверность выводов. Общая теория формирования признаковых пространств практически не разработана, однако совершенно очевидно, что она должна предусматривать ответы на следующие вопросы: каким признакам следует отдавать предпочтение на стадии сбора информации, какое количество их является достаточным и необходимым на различных стадиях математической обработки, указание способов упорядочивания признаков и т.п. Следует также отметить, что формирование признакового пространства является необходимой процедурой при любом виде логико-математической обработки геологической информации. В этом направлении при решении множества задач по распознаванию геологических объектов уже проделана большая работа [5,6] .

Процедуры, связанные с отбором признаков, являются довольно трудоемкими и весьма значительными по затратам времени. Особую тщательность, не нарушая, по возможности, принципа коллегиальности и всестороннего обсуждения правомерности введения каждого признака, следует проявить на начальном этапе сбора информации. При этом необходимо выработать единый подход ко всему имеющемуся материалу. Главное внимание здесь должно быть уделено уровню изученности (учет вида и масштабности съемочных и поисковых работ, стадии разведки или эксплуатации месторождений). На каждой из этих стадий добываются сведения различной степени детальности, поэтому характер признакового пространства зависит от того, что является важнее — детальность описания или общая характеристика. Первый путь неизбежно ведет к сокращению числа объектов и увеличению числа признаков, а второй — к сокращению признакового пространства и включению в сферу распознавания новых, мало изученных объектов. Совмещение этих тенденций, по-видимому, не имеет особого смысла, поскольку некоторые сочетания признаков в общей

таблице окажутся искусственными. Существенную помощь на этом этапе могут оказать информационно-поисковые системы, обеспечивающие предварительную унификацию и классификацию материала на перфокартах, магнитных лентах и т.п. В качестве самостоятельных оценок пригодности признакового пространства могут быть использованы различные числовые параметры, полученные в процессе статистической обработки материала.

Наиболее полная характеристика объектов исследования достигается различными способами, в соответствии с чем все пространство признаков подразделяется на ряд групп. Каждая группа характеризуется внутренним родством признаков и составлена только геологическими, петрографическими, геохимическими, геофизическими или какими-нибудь другими признаками, полученными в результате определенного вида исследования. Это подразделение позволяет легко установить преимущественную направленность и характер работ, выполненных на исследуемых объектах.

Структура признакового пространства может быть организована и по другому принципу: в качестве основы для выделения групп признаков выбираются разномасштабные геолого-структурные элементы или процессы, которые имеют прямое или косвенное отношение к наличию концентраций полезного компонента или обеспечивают теоретическую возможность его проявления. Именно такое подразделение признакового пространства нашло наиболее широкое применение: последовательно характеризуются регион, участок, рудное тело с особенностями внутреннего строения и условий залегания, вмещающие породы, околорудные образования, возрастные взаимоотношения и другие позиции.

На размер и структуру признакового пространства, как отмечалось ранее, решающее влияние оказывает целевое задание. Первые ограничения накладываются целеуказанием, обозначающим полезное ископаемое — железо, медь, золото, полиметаллы, фосфориты, бокситы и т.д. Это, естественно, требует привлечения в первую очередь тех признаков, которые обнаруживают связь именно с данным видом оруденения и характером его проявления (коренное, россыпное). Здесь в максимальной мере должны использоваться преимущества, приобретаемые геологами в результате узкой специализации и углубленного изучения объектов определенного класса. Именно профессиональный (а не формальный) подход не позволяет рассматривать

в системе одного признакового пространства и россыпные, и коренные месторождения какого-то полезного ископаемого, когда речь идет, например, о прогнозировании запасов. Другими словами, совсем не безразлично, какую природу имеют создавшиеся аномальные содержания полезного компонента, поскольку такие признаки как "характер плотика", "положение русел древних рек" и т.п. вряд ли имеют значение для прогноза и поиска золоторудных кварцевых жил. Данный момент является весьма важным, поскольку в конечном счете он предопределяет отсутствие в признаковом пространстве таких признаков, сочетание которых с оставшимися лишено геологического смысла.

В прогнозно-поисковых работах употребляются такие пространства признаков, которые позволяют производить их естественное группирование (объединение в подпространства: геологические, петрографические, минералогические, геохимические, геофизические и др.) Деление это не является безусловным, но для практических целей вполне понятно и приемлемо. Внутри каждого подпространства могут быть выделены еще более мелкие подразделения. Например, среди геологических признаков могут быть выделены: глобальные, региональные, локальные или тектонические, стратиграфические, магматические и другие группы признаков. В этой связи возникает вопрос, насколько равномерно должны быть охарактеризованы объекты признаками, принадлежащими к каждому из выделенных подпространств. Общим правилом в подобных случаях является соблюдение требования, согласно которому, каждый отбираемый признак "работал" бы не только на идентификацию объекта (как представителя определенного класса), но также на прямое или косвенное обеспечение решения целевого задания. Именно здесь наряду с профессиональными знаниями особую ценность приобретают специальные математические исследования, в простейшем случае — установление парных и частных корреляций того или иного характеристического признака с целевым.

Структура правильно организованного признакового пространства в общем виде должна быть иерархической, а для решения конкретных вопросов необходимо использовать те или иные ее части. Например, региональные признаки необходимы главным образом для оценки перспектив отдельных площадей на предмет обнаружения в них месторождений полезных ископаемых, а локальные — для оценки значи-

мости объектов. Применение классификационных графов для признаков преследует цель создания признакового пространства, однородного по внутренней структуре, уравнивания признаков по смысловой нагрузке. Это означает, что от процедур формальной обработки наиболее хорошо интерпретируемые результаты можно ожидать в том случае, если в признаковое пространство включены признаки преимущественно с одного уровня классификационного графа. Поэтому использование в признаковом пространстве взаимно перекрывающихся признаков, находящихся в соотношении соподчиненности, относится уже к области эксперимента, перспективы содержательной интерпретации результатов которого недостаточно ясны.

Следует иметь в виду, что в пределах одной группы могут оказаться признаки, разные по своей природе. В этом плане среди геологических признаков обращают на себя внимание возрастная характеристика объектов и наложенные вторичные процессы, а среди геофизических — параметрические, расчетные и обобщающие признаки. Возрастная характеристика геологических образований является весьма важным параметром при выполнении прогнозно-оценочных работ. Однако лишь цифры абсолютного возраста и то с известными оговорками можно считать однозначно установленными фактами. Обязательным требованием при датировании является использование анализов, выполненных по единой методике (по минералам или валовым пробам) желательно в одной лаборатории и по определенной шкале, что позволяет свести к минимуму влияние случайных факторов. Установление относительного геологического возраста объекта (массива, толщи, комплекса и т.д.) не является наблюдаемым фактом, а представляет собой конечную цель специальных работ: стратиграфических, палеонтологических или структурно-петрологических. Другими словами, относительный геологический возраст является целевым признаком, поскольку для его доказательства необходима своя категория фактов. Исходя из этого, в качестве признака в информационных таблицах может быть использована только унифицированная абсолютная геохронологическая характеристика, либо относительный возраст, подтвержденный формальным способом (как целевой признак).

Часто в качестве признаков вводится интенсивность проявления того или иного процесса (серицитизации, окварцевания, биотитизации и т.п.). Несмотря на то, что обычно дается порядковая оценка

("нет", "очень слабо", "слабо", "сильно" и др.), по-существу, это есть описание факта наличия определенного минерала, который развивается позже других. Отражением процесса в данном случае будет количественная или порядковая оценка содержания минерала. Другими словами, если в качестве признака фиксируется, например, "альбитизация", то наличие этого признака, выражаемое единицей, должно означать, что имеется только наложенный альбит (вторичный), а для подтверждения этого факта необходима особая система доказательств, основанная на использовании достоверных признаков замещения альбитом других минералов.

В геофизической группе признаков следует различать по крайней мере две разновидности: а) параметрические (получаемые при изучении физических полей и при лабораторных измерениях), б) расчетные (получаемые при обработке первичных материалов). При альтернативном решении вопроса предпочтение следует отдавать первой группе признаков, поскольку представители второй группы характеризуют, по-существу, какой-то вариант вычислений аппроксимационного характера. Именно первые признаки могут считаться однозначно установленными фактами (наличие разнознаковых локальных аномалий и их интенсивность на фоне гравитационного поля). Вычисленные характеристики геологического содержания (глубина залегания возмущающих масс, их форма, размеры и др.) являются относительными и значения их зависят от многих допущений по фоновому полю, вещественному составу возмущающих масс и их окружения и т.д. Большинство геофизических признаков представляется обычно несколькими градациями, что заставляет прибегать к формированию обобщающих признаков ("синусоидальный вид графика ΔT_g ", "спокойное поле", "линейный характер графика приращений" и т.п.).

Несколько особняком стоит группа признаков, получаемых при анализе геологических, тектонических, палеогеографических, гидрогеологических, геоморфологических и других видов карт, разрезов и схем. Информация, которую несут карты, имеет не только геометрический смысл, отражая специфику представленных на данной территории объектов. В известной мере она охватывает многие неметрические свойства объектов, представленные в закодированном виде. Другими словами, карта — это уже логически обработанный и минимизированный на основе личного опыта составителя (коллектива) ма-

териал, подчиненный выражению главной цели (совокупности целей). Поэтому, формируя признаковое пространство с привлечением данных, снятых с карт, следует всегда иметь в виду возможность появления обобщающих признаков. С данных позиций целесообразно, по-видимому, различать два вида графической информации. Первый, базирующийся на использовании преимущества априорных данных и сведений общегеологического характера, условно называется геологическим. Второй является чисто геометрическим, отражающим особенности строения региона в виде совокупности точек, линий или плоскостей [7]. Графический материал, например, широко использован А.М. Волковым [8], в работе которого на основе сейсморазведочных данных перед постановкой разведочного бурения оценивалась газоносность некоторых структур Березовского района Западной Сибири. Основной для распознавания здесь явились геометрические признаки, полученные при анализе различных структурных карт.

Способы задания признаков

В настоящее время не существует общепринятого определения признака — в большинстве случаев признак отождествляется со свойством, иногда в разряд признака возводится значение отдельных градаций какого-либо предмета, но имеется немало и таких работ, в которых вообще отсутствуют определения признака как категории, само собой разумеющейся.

Признаки в научно-познавательных и практических процессах необходимы главным образом или для целей идентификации объектов, или для выявления их специфики. В этом отношении наиболее общее определение дается Н.И. Кондаковым: "Признак — все то, в чем предметы, явления, сходны друг с другом или в чем они отличаются друг от друга" [9, стр. 416]. Использование формальных процедур обработки информации требует сужения и конкретизации понятия признака, под которым в дальнейшем понимается реализация способности объекта обладать (или не обладать) каким-то свойством или какой-то его частью. Различные состояния (градации) признака могут быть зафиксированы с помощью заранее выбранного алфавита — словесного, буквенного, графического, цифрового и т.п. Иначе говоря, признак объекта в качестве самостоятельной категории проявляется

только в том случае, если введена мера оценки свойства или состояния объекта.

В большинстве случаев признаки задаются в виде различной сложности высказываний об отношениях и свойствах объектов. По логической структуре в первом приближении такие суждения можно подразделить на две группы. Одна из них отвечает формуле " S есть P " и является удобной для организации целевых признаков. В таких суждениях предикат отражает сущность цели: "объект (S) есть месторождение (P)", "запасы месторождения (S) - I млн. тонн (P)" и т.д. Среди целевых признаков в задачах прогнозно-поискового профиля чаще других употребляется оценка месторождений и рудопроявлений по запасам. По этому признаку производится упорядочение объектов в процедурах обучения, а в процедурах распознавания определение целевого признака является конечной целью, то есть конечным результатом исследования. Естественно, что при прогнозировании запасов должна учитываться их категоричность, а при оценках рудопроявлений на предмет их принадлежности к разряду месторождений - способ подсчета запасов (геологический или геофизический). Очевидно, что нельзя в одну таблицу включать месторождения и рудопроявления, запасы одной части которых подсчитаны на основе геологических данных, а другой - на основе геофизических. Как правило, в качестве целевого признака выбирают количественные характеристики, однако, в принципе, целевым признаком может быть и дискретная величина (например, генезис, район локализации и т.п.).

Вторая, более обширная группа признаков, описывается формулой " aRc ", где a - объект (свойство), c - переменное суждение, R - переменная связка, отражающая отношение между " a " и " c ". Примером могут служить следующие признаки: "ореол рассеяния золота (a) меньше (R) 50 км^2 (c)", "интенсивность аномалии (a) колеблется в пределах (R) 400-800 гамм (c)" и т.д. По этой формуле целесообразнее формировать характеристические признаки, каждый из которых имеет потенциальную возможность перейти в разряд целевых при изменении постановки задачи. Так, например, в связи с проблемой поисков скрытого оруденения одним из главных целевых признаков может стать величина эродированности месторождений, которая позволит полнее раскрывать геометрическую (прост-

странственно-временную) сторону процессов мобилизации и концентрации полезных компонентов.

Типы признаков

В зависимости от целей исследования и способов обработки информации описание метрических свойств объектов может быть проведено на различных уровнях детальности: качественном, полуколичественном или количественном. Главная задача здесь заключается в том, чтобы выбрать числовую систему в максимальной степени соответствующую эмпирической, подобрать относительно простые и содержательно интерпретируемые отношения в рамках теорем представления и единственности [10,11] .

Если при характеристике объекта за единицу измерения принято свойство в целом, то получаются качественные признаки, отражающие способность объекта обладать или не обладать данным свойством (например, "магнитный" - "немагнитный"). Следующим шагом в детализации описаний объектов является введение так называемых полуколичественных признаков, получающихся благодаря делению свойства на крупные части - градации по границам, положение которых определяется, исходя из профессиональных соображений. Такие признаки отражают степень или интенсивность проявления свойства, поэтому их удобнее выражать целочисленными упорядоченными значениями. И, наконец, объекты описываются количественными признаками, которые всегда выражены числами и позволяют в ряде случаев получить наиболее объективный и точный вид описания, при этом детальность характеристики зависит, естественно, от выбранного масштаба измерения.

Для обработки качественных признаков целесообразнее применять дискретные методы. Количественные же признаки позволяют выполнять над собой более широкий спектр математических действий, независимо от того, взяты ли они в отдельности или в совокупности с другими количественными признаками. В результате получают выводимые (комплексные) признаки, содержательный смысл которых не всегда может быть точно определен, но в некоторых случаях они оказываются более информативными по сравнению с исходными признаками [12] .

С формальной точки зрения, признаки можно, по-видимому, разделить на две группы — непрерывные и дискретные. Первую группу составляют признаки, получаемые с помощью шкал отношений и интервалов. Во вторую группу входят собственно дискретные (булевы) признаки, а также α - и β -дискретные признаки, отличительные черты которых будут указаны ниже. Собственно дискретные признаки имеют логическую природу и полностью соответствуют качественным.

Представление количественного признака в дискретной форме производится путем разбиения всего диапазона его значений на конечное число интервалов, каждый из которых соответствует одной градации, занимающей вполне определенное место в их общем ряду. Стремление к минимальной потере информации, обеспечивающей цель, побуждает геологов вводить максимально возможное число градаций. Границы между градациями выбираются по разным мотивам, или исходя из формальных соображений, или из характера соотношений естественно складывающихся значений параметра, или из физической природы объекта и явления (наличие фазового перехода, критической точки и т.п.). В основу разбиения могут быть положены также графики частоты встречаемости или пределы, устанавливаемые эмпирическим путем. С тем, чтобы такие признаки отличить от булевых, назовем их α -дискретными. Последовательность градаций при образовании таких признаков вполне определенная, но границы между ними обычно условные. Числовое описание α -дискретных признаков производится в шкалах порядка.

Другая разновидность дискретных признаков является следствием детализации логических признаков путем введения дополнительных критериев, отражающих точечные (неметрические) свойства. В этом случае основное внимание уделяется самому факту проявления градаций, тогда как границы между ними остаются неизвестными. Более того, неочевидны ширина границ и последовательность (упорядоченность) таких градаций, количество которых зависит от детальности и полноты изученности объекта. Примером может служить текстурная характеристика руд, отражающая способ расположения минеральных агрегатов в пространстве: массивная, полосчатая, вкрапленная, прожилковая, брекчиевидная. Этим перечнем, конечно, не ограничивается множество текстур, однако предугадать появление новых (непромежуточных) разновидностей и место их в приводимом перечислении невозможно. С тем, чтобы такие признаки отличать

от группы собственно дискретных, назовем их β -дискретными. О характере границ между такими признаками ничего определенного сказать нельзя: по смысловому значению они могут отличаться, соприкасаться или даже частично перекрываться. Числовое описание этих признаков производится в шкалах наименований.

Таким образом, каждый из выделенных формальных типов признаков получается при исследовании свойств объектов с помощью различных шкал измерения. Это следует из приводимой ниже схемы, в которой всем типам признаков присвоены номера:

Свойства	Деление свойств и признаков по природе	Шкала измерения	Формальный тип признака	Номер типа
Метрические	Качественные	Булева	Собственно дискретный	1
	Полуколичественные	Порядка	α -дискретный	2
	Количественные	Отношений Интервалов	Непрерывный	3
Неметрические	Точечные	Наименований	β -дискретный	4

Кодирование признаков

Под кодированием понимается совокупность операций по приведению исходной логической и количественной информации к виду, удобному для математической обработки, т.е. представление имеющегося материала в числовой или порядковой форме.

А п р и о р и о е к о д и р о в а н и е. Условия кодирования качественной информации предопределяются, по существу, общепринятым порядком описания: выполнение признака обозначается единицей, невыполнение — нулем, а неизученность — прочерком. Для рационального описания наибольшую ценность имеют позитивные высказывания об объектах, т.е. целесообразнее отразить именно наличие какого-то свойства, а не его отсутствие. Это в полной мере относится и к геологическим объектам, которые во всем многообразии проявлений описываются преимущественно с позитивной стороны.

Например, характеризуя рудную жилу, экономнее сказать "имеется кварц", нежели "нет золота", "нет карбонатов", "нет сульфидов" и т.п. Условившись обозначать выполнение признаков через единицы, мы тем самым осуществляем профессиональную ориентацию кода. Формальное описание объекта в таком случае (при отсутствии цели и ограничений на размер признакового пространства) будет представлено в виде набора (слова), состоящего только из единиц. Подобным же образом будут выглядеть и формальные описания всех изолированно охарактеризованных объектов. Если теперь описать каждый из изученных объектов с точки зрения каждого свойства, хотя бы однажды встретившегося в индивидуальных описаниях, то размер **формальной** описания увеличится на число новых признаков, значения которых должны быть обозначены через нуль. Полученные таким образом описания представляются в виде таблиц, составление которых является главной целью первого этапа работ — этапа сбора и априорного кодирования информации.

При кодировании порядковых признаков необходимо вводить границы для формирования градаций. Число этих градаций обычно больше трех (например, "очень мелкие", "мелкие", "средние", "крупные" "очень крупные"), причем границы в большинстве своем являются условными. Такая информация по структуре является уже дискретной.

Количественная информация представляется в виде совокупности чисел, каждое из которых, если оно в пределах точности измерений отлично от соседнего, можно рассматривать как самостоятельную градацию. Для практических целей такая дробность является совершенно излишней, в связи с чем появляется необходимость введения более крупных градаций. Обычно эта операция выполняется чисто условно, с выбором границы по аналогии или исходя из личного опыта исследователя. По-видимому, кодирование количественной информации при достаточном числе объектов следует производить с учетом характера статистического распределения. Интересные исследования в этом плане выполнены Ю.Г.Шестаковым и др. [13], которые пришли к выводу о выборе разделяющих границ значений признаков таким образом, чтобы вероятность ошибочного отнесения объектов экзамена к двум из заданных классов были бы примерно одинаковы. В качестве разделяющих значений авторы предлагают использовать среднее арифметическое, моду или медиану.

Степень дискретности признаков понижается с повышением точности измерения свойств. С введением значительного (больше трех) числа градаций для значений признаков необходимо применять цифровой алфавит, который при надобности может быть переведен в нуль-единичный с помощью двух следующих приемов. Первый прием заключается в разбиении всех имеющихся градаций признака на две группы с последующим обозначением представителей одной из них только через единицы, а другой — только через нули. Вторым приемом сводится к выражению в нуль-единичном алфавите состояния каждой градации: выполнение какой-либо градации фиксируется единицей, а все остальные — обозначаются через нуль. Нетрудно заметить, что перекодирование в первом случае заключается в приведении количественных измерений к соотношению типа "много-мало", "часто-редко" и т.п., и, несомненно, связано с потерей информации. Во втором случае операция сводится к возведению каждой градации в ранг признака, что влечет за собой увеличение системы признаков на величину $\sum_{j=1}^k (z_j - 1)$, где k — количество признаков с градациями, а z_j — число градаций в j -м признаке ($z_j \geq 2$). Применение этих приемов получения "искусственных" булевых признаков связано с несомненной потерей информации.

Выбор алфавита и кода позволяет приводить исходные описательные таблицы (содержащие сведения об объектах в виде высказываний) к формальному виду. Вначале убирается классификационный признаковый граф и описательная таблица превращается в частично формализованную. Главное различие данных таблиц заключается в том, что в первой может меняться только порядок описываемых объектов, а порядок в системе признаков фиксирован. В частично формализованной таблице произвольно может меняться порядок и признаков, и объектов. Для получения формальной (полностью формализованной) таблицы необходимо закодировать значения признаков в выбранном алфавите. Отсюда следует, что профессиональная ориентация кода применяется на двух этапах: во-первых, при систематизации материала геологу приходится самому выбирать признаковые подпространства, и, во-вторых, решать вопрос о том, какие значения признака обозначить тем или иным элементом выбранного алфавита. Насколько квалифицированно будут выполнены эти две операции, настолько эффективной при прочих равных условиях будет содержательная

интерпретация полученных в процессе математической обработки результатов.

А л г о р и т м и ч е с к о е к о д и р о в а н и е . Алгоритмическое кодирование может осуществляться двумя путями. Часть алгоритмов обычно основывается на каком-либо формальном требовании, вытекающем, например, из условий наиболее эффективной работы выбранной процедуры распознавания. В качестве одного из таких требований может, например, выступать необходимость закодировать информацию таким образом, чтобы количество единиц и нулей в обрабатываемой таблице было бы примерно одинаковым [14]. Достигается это введением для каждого признака такой границы, относительно которой число бинарных значений должно уравниваться. В отличие от априорного кодирования алгоритмический подход позволяет автоматизировать поиск оптимального варианта методом перебора (возможно, неполного), однако при малом числе объектов он довольно легко осуществляется вручную.

Более сложное, специализированное кодирование проводится с учетом характера целевого задания и типа признака. Так, например, α -дискретные и β -дискретные признаки, измеряемые, как отмечалось, соответственно по шкалам порядка и наименований [10], кодируются по-разному, а в процедурах кодирования обязательно предусмотрено изменение структуры признаков в соответствии с характером цели. Поэтому данный способ кодирования, способствующий рациональному выбору решающего правила в процедурах распознавания, рассматривается подробно в следующей статье настоящего сборника .

Оценка существенности признаков

При описании предмета без связи с другими и безотносительно к цели (феноменологическое описание) все признаки равнозначны в смысле полезной информации о предмете, и такое описание трудно классифицировать как избыточное, поскольку вся совокупность сведений имеет структуру логического произведения и различение признаков по существенности лишено смысла.

Понятие существенности признака, употребляющееся в процедурах распознавания, имеет толкование, несколько отличное от общепринятого, в соответствии с которым существенными считаются при-

знаки, отражающие свойства и связи объекта, "без которых он существовать не может" [см.9, стр.507]. В логико-информационных исследованиях существенными считаются те признаки, которые в пределах фиксированного пространства наилучшим образом обеспечивают реализацию цели. Все остальные признаки следует считать несущественными и значительная их часть может быть исключена уже на стадии предварительной обработки информации. Другими словами, степень эмпирической существенности признака определяется силой очевидности его связи с обеспечением цели.

Понятно, что правильный подбор необходимых признаков сопряжен с рядом трудностей. Однако многолетний геологический опыт позволяет интуитивно оценивать отдельные группы признаков с точки зрения их значимости для решения прогнозных задач с учетом вида полезного ископаемого. Так, например, для оценки нефтегазоносности наиболее важными считаются структурные, фациальные, гидрогеологические, геохимические и гидрохимические группы признаков. Внутри геохимической группы особую ценность имеют соотношения окисных и закисных форм железа (продуктивным толщам присуща меньшая окисленность по сравнению с непродуктивными), наличие аутигенных форм железа (указывает на связь комплекса аутигенных минералов с вертикальной удаленностью пород от газоносного пласта) и др.

Следуя традиционным методам, геолог при решении задач поисково-разведочного характера в первую очередь обращает внимание на те признаки, которые чаще всего сопутствуют повышенным концентрациям полезного компонента. Совокупность их становится поисковым критерием, и вопрос об их существенности решается однозначно. В нуль-единичном обозначении информативность **н а л и ч и я** такого признака можно приблизительно оценить суммой единиц, а информативность его **о т с у т с т в и я** — количеством нулей. Это означает, что регистрация отсутствия свойства также важна, как и регистрация его наличия. Так как при математической обработке на первое место выступает структура числовых таблиц, то информативность не должна обязательно и полностью совпадать со значимостью признаков в смысле геологическом. Причем в одном и том же признаковом пространстве информативность признаков может меняться в зависимости от количества объектов (длины столбца). Другой причиной изменения информативности может быть изменение размера признако-

вого пространства и описание в пределах одного и того же признакового пространства геологических объектов (месторождений) из различных регионов. Изменение информативности признаков при переходе от одного региона к другому показано в работе [15] , в которой приводится сравнительная прогнозная оценка касситерит-сульфидных месторождений Тихоокеанского пояса и Приморья. Для месторождений первого региона установлен следующий порядок по **формальной** значимости признаков: факторы магматического контроля, морфология рудных тел, локальные рудовмещающие структуры и околорудные изменения, вмещающие породы, региональные геологические структуры. Для приморских месторождений на второе место перемещаются околорудные изменения, а на предпоследнее - морфология рудных тел.

Соотношение и интерпретация понятий "существенное" и "несущественное" в проблеме распознавания образов является основополагающим для определения формальной модели объекта. В этой связи совершенно справедливым является утверждение: " В самом общем случае ни про один признак нельзя сказать, что он всегда существенный или всегда несущественный. Этот вопрос может быть решен только после предварительного задания определенной цели описания... Один и тот же признак того же объекта может оказаться существенным в одной группе задач и несущественным - в другом"[16, стр.137] . Следовательно, выбираемые математические процедуры должны обеспечивать получение таких формальных оценок, которые при решении конкретной задачи в идеале своем приближались бы к оценкам геологической значимости, поскольку критерием применимости существующих алгоритмов распознавания остается практика. Рациональный выбор цели, правильная постановка задачи, использование доброкачественного и представительного материала и выбор соответствующей процедуры его обработки должны обеспечивать согласованность информативности признаков с их геологической значимостью.

Минимизация пространства признаков

Современные методы и техника измерений позволяют получить на геологических объектах до нескольких сотен параметров, однако, для целей распознавания удастся использовать далеко не весь перечень

признаков. При этом для первоначального рассмотрения и оценки выбирается максимально возможное количество признаков, которое по мнению специалистов-геологов прямо или косвенно оказывает влияние на обеспечение цели. В этой связи возникает необходимость в сокращении списка признаков и приходится решать вопросы о необходимом и достаточном числе признаков, а также о первоочередности их привлечения в новое пространство. Основанием для минимизации признакового пространства по содержательным и логическим соображениям служат представления о емкости, дробности, масштабности, детальности и взаимосвязи признаков. С этих позиций представляется очевидным, что "наиболее оптимальной должна быть система, сочетающая масштабированность признаков с учетом связей между ними" [17, стр. 10].

Естественное желание всесторонне охарактеризовать геологические объекты приводит в начальный период подготовки информации к составлению списков, в которых насчитывается до 200-300 признаков. Однако большинство из них исключается в дальнейшем по разным, главным образом, содержательным соображениям. Первым и очень важным шагом в ограничении признакового пространства является четкая и строгая формулировка целевого задания, что позволяет сузить круг описываемых объектов и довести количество используемых признаков до определенных границ. Дальнейший процесс уменьшения числа признаков идет через решение вопроса, насколько полно и глубоко должны быть охарактеризованы объекты для достижения указанной цели, причем в обработку включаются лишь те признаки, значения которых известны и имеют реальный смысл для всех рассматриваемых объектов. К этой же стадии минимизации относится замена двух признаков разностью или отношением. Этап профессиональной (доалгоритмической) подготовки материала к тематической обработке заканчивается, по существу, после составления общей таблицы, в которой сведения об объектах отражены с помощью двузначного или многозначного кодов.

Сокращение списка исходных признаков эквивалентно проектированию точек исходного пространства R_p в некоторые подпространства $R_{p'} \subset R_p$, где p' - количество набранных признаков [18]. А минимизация - это поиск такой подсистемы признаков, в пределах которой распознавание было бы наилучшим. При решении этого вопроса как в вероятностно-статистической, так и в детерминистской поста-

новках обязательно должны учитываться информационные оценки признаков (например, оценка разделяющей способности). Если независимые признаки ранжированы по величине информационной оценки, то искомым совокупностям можно, по-видимому, ограничить тем признаком, прибавление которого (а также всех последующих) не улучшает распознавание.

К сказанному следует добавить, что необходимость в минимизации пространства признаков довольно часто возникает по чисто техническим причинам, т.к. в основе многих из широко распространенных алгоритмов распознавания лежат трудоемкие процедуры, базирующиеся на логическом переборе признаков или их сочетаний.

Рассмотрение вопросов, связанных с предварительной обработкой исходных данных, показывает, что исходную информацию (как правило, разнохарактерную по своей природе) приходится неоднократно перерабатывать, систематизировать, а подчас и огрублять. Степень огрубления данных находится в сложной зависимости от вида и детальности логических формулировок (для качественных признаков), числа градаций (для порядковых и точечных признаков) и точности измерения (для количественных признаков). Примененный способ выявления основных логических типов признаков и выявление на этой основе внутренней структуры информационных таблиц позволяет судить о пригодности их для классификационных целей. При этом количественная оценка способности признаков обеспечивать цель оказывается возможной только в процессе математической обработки информации.

Л и т е р а т у р а

1. ДМИТРИЕВ А.Н., ЖУРАВЛЕВ Ю.И., КРЕНДЕЛЕВ Ф.П. Об одном принципе классификации и прогноза геологических объектов и явлений. — "Геология и геофизика", 1968, № 5, с.50—64.

2. ГЕОЛОГИЯ И МАТЕМАТИКА. (Задачи прогноза и распознавания в геологии, геохимии и геофизике). Новосибирск, "Наука", 1970.

224 с.

3. ГОЛЬДИН С.В. Одновременное распознавание образов в связи с задачей корреляции сейсмических волн. - В кн.: Второе Сибирское совещание по применению математических методов и ЭВМ в геологии и геофизике. Новосибирск, 1967, с.97-98.

4. ПРОХОРОВ В.Г. Опыт применения программы "Энтропия-I" для определения генетической принадлежности пирита. - В кн.: Второе Сибирское совещание по применению математических методов и ЭВМ в геологии и геофизике. Новосибирск, 1967, с.99-100.

5. КУРЦЕРАЙТЕ Ш.Д., ШЕСТАКОВ Ю.Г. Опыт количественной оценки перспектив магнетитового оруденения при крупномасштабном прогнозировании. - В кн.: Применения математических методов при геологических исследованиях в средней Сибири. Красноярск, 1973, с.216-221. (Материалы по геологии и полезным ископаемым Красноярского края, вып.9).

6. АЛЕШИН С.М. Выделение перспективных площадей для поисков железорудных месторождений Ангаро-Илимского и Ангаро-Катского типов. - В кн.: Тезисы докладов Всесоюзной конференции "Применение математических методов и ЭВМ при поиске полезных ископаемых". Новосибирск, 1973, с.127-129.

7. Математический метод палеотектонического анализа платформенных структур. Иркутск, 1969, 157 с. (Тр. ВЦ Иркутского гос. университета, вып. 3). - Авт.: ЛОХМАТОВ Г.И., ЕВДОКИМОВА В.Н., АЛАЕВ Г.Т., КИРГАНОВА Н.Н., РИВКИНД О.Л.

8. ВОЛКОВ А.М. Использование алгоритмов распознавания образов для разделения поднятий на газоносные и водоносные (на примере Березовского района Западной Сибири). - В кн.: Математические методы при геологических исследованиях в Западной Сибири. - В кн.: 1968, с.33-56. (Тр. ЗапСибНИГНИ, вып.18).

9. КОНДАКОВ Н.И. Логический словарь. М., "Наука", 1971. 655с.

10. СУППЕС П., ЗИНЕС Дж. Основы теории измерений. - В кн.: Психологические измерения. М., "Мир", 1967, с.9-110.

11. ЗАГОРУЙКО Н.Г. Методы распознавания и их применение. М., "Советское радио", 1972. 192 с.

12. ИВАНОВ Ю.Г., КОНОГОРОВ Г.С. Количественная оценка информативности поисковых признаков на экзогенное оруденение. - "Зов геология", № 2, 1971, с.115-125.

13. ШЕСТАКОВ Ю.Г., ВОЛОДИНА Э.Н., РОМАНОВА В.П. К вопросу кодирования числовой информации при распознавании образов. - В кн.: Второе Сибирское совещание по применению математических методов и ЭВМ в геологии и геофизике. Новосибирск, 1976, с.102-103.

14. Организация и обработка геологической информации с помощью ЭВМ на основе построения тупиковых тестов. - В кн.: Логико-информационные решения геологических задач. М., "Наука", 1975, с.83-128. - Авт.: ДМИТРИЕВ А.Н., КРЕНДЕЛЕВ Ф.П., БИШАЕВ А.А. и др.

15. КОНСТАНТИНОВ Р.М., ДМИТРИЕВ А.Н. Использование математических методов анализа геологических факторов, влияющих на масштабы оруденения (на примере месторождений касситерит-сульфидной формации). - "Геология рудных месторождений", 1970, № 2, с.56-64.

16. НЮБЕРГ Н.Д. О познавательных возможностях моделирования. - В кн.: Математическое моделирование жизненных процессов. М., "Мысль", 1968, с.136-151.

17. САДИКОВ М.А. Понятие "исключительности" при прогнозе месторождений. - В кн.: Тезисы докладов Всесоюзной конференции "Применение математических методов и ЭВМ при поиске полезных ископаемых". Новосибирск, 1973, с.9-12.

18. ГОЛЬДИН С.В., ПОПЛАВСКИЙ Н.Н. Построение нелинейной разделяющей поверхности при распознавании пересекающихся образов. - В кн.: Математические методы при геологических исследованиях в Западной Сибири. Тюмень, 1968, с.89-95. (Тр. ЗапСибНИГНИ, вып.18).

В.В.Бабич, Г.С.Федосеев

МЕТОД ЦЕЛЕВОГО КЛАССИФИЦИРОВАНИЯ И УПОРЯДОЧЕНИЯ ОБЪЕКТОВ ("КАСКАД-1")

Предлагаемый метод реализует один из подходов к распознаванию образов, основанный на процедуре "обучения с учителем" и позволяет частично формализовать и алгоритмизировать некоторые операции общей схемы распознавания (рис.1). Вводимая при бинарном описании объектов комплексная оценка информативности признаков содержащая количественную (частота встречаемости одного из эле-

ментов алфавита) и качественную (учет значений целевого признака) характеристики, позволяет осуществлять процедуры классифицирования и упорядочения объектов.

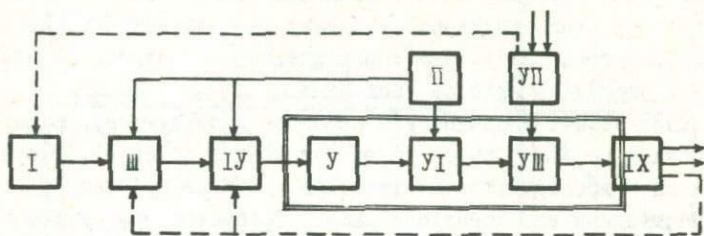


Рис. I. Общая схема распознавания.

I - исходная информация о множестве объектов; П - целевое задание; Ш - выбор эталонных объектов в соответствии с целевым заданием; IУ - классификация эталонных объектов (формирование классов); У - формирование "образов" выделенных классов; УI - выбор решающих правил; УП - информация о распознаваемых объектах (пробах); УШ - процедуры распознавания; IX - принятие решений. Пунктирными линиями показано использование получаемой информации в последующих решениях. Двойной линией ограничены блоки, реализуемые "Каскадом-I".

Исходная информация об объектах обучения (эталонах) и распознаваемых объектах (пробах), а также классификация эталонных объектов заранее задаются, а формирование "образов" (выявление "идеальных" объектов), процедуры распознавания и принятия решения производятся по определенным алгоритмам, выбор которых осуществляется с учетом видовой классификации прогнозно-поисковых задач [I]. Алгоритмы, вошедшие в "Каскад-I", строились таким образом, чтобы получаемые в процессе решения величины поддавались содержательной интерпретации.

При группировании объектов в классы по одному и тому же целевому свойству, многие геологические задачи прогнозного характера можно свести к трем видам: А - задачи классифицирования, Б - задачи упорядочения и С - задачи классифицирования с одновременным упорядочением, которые являются разновидностями более общей задачи распознавания.

Задача классифицирования формулируется следующим образом. Задано m объектов, охарактеризованных в пространстве n признаков

и разбитых по целевому признаку \mathcal{X}_{n+1} , имеющему порядковую шкалу, на t классов, образующих множество $K : K = \{K_1, K_2, \dots, K_t\}$, $K_i = \{S_{i_1}^i, \dots, S_{m_i}^i\}$, $i = 1, \dots, t$, и $\sum_{i=1}^t m_i = m$. Кроме того, имеется множество объектов $P = \{S_1, \dots, S_p\}$, охарактеризованных в пространстве \mathcal{L} признаков, для которых принадлежность к одному из заданных классов неизвестна. Требуется определить принадлежность каждого из этих объектов к одному из классов множества K .

Смысл задачи упорядочения сводится к следующему. Задан упорядоченный по \mathcal{X}_{n+1} набор из m эталонных объектов, охарактеризованных в пространстве \mathcal{L} признаков. Целевой признак \mathcal{X}_{n+1} имеет количественный или порядковый вид. Требуется для каждой пробы из P количественно оценить значение \mathcal{X}_{n+1} и найти тем самым ее место в упорядоченном наборе эталонов.

Формулировка задачи третьего вида такова. Задано m объектов, охарактеризованных в пространстве \mathcal{L} признаков и разбитых на t классов. При этом для каждого объекта из z первых классов ($1 \leq z \leq t-1$) **указана упорядоченность по значению \mathcal{X}_{n+1} , измеренного в количественной или порядковой шкале внутри класса.** Для оставшихся $t - z$ классов целевой признак имеет порядковый вид. Каждую пробу из набора P требуется отнести к одному из t классов, и если она отнеслась к одному из z классов, необходимо найти ее место в классе по значению \mathcal{X}_{n+1} . В случае отнесения пробы к одному из оставшихся ($t - z$) классов, определять ее место в классе не требуется.

Так как в каждом виде содержится две или более постановок (в зависимости от числа классов), то классификацию можно представить в виде двумерной схемы (рис.2). Схема показывает, что в задачах первого вида имеет место ступенчатое изменение значений целевого признака, т.е. для эталонных объектов должна быть указана их позиция с точностью до класса, а не реальные значения целевого признака. В задачах второго вида должно быть указано конкретное значение целевого признака или значение в шкале порядка. В задачах третьего вида реальные или порядковые значения целевого признака указываются лишь для объектов заранее обусловленных классов, для оставшихся объектов указывается только их принадлежность к классам.

Решение перечисленных видов задач, начинающееся с определения их вида, соответствует единой блок-схеме (рис.3). При этом


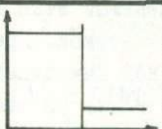
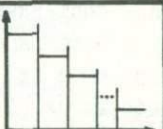
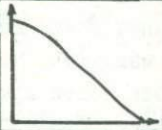
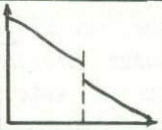
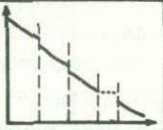
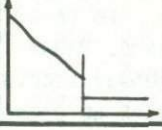
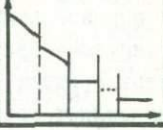
Вид задачи	Количество классов в обучении			
	1	2	...	N
A			
B			
C			



Рис. 2. Видовая классификация задач.

A - задачи классифицирования, B - задачи упорядочения, C - задачи классифицирования с одновременным упорядочением; 1 - порядковый целевой признак, 2 - количественный целевой признак, 3 - комбинированный целевой признак. Ось ординат - значения целевого признака, ось абсцисс - объекты. Условное разбиение на классы показано пунктирными линиями.

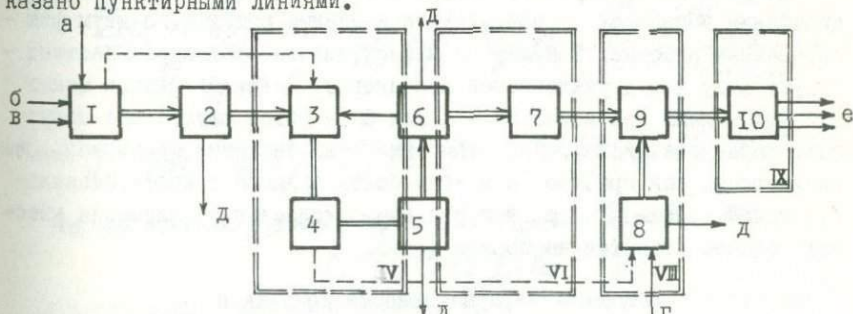


Рис. 3. Блок-схема решения прикладных задач.

1 - определение вида задачи, 2 - предварительная обработка информации, 3 - оценка информативности признаков и выбор оптимального варианта бинарного кодирования, 4 - кодирование информации, 5 - минимизация признакового пространства, 6 - минимизация набора эталонных объектов, 7 - оценка качества распознавания, 8 - отбраковка проб, 9 - распознавание проб, 10 - принятие решения. Внешние связи: а - целевое задание, б - информация об эталонах, в - таксономия эталонов, г - информация о пробах, д - отбракованная информация, е - результаты. Внутренние связи: основные (двойные линии), дополнительные (одинарные линии), вспомогательные (одинарные пунктирные линии); У, VI, VII, VIII, IX - блоки общей схемы распознавания (см. рис. 1).

необходимо иметь четкую формулировку целевого задания (с соответствующей информацией об эталонах) и классификацию объектов. Блоки I и 2 составляют так называемый этап доалгоритмической обработки, а остальные – алгоритмической. Предлагаемый метод может быть применен для обработки как качественной, так и количественной информации. В настоящей статье рассмотрение основных процедур метода ограничено лишь бинарным вариантом решения видовых задач в двухклассовой постановке, что позволяет наиболее полно раскрыть его содержательные послышки. Следует заметить, что "Каскад-I" базируется на семействе оценок информативности признаков [2,3], которые вычисляются путем попарного сравнения объектов между собой и оценкой каждого выявленного различия разностью соответствующих значений целевого признака. Различия в способах сравнения объектов придают оценкам информативности то или иное содержание, что позволяет привлекать каждую из оценок к решению лишь определенного вида задачи. Вид задачи является решающим и при выборе критериев минимизации признаков пространств.

Решение задач классифицирования

Описываемый ниже набор алгоритмов для решения задач классифицирования позволяет распределять испытуемые объекты по двум фиксированным классам. В принятой классификации эта задача соответствует виду А в двухклассовой постановке. Целевой признак может принимать лишь два значения – I или 0 (в общем случае это могут быть любые два числа). При этом единичные значения признака I_{n+1} приписываются, как правило, объектам более важного с профессиональной точки зрения класса, который при фиксировании положения классов обычно ставится на первое место.

Оценка информативности признаков

Количественная оценка индивидуальной значимости признака при разделении заданных классов в задаче классифицирования производится при помощи процедуры попарного сравнения объектов первого класса с объектами второго. Существенность выявленных различий в описании объектов разных классов по j -му признаку ($1 \leq j \leq n$) оценивается разностью соответствующих значений целевого признака по следующей схеме:

I	2	3
I	0	+I
0	I	-I
I	I	0
0	0	0

, где I - значение j-го признака у объектов I класса;
 2 - значение j-го признака у объектов II класса;
 3 - существенность различия.

Оценка информативности признака получается путем алгебраического суммирования результатов всех попарных сравнений:

$$\tilde{P}_j^v = \sum_{i=1}^{m_1} \sum_{i'=1}^{m_2} (t_{ij} - t'_{ij})(f_i - f_{i'})$$

где m_1 - число объектов в I классе; m_2 - число объектов во II классе; t_{ij} и t'_{ij} - значения j-го признака у i-го и i'-го объектов; f_i и $f_{i'}$ - значения целевого признака у объектов I и II классов.

Приведем данное выражение к виду, более удобному для содержательной интерпретации, введя нормировку на величину $m_1 \cdot m_2$:

$$\begin{aligned} P_j^v &= \frac{\tilde{P}_j^v}{m_1 \cdot m_2} = \frac{\sum_{i=1}^{m_1} \sum_{i'=1}^{m_2} (t_{ij} - t'_{ij})(f_i - f_{i'})}{m_1 \cdot m_2} = \frac{\sum_{i=1}^{m_1} \sum_{i'=1}^{m_2} (t_{ij} \cdot f_i - t'_{ij} \cdot f_{i'} - t_{ij} \cdot f_{i'} + t'_{ij} \cdot f_i)}{m_1 \cdot m_2} = \\ &= \frac{m_2 \cdot \sum_{i=1}^{m_1} t_{ij} \cdot f_i}{m_1 \cdot m_2} - \frac{\sum_{i=1}^{m_1} f_i}{m_1} \cdot \frac{\sum_{i'=1}^{m_2} t'_{ij}}{m_2} - \frac{\sum_{i=1}^{m_1} t_{ij}}{m_1} \cdot \frac{\sum_{i'=1}^{m_2} f_{i'}}{m_2} + \frac{m_1 \cdot \sum_{i'=1}^{m_2} t'_{ij} \cdot f_{i'}}{m_1 \cdot m_2} = \\ &= \frac{\sum_{i=1}^{m_1} t_{ij} \cdot f_i \cdot \sum_{i'=1}^{m_2} t'_{ij} \cdot m_1}{\sum_{i=1}^{m_1} t_{ij} \cdot m_1^2} - f_{cp}^I \cdot P_j^{\bar{u}} - P_j^I \cdot f_{cp}^{\bar{u}} + \frac{\sum_{i'=1}^{m_2} t'_{ij} \cdot f_{i'} \cdot \sum_{i=1}^{m_1} t_{ij} \cdot m_2}{\sum_{i'=1}^{m_2} t'_{ij} \cdot m_2^2} = \\ &= P_j^I \cdot f_{cp1}^I - f_{cp}^I \cdot P_j^{\bar{u}} - P_j^I \cdot f_{cp}^{\bar{u}} + f_{cp1}^{\bar{u}} \cdot P_j^{\bar{u}} = P_j^I (f_{cp1}^I - f_{cp}^{\bar{u}}) - P_j^{\bar{u}} (f_{cp}^I - f_{cp1}^{\bar{u}}) \end{aligned}$$

Так как в рассматриваемом случае $f_{cp1}^I = f_{cp}^I$ и $f_{cp1}^{\bar{u}} = f_{cp}^{\bar{u}}$, то

$$P_j^v = (P_j^I - P_j^{\bar{u}})(f_{cp}^I - f_{cp}^{\bar{u}})$$

В приведенных выше вычислениях: P_j^I и $P_j^{\bar{u}}$ - частоты встречаемости символа "I" в j-ом признаке, соответственно, в I и II классах; f_{cp}^I и $f_{cp}^{\bar{u}}$ - средние значения целевого признака для объектов I и II классов; f_{cp1}^I и $f_{cp1}^{\bar{u}}$ - средние значения целевого признака для объектов I и II классов, вычисленные лишь с учетом объектов, имеющих в j-ом признаке символ "1".

Сопоставление полученной величины с некоторыми близкими по смыслу оценками, приведенными в [4] показывает, что P_j^v является мерой связи бинарного характеристического признака с бинарным целевым и связана, например, с коэффициентом парной корреляции z следующим соотношением:

$$P_j^v = z / \sqrt{\frac{(A)(\alpha)}{(B)(\beta)}}$$

где (A) и (α) — соответственно числа "совпадающих" и "различающихся" значений по целевому и характеристическому признакам в обоих классах, (B) и (β) — соответственно числа выполнений и невыполнений объектами обоих классов j -го признака.

Значения оценки информативности признаков меняются в пределах $+1 \geq P_j^v \geq -1$. Абсолютная ее величина отражает межклассовую разделяющую способность признаков. Максимальное значение абсолютной величины P_j^v , равное единице, получают признаки, совпадающие по структуре с целевым, или зеркальные по отношению к нему (знак оценки P_j^v указывает на более "важное", т.е. чаще встречающееся в объектах первого класса, значение признака). Нулевую оценку получают признаки, состоящие из одних нулей или единиц, или же признаки с $\mathcal{G}_j^I = \mathcal{G}_j^K$. При этом на информативность признака не влияет конкретное распределение значений признака внутри классов. Это свойство P_j^v соответствует содержанию задачи, поскольку различать объекты внутри классов не требуется.

Отметим еще одно свойство P_j^v . Оценки информативности P_j^v для столбцов вида $(t_1, t_2, \dots, t_{m_1}, t'_1, t'_2, \dots, t'_{m_2})$ и $(\bar{t}_1, \bar{t}_2, \dots, \bar{t}_{m_1}, \bar{t}'_1, \bar{t}'_2, \dots, \bar{t}'_{m_2})$, где "-" — знак логического отрицания, а "'" — знак транспонирования, равны по абсолютной величине и отличаются лишь знаком, т.е. замена в любом признаке всех значений на противоположные не влечет за собой изменения его межклассовой разделяющей способности, а приводит к смене более "важного" значения кода на противоположное. Таблицы, составленные из столбцов с $P_j^v \geq 0$ в дальнейшем называются таблицами с переориентированным кодом.

Кодирование информации

Описываемые ниже две процедуры алгоритмического кодирования разработаны с учетом вида решаемой задачи и типов признаков^{*)}. Они получили условные названия способов "скользящей границы" и "перебора".

Способом "скользящей границы" возможные варианты кодирования находятся путем последовательного перемещения границы (А), разделяющей упорядоченный по возрастанию (или убыванию) ряд значений признака на две подсовокупности. При каждом новом положении границы всем значениям, находящимся по одну сторону, присваивается символ "1", а находящимся по другую — "0". Таким образом, всех возможных вариантов кодирования (V_s) будет: $V_s = \nu - 1$, где ν — число различающихся значений признака. Например, для признака j при фиксированном ряде значений 3-4-6-7-8 возникают разные варианты (I), а применение каждого варианта кодирования к исходному j -му признаку приводит к различным результатам (II):

ряд значений	варианты кодирования			
	I	II	III	IV
3	I	I	I	I
4	O	I	I	I
6	O	O	I	I
7	O	O	O	I
8	O	O	O	O

признак	вид признака			
	j_I	j_{II}	j_{III}	j_{IV}
8	O	O	O	O
6	O	O	I	I
4	O	I	I	I
7	O	O	O	I
3	I	I	I	I
4	O	I	I	I

Способ "перебора" не требует фиксирования значений признака в определенный ряд. Различные варианты кодирования признака получаются при последовательном присвоении одного из бинарных символов каждому из возможных сочетаний по $1, 2, \dots, \nu - 1$. При таком методе кодирования общее число вариантов равно:

$$V_p = C_{\nu}^1 + C_{\nu}^2 + \dots + C_{\nu}^{\nu-2} + C_{\nu}^{\nu-1} = 2^{\nu} - 2$$

^{*)} См. статью Г.С.Федосеева в настоящем сборнике.

Легко проверить, что группа вариантов C_5^1 и C_5^{5-1} , C_5^2 и C_5^{5-2} и т.д. обладают свойством инверсионности, т.е. отличаются друг от друга тем, что все символы "I" заменяются символами "O" и наоборот. Подобная инверсия признака не влияет на конечные результаты решения, поэтому фактических вариантов кодирования будет 2^{5-1} . Для приведенного выше j -го признака ($\nu=5$) имеется 15 возможных вариантов кодирования (I), применение которых меняет соответствующим образом структуру признака (II):

Ряд значений	Варианты кодирования														
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
6	I	O	O	O	O	I	I	I	I	O	O	O	O	O	O
7	O	I	O	O	O	I	O	O	O	I	I	I	O	O	O
4	O	O	I	O	O	O	I	O	O	I	O	O	I	I	O
3	O	O	O	I	O	O	O	I	O	O	O	I	I	O	I
8	O	O	O	O	I	O	O	O	I	O	I	O	I	O	I
C_5^1					C_5^2										

Признак	j_1	j_2	j_3	j_4	j_5	j_6	j_7	j_8	j_9	j_{10}	j_{11}	j_{12}	j_{13}	j_{14}	j_{15}
	8	O	O	O	O	I	O	O	O	I	O	O	I	O	I
6	I	O	O	O	O	I	I	I	I	O	O	O	O	O	O
4	O	O	I	O	O	O	I	O	O	I	O	O	I	I	O
7	O	I	O	O	O	I	O	O	O	I	I	I	O	O	O
3	O	O	O	I	O	O	O	I	O	O	I	O	I	O	I
4	O	O	I	O	O	O	I	O	O	I	O	O	I	I	O

Аналогичные результаты можно получить и способом "скользящей границы", если последовательно применять его для рядов значений данного признака, включающих все возможные перестановки.

С позиций описанных способов кодирования качественный тип признаков является наиболее простым для этой процедуры. В этом случае оба способа дают однозначный результат, т.к. возможен только один вариант кодирования, при котором одно из значений признака обозначается через "O", а другое - через "I". Кодирование качественных признаков в пробах также не вызывает затруднений:

каждое значение признака у проб заменяется тем символом, которым оно обозначено в эталонах.

Количественные и порядковые признаки следует кодировать первым способом, поскольку значения этих признаков всегда можно расположить в ряд, отражающий увеличение (или уменьшение) степени проявленности свойства на объектах. При этом, где бы не прошла граница А в принятой последовательности значений признака у эталонов, любое значение в пробе, если даже оно не встречалось у эталонов, всегда определится относительно границы кодирования. Исключение составляет случай, когда значение признака у пробы попадает в область границы кодирования. В такой ситуации необходимо найти среднее арифметическое между ближайшими к границе значениями и полученную величину считать за границу А.

При кодировании точечных признаков способ "скользящей границы" не применим, поскольку градации этих признаков не отражают степень проявленности свойства на объектах и направленной фиксированной последовательности значений составить невозможно. В этой связи для кодирования точечных признаков используется способ "перебора". При этом, если у пробы признак имеет значение, встречающееся у объектов обучения, вопрос об обозначении его одним из символов в рамках принятого кода решается однозначно. В противном случае следует объединить "аномальное" значение с каким-либо значением данного признака у эталонов. Если же такое объединение представляется невозможным, то следует распознавать пробу в признаковом пространстве без данного признака, или отказа от ее распознавания.

Вопрос о целесообразности того или иного варианта при зафиксированном способе кодирования решается при вычислении для каждого из них оценки P_j^v . Оптимальным считается тот вариант, при котором признак получает максимальную (по модулю) оценку.

Минимизация признакового пространства

Введем количественную характеристику каждого объекта из эталонной выборки, вычисляемую в таблицах с переориентированным кодом по формуле:

$$J [P_j^v]_{s_i} = \sum_{j=1}^n t_{ij} \cdot P_j^v \quad i = 1, 2, \dots, m_1; j = 1, 2, \dots, m_2$$

В дальнейшем будем называть величину $J[P_j^v]_{s_i}$ строчечной нагрузкой. Сопоставим каждому набору признаков $X = \{x_{j_1}, \dots, x_{j_\ell}\}$, $1 \leq \ell \leq n$, число $\rho^\ell(X) = \rho^\ell$, которое служит критерием минимизации и представляет собой разность между минимальной строчечной нагрузкой среди объектов I класса и максимальной - среди объектов II класса, вычисленных по набору признаков X : $\rho^\ell = J[P_j^v]_{\min}^I - J[P_j^v]_{\max}^{II}$. При заданном способе порождения наборов X минимизированным является такой набор X из ℓ признаков, при котором оценка ρ^ℓ достигает максимальной величины. В частном случае этому требованию может удовлетворять исходное признаковое пространство.

М и н и м и з а ц и я по принципу "упорядочения признаков по информационным весам". Признаки располагаются в порядке убывания абсолютной величины их информационных весов, и в пространстве каждого поднабора $\{N_\ell\}$ (при $\ell = 1, 2, \dots, n$) подсчитываются строчечные нагрузки $J[P_i^v]$ для каждого объекта обучения. Выбирается тот поднабор признаков, который обеспечивает наилучшее разделение классов (ρ^ℓ - максимальна).

М и н и м и з а ц и я по принципу "взаимодополняемости признаков". Из полного пространства n удаляется первый признак и вычисляется величина ρ_1^{n-1} . Затем этот признак возвращается в признаковое пространство и удаляется следующий, вычисляется величина ρ_2^{n-1} и т.д. Эта процедура продолжается до тех пор, пока не будет удален, а затем возвращен n -ый признак (ρ_n^{n-1}). При сравнении всех ρ_j^{n-1} с величиной межклассовой растяжки на полном признаковом пространстве ρ^n может иметь место несколько вариантов (табл. I).

Вариант	Соотношение величин межклассовой растяжки		
	$\rho_j^{n-1} > \rho^n$	$\rho_j^{n-1} = \rho^n$	$\rho_j^{n-1} < \rho^n$
1	+	+	+
2	+	+	-
3	+	-	+
4	+	-	-
5	-	+	+
6	-	+	-
7	-	-	+

Таблица I
Варианты при минимизации пространства признаков

Примечание: "+" - наличие ситуации с указанным соотношением; "-" - отсутствие ситуации с указанным соотношением.

ж) Близкий в процедурном отношении алгоритм описан в работе [5].

При минимизации в первых четырех случаях из признакового пространства исключается признак, при удалении которого максимально улучшилось разделение классов. Если таких признаков несколько, удаляется любой из них. В пятом и шестом случаях удаляется любой из признаков, при выбрасывании которых $\rho_j^{n-1} = \rho^n$. Наконец, в седьмом случае из пространства исключается признак, при удалении которого разделение классов ухудшается в наименьшей степени. Если таких признаков несколько, удаляется любой из них. В результате исходное признаковое пространство, состоящее из n признаков, приводится к пространству размерности $n-1$. С полученным подпространством производятся аналогичные преобразования и его размерность последовательно понижается до $n-2, n-3, \dots, n-(n-1)$. В качестве минимизированного выбирается поднабор признаков, обладающий максимальной величиной ρ^e .

М и н и м и з а ц и я по принципу "полного перебора признаков". Там, где позволяет объем информации, минимизация признакового пространства по указанному критерию может быть произведена путем полного перебора всех поднаборов признаков.

Анализ эталонной выборки объектов

Выше отмечалось, что знак P_j^v определяется процедурой подсчета информативности признаков: если $\mathcal{G}_j^I > \mathcal{G}_j^II$, то оценка имеет положительный знак и наоборот. С учетом знака для эталонных классов можно составить обобщающие строки, которые задают две полярные точки n -мерного гиперкуба. Данные строки являются описаниями специфических объектов, которые в дальнейшем называются "идеальными" и обозначаются соответственно через R_I и R_{II} : $R_I = (t_1, \dots, t_n)$, $R_{II} = (t'_1, \dots, t'_n)$, $j = 1, \dots, n$; $t_j = 1, t'_j = 0$ при $P_j^v > 0$ и $t_j = 0, t'_j = 1$ при $P_j^v < 0$.

При переориентации кода у признаков с $P_j^v < 0$ описание "идеального" объекта R_I приобретает вид единичной строки, а описание R_{II} — нулевой, т.е. R_{II} помещается в начало координат, а R_I — в наиболее удаленную от нее вершину гиперкуба. Естественно, что процедура переориентации кода у части признаков не нарушает первоначальной полярности данных точек. "Идеальные" объекты являются экстремальными в информативном смысле: $J[P_j^v]_{R_I} = \sum_{j=1}^n P_j^v$ и $J[P_j^v]_{R_{II}} = 0$,

а строчечная нагрузка реальных объектов изменяется в диапазоне $J[p_j^v]_{R_I} \geq J[p_j^v]_{s_i} \geq J[p_j^v]_{R_{II}}$ и служит оценкой их близости к "идеальным" объектам. Следовательно, по положению точек $J[p_j^v]_{s_i}$ на отрезке $R_I R_{II}$ можно судить о соотношении эталонов в классах и степени различия между классами ("контрастности"). Граница K , проходящая через точку $J[p_j^v]_{R_I}/2$ на прямой $R_I R_{II}$ является границей равного сходства эталонных объектов с "идеальными" объектами обоих классов, соотношение которых можно свести к четырем случаям (рис.4).

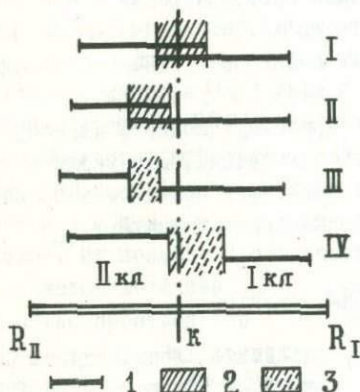


Рис. 4. Расположение эталонных классов относительно границы равного сходства (K) с "идеальными" объектами. I - диапазоны изменения величины $J[p_j^v]$ в эталонных классах; 2 - область пересечения классов ($\rho^e < 0$); 3 - величина растяжки классов ($\rho^e > 0$).

В первом случае в обоих классах присутствуют объекты, обладающие большим сходством с "идеальными" объектами противоположного класса, в силу чего классы пересекаются. Во втором случае лишь в одном классе присутствуют подобные объекты, поэтому классы также пересекаются. Третий случай отличается от предыдущего тем, что классы разделились, хотя один из них заходит за границу равного сходства. Наконец, последний случай в наибольшей степени отвечает требованию "контрастности".

Анализ, проведенный по описанной схеме для эталонных классов на минимизированном признаковом пространстве, позволяет выявить объекты, понижающие "контрастность", которые, если они являются причиной пересечения классов, могут быть удалены из материала

обучения. Если по указанной причине один или несколько объектов удалены из эталонной выборки, оставшаяся информация должна быть вновь приведена к бинарному виду и минимизирована, после чего повторно проводится анализ материала обучения. Этот процедурный цикл повторяется до тех пор, пока исходные классы не разделятся, или же область пересечения классов не достигнет некоторого заранее установленного предела. Естественно, что удаление объектов из обучающей выборки приводит к необходимости вводить некоторые ограничения к допуску проб на распознавание (см. "Отбраковка проб").

Внешний экзамен

О надежности классифицирования проб обычно судят по соотношению правильных и неправильных ответов при классифицировании объектов с заранее известной принадлежностью к заданным классам (объектов экзамена), причем объекты экзамена задаются независимо от материала обучения. В "Каскаде-I" реализован иной способ внешнего экзамена: в качестве экзаменуемых объектов выступают эталонные объекты. Процедура построена таким образом, что каждый из эталонных объектов поочередно удаляется из обучающей выборки и ставится в положение пробы. В оставшейся таблице заново производится бинарное кодирование и оценка информативности признаков. Классифицирование осуществляется в объеме того минимизированного признакового пространства, в котором предполагается классифицировать пробы. Правильное классифицирование оценивается единицей, неправильное - нулем. попадание в зону неопределенности (отказ от классифицирования) - 0,5. Окончательная оценка надежности классифицирования вычисляется по формуле: $B_{кл.} = \sum_{i=1}^m \cdot B_{i,кл.} \cdot 100 / m$. (m - число объектов экзамена).

Отбраковка проб

К процедуре классифицирования, на имеющемся материале обучения могут быть допущены далеко не все пробы, тем более, что алгоритм предусматривает удаление некоторых эталонных объектов. Одним из способов контроля за допуском или отбражкой проб является логический анализ оставшегося материала обучения, основанный

на том, что информационную бинарную таблицу при двухклассовой постановке можно привести к блоковому виду:

Классы	Б Л О К И				
	I II	A	B	C ₁	C ₂

В блок А входят отождествляющие межклассовые признаки, каждый из которых представлен только одним символом бинарного алфавита ("1" или "0") [6]. Совокупность таких признаков представляет из себя так называемый критерий общности – своеобразную меру однородности описания объектов обучения. Специфика данного блока заключается в том, что он позволяет классифицировать пробы на допустимые и недопустимые к распознаванию на данном материале обучения. Строго говоря, все пробы должны полностью удовлетворять критерию общности, однако в практических задачах это требование, как правило, не соблюдается. Математической оценки соотношения признаков данного блока с общим количеством признаков не существует. Определенные ограничения на размер критерия общности накладываются профессионалами, исходя из содержательных соображений, ими же должен решаться вопрос о допуске проб к распознаванию при частичном выполнении критерия общности.

Блок В включает в себя "сильные поисковые" признаки, т.е. такие, которые в первом классе представлены одним символом, а во втором – другим. Его можно назвать блоком внутриклассовых критериев общности. В реальных таблицах этот блок, как правило, отсутствует. Если же признаки этого типа имеются в таблице, причем в достаточно большом количестве, то анализ проб по блоку В позволяет либо сразу сделать заключение об их принадлежности к классам, либо отказаться от распознавания (в случае частичного выполнения). Таким образом, блок В, с одной стороны, выполняет роль дополнительного контроля, а с другой – обладает классифицирующей способностью. Ход рассуждений при анализе проб по данному блоку выражается следующей схемой:

I кл.	II кл.	Принимаемое решение
+	-	Проба принадлежит I классу
-	+	Проба принадлежит II классу
±	±	Отказ от распознавания

Примечание: "+" - полное выполнение пробой одного из внутриклассовых критериев общности; "-" - полное невыполнение пробой одного из внутриклассовых критериев общности; "±" - частичное выполнение пробой внутриклассовых критериев общности.

Блок С (блок "слабых поисковых" признаков) состоит из признаков, имеющих внутри одного класса отождествляющую структуру, а внутри второго - произвольную неоттождествляющую. Он разбивается на два подблока - C_1 , в который входят признаки, отождествляющие объекты первого класса, и C_2 , состоящие из признаков, отождествляющих объекты второго класса. Контролирующая и классифицирующая способность блока С выражена не так отчетливо, как у предыдущих блоков, но анализ проб по "слабым поисковым" признакам позволяет в ряде случаев отказаться от распознавания пробы, либо отнести ее к одному из классов, или указать необходимость привлечения к распознаванию оставшихся признаков. Ситуации, возможные при логическом анализе упорядоченной информационной таблицы по подблокам C_1 и C_2 отражаются схемой (табл.2).

Таблица 2
Схема логического анализа исходной информации

Подблок	Класс	Варианты соотношения описаний проб и эталонов															
		I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
C_1	I	+	+	+	+	-	-	-	-	±	±	±	±	±	±	±	±
	II	±	±	±	±	±	±	±	±	+	±	+	+	±	±	±	±
C_2	I	+	±	±	±	±	±	±	±	±	±	±	±	+	±	±	±
	II	±	+	-	±	+	-	±	±	+	+	-	±	±	±	-	±
π		а	г	а	в	б	г	в	б	в	в	в	в	в	в	в	в

Примечание: "+" - полное совпадение описаний проб и эталона, "-" - полное несовпадение описаний пробы и эталона, "±" - частичное совпадение описаний пробы и эталона. π - принимаемое решение: а-проба принадлежит I классу, б-проба принадлежит II классу, в-привлечение блока Д, г-отказ от распознавания.

Наконец, блок Д представлен характеристическими признаками, чередование нулей и единиц в которых не подчиняется вышеописанным закономерностям. Признаки этого блока привлекаются к совместной с признаками блока В и С обработке в том случае, если предварительный логический анализ упорядоченной информационной таблицы не позволил сделать конкретных выводов о пробе.

Классифицирование проб

Для каждой пробы с учетом полученных оценок информативности признаков вычисляются строчечные нагрузки $\Sigma[P_j^v]$ в объеме минимизированного признакового пространства. Вывод о принадлежности пробы к определенному классу или отказ от классифицирования осуществляются в зависимости от положения пробы на отрезке $R_I R_{II}$. Классифицирование производится с учетом трех возможных вариантов разделения эталонных классов (рис.5).

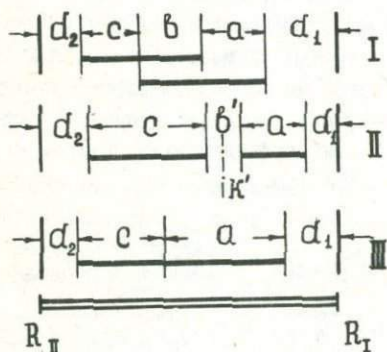


Рис. 5. Основные варианты взаимного расположения эталонных классов. Зоны принадлежности: a - I класс; c - II класс; b, b', d и d_2 - зоны неопределенности; k' - граница равного сходства с крайними объектами реальных классов.

В первом случае классы пересеклись, что обусловило наличие здесь зоны неопределенности $b = |\rho^e|$, при попадании в которую следует отказ от классифицирования пробы. Отнесение к I или II классам производится при попадании пробы соответственно в зоны a и c . Попадание проб в зоны d_1 и d_2 свидетельствует о том, что они обладают большим сходством с "идеальными" объектами классов, чем любой из эталонов, тем не менее эти пробы выходят за рамки материа-

да обучения. Поэтому, при строгом подходе следует, по-видимому, рассматривать зоны d_1 и d_2 как зоны неопределенности, хотя можно говорить о предпочтительной близости пробы к тому или иному классу. Во втором случае зона неопределенности b' обусловлена "растяжкой" классов ($b' = \rho^e$). Проведя границу k' через точку **равного** сходства с крайними объектами противоположных классов, **можно** по положению пробы относительно этой границы судить о ее предпочтительной близости к одному из эталонных классов, хотя при строгом подходе зону b' по аналогии с d_1 и d_2 следует рассматривать как область отказа от распознавания. Последний вариант является частным случаем предыдущих, при котором $b = b' = 0$.

Описанная процедура классифицирования допускает наличие пропусков в распознаваемом объекте. В этом случае для пробы вычисляются две величины строчечной нагрузки: одна - при замене пропусков более "важными" значениями признаков ($\mathcal{J}[P_j^v]_{\max}$), другая - при замене пропусков менее "важными" значениями ($\mathcal{J}[P_j^v]_{\min}$). Интервал значений строчечных нагрузок от $\mathcal{J}[P_j^v]_{\min}$ до $\mathcal{J}[P_j^v]_{\max}$ включает все возможные варианты заполнения пропусков и, в зависимости от расположения этих точек, позволяет делать вывод о принадлежности пробы к одному из классов, или о предпочтительной близости к одному из классов, либо отказаться от распознавания.

Решение задач упорядочения

При решении задач упорядочения нет необходимости учитывать априорное разбиение объектов на классы. Математические процедуры в этом случае направлены на получение максимального соответствия расчетных оценок для объектов (строчечных нагрузок) значениям целевого признака последних. В рамках "Каскада-1" это достигается несколькими путями, различающимися способами подсчета информативности признаков.

Оценки информативности признаков

Информативность признаков вычисляется путем попарного сравнения эталонных объектов с учетом различий в значениях целевого признака [7]. Различия в способах сравнения объектов определяют специфику получаемых оценок, поскольку придадут им различное смыс-

ловое содержание. В целом же все оценки учитывают упорядоченность объектов по целевому признаку и теряют смысл, если упорядочение эталонов не произведено.

Оценка φ_j . Производится сравнение каждого объекта с нижестоящими объектами. По аналогии с оценкой P_j^v каждое выявленное различие оценивается разностью соответствующих им значений целевого признака, после чего результаты сравнения суммируются и делятся на m^2 .

$$\varphi_j = \frac{\sum_{i=1}^m \sum_{i'=1}^m (t_{ij} - t'_{i'j})(f_i - f_{i'})}{m^2}$$

После преобразований получим:

$$\varphi_j = P_j (f_{cp.1} - f_{cp.}),$$

где P_j - частота встречаемости символа "I" в j -ом признаке; $f_{cp.1}$ - среднее значение целевого признака для объектов, выполнивших j -ый признак (с символом "I"); $f_{cp.}$ - среднее значение целевого признака для всех объектов набора; m - число объектов.

Если в обрабатываемой таблице эталонные объекты разбить на два условных класса по границе $f_{cp.}$, то максимальную (по модулю) оценку получают "сильные поисковые" признаки блока В, а минимальную - отождествляющие признаки блока А. Среди признаков, соответствующих по структуре блоку С (в таблице с переориентированным кодом) большую информативность при равном количестве единиц получает тот признак, в котором единицы соответствуют объектам с более высокими значениями целевого признака, а именно признаки, в которых величина $(f_{cp.1} - f_{cp.})$ больше. В признаках блока Д информативность также предопределяется не только количеством единиц, но и их положением внутри условных классов. Таким образом, в отличие от P_j^v оценка информативности φ_j учитывает не только количественное соотношение единиц и нулей в столбце, но и характер их расположения.

Оценка φ_j^* . Первый объект из упорядоченного набора сравнивается со всеми нижестоящими, затем первые два объекта сравниваются со всеми нижестоящими, затем первые три объекта и т.д. вплоть до m объектов. Каждый промежуточный результат нормируется на величину $(m-k)k$, где $k=1, \dots, m$ - число объектов, подлежащих сравнению с нижестоящими. Сумма всех результатов сравнения является окончательной оценкой информативности признаков:

$$\varphi_j^* = \frac{\sum_{k=1}^m \frac{\sum_{i=1}^k \sum_{i'=k}^m (t_{ij} - t_{i'j})(f_i - f_{i'})}{k(m-k)}}{k(m-k)}$$

После преобразования получим:

$$\begin{aligned} \varphi_j^* &= \sum_{k=1}^m (\mathcal{G}_j^{(k)} - \mathcal{G}_j^{(m-k)})(f_{cp}^{(k)} - f_{cp}^{(m-k)}) + \mathcal{G}_j^{(k)}(f_{cp.1}^{(k)} - f_{cp.}^{(k)}) + \mathcal{G}_j^{(m-k)}(f_{cp.1}^{(m-k)} - f_{cp.}^{(m-k)}) = \\ &= \sum_{k=1}^m (P_j^{\vee(k, m-k)} + \varphi_j^{(k)} + \varphi_j^{(m-k)}), \end{aligned}$$

где $P_j^{\vee(k, m-k)}$ - оценка информативности P_j^{\vee} , вычисленная для классов, один из которых представлен k первыми объектами, а другой - всеми нижестоящими; $\varphi_j^{(k)}$ и $\varphi_j^{(m-k)}$ - оценка информативности φ_j , вычисленная, соответственно, для k первых объектов и $(m-k)$ нижестоящих

Как следует из окончательной формулы, оценка φ_j^* включает в себя как оценку P_j^{\vee} , так и φ_j , вычисленные для всех возможных разбиений указанным способом набора объектов на два формальных класса (k и $m-k$). Так же, как и в предыдущем случае, оценка φ_j^* минимальна и равна нулю для отождествляющих столбцов и максимальна по абсолютной величине для признаков из блока В относительно границы f_{cp} . В признаках из блоков С и Д большую оценку (в таблице с переориентированным кодом) получают те признаки, в которых единицы соответствуют объектам с более высокими значениями целевого признака. Следовательно, оценка φ_j^* учитывает количественное соотношение единиц и нулей и особенности их расположения в столбце.

Оценка $\tilde{\varphi}_j$. Вычисление этой оценки производится аналогично φ_j , но результат сравнения каждого i -го объекта со всеми нижестоящими делится на величину $(m-i)$:

$$\tilde{\varphi}_j = \sum_{i=1}^m \frac{\sum_{i'=i}^m (t_{ij} - t_{i'j})(f_i - f_{i'})}{m-i}$$

Преобразовав эту формулу, получаем:

$$\begin{aligned} \tilde{\varphi}_j &= \sum_{i=1}^m (\mathcal{G}_j^{(i)} - \mathcal{G}_j^{(m-i)})(f_{cp}^{(i)} - f_{cp}^{(m-i)}) + \mathcal{G}_j^{(m-i)}(f_{cp.1}^{(m-i)} - f_{cp.}^{(m-i)}) = \\ &= \sum_{i=1}^m (P_j^{\vee(i, m-i)} + \varphi_j^{(m-i)}), \end{aligned}$$

где $P_j^{\vee(i, m-i)}$ - оценка информативности P_j^{\vee} , вычисленная для классов, один из которых $(m-i)$ представлен i -м объектом, а другой - всеми нижестоящими; $\varphi_j^{(m-i)}$ - оценка информативности φ_j , вычисленная для классов, представленных $(m-i)$ объектами.

Кодирование информации

Бинарное кодирование исходной информации при решении задач упорядочения производится по описанной ранее схеме. Критерием оптимальности того или иного варианта кодирования служит абсолютная величина одной из трех оценок φ_j, φ_j^* и $\tilde{\varphi}_j$, выбранной для решения поставленной задачи.

Минимизация признакового пространства

В качестве критерия оптимальности при минимизации принимается степень соответствия значений строчечных нагрузок $J[\varphi]$ (где φ — любая из трех оценок $\varphi_j, \varphi_j^*, \tilde{\varphi}_j$) значениям целевого признака τ_{n+1} .

Строчечные нагрузки $J[\varphi]$ вычисляются по таким же правилам, что и в задаче классифицирования и формула их вычисления для таблиц с перекодированным кодом имеет вид: $J[\varphi] = \sum_{j=1}^n t_{ij} \cdot \varphi_j$

В целях сравнения, значения τ_{n+1} и $J[\varphi]$ нормируются по max-min:

$$f_i^* = \frac{f_i - f_{\min}}{f_{\max} - f_{\min}} \quad \text{и} \quad J[\varphi]_i^* = \frac{J[\varphi]_i - J[\varphi]_{\min}}{J[\varphi]_{\max} - J[\varphi]_{\min}} \quad i = 1, 2, \dots, m$$

Подобная нормировка позволяет привести разномасштабные величины к интервалу значений от 0 до 1 и сравнивать их соотношения внутри фиксированного интервала. Отклонение величин строчечных нагрузок от соответствующих значений целевого признака можно охарактеризовать коэффициентом, который определяет зону несоответствия нормированных f_i^* и $J[\varphi]_i^*$ для каждой пары значений. Коэффициент несоответствия ψ_i^e , определяющий ширину зон несоответствия, вычисляется по формуле:

$$\psi_i^e = |f_i^* - J[\varphi]_i^*|, \quad \text{где} \quad i = 1, 2, \dots, m, \quad e = \{1, 2, \dots, n\}$$

При этом возможны два общих случая:

1) все зоны несоответствия разобщены,

2) некоторые зоны несоответствия соприкасаются или имеют области (диапазоны) пересечения (рис.6). В первом случае критерием

минимизации является величина $\lambda_1 = \frac{\psi_{i \max}^e}{2}$, а во втором —

$$\lambda_2 = \frac{(\psi_i^e \cup \psi_{i+1}^e)_{\max}}{2}$$

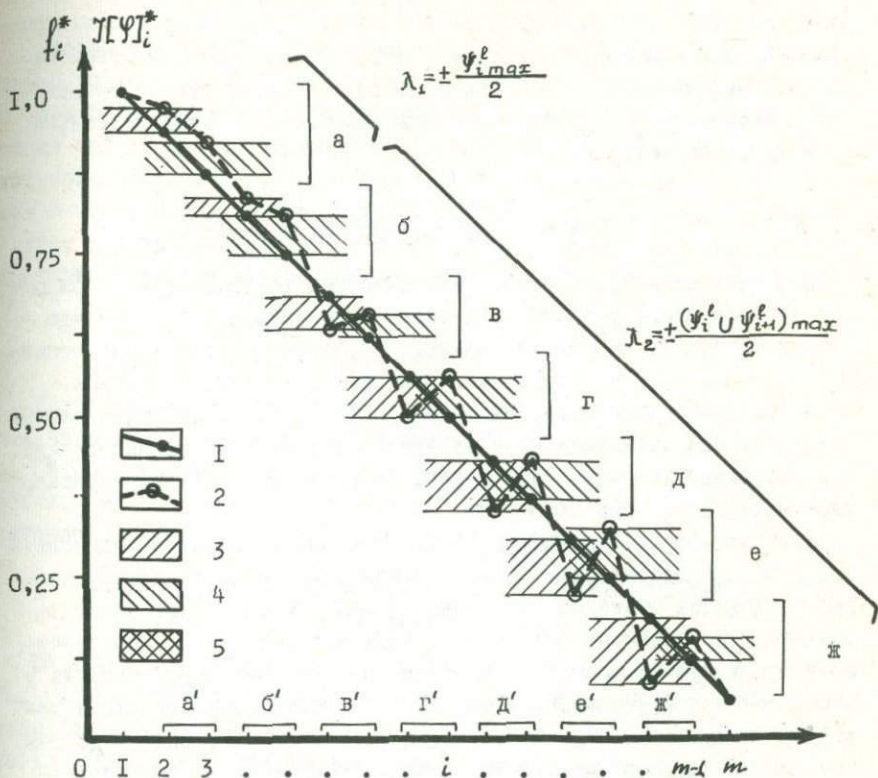


Рис. 6. Соотношение зон несоответствия для попарно сгруппированных объектов.

1 - вариационная кривая f_i^* ; 2 - вариационная кривая $\Pi\Pi_i^*$; 3 - зона несоответствия для первого объекта каждой пары; 4 - зона несоответствия для второго объекта каждой пары; 5 - диапазоны совмещения этих зон; а, б, ..., ж - возможные варианты соотношения ψ_i^l и ψ_{i+1}^l для выбранных пар объектов а', б', ..., ж'; λ_1 и λ_2 - ошибки определения значений целевого признака.

Таким образом, процедура минимизации в задачах упорядочения сводится к поиску такого поднабора признаков, для которого величина, определяющая в конечном счете точность оценки целевого признака у проб, минимальна, независимо от того, определяется ли она единой зоной несоответствия, или же дизъюнкцией накладывающихся, а также соприкасающихся зон.

Анализ эталонной выборки объектов

Для выяснения степени влияния каждого объекта эталонной выборки на результаты упорядочения также используется понятие "идеальных" объектов (R_I^u и R_{II}^u), однако смысловое их содержание отличается от R_I и R_{II} , хотя вычисляются они аналогичным образом. Особенность R_I^u и R_{II}^u состоит в том, что они являются "идеальными" не для заранее заданных классов (как в задаче классифицирования), а для формальных, выделение которых зависит от применения оценок информативности признаков.

Величины строчечных нагрузок для эталонных объектов, вычисленные в объеме минимизированного признакового пространства, указывают положение эталонов на отрезке R_I^u и R_{II}^u и позволяют выявлять объекты, определяющие величину λ . Если эта величина не обеспечивает требуемую точность определения значений целевого признака у проб, такие объекты необходимо изъять из обучающей выборки и повторить этап обучения. Этот процедурный цикл повторяется до тех пор, пока исходная информация не будет обеспечивать необходимую точность решения задачи. Удаление некоторых эталонов из обучающей выборки приводит одновременно к изменению ограничений допуска проб к распознаванию (см. "Отбраковка проб").

Внешний экзамен

Внешний экзамен производится по описанной выше методике путем последовательного распознавания каждого удаленного эталонного объекта f_i^* на минимизированном пространстве признаков. Если значение f_i^* у объекта экзамена попадает в интервал $\mathcal{J}[\varphi]_{i \pm \lambda}$, то такое распознавание оценивается "1", в противном случае - "0".

Величина $B_{yn} = \frac{\sum_{i=1}^m \delta_i, \text{ у.п.}}{m} \cdot 100\%$ дает оценку качества распознавания (m - число эталонов)

Отбраковка проб

Отбраковка проб в задачах упорядочения производится по единственно возможному в данном случае блоку отождествляющих признаков. С формальных позиций, к распознаванию на имеющемся материале обучения могут быть допущены лишь пробы, полностью удовлетворяющие критерию общности, однако следует учитывать профессиональное мнение о важности каждого признака, вошедшего в этот блок. В этом случае к распознаванию могут быть допущены пробы, выполняющие критерий общности лишь частично.

Упорядочение проб

Упорядочение проб производится по процедуре подсчета строчечных нагрузок в объеме минимизированного признакового пространства [8]. Ожидаемое значение целевого признака распознаваемого объекта в соответствии с получаемой величиной $J[\varphi]_i^{*пр.}$ лежит в интервале $J[\varphi]_i^{*пр.} \pm \Delta$ при надежности распознавания равном $\epsilon_{уп.}$.

Решение задач классифицирования с одновременным упорядочением

Задачи данного вида могут, вообще говоря, решаться последовательно с помощью описанных выше алгоритмов классифицирования и упорядочения. Однако существуют способы одноактного их решения, когда обе операции выполняются одновременно. Обращаться к этим способам целесообразно тогда, когда они дают вполне удовлетворительные результаты на этапе обучения, в противном случае следует применять последовательное решение задачи.

Оценки информативности признаков

Для решения описываемого вида задачи предлагаются две оценки информативности признаков, отличающиеся способом сравнения объектов разных классов [7, 8].

Оценка φ_j° . Производится сравнение каждого объекта первого класса с каждым объектом второго по j -му признаку и каждое выделенное различие оценивается разностью значений целевого признака. Суммарный результат делится на величину $m_1 \cdot m_2$ - число всех сравнений:

$$\varphi_j^{\circ} = \frac{\sum_{i=1}^{m_1} \sum_{i'=1}^{m_2} (t_{ij} - t'_{i'j})(f_i - f'_{i'})}{m_1 m_2}$$

Преобразовав формулу, получим:

$$\varphi_j^{\circ} = \mathcal{P}_j^I \cdot f_{cp.I}^I - f_{cp.I}^I \cdot \mathcal{P}_j^II - \mathcal{P}_j^I \cdot f_{cp.I}^{II} + f_{cp.I}^{II} \cdot \mathcal{P}_j^II$$

Поскольку $f_{cp.I}^{II} = f_{cp.I}^I$, то, как нетрудно убедиться:

$$\varphi_j^{\circ} = (\mathcal{P}_j^I - \mathcal{P}_j^{II})(f_{cp.I}^I - f_{cp.I}^{II}) + \mathcal{P}_j^I (f_{cp.I}^I - f_{cp.I}^I) = P_j^V + \varphi_j^I$$

где φ_j^I - оценка φ_j для объектов I класса.

Легко видеть, что оценка φ_j° есть обобщение оценок P_j^V и φ_j^I , совпадающая с первой при $f_{cp.I}^I = f_{cp.I}^{II}$, а со второй - при $f_{cp.I}^I = f_{cp.I}^I$ или $\mathcal{P}_j^I = \mathcal{P}_j^{II}$. В общем же случае φ_j° учитывает как классифицирующую, так и упорядочивающую способности признака.

Оценка $\hat{\varphi}_j$. Эта оценка получается при последовательном сравнении сначала первого объекта I класса со всеми объектами II, затем первых двух объектов, первых трех и т.д. до полного объема первого класса. Каждый результат делится на число сравнений, окончательная оценка информативности получается путем суммирования всех промежуточных результатов:

$$\hat{\varphi}_j = \sum_{\kappa=1}^{m_1} \frac{\sum_{i=1}^{\kappa} \sum_{i'=1}^{m_2} (t_{ij} - t'_{i'j})(f_i - f'_{i'})}{\kappa \cdot m_2}$$

После соответствующих преобразований получим:

$$\hat{\varphi}_j = \sum_{\kappa=1}^{m_1} (\mathcal{P}_j^{(\kappa)} - \mathcal{P}_j^{II}) (f_{cp}^{(\kappa)} - f_{cp}^{II}) + \mathcal{P}_j^{(\kappa)} (f_{cp.I}^{(\kappa)} - f_{cp.I}^{(\kappa)}) = \sum_{\kappa=1}^{m_1} (P_j^{V(\kappa,II)} + \varphi_j^{(\kappa)})$$

Как следует из окончательной формулы, оценка $\hat{\psi}_j$, подобно ψ_j^0 , учитывает одновременно как классифицирующую, так и упорядочивающую способности признака.

Кодирование информации

Бинарное кодирование исходной информации при решении задачи классифицирования с одновременным упорядочением производится по описанной ранее схеме. Критерием оптимальности того или иного варианта кодирования служит абсолютная величина одной из двух оценок ψ_j^0 и $\hat{\psi}_j$, выбранной для решения поставленной задачи.

Минимизация признакового пространства

Критерий оптимальности при минимизации признакового пространства складывается в данном случае из одновременного учета двух величин — ρ^e и λ , причем последняя величина вычисляется лишь для эталонов первого класса. Вопрос о предпочтительной важности какой-либо из этих величин требует, очевидно, самостоятельного изучения. Однако, исходя из наиболее общих соображений, можно полагать, что вначале следует ориентироваться на величину ρ^e , а λ использовать в качестве контролирующей величины. Таким образом, задачу следует решать данным способом лишь в том случае, если $\rho^e > 0$, а λ при этом обеспечивает необходимую точность оценки целевого признака. Во всех остальных случаях задачу следует решать с помощью последовательного применения алгоритмов классифицирования и упорядочения.

Анализ эталонной выборки объектов

Анализ расположения эталонов на прямой R_1R_2 по величинам строчечных нагрузок в объеме минимизированного признакового пространства позволяет, с одной стороны, выявлять объекты, нарушающие "контрастность" классов, а с другой — объекты, несоответствие которых целевому признаку обуславливает большую ошибку при упорядочении. Удаление таких объектов из эталонной выборки с последующим переобучением позволяет увеличить "контрастность" классов и уменьшить величину λ .

Внешний экзамен

При последовательном распознавании эталонных объектов после их удаления из обучающей выборки вычисляются две величины: оценка качества межклассового распознавания $\mathcal{E}_{кл}$ и оценка точности внутриклассового распознавания (упорядочения) $\mathcal{E}_{уп}$, которые соответственно характеризуют качество распознавания проб при классифицировании и упорядочении (в случае отнесения пробы к первому классу).

Отбраковка проб

Процедура отбраковки проб полностью аналогична таковой в задачах классифицирования. Если в результате сопоставления пробы с блоковой таблицей материала обучения проба допущена к распознаванию и получен ответ о ее принадлежности ко II классу, процесс распознавания прекращается. Если проба допущена к распознаванию и по результатам логического анализа отнесена к I классу, то в объеме минимизированного пространства для нее оценивается значение целевого признака. Если же в результате анализа не было получено альтернативного ответа о классовой принадлежности, то проба по процедуре подсчета строчечных нагрузок классифицируется и одновременно упорядочивается в объеме минимизированного признакового пространства. Пробы, отсеянные при блоковом анализе, к распознаванию не допускаются.

Классифицирование и упорядочение проб

Для каждой пробы в объеме минимизированного признакового пространства подсчитываются строчечные нагрузки. Качество и надежность распознавания проб в зависимости от их отнесения к тому или иному эталонному классу определяются величинами $\mathcal{E}_{кл}$, $\mathcal{E}_{уп}$ и λ .

ж ж ж

Общей особенностью алгоритмов из серии "Каскад", обеспечивающих целевое классифицирование и упорядочение объектов, является инвариантность относительно направленности априорного кодирования.

ния и вовлечение в обработку значений целевого признака. Последняя особенность позволяет учитывать специфику решаемых задач и получать на одном и том же исходном материале различные результаты при изменении постановки геологической задачи. Естественно, что применение описанного метода может быть успешным при наличии профессионального контроля за качеством исходного материала и при соответствии применяемого алгоритма содержанию задачи. Большинство алгоритмов этого метода имеют программную реализацию на ЭВМ типа БЭСМ-6 и М-220. Примеры практического применения "Каскада" приведены в некоторых статьях настоящего сборника.

Авторы считают своим приятным долгом выразить благодарность В.О.Красавчикову, оказавшему большую помощь в математическом оформлении предлагаемого метода, а также Г.С.Лбову и В.Н.Елкиной, сделавшим ряд ценных критических замечаний при прочтении статьи в рукописи.

Л и т е р а т у р а

1. БАБИЧ В.В., ФЕДОСЕЕВ Г.С. Типизация прогнозно-поисковых задач и некоторые подходы к их решению. - В кн.: Состояние и направление исследований по металлогении траппов. Красноярск, 1974, с.118 - 119.

2. БАБИЧ В.В., КРАСАВЧИКОВ В.О., ФЕДОСЕЕВ Г.С. Комплексная оценка информативности признаков при решении прогнозных задач. - В кн.: Состояние и направление исследований по металлогении траппов. Красноярск, 1974, с.117-118.

3. БАБИЧ В.В., КРАСАВЧИКОВ В.О., ФЕДОСЕЕВ Г.С. Программы метода суммарного учета мер приуроченности и согласования. - В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.122-126.

4. ШАРАПОВ И.П. Измерение силы связи между качественными признаками в геологии.-В кн.: Научные труды Пермского научно-исследовательского угольного института (сборник УИИ). Пермь, 1965, с.21-33.

5. T.Marill , D.M.Green. On the effectiveness of receptors in recognitions systems.-IEEE"Trans on information theory," v. IT - 9, January, 1963, pp. 11-17.

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

7. КАНДЫБА В.И. Программа "Оценка информативности признаков". - В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.127-130.

8. БАБИЧ В.В., Программа "Подсчет строчечных нагрузок по заданной информативности для бинарных таблиц". - В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.135-142.

А.А.Бишаев

МЕТОД "ЦЕЛЕВАЯ ИТЕРАЦИОННАЯ КЛАССИФИКАЦИЯ" ("ЦИКЛ")

При решении геологических, технических, медицинских и некоторых других задач методами автоматической классификации и распознавания образов очень важными являются вопросы выбора информативной системы признаков (и.с.п.), а также решающего правила, в соответствии с получаемой системой признаков [1-6]. Эти вопросы осложняются тем, что в практических задачах прогнозно-поискового профиля исходная информация часто задается разнородными по своей природе признаками (качественными и количественными), имеющими неизвестные значения (прочерки) на некоторых объектах. Следует также отметить, что большое число алгоритмов, особенно статистического характера, осуществляет выбор и.с.п., на основе индивидуальной оценки информативности каждого признака в предположении о его связи, независимой от других признаков, со значениями целевого признака^ж). Во многих задачах, как, например, показано в [3], этот подход не дает достаточно хорошего результата. Поэтому естественно полагать, что необходимо построение комплекса процедур для нахождения информативной системы взаимосвязанных в целевом отношении разнородных косвенных признаков, и далее, для принятия

ж) Целевым мы называем признак, по которому устанавливается исходная классификация эталонных объектов.

решений по этим признакам. Необходимость такого комплекса процедур, основанных на использовании элементов различия и сходства объектов в признаковом пространстве, и полезность их использования отчетливо были вскрыты в развитии аппарата тестового подхода и его применений [7-10] .

В рассматриваемом здесь методе большое внимание уделяется этапу выбора такой и.с.п., которая позволяет добиваться либо разделения эталонных объектов, либо объединения, либо пропорционального совмещения того и другого в соответствии с исходной целевой классификацией этих объектов, устанавливаемой на основе их различий и сходств по целевому признаку [5,10,11] . То есть, под и.с.п. здесь понимается такое подпространство исходных характеристических признаков, которое на основе простых сравнительных способов принятия решений (по мерам различия и сходства), в конечном итоге дает оценки значений целевого признака на эталонных объектах, близкие или совпадающие с заданными целевыми значениями.

Способы выбора информативной системы зависимых косвенных признаков, когда по каждому из них невозможно получить классификацию эталонных объектов близкую к целевой, разнообразные и трудоемкие. Чаще всего они связаны с перебором всех возможных вариантов сочетаний признаков и выбором из них лучшего, или с построением приближенных способов поиска информативной системы признаков. Применяемые для решения этой задачи методы базируются либо на логико-эвристических, либо на статистических процедурах [2,5,9,10,12-14] . В этих случаях информативность отдельного признака для получения целевой классификации эталонных объектов оценивается "коллективно", то есть зависит от "дополнительности" данного признака в сочетании с другими.

В методе "Цикл" получение и.с.п. проводится, во-первых, при ослаблении ограничений на исходные данные как в информационном (признаки разного вида, "прочерки"), так и в статистическом плане (без требований о независимости признаков, законе распределения, числе объектов и признаков); во-вторых, без полного перебора, путем последовательных приближений с помощью итерационных процедур [15,16] ; в-третьих, с целевым учетом как индивидуальной "силы" признаков, так и "дополнительности" их в сочетаниях. Большинство

известных алгоритмов распознавания не удовлетворяет этим условиям, либо удовлетворяет частично. В этом смысле перечисленные особенности метода рассматриваются в качестве некоторых исходных условий, важных в практическом отношении, т.е. позволяющих расширить класс решаемых практических задач. Можно отметить, [13], где совокупность исходных условий сравнима с исходными условиями данного метода, хотя сами методы существенно различны. Рассматриваемый метод, как при выборе и.с.п., так и в правилах принятия решений, базируется на вычислении мер различия и сходства между объектами в пространстве целевого и характеристических признаков. Содержательное обоснование метода вытекает из следующих требований к целевой классификации эталонных объектов по и.с.п.:

а) Любые два объекта, с заданной мерой различия по целевому признаку, должны иметь такую же или большую^{ж)} меру различия по информативной системе характеристических признаков. Большая мера различия означает, что объекты одного класса (объекты с одинаковыми или "близкими" значениями целевого признака) в этом случае могут быть существенно неоднородными, т.е. объекты с одинаковыми значениями целевого признака могут существенно различаться по такой и.с.п., так как сходства между ними не требуется.

б) Любые два объекта с заданной мерой сходства по целевому признаку, должны иметь такую же или большую меру сходства по и.с.п. Это означает, что объекты разных классов могут быть "похожими", так как их различие специально не требуется, зато не должны существенно различаться объекты одного и того же класса.

в) Совмещение требований (а) и (б), т.е. их одновременное выполнение. Это такое требование, когда меры сходства эталонных объектов внутри каждого класса и меры их различия в разных классах должны быть максимальными.

Значения информационных весов характеристических признаков, которые нужно найти для максимального приближения к каждому из перечисленных требований в целевой классификации эталонных объектов, будем называть: а) разделяющими и обозначать $(R_j)_n$; б) объединяющими — $(Q_j)_n$; в) совмещенными — $(P_j)_n$, где n — число

ж) Отклонение меры в "большую" сторону в этом случае считается несущественным для целевой классификации.

признаков. Для количественной оценки этих требований в методе "Цикл" вначале, при условии равноценности характеристических признаков, вычисляются меры сходства и различия для всех пар объектов по отдельным признакам, а затем по их общему списку. В дальнейшем, на основании этих мер и мер различия и сходства по целевому признаку, задаются начальные значения целевых требований и производится количественная оценка информативности каждого признака, характеризующая его как индивидуальную, так и коллективную существенность относительно исходной целевой классификации объектов в первом приближении. Затем полученная в первом приближении система признаков используется для нахождения системы признаков в следующем приближении и так далее, то есть строится итерационная последовательность. В соответствии с разными (а, б, в) требованиями к целевой классификации объектов возможен выбор и с.п.трех видов, согласно которым в методе формулируются три правила принятия решений на основе функций, зависящих от расстояний между объектами. Ниже дается алгоритмическое описание метода "Цикл" с интерпретацией основных моментов его функционирования.

Задание исходной информации и вычисление мер различия и сходства

Исходная информация представляется в виде числовой $m \times (n+1)$ матрицы значений $T = (t_{ij})$, ($i = 1, \dots, m$; $j = 1, \dots, n+1$) множества признаков (столбцов) $X = x_1, \dots, x_{n+1}$ на множестве объектов (строк) $S = s_1, \dots, s_m$, где x_{n+1} обозначает целевой признак. Задаются начальные значения (нулевое приближение) одного из трех видов $(R_j)_n$, $(Q_j)_n$, $(P_j)_n$ информационных весов характеристических признаков, например, $P_1^{(0)} = P_2^{(0)} = \dots = P_n^{(0)}$, где $\sum_{j=1}^n P_j^{(0)} = 1$ - условие нормы - ровки. Для целевого признака считается, что $P_{n+1} = 1$. Совместное рассмотрение информации различной природы в соответствии с разными видами шкал признаков [17], а также признаков, содержащих "прочерки", является существенной особенностью метода. В исходной информации могут быть представлены одновременно признаки трех видов:

I. Качественные, значения которых соответствуют шкале наименований.

2. "Относительно количественные", значения которых соответствуют шкале интервалов (относительная температура по Цельсию и т.п.). К тому же виду условно относятся "порядковые признаки", значения которых заданы на шкале порядка, например, твердость минералов по шкале Мооса.

3. "Абсолютно количественные", значения которых соответствуют шкале отношений (абсолютный возраст и т.п.).

Для таблицы $[m \times (n+1)]$ в соответствии с разными видами признаков на каждом из них (x_j) , $j=1, \dots, n+1$. для каждой пары эталонных объектов (s_i', s_i'') , $i'=1, \dots, m-1$, $i''=2, \dots, m$; $i'' > i'$, $k=1, \dots, m$ - номера пар, где $(m = C_m^2)$, вводятся способы вычисления мер различия (d_{kj}) и дополнительных к ним мер сходства (g_{kj}) , например:

1). Для качественных признаков

$$d_{kj}(1) = \{ 0 \text{ при } t_{i'j} = t_{i''j}; \quad 1 \text{ при } t_{i'j} \neq t_{i''j} \}; \quad (1)$$

2). Для "относительно количественных" признаков

$$d_{kj}(2) = \{ |t_{i'j} - t_{i''j}| / \max_{i=1, \dots, m} t_{ij} - \min_{i=1, \dots, m} t_{ij} \}; \quad (2)$$

где $t_{i'j}$ и $t_{i''j}$ - действительные числа любого знака, (подобные (1) и (2) способы вычисления мер различия или сходства можно найти у разных авторов, например, в [18]).

3). Для "абсолютно-количественных" признаков вычислим величину $\delta_{kj} = \max_{i=1, \dots, m} t_{ij} / \max(t_{i'j}, t_{i''j})$, тогда $d_{kj}(3) = \{ 0 \text{ при } t_{i'j} = t_{i''j} = 0; \quad d_{kj}(2) \times \delta_{kj} \}; \quad (3)$

где $t_{i'j}$ и $t_{i''j}$ - действительные числа одного знака. Таким образом, эта мера построена на отношении между значениями признака, в то время как предыдущие зависят от разности.

Если по какому-либо признаку один или оба из сравниваемых объектов имеют неизвестное значение ("прочерк"), то во всех рассмотренных способах вычисления мер различия (d_{kj}) и сходства (g_{kj}) в этих случаях принимается $d_{kj} = 0$, $g_{kj} = 0$.

С учетом начальных значений информационных весов $(P_j^{(0)})_n$ для мер различия и сходства по каждому признаку будем иметь значения $(P_j^{(0)} \cdot d_{kj})$ и $(P_j^{(0)} \cdot g_{kj})$, откуда вычисляются начальная мера различия (нормированное расстояние) $\rho_k^{(0)}$ и начальная мера сходства для каждой пары эталонных объектов по всей системе характери-

стических признаков:

$$\rho_k^{(0)} = \sum_{j=1}^n P_j^{(0)} \cdot d_{kj}; \quad \lambda_k^{(0)} = \sum_{j=1}^n P_j^{(0)} \cdot g_{kj} = 1 - \rho_k^{(0)} \text{ где } 0 \leq (\rho_k^{(0)}, \lambda_k^{(0)}) \leq 1. \quad (4)$$

При использовании начальных весов ($R_j^{(0)}$) или ($Q_j^{(0)}$) вычисляются соответственно только меры различия ($\rho_k^{(0)}$) или сходства ($\lambda_k^{(0)}$), которые будут равны найденным по (4) в силу условия начальной равноценности признаков.

Постановка и решение задачи выбора и.с.п.

Вычислив по (1,2,3,4) целевые ($d_{k, n+1}$) и характеристические ($\rho_k^{(0)}$) меры различия, а также дополнительные к ним меры сходства ($g_{k, n+1}, \lambda_k^{(0)}$), сформулируем количественную оценку двух целевых требований ($D_k^{(0)}, G_k^{(0)}$) к начальной системе признаков соответственно для разделения и объединения каждой пары эталонных объектов:

$$D_k^{(0)} = d_{k, n+1} / \rho_k^{(0)}; \quad G_k^{(0)} = g_{k, n+1} / \lambda_k^{(0)} \quad (5)$$

Если исходная система признаков удовлетворяет рассмотренным выше (а,б,в) требованиям к и.с.п., то полученные по (5) оценки (целевые требования) будут следующими:

а) $D_k^{(0)} \leq 1$, б) $G_k^{(0)} \leq 1$, в) $D_k^{(0)} = G_k^{(0)} = 1$, ($k = 1, \dots, m$). Численно эти целевые требования выражают "степень неинформативности" заданной системы признаков для целевого разделения или объединения одной пары эталонных объектов. Так как отношения (5) могут быть сколь угодно велики при $\rho_k^{(0)} \ll d_{k, n+1}$ или $\lambda_k^{(0)} \ll g_{k, n+1}$ (например, $d_{k, n+1} \neq 0, \rho_k^{(0)} \rightarrow 0$ или $g_{k, n+1} \neq 0, \lambda_k^{(0)} \rightarrow 0$), то в таких случаях заданная система признаков считается полностью неинформативной для целевого разделения или объединения двух объектов и необходимо изменить либо эталонную выборку (удалением одного из этих объектов или обоих одновременно), либо исходную систему признаков (путем включения новых признаков, а также частичной или полной замены старых). Поэтому область определения начальных целевых требований задается неравенством: $0 \leq (D_k^{(0)}, G_k^{(0)}) \leq C$, C - ограничительное условие, например, $C = 2n'$ где n' - число признаков с ненулевым весом. Поскольку $g_{k, n+1} = 1 - d_{k, n+1}$; $\lambda_k^{(0)} = 1 - \rho_k^{(0)}$, то для любого $k = 1, \dots, m$ между $D_k^{(0)}$ и $G_k^{(0)}$ (5) в зависимости от значений информационных весов ($P_j^{(0)}$) всегда имеет место одно из следующих соотношений:

$$а) D_{\kappa}^{(o)} < 1 \iff G_{\kappa}^{(o)} > 1 ; \quad б) D_{\kappa}^{(o)} > 1 \iff G_{\kappa}^{(o)} < 1 ; \quad в) D_{\kappa}^{(o)} = 1 \iff G_{\kappa}^{(o)} = 1, \quad (6)$$

где знак \iff употребляется вместо слов "тогда и только тогда, когда...". Рассмотрим три варианта комбинаций из этих соотношений на всем множестве пар целевых требований ($\kappa = 1, \dots, M$), при которых будем считать систему признаков полностью информативной.

1) для целевого разделения эталонных объектов, если выполнены только (а) и (в) (т.е. $D_{\kappa}^{(o)} \leq 1$);

2) для целевого объединения этих объектов при выполнении только (б) и (в), т.е. $G_{\kappa}^{(o)} \leq 1$;

3) для целевого пропорционального совмещения разделения и объединения эталонных объектов при выполнении только (в), т.е. $D_{\kappa}^{(o)} = G_{\kappa}^{(o)} = 1$.

Поэтому зададим три варианта мер $(J_1^{(o)}; J_2^{(o)}, J_3^{(o)})$ для оценки информативности начальной системы признаков соответственно с разделяющими $(R_j^{(o)})$, объединяющими $(Q_j^{(o)})$ и совмещенными $(P_j^{(o)})_n$ весами.

$$1) J_1^{(o)} = M / \sum_{\kappa=1}^M A_{\kappa}^{(o)} ; \quad 2) J_2^{(o)} = M / \sum_{\kappa=1}^M B_{\kappa}^{(o)} ; \quad 3) J_3^{(o)} = (J_1^{(o)} + J_2^{(o)}) / 2 \quad (7.1-7.3)$$

где $A_{\kappa}^{(o)} = \{ 1 \text{ при } D_{\kappa}^{(o)} \leq 1, D_{\kappa}^{(o)} \}$; $B_{\kappa}^{(o)} = \{ 1 \text{ при } G_{\kappa}^{(o)} \leq 1, G_{\kappa}^{(o)} \}$.

Очевидно, что область определения данных мер информативности

$$- 0 \leq (J_1^{(o)}, J_2^{(o)}, J_3^{(o)}) \leq 1, \quad \text{где } J - \text{ полностью информативная система признаков.}$$

Задача выбора и.с.п. состоит в том, чтобы для каждой меры $J_1^{(o)}, J_2^{(o)}, J_3^{(o)}$ построить правило нахождения значений $(R_j^{(e)}, Q_j^{(e)}, P_j^{(e)})_n$, дающих ее максимум, т.е. найти систему признаков, называемую информативной, с таким набором значений информационных весов $(R_j^{(e)}, Q_j^{(e)}, P_j^{(e)})_n$, который дает целевые требования к этой системе максимально приближающиеся к случаям, соответствующим полностью информативной системе признаков. Это приближение оценивается в среднем по всем парам эталонных объектов в соответствии со значениями мер $J_1^{(e)}, J_2^{(e)}, J_3^{(e)}$.

Нахождение первого приближения к и.с.п.

Рассмотрим правило нахождения информационных весов $(P_j^{(1)})_n$, дающих первое приближение к полностью информативной системе признаков в третьем варианте пропорционального совмещения целевых требований разделять и объединять объекты.

Согласно (5,6) целевое требование $D_{\kappa}^{(0)} > 1$ показывает, во сколько раз $\rho_{\kappa}^{(0)}$ должна быть больше, чтобы система признаков стала полностью информативной для целевого разделения данной пары эталонных объектов; $D_{\kappa}^{(0)} < 1$ означает, во сколько раз $\rho_{\kappa}^{(0)}$ может быть меньше, чтобы система признаков все еще оставалась полностью информативной для целевого разделения двух объектов. То же самое можно сказать относительно $G_{\kappa}^{(0)}$ в смысле целевого объединения пары эталонных объектов. В связи с этим будем использовать оба значения целевых требований $D_{\kappa}^{(0)}$, $G_{\kappa}^{(0)}$ в качестве коэффициентов "поощрения" или "наказания" признаков при поиске более информативной системы. Учитывая значения целевых требований $D_{\kappa}^{(0)}$, $G_{\kappa}^{(0)}$ введем разделяющие $z_{\kappa j}^{(1)}$ и объединяющие $q_{\kappa j}^{(1)}$ целевые нагрузки каждого признака в каждой паре эталонных объектов для нахождения новой системы признаков (в первом приближении к искомой и.с.п.):

$$z_{\kappa j}^{(1)} = P_j^{(0)} \cdot a_{\kappa j} \cdot D_{\kappa}^{(0)} ; \quad q_{\kappa j}^{(1)} = P_j^{(0)} \cdot g_{\kappa j} \cdot G_{\kappa}^{(0)} . \quad (8)$$

Если находить суммарные целевые нагрузки каждого признака отдельно для разделения и объединения всего множества пар эталонных объектов:

$$z_j^{(1)} = \sum_{\kappa=1}^M z_{\kappa j}^{(1)} ; \quad q_j^{(1)} = \sum_{\kappa=1}^M q_{\kappa j}^{(1)} , \quad (9)$$

то после нормировки из них получим разделяющие $(R_j^{(1)})_n$ или объединяющие $(Q_j^{(1)})_n$ целевые веса признаков, дающие первое приближение целевых требований к соотношениям $(6^A$ и $6^B)$ или $(6^B$ и $6^A)$:

$$R_j^{(1)} = z_j^{(1)} / \sum_{j=1}^n z_j^{(1)} ; \quad Q_j^{(1)} = q_j^{(1)} / \sum_{j=1}^n q_j^{(1)} \quad (10)$$

Найдем совместную разделяющую и объединяющую целевую нагрузку признака для κ -ой пары объектов:

$$B_{\kappa j}^{(1)} = z_{\kappa j}^{(1)} + q_{\kappa j}^{(1)} . \quad (II)$$

Такая целевая нагрузка признака выражает его "весомость" для совместного разделения и объединения данной пары объектов. Логично

чески она показывает, что при получении нового информационного веса значимость признака должна быть увеличена или уменьшена при поощрении и наказании его целевыми требованиями $D_{\kappa}^{(0)}, G_{\kappa}^{(0)}$ пропорционально мерам различия $d_{\kappa j}$ и сходства $g_{\kappa j}$. При этом $d_{\kappa j}, g_{\kappa j}$ учитывают индивидуальную силу признака, а вместе с $D_{\kappa}^{(0)}, G_{\kappa}^{(0)}$ учитывается и целевая "дополнительность" признака для совместного разделения и объединения каждой пары объектов в сочетании с другими признаками. Ясно, что из совместной нагрузки признака может быть определена индивидуально-коллективная оценка его информативности для κ -ой пары объектов:

$$P_{\kappa j}^{(1)} = G_{\kappa j}^{(1)} / \sum_{j=1}^n G_{\kappa j}^{(1)}. \quad (12)$$

Вычислим суммарную целевую нагрузку каждого признака для совместного разделения и объединения всех пар эталонных объектов:

$$G_j^{(1)} = \sum_{\kappa=1}^m G_{\kappa j}^{(1)} \quad (13)$$

После нормировки получим целевой информационный вес каждого признака в первом приближении:

$$P_j^{(1)} = G_j^{(1)} / \sum_{j=1}^n G_j^{(1)} \quad (14)$$

Очевидно, что $\sum_{j=1}^n P_j^{(1)} = 1$. Эти информационные веса задают новую систему признаков, в среднем приближающую целевые требования $D_{\kappa j}^{(1)}, G_{\kappa j}^{(1)}$ ($\kappa=1, \dots, m$) к полностью информативной системе, т.е. к соотношению (G^B) , а интерпретация этих весов такая же, как и описанных выше целевых нагрузок признаков для отдельных пар эталонных объектов. Случай $P_j^{(1)} < \beta$ (β - порог значимого веса) показывает, что данный признак можно минимизировать (считать отсутствующим) в новой системе.

Классифицирующий оператор K и его основные свойства

Подставляя $(P_j^{(1)})_n$ вместо $(P_j^{(0)})_n$ в формулы (4,5,7) найдем меры различия $\rho_{\kappa}^{(1)}$ и сходства $\lambda_{\kappa}^{(1)}$ и соответствующие им целевые требования $D_{\kappa}^{(1)}, G_{\kappa}^{(1)}$ к системе признаков с информационными весами в первом приближении к выбираемой и.с.п., то есть имеет место последовательность операций, которая каждой паре целевых требований $(D_{\kappa}^{(0)}, G_{\kappa}^{(0)})$ к начальной системе признаков $(P_j^{(0)})_n$ однозначно ста-

вит в соответствие другую пару целевых требований $(D_{\kappa}^{(1)}, G_{\kappa}^{(1)})$ к новой системе признаков $(P_j^{(1)})_n$. Эту последовательность операций будем называть классифицирующим оператором K .

Рассмотрим основные свойства этого оператора:

1) Возьмем одну пару эталонных объектов, для которой выполнено соотношение (6^B) , т.е. $D_{\kappa}^{(0)} = G_{\kappa}^{(0)} = 1$, и выясним действие оператора K . При нахождении новых информационных весов признаков в операции (8) имеем: $z_{\kappa j}^{(1)} = P_j^{(0)} \cdot d_{\kappa j} \cdot 1$; $q_{\kappa j}^{(1)} = P_j^{(0)} \cdot g_{\kappa j} \cdot 1$

Если для каждого признака согласно (II) найти совместные нагрузки $G_{\kappa j}^{(1)} = z_{\kappa j}^{(1)} + q_{\kappa j}^{(1)} = P_j^{(0)} \cdot [d_{\kappa j} + (1 - d_{\kappa j})] = P_j^{(0)}$ и по (I2) определить новый информационный вес $P_{\kappa j}^{(1)}$, то получим

$$P_{\kappa j}^{(1)} = P_j^{(0)} / \sum_{j=1}^n P_j^{(0)} = P_j^{(0)}$$

Это значит, что при условии (6^B) начальные информационные веса не меняются. Поэтому результатом действия оператора в этом случае будет пара $D_{\kappa}^{(1)}, G_{\kappa}^{(1)}$, где $D_{\kappa}^{(1)} = D_{\kappa}^{(0)} = 1$ и $G_{\kappa}^{(1)} = G_{\kappa}^{(0)} = 1$. Это означает, что

$$D_{\kappa}^{(0)} = D_{\kappa}^{(1)} = 1 = G_{\kappa}^{(0)} = G_{\kappa}^{(1)} \quad (15)$$

при $(P_j^{(0)})_n = (P_{\kappa j}^{(0)})_n$. Таким образом, при условии (6^B) полностью информативная система признаков для совместного целевого разделения и объединения пары объектов в результате действия оператора K не изменяется, также как не меняются целевые требования к ней.

2). Пусть для κ -ой пары объектов выполнено соотношение (6^B) , т.е. $D_{\kappa}^{(0)} \geq 1$ и $G_{\kappa}^{(0)} < 1$. Тогда при нахождении информационных весов признаков $P_{\kappa j}^{(1)}$ в операции (8) $D_{\kappa}^{(0)}$ пропорционально поощряет различающие, а $G_{\kappa}^{(0)}$ наказывает объединяющие целевые нагрузки каждого признака. Поэтому, согласно (II), для совместной целевой нагрузки любого признака будем иметь:

$$G_{\kappa j}^{(1)} = z_{\kappa j}^{(1)} + q_{\kappa j}^{(1)} = P_j^{(0)} \cdot [d_{\kappa j} \cdot D_{\kappa}^{(0)} + g_{\kappa j} \cdot G_{\kappa}^{(0)}] \quad (16)$$

Отсюда получим информационный вес признака для случая двух объектов

$$P_{\kappa j}^{(1)} = G_{\kappa j}^{(1)} / \sum_{j=1}^n G_{\kappa j}^{(1)}$$

При сравнении по (16) выражений в скобках у любых двух признаков x_j' , x_j'' , для которых $d_{\kappa j}' > d_{\kappa j}''$, всегда будет выполнено неравенство

$$[d_{\kappa j}' \cdot D_{\kappa}^{(0)} + g_{\kappa j}' \cdot G_{\kappa}^{(0)}] > [d_{\kappa j}'' \cdot D_{\kappa}^{(0)} + g_{\kappa j}'' \cdot G_{\kappa}^{(0)}],$$

поскольку $D_k^{(0)} > 1 > G_k^{(0)}$. Так как эти выражения являются множителями к начальным весам $P_j^{(0)}$, $P_j^{(1)}$ при вычислении совместных целевых нагрузок, а также информационных весов $P_{kj}^{(1)}$, $P_{kj}^{(0)}$, то верно следующее: новый информационный вес каждого признака $P_{kj}^{(1)}$, $P_{kj}^{(0)}$ будет иметь изменения (множитель) по отношению к начальному весу в прямой зависимости от значения $(d_{kj} \cdot D_k^{(0)})$. Этот факт показывает, что при любом начальном наборе $(P_j^{(0)})_n$, удовлетворяющем соотношения (6^б) и при наличии дифференциации в значениях $(d_{kj} \cdot D_k^{(0)})$ выполняется неравенство:

$$\rho_k^{(1)} = \sum_{j=1}^n P_{kj}^{(1)} \cdot d_{kj} > \sum_{j=1}^n P_j^{(0)} \cdot d_{kj} = \rho_k^{(0)}, \quad (18)$$

из чего в соответствии с (5, 6^б) следует:

$$d_{k, n+1} / \rho_k^{(0)} > d_{k, n+1} / \rho_k^{(1)}; \quad D_k^{(0)} > D_k^{(1)}, \quad G_k^{(0)} < G_k^{(1)} \quad (19)$$

Таким образом, соотношения (19) показывают, что в результате действия оператора K , при условии неполной информативности начальной системы признаков для целевого разделения двух эталонных объектов и дифференциации в значениях $d_{kj} \cdot D_k^{(0)}$ ($j=1, \dots, n$) новая система признаков приближается к полностью информативной для целевого разделения этих объектов (т.е. к выполнению условий (6^в) и или (6^а)).

3). Пусть для k -ой пары объектов выполнено соотношение $D_k^{(0)} < 1$ и $G_k^{(0)} > 1$. Тогда все предыдущие выводы, сделанные относительно $D_k^{(0)}$ и $D_k^{(1)}$, будут верны и для $G_k^{(0)}$ и $G_k^{(1)}$ в силу симметричности как самих величин D_k и G_k , так и действия над ними оператора K . Поэтому, исходя из (6^а), аналогично (19), получим:

$$G_k^{(0)} > G_k^{(1)} \quad \text{и} \quad D_k^{(0)} < D_k^{(1)} \quad (20)$$

Таким образом, в результате действия оператора K , при условии неполной информативности начальной системы признаков для целевого объединения двух эталонных объектов и дифференциации в значениях $d_{kj} \cdot G_k^{(0)}$ ($j=1, \dots, n$) новая система признаков приближается к полностью информативной для целевого объединения этих объектов, т.е. к выполнению соотношений (6^в) или (6^б).

Из учета трех рассматриваемых свойств оператора K видно, что при выборе и.с.п. в первом приближении $(P_{kj}^{(1)})_n$ для одновремен-

ного разделения и объединения отдельной пары эталонных объектов, в соответствии с целевыми требованиями $D_{\kappa}^{(0)}$ и $G_{\kappa}^{(0)}$, в определенных условиях этот оператор всегда уменьшает целевое требование, большее единицы и увеличивает меньшее единицы. Поскольку в разбираемом нами случае оба целевых требования, контролируя выбор и.с.п., действуют совместно, а "поощрение" и "наказание" признаков этими требованиями можно сделать сколь угодно осторожными (например, путем извлечения корня h -ой степени из значения совместной целевой нагрузки (II) каждого признака, что приводит к менее резкому изменению новых информационных весов (I2) в сравнении с начальным $(P_j^{(0)})_{\kappa}$, то каждое из новых целевых требований $(D_{\kappa}^{(i)}, G_{\kappa}^{(i)})$ согласно (19) и (20) уменьшаясь и увеличиваясь будут приближаться к значению I. Средняя величина $(D_{\kappa}^{(i)} + G_{\kappa}^{(i)})/2$, вследствие (4) и (5), также будет уменьшаться, приближаясь к значению I, а в случае $D_{\kappa}^{(0)} = 1 = G_{\kappa}^{(0)}$ выполняется (15) - первое свойство оператора K .

Распространим наше рассуждение на все множество пар эталонных объектов, для которых в соответствии с $(6^A, 6^B, 6^C)$ возможны все случаи сочетания трех соотношений из двух целевых требований. В сравнении с весами $(P_{\kappa j}^{(i)})$ для двух объектов (I2), при поиске целевых информационных весов признаков $(P_j^{(i)})_{\kappa}$ в этом случае добавляется лишь операция суммирования (I3) совместных целевых нагрузок по всем парам объектов. Поэтому условие полной информативности системы признаков для двух объектов (6^B) распространится на все множество их пар, и, согласно (19,20), в определенных условиях имеет место:

$$J_3^{(0)} < J_3^{(i)} \leq 1 \quad (21)$$

Построение итерационной процедуры выбора и.с.п.

Поскольку область значений $(D_{\kappa}^{(i)}, G_{\kappa}^{(i)})$ оператора совпадает с областью его определения $(D_{\kappa}^{(0)}, G_{\kappa}^{(0)})$ естественным способом нахождения окончательной информативной системы признаков для целевой классификации объектов является итерационный процесс: $K(D_{\kappa}^{(r)}, G_{\kappa}^{(r)}) = (D_{\kappa}^{(r+1)}, G_{\kappa}^{(r+1)})$, $r = 1, \dots, \ell, \ell+1$ и т.д. Сходимость этого процесса подтверждена на множестве численных машинных эксперимен-

тов: (по программам для ЭВМ М-222 и БЭСМ-6) как с модельными числовыми матрицами, так и содержащими реальную геологическую информацию. Отметим, что строгое доказательство сходимости несколько затруднено в связи с нелинейным преобразованием в этом процессе (5) и может следовать из того, что значения целевых требований, во-первых, ограничены, во-вторых, уменьшаются при действии оператора K на каждую отдельно взятую пару эталонных объектов (19, 20), а на множестве пар это уменьшение оценивается в среднем мерой информативности $J_3^{(r)}$ (7).

На каждой итерации фиксируется новая (несколько измененная) система признаков и при соответствующих условиях происходит некоторое увеличение меры общей целевой информативности этой системы в сравнении с начальной системой признаков. Последовательность таких итераций заканчивается на ℓ -м приближении, когда с заданной точностью $P_j^{(\ell)} = P_j^{(\ell+1)} \pm \varepsilon$, $j = 1, \dots, n$ и, следовательно, $J_3^{(\ell)} \approx J_3^{(\ell+1)}$. При этом происходит минимизация пространства признаков (зануление признаков с весом $P_j < \beta$, где, например, $\beta \doteq = 0,05 \cdot 1/n'$ (n' - число ненулевых признаков)). Систему признаков с информационными весами ($P_j^{(\ell)}$) $_n$ будем считать достаточно информативной для совместного целевого разделения и объединения (классификации) эталонных объектов, если мера информативности этой системы $J_3^{(\ell)} > \chi$ (χ - порог, выбираемый в зависимости от требуемой точности приближения к полностью информативной системе признаков).

Распознавание объектов-проб

В связи с тем, что получение и.с.п. в данном методе базируется на приведении мер различия и мер сходства между объектами по характеристическим признакам в соответствии с таковыми по целевому признаку, здесь, согласно трем вариантам выбора и.с.п. (7), рассматриваются три простых решающих правила на основе функций близости.

Пусть на множестве эталонных объектов S_1, \dots, S_m получена и.с.п. определенного вида (7.1, 7.2, 7.3) и требуется провести распознавание некоторого объекта-пробы S^n .

Правило аналогии или позитивная форма
принятия решений

Если получена разделяющая и.с.п. с информационными весами $(R_j^{(e)})_n$ (IO), тогда решение о принадлежности пробы принимается по ближайшему эталонному объекту (по аналогии). То есть для качественного целевого признака определяется, на какой из объектов в эталонных классах (образах) "похожа" данная проба, а для количественного целевого признака — рядом с каким из объектов в эталонной последовательности классов* и в каком диапазоне целевых значений можно указать место пробы. Для этого по формулам (I-4) у каждой пары эталон-проба вычисляется мера сходства:

$$\lambda_i^n = \sum_{j=1}^n R_j^{(e)} \cdot g_{ij}^n, \quad (i = 1, \dots, m) \quad (22)$$

Рассмотрим решающее правило относительно любых двух классов эталонных объектов $S'_1, \dots, S'_m; S''_1, \dots, S''_m$; ($m' + m'' = m$). Целевое расстояние между объектами разных классов $d_{n+1}(S', S'')$ будем считать "ценой" принятия решения по аналогии. В каждом из эталонных классов найдем максимальные меры сходства, фиксирующие ближайшие аналоги $S_{i^*}^n, S_{i^{**}}^n$ пробы S^n ,

$$\max_{i=1, \dots, m'} \lambda_i^n = \lambda_{i^*}^n (S_{i^*}^n); \quad \max_{i'=m'+1, \dots, m} \lambda_{i'}^n = \lambda_{i^{**}}^n (S_{i^{**}}^n) \quad (23)$$

и назовем их коэффициентами принадлежности по "анalogии". Эти величины оценивают максимально возможную "похожесть" пробы на соответствующие классы. Для количественного целевого признака они указывают эталоны, относительно которых у пробы возможны минимальные диапазоны нормированных отклонений по цели, равные

$$\pm (1 - \lambda_{i^*}^n); \quad \pm (\lambda - \lambda_{i^{**}}^n)$$

Наибольший из коэффициентов принадлежности фиксирует для данной пробы S^n ближайший эталонный класс, например, S' . Принадлежность пробы к этому классу будем считать "уверенной", если $\lambda_{i^*}^n \geq \varphi$ (φ — порог "уверенной" принадлежности), $\varphi = \frac{1}{2} \bar{g}'_{n+1}$, см. (27). Сравнительное решение принимается по разности этих максимальных мер сходства (по коэффициенту предпочтения):

$$\Delta^n(S', S'') = \lambda_{i^*}^n - \lambda_{i^{**}}^n \quad (24)$$

*) Класс может быть представлен и одним объектом.

Если проба ближе к классу S' , то $\Delta^n > 0$, в противном случае $\Delta^n < 0$. Значение $|\Delta^n|$ оценивает относительную "контрастность" в решении, где $|\Delta^n| \geq \eta$ - область принятия "уверенных" решений, а $|\Delta^n| < \eta$ - область приближенного решения или отказа от решения (η - фиксированный порог, например, выбирается $\eta = \frac{1}{2} d_{n+1}(S', S'')$).

В случае наличия более двух классов эталонных объектов с различной "ценой" (d_{n+1}) сравнительное (попарное) принятие решений рекомендуется начинать с классов, имеющих $(\lambda_{i*}^n \cdot \lambda_{i**}^n) = \text{maximum}$. Если $|\Delta^n| \geq \eta$, то получаемая "цена" решения $d_{n+1}(S', S'')$ будет минимальной и соответствовать максимально возможной детальности (точности) этого решения.

Правило исключения или негативная форма принятия решений

Если получена объединяющая и.с.п. с информационными весами $(Q_j)_n$ (IO), тогда можно принимать решения о невключении пробы в тот или иной класс (образ). Для этого на основе (I-4) по всем парам эталон-проба вычисляются расстояния:

$$\rho_i^n = \sum_{j=1}^n Q_j \cdot d_{ij}^n, \quad (i = 1, \dots, m) \quad (25)$$

На каждом из эталонных классов (S' , S'') определяется коэффициент исключения пробы как среднее расстояние пробы до класса, взятое с обратным знаком:

$$-\bar{\rho}'^n = -\left(\sum_{i=1}^{m'} \rho_i^n / m'\right); \quad -\bar{\rho}''^n = -\left(\sum_{i=m'+1}^m \rho_i^n / m''\right) \quad (26)$$

Пусть (ξ) - порог исключения. Тогда если $-\bar{\rho}'^n \leq -\xi'$ или $-\bar{\rho}''^n \leq -\xi''$, то принимается решение о невключении пробы в соответствующие классы, в противном случае ($-\bar{\rho}'^n > -\xi'$ или $-\bar{\rho}''^n > -\xi''$) решение принимается приближенно, или делается отказ от решения. То есть, чем больше коэффициент исключения, тем больше возможная определенность негативного решения.

Для количественного целевого признака это означает, что диапазон изменения возможных нормированных отклонений по цели между пробой и эталонным классом сужается и лежит в пределах от $\bar{\rho}'^n$ или $\bar{\rho}''^n$ до I. Среднюю меру сходства эталонов по целевому признаку внутри данного класса S' , в котором $M' = C_{m'}^2$ - число пар срав-

ниваемых объектов, будем считать "ценой" принятия решения по исключительности:

$$\bar{g}'_{n+1} = \sum_{k=1}^{M'} g_{k,n+1} / M' \quad (27)$$

Для нескольких эталонных классов порядок оценки пробы на "исключение" устанавливается по возрастанию: $-\bar{p}'^n < -\bar{p}''^n < \dots$

Отметим, что порог "уверенного" исключения из класса предполагается брать следующим: $-\xi' = \frac{1}{2} \bar{g}' - 1$.

Правило пропорциональности, или совмещенная форма принятия решений

В случае, когда задана и.с.п. для разделения и объединения эталонных объектов пропорционально их целевым различиям $d_{k,n+1}$ и сходствам $g_{k,n+1}$, рассмотренные выше правила принятия решений (23,24,26) совмещаются естественным способом:

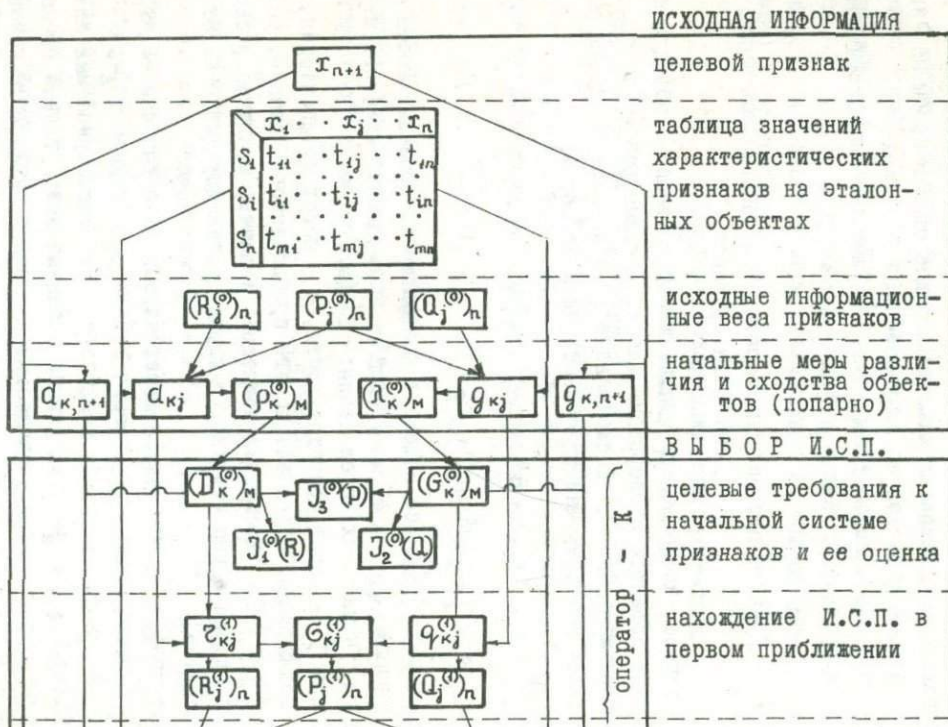
$$V'^n = \lambda_{i^*}^n (S'_{i^*}) - \bar{p}'^n; \quad V''^n = \lambda_{i^{**}}^n (S''_{i^{**}}) - \bar{p}''^n \quad (28)$$

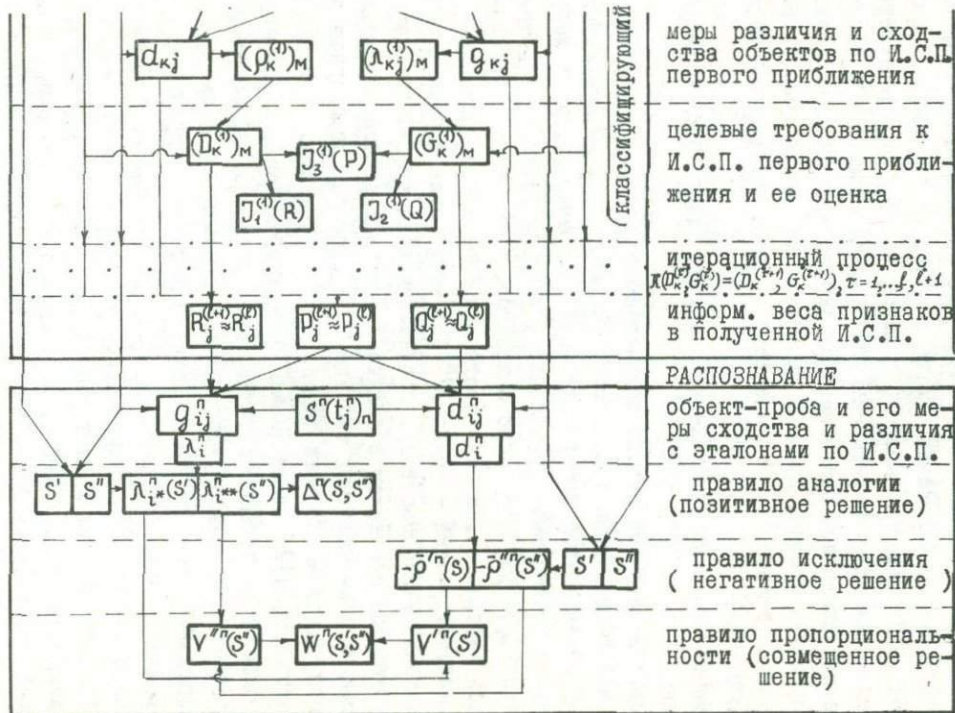
$$W^n(S', S'') = (V'^n - V''^n) / 2 \quad (29)$$

Здесь V'^n и V''^n - совмещенные коэффициенты принадлежности пробы к классам S' и S'' , $W^n(S', S'')$ - совмещенный коэффициент предпочтения. Область их определения: $-1 \leq (V'^n, V''^n, W^{(S', S'')^n}) \leq 1$

Пусть ξ - порог принадлежности, η - порог предпочтения, ξ - порог исключения. Установим пороги приближенного решения, равные $(\frac{1}{2}\xi, \frac{1}{2}\eta, -\frac{1}{2}\xi)$. Рассмотрим последовательность принятия решения, включая приближенные решения, в случае более двух классов эталонных объектов.

Все коэффициенты принадлежности пробы p^n к классам S', S'', S''', \dots располагаются в порядке убывания $V'^n \geq V''^n \geq \dots$. Если $-\frac{\xi}{2} \geq V'^n = \max$, то это означает, что проба исключается из рассматриваемых классов. Если $-\frac{\xi}{2} < V'^n < \frac{\xi}{2}$, то принятие решения относительно принадлежности пробы к классу считается неопределенным. И, наконец, для всех $V'^n \geq V''^n \geq \dots \geq \frac{\xi}{2}$ найдем последовательно $W^n(S', S''), W^n(S', S'''), \dots$. Если $W^n(S', S'') < \frac{1}{2}\eta$, то считается, что проба занимает промежуточное положение между этими классами, а при $W^n(S', S'') \geq \frac{1}{2}\eta$ она классифицируется в S' .





ФУНКЦИОНАЛЬНАЯ БЛОК - СХЕМА АЛГОРИТМА "ЦИКЛ"

Прежде чем проводить распознавание "неизвестных" объектов (проб) с помощью предлагаемых правил принятия решений рекомендуется провести распознавание по заданной системе признаков самих эталонных объектов и тем самым, проверив правильность или неправильность их распознавания, окончательно оценить систему признаков. Процедуру распознавания, когда в качестве проб взяты эталонные объекты, а сравнение каждого объекта на себя не учитывается в принятии решений, мы называем "внутренним экзаменом". Ошибка в случае неправильного распознавания любого эталона будет равна "цене" принятия решения по нему относительно своего класса. Неопределенное решение (отказ) считается равным половине соответствующей ошибки.

Общая функциональная блок-схема метода "Цикл" представлена на рисунке.

На алгоритмической основе предлагаемого метода в машинном языке "Альфа" реализована программа "Цикл-1" для ЭВМ М-222 и БЭСМ-6. В общем случае при работе этой программы на этапе выбора и.с.п. вычислительная трудоемкость линейно растет с увеличением n и C_m^2 , где n - число признаков, а m - число объектов в исходной информации. На базе программы "Цикл-1" подготовлена ее упрощенная модификация - "Цикл-2" (без вычисления попарных мер различия и сходства объектов), предназначенная для выбора информативной системы признаков и получения целевых оценок объектов непосредственно на исходной таблице T . В этом случае с помощью целевого признака может быть задано два и более классов эталонных объектов, причем классы должны быть упорядочены, а допустимыми для шкал всех признаков считаются линейные преобразования вида $x' = ax + b$. В начале работы этой программы значения $(t_{ij})_{m,n+1}$ всех признаков линейно преобразуются к такому виду $(t'_{i,j})_{m,n+1}$, которые лежат в диапазоне $0 \leq (t'_{i,j}) \leq 1$, а все парные линейные связи целевого признака с характеристическими - одного знака (то есть для каждого из характеристических признаков фиксируется направленность кода). Затем, с помощью описанного выше классифицирующего оператора K , в который вместо мер различия $(d_{k,j})$ или сходства $(1 - d_{k,j})$ подставляются значения $t'_{i,j}$ или $(1 - t'_{i,j})$ делаются последовательные приближения к и.с.п. и к получению требуемых целевых оценок эталонных объектов. Хотя оптимальное решение этой задачи может оказаться и невозможным при учете **ТОЛЬ-**

ко парных линейных связей с целевым признаком для фиксации направленности кода характеристических признаков, но известен ряд задач, для которых учет парных связей достаточен даже без процедуры выбора и.с.п. Поэтому применение программы "Цикл-2" в определенных случаях будет весьма полезным, имея в виду и уменьшение вычислительной трудоемкости при выборе и.с.п., а также наглядность получаемых целевых оценок объектов.

✂ ✂ ✂

В заключение дадим краткую характеристику особенностей метода "Целевая итерационная классификация" в решении некоторых основных задач автоматического распознавания образов.

1. С помощью метода "Цикл" проводится целевая обработка информации различной природы, т.е. на множестве исходных объектов обрабатывается непосредственно информация, заданная значениями признаков с различным видом шкал, где цель определяется целевым признаком, а значения признаков на некоторых объектах могут быть неизвестны.

2. В решении наиболее трудоемкой задачи выбора информативной системы признаков (с количественной оценкой каждого из них в этой системе) реализуется способ последовательных приближений относительно фиксированной цели. В последовательных оценках информативностей признаков учитываются как их индивидуальные, так и коллективные особенности в системе для получения требуемых целевых соотношений между объектами в пространстве информативных характеристических признаков. Рассматриваются три различных варианта решения этой задачи: а) требуется только различать объекты разных классов, а сходство объектов в пределах каждого из них необязательно; б) требуется только объединять объекты в пределах каждого класса, а различие объектов разных классов необязательно; в) требуется разделять объекты разных классов и объединять их внутри каждого класса. При этом в машинной реализации метода предусматривается регулирование роли индивидуальных и коллективных особенностей признаков, скорости сходимости итерационного процесса, а также возможность исключения части объектов, для которых не может быть найдена среди исходных признаков информативная

система. С помощью процедуры выбора и.с.п. методом "Цикл" решается важная в практическом отношении задача нахождения минимального или близкого к нему набора признаков, обладающего заданным целевым свойством.

3. Реализуются три принципа принятия решений при распознавании объектов-проб с неизвестными значениями целевого признака, в строгом соответствии с тремя вариантами выбора и.с.п.: а) правило аналогии, если получена и.с.п. для целевого разделения эталонных объектов; б) правило исключения, если получена и.с.п. для целевого объединения эталонных объектов; в) правило пропорциональности, при и.с.п., совмещающей целевое разделение и объединение эталонов. Будучи рассмотренной как задача распознавания, данным методом решается и задача заполнения пропусков в некоторой информационной таблице. В этом случае каждый признак с пропусками последовательно рассматривается в качестве целевого признака.

Автор благодарит А.Н.Дмитриева, Ю.И.Журавлева, Н.Г.Загоруйко, Ю.Л.Васильева, Г.С.Лбова за обсуждение работы и ряд полезных замечаний.

Л и т е р а т у р а

1. 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.1-14.

2. ЗАГОРУЙКО Н.Г. Методы распознавания и их применение. М., "Советское радио", 1972, 192 с.

3. ВОРОНИН Ю.А. и др. Геология и математика (Задачи распознавания в геологии, геохимии и геофизике). Под ред. Фотиади Э.Э., Новосибирск, "Наука", 1970, с.1-224.

4. ДОРОФЕЮК А.А. Алгоритмы автоматической классификации (обзор). - "Автоматика и телемеханика", 1972, № 12, с.78-113.

5. БИШАЕВ А.А. Итерационный способ нахождения информативной системы признаков для целевой классификации объектов. - В кн.: III Всесоюзная конференция по проблемам теоретической кибернетики (тезисы докладов). Новосибирск, 1974, с.185-187.

6. ВАПНИК В.Н., ЧЕРВОНЕНКИС А.Я. Теория распознавания образов (статистические проблемы обучения), М., "Наука", 1974, с.1-416.

7. ЖУРАВЛЕВ Ю.И., НИКИФОРОВ В.В. Алгоритмы распознавания, основанные на вычислении оценок. - "Кибернетика", 1971, № 3, Киев, с.1-11.

8. ЯБЛОНСКИЙ С.В., ДЕМИДОВА Н.Г., КОНСТАНТИНОВ Р.М. и др. Тестовый подход к количественной оценке геолого-структурных факторов и масштабов оруденения. - "Геология рудных месторождений", 1971, № 2.

9. ДМИТРИЕВ А.Н., КРЕНДЕЛЕВ Ф.П., БИШАЕВ А.А., СМЕРТИН Е.А., ШТАТНОВА Т.И. Организация и обработка геологической информации с помощью ЭВМ на основе построения тупиковых тестов. - В кн.: Логико-информационные решения геологических задач. М., "Наука", 1975, с.83-128.

10. БИШАЕВ А.А., КАРБЫШЕВ В.Д. Методика решения задач классификации геологических объектов с помощью ЭВМ (на примере дифференцированных трапповых интрузий). - В кн.: Тезисы семинара "Применение математических методов и ЭВМ в геологии", Алма-Ата, 1974, с.192-195.

11. БИШАЕВ А.А. Метод нахождения целевой информативности признаков. - В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии (тезисы докладов к конференции). Новосибирск, 1973, с.55-56.

12. БОНГАРД М.М. Проблема узнавания. М., "Наука", 1967. 314 с.

13. ЛЕОВ Г.С., КОЖКОВ В.И., МАНОХИН А.Н. Об одном алгоритме распознавания в пространстве разнотипных признаков. - В кн.: Вычислительные системы. Вып.55, Новосибирск, 1973, с.108-111.

14. ЖУРАВЛЕВ Ю.И. Экстремальные задачи, возникающие при обосновании эвристических процедур. - В кн.: Проблемы прикладной математики и механики. М., "Наука", 1971.

15. КОЛЛАТЦ Л. Функциональный анализ и вычислительная математика. М., "Мир", 1969. 421 с.

16. ВАСИЛЬЕВ Ю.Л., ДМИТРИЕВ А.Н. Спектральный подход к сравнению объектов, охарактеризованных набором признаков. - "ДАН СССР", 1972, т.206, № 6, с.1309-1312.

17. СУПЕС П., ЗИНЕС Дж. Основы теории измерений. - В кн.: Психологические измерения. М., "Мир", 1969, с.1-110.

18. ВОРОНИН Ю.А. Введение мер сходства и связи для решения геолого-геофизических задач. - "ДАН СССР", 1971, т.799, № 5, с. с.1011-1014.

С.В.Макаров, Е.А.Смертин

ЦЕНТРИРОВАННАЯ КАЧЕЛЬНАЯ ПРОЦЕДУРА ДЛЯ НАХОЖДЕНИЯ СОГЛАСОВАННОЙ СИСТЕМЫ ИНФОРМАЦИОННЫХ ВЕСОВ

Работа нацелена на дальнейшее расширение применимости и развитие новых математических процедур в методе согласованных оценок строк и столбцов таблиц в анализе геологической информации. В частности, проводится сопоставительный анализ метода главных компонент и качельных процедур метода согласованных оценок, а также излагается новая центрированная качельная процедура для нахождения согласованной системы информационных весов. Большую помощь в ходе написания работы оказал А.Н.Дмитриев, которому авторы признательны за ряд ценных предложений.

Пусть дана таблица наблюдений X размера $m \times n$, в которой элемент x_{ik} есть значение k -ого признака на i -том объекте ($i = 1, \dots, m$; $k = 1, \dots, n$). Для решения задач диагностики и классификации важно уметь оценивать информационные веса признаков и объектов таблицы X , руководствуясь каким-либо принципом. В частности, знание весов столбцов позволяет выделить подсистему наиболее важных признаков и тем понизить размерность задачи. Существует много способов задания информационных весов [1-3]. Один из них вытекает из метода главных компонент (МГК), который сводится в основном к разложению векторов-наблюдений по собственным векторам (с.в.) ковариационной матрицы (к.м.), полученной для таблицы наблюдений X . В качестве веса строки (объекта) можно применять квадрат величины проекции нормированной и центрированной строки на главный с.в. Аналогично можно задать веса признаков, обработав по МГК транспонированную таблицу наблюдений. Перемена ролей строк и столбцов в интерпретации многомерных наблюдений выражает известный дуализм, отмеченный в [4]. Нас будет ин-

тересовать согласование весов объектов и признаков, получаемых как решение двойственных задач, посредством специального объединения этих задач.

Метод главных компонент

Сначала рассмотрим МГК. Введем обозначения:

\bar{x}_i - среднее значение элементов i -той строки;

σ_i^2 - дисперсия элементов i -той строки;

\bar{x}_k, τ_k^2 - то же для k -ого столбца;

u_{ik}, v_{ik} - нормированные (по строке и по столбцу) величины:

$$u_{ik} = \frac{x_{ik} - \bar{x}_i}{\sigma_i} ; \quad v_{ik} = \frac{x_{ik} - \bar{x}_k}{\tau_k} .$$

Вообще говоря, предполагается, что все необходимые усреднения выполняются с весами, которые заранее не известны и подлежат определению в процессе решения задачи (или задаются произвольно).

Ниже в матричной записи приведены операторы.

F, H - операторы центрирования и B, D - операторы нормирования значений весов строк и столбцов соответственно:

$$F = \left[\begin{array}{cccc} 1 - \frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ \frac{1}{n} & 1 - \frac{1}{n} & \dots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & -\frac{1}{n} & \dots & 1 - \frac{1}{n} \end{array} \right] ; \quad H = \left[\begin{array}{cccc} 1 - \frac{1}{m} & -\frac{1}{m} & \dots & -\frac{1}{m} \\ -\frac{1}{m} & 1 - \frac{1}{m} & \dots & -\frac{1}{m} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m} & -\frac{1}{m} & \dots & 1 - \frac{1}{m} \end{array} \right] ; \quad (1)$$

$$B = \text{diag} \sigma_i^{-1} = \begin{bmatrix} \sigma_1^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_m^{-1} \end{bmatrix} ; \quad D = \text{diag} \tau_k^{-1} = \begin{bmatrix} \tau_1^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \tau_n^{-1} \end{bmatrix} \quad (2)$$

Стандартизованные таблицы U, V имеют вид

$$U = \|u_{ik}\| = B \cdot X \cdot F ; \quad V = \|v_{ik}\| = H \cdot X \cdot D \quad (3)$$

К.М. таблиц X и X^T получаются по формулам:

$$\begin{aligned} \text{Cov } X &= V^T \cdot V = D^T \cdot X^T \cdot H^T \cdot H \cdot X \cdot D, \\ \text{Cov } X^T &= U \cdot U^T = B \cdot X \cdot F \cdot F^T \cdot X^T \cdot B^T. \end{aligned} \quad (4)$$

Основные уравнения МГК имеют вид:

$$U \cdot U^T \cdot \alpha = \lambda_\alpha \cdot \alpha, \quad (5)$$

$$V^T \cdot V \cdot \beta = \lambda_\beta \cdot \beta. \quad (6)$$

(α, β - с.в.; $\lambda_\alpha, \lambda_\beta$ - собственные числа).

Уравнения (5) и (6) решаются независимо. Как известно из [4] на главных с.в. α, β достигаются максимальные (относительно всевозможных нормированных векторов α, β) значения функционалов:

$$P = \sum_{k=1}^n \left(\sum_{i=1}^m \alpha_i \cdot u_{ik} \right)^2, \quad (7)$$

$$Q = \sum_{i=1}^m \left(\sum_{k=1}^n \beta_k \cdot v_{ik} \right)^2, \quad (8)$$

причем
$$P_{ext} = \sum_{k=1}^n \left(\sum_{i=1}^m \alpha_i \cdot u_{ik} \right)^2 = \lambda_\alpha \quad (9)$$

$$Q_{ext} = \sum_{i=1}^m \left(\sum_{k=1}^n \beta_k \cdot v_{ik} \right)^2 = \lambda_\beta \quad (10)$$

Слагаемые в (9), (10), стоящие под знаками внешних сумм, суть квадраты первых главных компонент. Они определяют веса строк ($a_1^2, a_2^2, \dots, a_m^2$) и веса столбцов ($b_1^2, b_2^2, \dots, b_n^2$):

$$a_i^2 = \mu^{-2} \cdot \left(\sum_{k=1}^n \beta_k \cdot v_{ik} \right)^2 \quad (i = 1, 2, \dots, m), \quad (11)$$

$$b_k^2 = \nu^{-2} \cdot \left(\sum_{i=1}^m \alpha_i \cdot u_{ik} \right)^2 \quad (k = 1, 2, \dots, n). \quad (12)$$

Здесь и всюду ниже μ, ν - нормирующие множители. В силу (9), (10) имеем: $\mu^2 = \lambda_\beta, \quad \nu^2 = \lambda_\alpha$

Из (11), (12) следует:

$$a_i = \mu^{-1} \cdot \sum_{k=1}^n \beta_k \cdot v_{ik}, \quad (i = 1, 2, \dots, m) \quad (13)$$

$$b_k = \nu^{-1} \cdot \sum_{i=1}^m \alpha_i \cdot u_{ik}, \quad (k = 1, 2, \dots, n) \quad (14)$$

В матричной записи:

$$\left. \begin{aligned} \mu \cdot a &= V \cdot \beta, \\ \nu \cdot b &= U^T \cdot \alpha \end{aligned} \right\} \quad (15)$$

Замечания о стандартизации и сравнение МГК и КП,

Если бы матрицы U, V совпадали (т.е. $U=V=Y$, где Y - об- щее значение U и V), то можно было бы дополнительно положить

$$\alpha_i = \alpha_i \quad (i=1, 2, \dots, m), \quad \beta_k = \beta_k \quad (k=1, 2, \dots, n), \quad (I7)$$

после чего мы получили бы систему

$$\left. \begin{aligned} \mu \cdot \alpha &= Y \cdot \beta, \\ \nu \cdot \beta &= Y^T \cdot \alpha, \end{aligned} \right\}, \quad (I8)$$

положительное решение которой можно принять за набор весов, связанных между собой симметричными уравнениями (I8).

Именно такая ситуация возникает, когда таблица наблюдений X не подвергается стандартизации с помощью операторов F, H, B, D . Тогда $Y=X$ и, следовательно, (I8) переходит в (I9):

$$\left. \begin{aligned} \mu \cdot \alpha &= X \cdot \beta, \\ \nu \cdot \beta &= X^T \cdot \alpha, \end{aligned} \right\}, \quad (I9)$$

откуда следует ($\lambda = \mu \cdot \nu$):

$$\left. \begin{aligned} X \cdot X^T \cdot \alpha &= \lambda \cdot \alpha \\ X^T \cdot X \cdot \beta &= \lambda \cdot \beta \end{aligned} \right\}. \quad (20)$$

Отметим сходство между (20) и системой МГК (5-6).

Если все элементы исходной таблицы X неотрицательные, то, как показано в [5], система (20) имеет неотрицательное решение $\alpha_i \geq 0, \beta_k \geq 0$, и в качестве весов могут быть взяты компоненты векторов α, β . В работе [3] описана качельная процедура (КП) для вычисления α, β , представляющая собой итерационный процесс решения системы (I9), и в этой же работе сформулированы условия сходимости КП к решению спектральной задачи (20). Получаемые с помощью КП веса α, β согласованы между собой в том смысле, что они связаны посредством "симметричной" системы (I9), а вытекающая из (I9) система (20) **представляет** собой объединение двойственных задач типа (5) и (6). Однако, при всей простоте и привлекательности такой способ получения согласованных весов обладает существенным недостатком инвариантности: ввиду отсутствия центрирования результат вычисления весов зависит от выбора нульпунктов признаков. Кроме того, аналогия КП с МГК - неполная,

поскольку в уравнениях МГК фигурируют стандартизованные, а в КП-нецентрированные таблицы, причем к.м. U, U^T и V^T, V не связаны между собой транспонированием (в отличие от матриц $X \cdot X^T$ и $X^T \cdot X$ для КП). Поэтому желательно видоизменить методику, отправляясь от центрированных таблиц наблюдений и сохраняя, по возможности, аналогию с МГК, а также характерную для КП симметричность относительно двойственных задач.

З а м е ч а н и е 1. Если ограничиться только условием (I7), но не требовать, чтобы выполнялось равенство $U = V$, то мы получим систему ($U \neq V$)

$$\left. \begin{aligned} \mu \cdot \alpha &= V \cdot \beta \\ \nu \cdot \beta &= U^T \cdot \alpha \end{aligned} \right\}, \quad (21)$$

которая не имеет нетривиальных решений. Это следует из того, что операторы центрирования F, H обладают свойством:

$$F^2 = F, \quad H^2 = H. \quad (22)$$

З а м е ч а н и е 2. МГК не дает весов непосредственно. Веса вводятся как функции собственных векторов (и наблюдений X); см. (II), (I2). В то же время решение уравнений (20) можно рассматривать как веса, ибо $\alpha \geq 0, \beta \geq 0$. При построении центрированного варианта КП мы сохраним эту особенность.

Введем обобщенные операторы центрирования и нормирования:

$$F = \begin{bmatrix} 1 - \beta_1 & -\beta_1 & \dots & -\beta_1 \\ -\beta_2 & 1 - \beta_2 & \dots & -\beta_2 \\ \dots & \dots & \dots & \dots \\ -\beta_n & -\beta_n & \dots & 1 - \beta_n \end{bmatrix}, \quad H = \begin{bmatrix} 1 - \alpha_1 & -\alpha_2 & \dots & -\alpha_m \\ -\alpha_1 & 1 - \alpha_2 & \dots & -\alpha_m \\ \dots & \dots & \dots & \dots \\ -\alpha_1 & -\alpha_2 & \dots & 1 - \alpha_m \end{bmatrix} \quad (23)$$

(по диагонали - величины вида $1 - \beta_k, 1 - \alpha_i$; вне диагонали - вида $-\beta_k, -\alpha_i$), F, H соответствуют усреднению с весами α_i, β_k .

$$B = \begin{bmatrix} b_1^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & b_m^2 \end{bmatrix}, \quad D = \begin{bmatrix} d_1^2 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & d_n^2 \end{bmatrix} \quad (24)$$

Числа b_i^2, d_k^2 суть масштабные коэффициенты, назначаемые на основе каких-либо "внешних" соображений. В простейшем случае $b_i^2 \equiv d_k^2 = 1$.

После принятия таких операторов "первичной обработки" наблюдений, матрицы U, V выражаются через исходную таблицу X , как и прежде, по формулам:

$$U = B \cdot X \cdot F, \quad V = H \cdot X \cdot D$$

Центрированная качельная процедура (ЦКП)

Рассмотрим систему уравнений:

$$\left. \begin{aligned} \mu \cdot \alpha &= \psi(U; V) \cdot \beta, \\ \nu \cdot \beta &= \theta^T(U, V) \cdot \alpha \end{aligned} \right\} \quad (25)$$

Матрицы ψ, θ (тех же размеров, что и X) суть функции центрированной таблицы наблюдений X ; μ, ν — нормированные множители, подбираемые так, чтобы выполнялась нормировка $\|\alpha\| = \|\beta\| = 1$.

Из (25) следует (полагая $\lambda = \mu \cdot \nu$) новая система

$$\left. \begin{aligned} \psi \cdot \theta^T \cdot \alpha &= \lambda \cdot \alpha, \\ \theta^T \cdot \psi \cdot \beta &= \lambda \cdot \beta. \end{aligned} \right\} \quad (26)$$

Нас интересует только такое ее решение, которое можно принять за весовые векторы: $\alpha \geq 0, \beta \geq 0, \|\alpha\| = \|\beta\| = 1$

Такое решение можно гарантировать, если $\psi \cdot \theta^T$ (а также $\theta^T \cdot \psi$) — матрица Фробениуса (т.е. полуположительная и неразложимая матрица), имеющая неотрицательный собственный вектор, соответствующий наибольшему собственному значению λ ("вектор Фробениуса") (См. [5]).

Конкретизируем матричные функции ψ, θ так, чтобы $\psi \cdot \theta^T$ была матрица Фробениуса.

1) Пусть $\psi = \theta$ и элементы ψ_{ik} матрицы ψ (а также θ) суть величины

$$\psi_{ik} = \sqrt{u_{ik}^2 + v_{ik}^2} \quad (27)$$

При таком выборе ψ, θ уравнения (25) принимают вид, симметричный относительно транспонирования (т.е. относительно перемены ролей между объектами и признаками), подобный (19).

$$\left. \begin{aligned} \mu \cdot \alpha &= \psi \cdot \beta \\ \nu \cdot \beta &= \psi^T \cdot \alpha \end{aligned} \right\} \quad (28)$$

Название "Качельная процедура" оправдано тем, что систему (28) можно решать методом итерации, начиная с какого-либо начального значения α (или β), получая на каждом шаге положительный вектор.

2) Пусть $\psi_{ik} = (x_{ik} - \bar{x}_i)^2$, а $\theta_{ik} = (x_{ik} - \bar{x}_k)^2$

Тогда система уравнений (25) принимает вид

$$\left. \begin{aligned} \mu \cdot \alpha &= U \cdot U^T \cdot \beta, \\ \nu \cdot \beta &= V^T \cdot V \cdot \alpha \end{aligned} \right\}, \quad (29)$$

где $U = X \cdot F$, $V = HX$, а F и H заданы (1).

3) При выборе $\psi_{ik} = (x_{ik} - \bar{x}_k)^2$, $\theta_{ik} = (x_{ik} - \bar{x}_i)^2$ система (25) принимает вид

$$\left. \begin{aligned} \mu \alpha &= V \cdot V^T \cdot \beta \\ \nu \beta &= U^T \cdot U \cdot \alpha \end{aligned} \right\} \quad (30)$$

где U и V заданы в (3).

Процедуры, заданные в формулах (29), (30), запрограммированы и по ним имеются численные решения (см. наст. сборник стр. 160), выполненные с помощью нижеследующей процедуры распознавания [6].

Распознавание с помощью ЦКП

Пусть имеются две группы объектов распознавания $S_1^1, \dots, S_{m_1}^1$ и $S_1^2, \dots, S_{n_2}^2$. Из их описаний обычным путем составим таблицы X_1 и X_2 . Пусть $\rho^\ell = (\rho_1^\ell, \dots, \rho_n^\ell)$ вектор нагрузок столбцов таблицы X_ℓ , $\ell = 1, 2$, подсчитанных по ЦКП. Для произвольной строки $S = (t_1, \dots, t_n)$ определим

$$\bar{z}(\rho^\ell, S) = \sum_{j=1}^n (t_j - \bar{x}_j^\ell)^2 \cdot \rho_j^\ell \quad \text{и} \quad \bar{R}(\rho^1, \rho^2, S) = \frac{z(\rho^1, S)}{z(\rho^2, S)}, \quad (z(\rho^2, S) \neq 0)$$

где \bar{x}_j^ℓ - среднее арифметическое j -го столбца таблицы X_ℓ .

Положим $\alpha_1 = \max_{S_j^1 \in X_1} \bar{R}(\rho^1, \rho^2, S_j^1)$, $\alpha_2 = \min_{S_j^2 \in X_2} \bar{R}(\rho^1, \rho^2, S_j^2)$ и сформулируем

при $\alpha_1 < \alpha_2$ решающее правило:

а) при $\bar{R}(\rho^1, \rho^2, S) \leq \alpha_1$ S относится к классу, представленному таблицей X_1 ;

б) при $\bar{R}(\rho^1, \rho^2, S) \geq \alpha_2$ S относится к классу, представленному X_2 ;

в) при $\alpha_2 > \bar{R}(\rho^1, \rho^2, S) > \alpha_1$ S не распознается.

Если $\alpha_1 \geq \alpha_2$, то решающее правило таково:

- а) при $\bar{R} < \alpha_2$ S относится к X_1 ;
- б) при $\bar{R} > \alpha_1$ S относится к X_2 ;
- в) при $\alpha_1 \geq \bar{R} \geq \alpha_2$ S не распознается.

Т.е. в том случае, если $\alpha_1 = \alpha_2 = \alpha$, то:

- а) при $\bar{R} < \alpha$ S относится к X_1 ;
- б) при $\bar{R} > \alpha$ S относится к X_2 ;
- в) при $\bar{R} = \alpha$ S не распознается.

Содержательно величина $\bar{z}(\rho^{\ell}, s)$ отражает удаленность строки S от "средней строки" таблицы X_{ℓ} , $\ell = 1, 2$. Чем выше значение $\bar{z}(\rho^{\ell}, s)$,

тем дальше строка S отстоит от строки, составленной из значений \bar{x}_j^{ℓ} . Причем коэффициенты ρ_j^{ℓ} характеризуют величину рассеяния значений j -го столбца относительно их среднего арифметического \bar{x}_j^{ℓ} ; так, что чем выше это рассеяние, тем больше значение ρ_j^{ℓ} в ряду $\rho_1^{\ell}, \rho_2^{\ell}, \dots, \rho_n^{\ell}$. Или можно сказать, что чем выше концентрация значений j -го столбца вокруг среднего, тем меньше значение ρ_j^{ℓ} в ряду $\rho_1^{\ell}, \dots, \rho_n^{\ell}$. То же самое можно сказать о величине $\bar{z}(\rho^{\ell}, s)$: чем выше концентрация значений компонент строки S относительно соответствующих компонент средней строки таблицы X^{ℓ} , тем меньше значение $\bar{z}(\rho^{\ell}, s)$. Таким образом, по величине $\bar{z}(\rho^{\ell}, s)$ вокруг строки $(\bar{x}_1^{\ell}, \dots, \bar{x}_n^{\ell})$ формируется некоторое "ядро", которое состоит из таких строк S , которые характеризуются достаточно малым значением величины $\bar{z}(\rho^{\ell}, s)$. Чем дальше отстоят строки S от центра ядра $(\bar{x}_1^{\ell}, \dots, \bar{x}_n^{\ell})$, тем большим значением величины $\bar{z}(\rho^{\ell}, s)$ они характеризуются. Потому ясно, что чем меньше значение отношения $\bar{R}(\rho^1, \rho^2, s)$, тем больше строка S тяготеет к таблице X_1 и наоборот, чем больше значение этого отношения, тем больше S тяготеет к X_2 . Отсюда естественно ожидать, что $\alpha_1 \leq \alpha_2$ или же, если $\alpha_1 > \alpha_2$, то число строк S таблиц X_1 и X_2 , для которых $\alpha_1 \geq \bar{R}(\rho^1, \rho^2, s) \geq \alpha_2$ мало по сравнению с общим числом строк таблиц X_1 и X_2 . Такие соображения предлагаются в качестве обоснования выбора процедуры распознавания.

Подытожим результаты работы.

1. Теоретическое сравнение процедур МГК и КП выявило их сходство.
2. Выявлен недостаток инвариантности весов КП в смысле зависимости весов от выбора нульпунктов признаков.

3. Построение ЦКП позволяет, сохраняя достоинства КП, избежать недостатка инвариантности и прийти к весам, которые зависят от среднего арифметического по признаку.
4. Рассмотрение практических примеров показало, что ЦКП позволяет получить дополнительную информацию о таблице.
5. После центрирования и нормирования матрицы исходных данных отдельно по строкам и отдельно по столбцам получаем две матрицы. Пока неизвестно, будет ли сходимость качельной процедуры в этом случае. Авторами этот вопрос исследуется в другой работе.

Л и т е р а т у р а

1. 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.1-14.
2. СМЕРТИН Е.А. Δ -тесты в задачах тестового распознавания. - В кн.: Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск, 1973, с.65-66.
3. ВАСИЛЬЕВ Ю.Л., ДМИТРИЕВ А.Н. Спектральный подход к сравнению объектов, охарактеризованных набором признаков. - "ДАН СССР", 1972, т.206, № 6, с.1309-1312.
4. АЙВАЗЯН С.А., БЕЖАЕВА З.И., СТАРОВЕРОВ О.В. Классификация многомерных наблюдений. М., "Статистика", 1974. 240 с.
5. ЛАНКАСТЕР К. Математическая экономика. М., "Сов.радио", 1972. 464 с.
6. ДМИТРИЕВ А.Н., КРАСАВЧИКОВ В.О. Программа метода согласованных оценок. - В кн.: Логико-математическая обработка геологической информации, Новосибирск, 1975, с.6-13.

В.В. Бабич, Г.С. Федосеев

ПРОГНОЗНАЯ ОЦЕНКА ЖЕЛЕЗОРУДНЫХ ОБЪЕКТОВ
КОНДОМСКОГО РАЙОНА ГОРНОЙ ШОРИИ

Кондомская группа магнетитовых месторождений представлена двумя основными рудными узлами – Таштагольско-Кочуринским и Шерегешевским. В тектоническом отношении рудные узлы приурочены к Кондомской грабен-синклинали, выполненной в основном осадочно-вулканогенными среднекембрийскими образованиями (мундыбашская свита), прорванными сиенитами, пироксеновыми порфиритами и гранитами. От соседних территорий, сложенных более древними (Pt-Gr) кремнисто-карбонатными толщами, грабен-синклиналь отделена Западным и Восточным разломами. В Кондомском районе выполнен большой объем не только разведочных, но и эксплуатационных работ, однако некоторые участки нуждаются в дополнительном уточнении перспективности на железоруднение. Для решения этой задачи была предпринята попытка привлечь математические средства обработки имеющейся по данному району геолого-геофизической информации*.

Краткая характеристика объектов
и блок-схема решения задачи

Все привлеченные к обработке объекты, под которыми здесь понимаются участки, включающие одно или несколько реальных или предполагаемых рудных тел и представляющие собой вертикальные призмы, размеры которых определяются проекцией рудных тел на дневную поверхность и нижним горизонтом проведенных буровых работ, сведены в три группы – месторождения, рудопроявления и пробы (рис. I). Основой такого деления является значение целе-

*/ Исходная информация представлена геологами Шалымской экспедиции ЗСТУ М.П. Тараймовичем, В.П. Орловым и др.

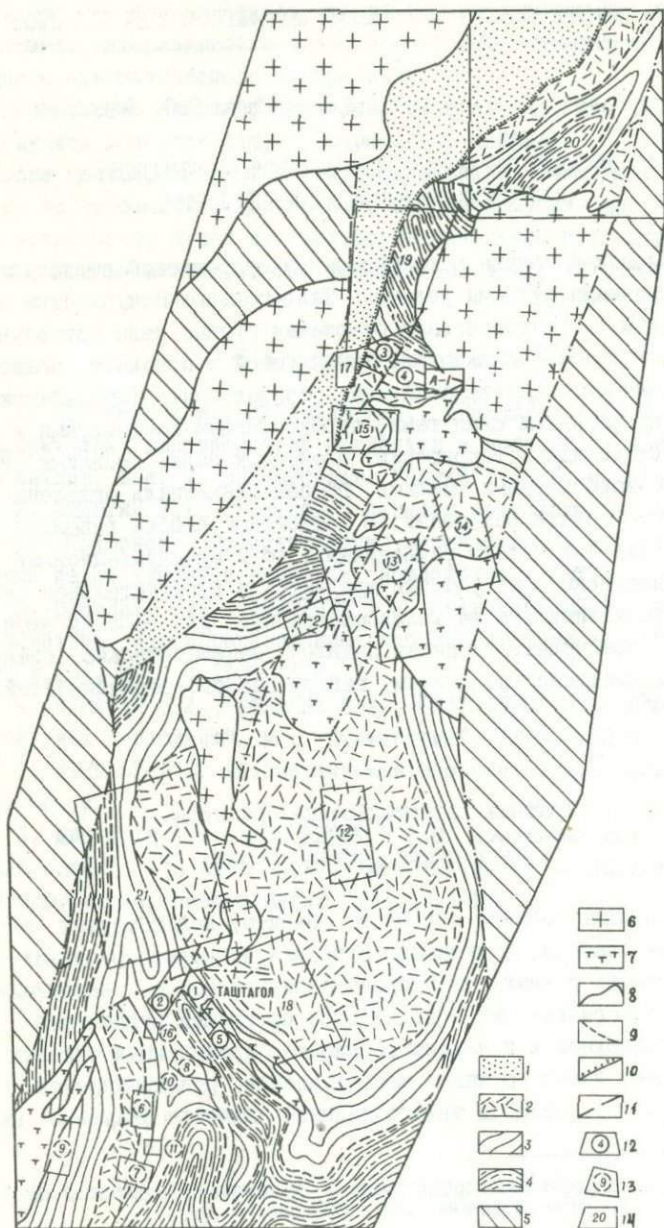


Рис. 1. Схематическая геологическая карта Кондомского железорудного района (по данным М.П. Тараймовича, В.П. Орлова, Е.Р. Фролова).

I - песчаники, сланцы, гравелиты, конгломераты (O_1); 2 - надрудная толща: порфириты средне-основного состава, туфы трахитовых порфиров и альбитофиров ($C_{m_2} m_{n_3}$); 3 - рудовмещающая толща: туфы и туффиты трахитовых порфиров, туфопесчаники, туфоалевролиты, известняки, мергели ($C_{m_2} m_{n_2}$); 4 - подрудная толща: эффузивы, туфы и туфобрекчи андезито-базальтового состава, метаморфические сланцы, туфопесчаники, конгломераты ($C_{m_2} m_{n_2}$); 5 - метаморфические амфиболовые и кремнистые сланцы, микрокварциты с линзами известняков и доломитов ($Pt_3-C_{m_4}$); 6 - граниты, гранодиориты, граносиениты, реже - диориты (D_2); 7 - сиениты (C_{m_2}); 8 - границы интрузивов; 9 - литологические границы; 10 - несогласное залегание; II - тектонические нарушения; I2 - объекты обучения I класса: Таштагол Восточный (1), Таштагол Западный (2), Подрусловый (3), Новый Шерегеш (4) и Юго-Восточный (5); I3 - объекты обучения II класса: 1-й Кочуринский (6), 2 и 3-ий Кочуринские (7), 6-ой Кочуринский (8), 7-ой Кочуринский (9), 5-ый Кочуринский (10), 4-ый Кочуринский (II), Черничный (12), Каменушинский (13), Таензинский (14), Кварцитный (15) и Северо-Кочуринский (16); I4 - объекты-пробы: Юго-Западный Шерегеш (17), Восточно-Таштагольский (18), Северный Шерегеш (19), Кедровка (20) и Тургеневка (21).

вого признака (запасы), которое для месторождений (I класс) укладывается в пределы 40-215 млн. тонн, для рудопроявлений (II класс) - не превышает 10-12 млн. тонн, а для проб - неизвестно.

Рудовмещающими породами в месторождениях являются в основном туфы и туффиты щелочного состава с прослоями известняков. Иногда они прорваны штоками сиенитов и пироксеновых порфиритов. Из дайковых пород развиты микросиениты и лампрофиры. Рудные зоны (мощностью 40-300 м) содержат от одного до восьми рудных тел. В пределах некоторых месторождений рудные зоны выходят на дневную поверхность, в большинстве случаев они прослежены скважинами по простиранию и падению. Рудным телам присущи пластообразные и линзообразные формы и меняющиеся в широких пределах размеры: по простиранию - от 150 до 1000 метров, по падению - от 300 до 1000 метров и мощность от 15 до 45 метров. Из других форм отмечаются крутостоящие штокверки, трубообразные тела и гнезда. Руды преимущественно магнетитовые с примесью гематита. Околорудные изменения выражаются в эпидотизации, скарнировании, хлоритизации, ороговиковании, пропицитизации и фельдшпатизации. Вид магнитных полей обусловлен характером залегания рудной зоны. В местах ее выхода на дневную по-

верхность интенсивность аномалий достигает 60 000 гамм. Залегание зоны на глубине обуславливает аномалии интенсивностью до 2000–4000 гамм, а на Подрусловом и Юго-Восточном месторождениях – до 200–600 гамм.

Рудопроявления размещаются главным образом в крыльях синклиналей и антиклиналей, а единичные объекты приурочены к мульдам складок или зонам трещиноватости. Вмещающие породы – туффиты, туфопесчаники, туфоалевролиты с постепенными переходами в типично осадочные образования (песчаники, алевролиты, мергелистые сланцы и известняки). Иногда в роли вмещающих выступают пропилитизированные окварцеванные туфы трахитовых порфиритов мундыбашской свиты. Интрузивные породы представлены в основном дайковой фацией: диабазовые и пироксен-роговообманковые порфириты, микросиениты, сиенит-порфиры, щелочные кварцевые порфиры и лампрофиры. На некоторых участках довольно широко развиты сиениты, контактовые изменения при этом выражаются в ороговикании, калишпатизации и серицитизации. Интенсивность магнитного поля над объектами с обнажающимися рудными телами обычно равна 1000–5000 гамм, поднимаясь в единичных случаях до 30000 гамм. На объектах с гнездообразным и штокверковым магнетитовым оруденением напряженность поля снижается до 500–900 гамм.

Большинство участков, составляющих пробы, располагается в пределах Шерегешевского рудного поля, и только Восточно-Таштагольский и Тургеневский находятся на северном фланге Таштагольско-Кочуринского рудного поля. Объекты этой группы приурочены к крыльям складок второго порядка, реже они располагаются в ядрах и мульдах складок. Разделение вмещающих пород в значительной степени условное вследствие значительной их эродированности и перекрытия ордовикскими песчано-сланцевыми отложениями, мощность которых иногда достигает 600 метров. Местными геологами выделяется три толщи: нижняя, существенно терригенная (туфогенные песчаники и алевролиты); средняя, терригенно-карбонатная с меняющейся по объему примесью щелочной и субщелочной пирокластике; верхняя, существенно вулканогенная (основные и средние эффузивы). Из магматических образований отмечаются мелкие тела сиенитов, субвулканических порфиритов, щелочных гранитов, гранодиоритов и габбро, дайки лампрофиров, диабазов, диорит-порфиритов, гранит-порфиров и гранит-аплитов. Контактные

изменения проявились в виде ороговивания, скарнирования, щелочного метасоматоза и сульфидизации. Сведения о рудных телах проб ограничены, известно лишь вкрапленное и прожилково-вкрапленное магнетитовое оруденение. Интенсивность магнитных аномалий над объектами-пробами весьма разнообразна: от 150-400 до 1500-4500 с повышением до 15000-20000 гамм.

Решаемая задача формулировалась следующим образом: на основании сравнительного анализа информации о месторождениях, рудопроявлениях и пробах выяснить, к какому из первых двух классов принадлежит каждая проба, а затем ориентировочно оценить значение целевого признака у проб, отнесенных к первому классу. В принятой в работе [1] терминологии это означает, что необходимо провести классифицирование проб с последующим их упорядочением. Блок-схема решения задачи приведена на рисунке 2.

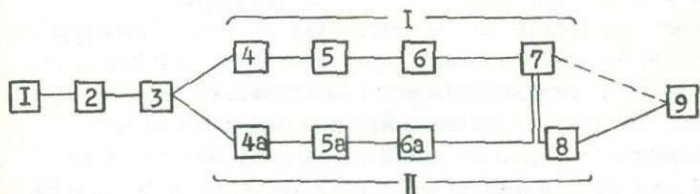


Рис. 2. Блок-схема решения задач классифицирования (I) и упорядочения (II)

I - постановка задачи; 2 - доалгоритмическая обработка информации; 3 - уточнение постановки задачи; 4, 4a - бинарное кодирование информации и получение оценок информативности признаков; 5, 5a - логический анализ закодированной информации (исследование блоковой структуры таблиц); 6, 6a - минимизация признаков пространства (выбор оптимальной подсистемы признаков); 7 - классифицирование проб; 8 - оценка целевого признака (запасов для проб, отнесенных к классу месторождений); 9 - анализ полученных результатов.

Составление паспорта и предварительная минимизация пространства признаков

Структура паспорта, фрагмент которого приводится в таблице I, определилась после тщательного анализа всего фактического материала по объектам Кондомской группы. Она отражает стремление к всесторонней, полной и однородной характеристике объектов, сведения о которых распределены по пяти информационным массивам. Первый массив, содержащий признаки географо-экономического характера, не привлекался для математической обработки. В остальных массивах приводятся сведения об условиях залегания рудных тел, их морфологии, составе горных пород и руд, локальной и региональной тектонике, геофизике и т.д. Группы признаков имеют сквозную нумерацию, а каждому признаку присваивается индекс, указывающий на группу (первое число), номер кодируемого сообщения (второе число) и градацию (третье число — "степень"). Последняя графа отражает тип признака: 1 — качественные, 2 — полуколичественные (порядковые); 3 — количественные; 4 — точечные. Исключение из общего списка некоторых признаков объясняется тем, что они не использовались при математической обработке по разным причинам (неоднозначность в содержательной интерпретации, обилие прочерков, дублирование или взаимоисключение и т.п.). В результате предварительной логической обработки число признаков сократилось с 169 до 85. Оставшиеся признаки были переведены в бинарный вид в соответствии с методикой алгоритмического кодирования [1], упорядочены и сведены в таблицу 2, которая в наибольшей степени удовлетворяет требованиям для формальной обработки информации.

Нетрудно заметить, что в упорядоченной таблице отчетливо проявляется блоковая структура. Блок А (межклассовый критерий общности) содержит признаки, каждый из которых представлен только одним элементом бинарного алфавита (или нулями, или единицами). В идеальном случае все пробы, предъявленные к распознаванию, должны отвечать критерию общности. Однако в практических задачах только немногие пробы удовлетворяют этому условию. В данной задаче, например, таким исключением является Юго-Западный Шерегеш, который проходит по всему набору признаков. Для Восточно-Таштагольского объекта выполняется только 6 при-

Т а б л и ц а I

Фрагмент паспорта ШЭ-72

№ пп	Описание группы признаков	Кодируемые сообщения	Признаки	
			Индекс	Тип
I	2	3	4	5
7	Условия залегания рудных тел	I. Падение:		
		а) вертикальное (80 - 90°)	7-1 ^I	I
		б) крутое (45-80°)	7-1 ²	I
		в) пологое (0-45°)	7-1 ³	I
		г) сложное	7-1 ⁴	I
		2. Ориентировка (по отношению к региональным структурам):		
8	Пластообразная форма рудных тел	а) продольная	7-2 ^I	I
		б) поперечная	7-2 ²	I
		в) диагональная	7-2 ³	I
9	Выход рудной залежи на поверхность		8-2	I
10	Околорудные изменения	I. Скарнирование	9-1	I
II	Вторичные изменения	2. Эпидотизация	10-1	2
		3. Гематитизация	10-2	2
		1. Карбонатизация	10-3	2
		2. Хлоритизация	II-1	2
		3. Сульфидизация	II-2	2
		4. Альбитизация	II-3	2
		5. Калишпатизация	II-4	2
I2	Наличие ордовика	6. Окварцевание	II-5	2
		7. Мраморизация	II-6	2
			II-7	2
I4	Развитие интрузий	I2-2	I	
		I. Сиениты:		
		а) по данным бурения	I4-1 ^I	2
		б) по выходам на поверхность	I4-1 ²	2
		2. Гранитоиды:		
		а) по данным бурения -	I4-2 ^I	2
		б) по выходам на поверхность	I4-2 ²	2

I	2	3	4	5
15	Положение объекта в структурах III порядка	3. Габброиды:		
		а) по данным бурения	I4-3 ^I	2
		б) по выходам на поверхность	I4-3 ²	2
16	Состав рудовмещающей толщи (в % от площади участка месторождения)	1. Рудное тело находится в пределах моноклинали (крыло складки)	I5-1	1
		2. Рудное тело находится в пределах периклинали (замок складки)	I5-2	1
		3. Рудное тело находится в пределах флексурного перегиба	I5-3	1
17	Фациальная изменчивость рудовмещающей толщи	1. Средние и основные эффузивы и их туфы:		
		а) порфириты	I6-1 ^I	2
		б) туфы	I6-1 ²	2
		2. Туфы трахитовых порфиров	I6-2	2
18	Кoeffициент эксплозивности рудовмещающей толщи (объемный)	3. Терригенные породы (туфопесчаники, туффиты, алевролиты и т.д.)	I6-3	2
		4. Известняки и другие карбонатные породы	I6-4	2
			I7	2
19	Метаморфизм рудовмещающей толщи	$K_9 = \frac{\text{туфы}}{\text{эффузивы}}$	I8	2
20	Состав подрудной толщи (в %)	1. Региональный	I9-1	1
		2. Контактный метасоматоз	I9-2	2
21	Состав подрудной толщи (в %)	1. Эффузивы и пирокласты среднего и основного составов:		
		а) андезиты и базальты	20-1 ^I	2
		б) туфы и туфобрекчии	20-1 ²	2
		2. Туфы трахитовых порфиров	20-2	2
		3. Туфопесчаники	20-3	2

I	2	3	4	5
21	Фациальная изменчивость подрудной толщи	4. Содержание железа (Fe_2O_3)	20-4	2
			21	2
22	Кoeffициент эксплозивности подрудной толщи		22	2
23	Региональный метаморфизм подрудной толщи		23	1
24	Состав нарудной толщи (в % от площади участка месторождения)	1. Средние и основные эффузивы и их туфы	24-1	2
		2. Туфы и туффиты трахитовых порфиров и альбитофиров	24-2	2
		3. Туфопесчаники и алевролиты	24-3	2
26	Пространственное соотношение рудных тел и известняков	1. Руда внутри известняков	26-1	1
		2. Руда фациально замещает известняки	26-2	1
		3. Руда по восстановил от известняков	26-3	1
27	Плавный характер горизонтального градиента		27	1
29	Интенсивность аномалий рудного участка		29	1
33	Соотношение текстур (в баллах)	1. Массивная	33-1	2
		2. Полосчатая	33-2	2
		3. Прожилковая	33-3	2
		4. Брекчиевидная, пятнистая и вкрапленная	33-4	2
34	Структура (зернистость)	1. Мелко- и тонкозернистые (≤ 1 мм)	34-1	1
		2. Средне- и крупнозернистые (> 1 мм)	34-2	1
37	Взаимоотношение рудных тел с вмещающими толщами	1. Согласное	37-1	1
		2. Несогласное	37-2	1
		3. Сложное	37-3	1

Продолжение таблицы I

I	2	3	4	5
38	Тип структуры I порядка (грабен-синклиналь)		38-I	I
40	Положение объекта в структурах II порядка	1. В крыльях антиклиналей и синклиналей 2. В ядрах антиклиналей 3. В замковых частях антиклиналей и синклиналей	40-I 40-2 40-3	I I I
47	Совмещение в плане локальных магнитных и гравитационных аномалий	1. Полное 2. Частичное 3. Отсутствие	47-I 47-2 47-3	I I I
48	Геохимический характер рудовмещающих толщ		48	I
49	Наличие магматических пород в регионе	1. Кислые 2. Средние 3. Основные	49-I 49-2 49-3	I I I
50	Возраст интрузивных образований относительно оруденения	1. Кислые: а) дорудные б) пострудные 2. Средние: а) дорудные б) пострудные 3. Основные: а) дорудные б) пострудные	50-I ^I 50-I ² 50-2 ^I 50-2 ² 50-3 ^I 50-3 ²	I I I I I I
52	Дайковые образования	1. Кислые: а) пострудные 2. Средние: а) дорудные б) пострудные 3. Основные: а) дорудные б) пострудные	52-I ² 52-2 ^I 52-2 ² 52-3 ^I 52-3 ²	I I I I I

наков критерия общности и нет возможности предварительно оценить степень выполнимости остальных 12 признаков, что естественно приводит к снижению достоверности результатов его распознавания. Северный Шерегеш, Кедровка и Тургеневка включены в дальнейшую обработку условно, исходя из предположения выполнимости блока отождествления.

Алгоритмическая обработка данных

Обработка исходных данных производилась по алгоритму "Каскад-1" [1,2] с использованием программ для ЭВМ М-222 и БЭСМ-6 [3,4,5]. Первая часть задачи (классифицирование проб) относится к виду А-2 и потому решалась с использованием оценки \tilde{P}_j^y , величины которой, полученные для каждого признака при бинарном кодировании, приведены в последней строке таблицы 2. Признаки блока А, имеющие нулевую оценку и не обладающие межклассовой разделительной способностью в сфере имеющегося материала обучения, были исключены из дальнейшей обработки. Принятые к обработке 67 признаков упорядочивались по абсолютной величине оценки \tilde{P}_j^y , а затем для всей последовательности поднаборов признаков вычислялись величины межклассовых различий ρ^e (рис. 3), что позволило произвести минимизацию признакового пространства. В результате минимизации исходное пространство было сокращено до 31 признака.

Результаты классифицирования проб в объеме минимизированного пространства признаков (рис. 4) показали, что наличие прощеров в описаниях распознаваемых объектов существенно затрудняет принятие окончательного решения по их принадлежности к тому или иному классу. Во всяком случае, уверенно можно говорить об отнесении ко II классу Юго-Западного Шерегеша (17). Что касается классифицирования оставшихся объектов, то оно является в значительной мере условным: решение о принадлежности Верхне-Таштагольского (18) и Северного Шерегеша (19) к первому классу, а Кедровки (20) и Тургеневки (21) - ко второму принимается с учетом положения середины интервала

$$J[\tilde{P}_j^y]_{\max} \div J[\tilde{P}_j^y]_{\min}$$

Пробы, отнесенные к классу месторождений, обрабатывались по алгоритму решения задач упорядочения с привлечением оценки \tilde{P}_j^y . После выбора оптимальных вариантов кодирования всех исход-

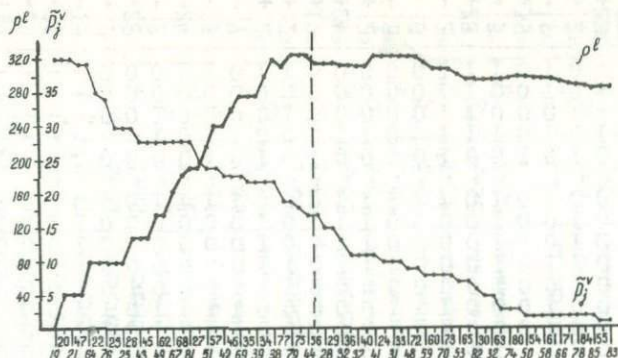


Рис. 3. График минимизации способом упорядочения признаков по информационным весам (P_i^v) при разделении классов (ρ^e). Пунктиром обозначена граница минимизированного набора.

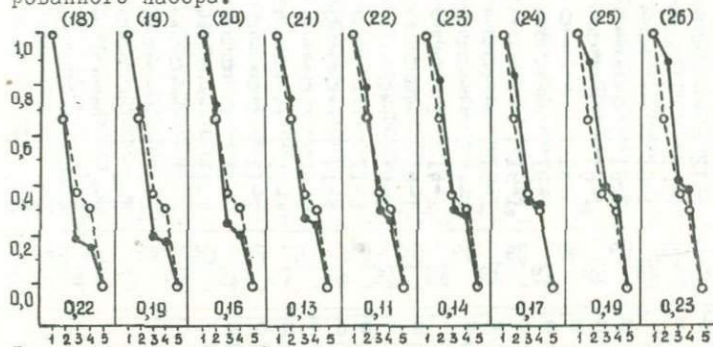


Рис. 5. Фрагмент графика для минимизации признакового пространства при упорядочении объектов. В скобках — количество признаков в наборе, внизу — значение величины λ

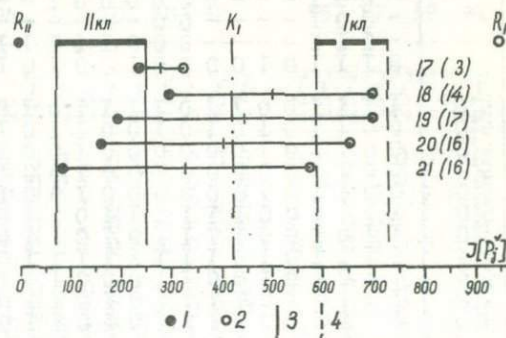


Рис. 4. Результаты классифицирования проб (17-21). В скобках указано число признаков с прочерками. 1 — строчечные нагрузки реальных описаний, 2 — строчечные нагрузки при замене прочерков высокоинформативными значениями, 3 — границы интервала значений $J[P_i^v]$ для объектов II класса, 4 — границы интервала значений $J[P_i^v]$ для объектов I класса.

ных признаков и расположения их в порядке убывания по абсолютной величине оценки $\tilde{\varphi}_j$, для каждого поднабора был вычислен коэффициент точности определения целевого признака - λ . По минимальному значению этого коэффициента определен минимизированный набор из 22 признаков ($\lambda = 0,11$), при котором оценки $J[\tilde{\varphi}_i]$ в наибольшей степени соответствуют целевому признаку (рис. 5).

Т а б л и ц а 3

Оценка перспективных запасов по величине $J[\tilde{\varphi}_j]$

$\begin{matrix} n \\ m \end{matrix}$	52	59	60	35	75	64	47	55	37	72	76	77	62	36	82	40	42	43	45	49	57	58	$J[\tilde{\varphi}_j]$	x_{n+1}
I	0	I	I	0	I	I	I	I	0	I	I	0	0	I	0	I	0	0	I	I	I		I88I	2I6
2	0	0	I	0	I	0	0	0	0	I	0	0	0	I	0	I	0	0	I	0	0		I606	I60
3	I	0	0	I	I	I	0	0	0	I	0	0	0	I	0	0	0	I	I	0	0		68I	I05
4	I	0	0	I	0	0	0	0	0	I	0	0	0	I	0	0	0	I	0	0	0		620	95
5	I	0	0	I	0	0	0	0	0	I	0	0	0	I	0	0	0	I	I	I		0	40	
I8	I	0	0	I	0	0	-	0	0	I	0	0	I	0	I	0	-	-	-	-	I	I	0-340	40-70
I9	0	0	0	I	-	0	-	I	0	-	0	0	I	0	I	0	-	-	-	0	0		466-1039	80-140
$\tilde{\varphi}_j$	-I82	+I82	+I82	-I82	+I73	+I72	+I16	+I16	+I16	+I16	+I16	+I16	+I12	+66	+66	+56	+56	+56	+56	+56	+56			

Примечание: n - номер признака; x_{n+1} - целевой признак; m - номер объекта; $\tilde{\varphi}_j$ - оценка информативности признаков; $J[\tilde{\varphi}_j]$ - оценка строк.

При оценке прогнозных запасов (табл. 3), в силу наличия прочерков у проб, для каждой из них вычислялись минимальная и максимальная оценки строчечных нагрузок, исходя из предположения, что вначале признаки вместо прочерков имеют малоинформативные значения ("0" - если информационный вес $\tilde{\varphi}_j$ положительная величина и "I" - если отрицательная), а затем - высокоинформативные ("0" - если информационный вес признака - отрицательная величина и "I" - если положительная).

Диапазон значений $J[\tilde{\varphi}_j]_{\max} \div J[\tilde{\varphi}_j]_{\min}$ позволяет судить о предполагаемых запасах магнетитовых руд, которые для Восточно-Таштагольского участка оцениваются интервалом 40-70 млн. тонн, а для Северного Шерегеша - 80-140 млн. тонн.

Изложенные результаты решения задачи оценки перспективности на железооруденение объектов Кондомского района Горной Шории позволяют сделать следующие выводы:

I. Привлеченный для обработки геолого-геофизических данных алгоритм "Каскад-I" показал хорошие результаты на этапе обучения. Заданные на обучение классы месторождений и рудопроявлений

полностью разделились, а при упорядочении эталонных объектов было получено очень близкое соответствие значениям целевого признака.

2. При подготовке исходного материала к алгоритмической обработке необходимо стремиться к однородности описания эталонных объектов и проб, поскольку наличие в пробах прочерков по некоторым признакам приводит, во-первых, к ослаблению контроля за их допуском к распознаванию на имеющемся материале обучения, а во-вторых, — к повышению неопределенности при классифицировании и оценке значения целевого признака (запасов).

3. Как показала алгоритмическая обработка, из пяти предъявленных на распознавание объектов Кондомского района наиболее перспективными на железоруднение следует считать Восточно-Таштагольский участок и Северный Шерегеш.

Л и т е р а т у р а

1. БАБИЧ В.В., ФЕДОСЕЕВ Г.С. Метод целевого классифицирования и упорядочения объектов ("Каскад-1"). См. настоящий сборник, с. 42.

2. БАБИЧ В.В., КРАСАВЧИКОВ В.О., ФЕДОСЕЕВ Г.С. Программы метода суммарного учета мер приуроченности и согласования. — В кн.: Логико-математическая обработка геологической информации (оперативно-информационный материал). Новосибирск, 1975, с. 122-126.

3. КАНДЫБА В.Н. Программа "Оценка информативности признаков". — В кн.: Логико-математическая обработка геологической информации (оперативно-информационный материал). Новосибирск, 1975, с. 127-131.

4. БАБИЧ В.В. Программа "Подсчет строчечных нагрузок по заданной информативности для бинарных таблиц". — В кн.: Логико-математическая обработка геологической информации (оперативно-информационный материал). Новосибирск, 1975, с. 135-142.

5. СОКОЛОВ А.Д. Программа "Оптимальное бинарное кодирование признаков, подсчет строчечных нагрузок и минимизация". — В кн.: Логико-математическая обработка геологической информации

(оперативно-информационный материал). Новосибирск, 1975, с.151-158.

В.Д. Карбышев, А.А. Бишаев

ПРОГНОЗИРОВАНИЕ ПРОДУКТИВНОСТИ ДИФФЕРЕНЦИРОВАННЫХ ТРАПОВЫХ ИНТРУЗИЙ СЕВЕРА СИБИРСКОЙ ПЛАТФОРМЫ

Несмотря на открытие Талнахского рудного узла, на севере Сибирской платформы, в том числе и в Норильском районе, интенсивные поиски месторождений медно-никелевых руд продолжают. Масштабы проявления медно-никелевого оруденения, связанного с траповым магматизмом, различны — от промышленных концентраций до исчезающе малых количеств полезных компонентов.

Эмпирически давно установлена связь сульфидного медно-никелевого оруденения с дифференцированными траповыми интрузиями, обладающими определенным набором признаков: высокой степенью дифференцированности, присутствием в строении пикритовых и такситовых габбро-долеритов и т.д. Но эти же признаки нередко характерны и для нерудных интрузий. Поэтому определяющим прямым поисковым признаком служит наличие в породах сульфидного медно-никелевого оруденения (вкрапленного или прожилково-вкрапленного), что является решающим фактором для постановки дальнейших поисково-разведочных работ. Дело значительно усложняется, когда необходимо определить перспективность рудоносных интрузий, т.е. произвести разбраковку интрузий, в связи с которыми обнаружено сульфидное медно-никелевое оруденение. Отсутствие прямых поисковых признаков в этом случае заставляет исследователей искать косвенные поисковые признаки (геолого-структурные, петрохимические, геохимические, минералого-петрографические и т.д.), и обращаться к математическим средствам решения этой задачи.

Первый опыт применения логико-математических методов (тестовый подход) для обработки информации по дифференцированным

трапзовым интрузиям освещен в работах [1,2]. В настоящей статье приводятся результаты решения задачи прогнозирования продуктивности дифференцированных трапзовых интрузий, полученные по более широкому классу этих объектов на базе новых вычислительных приемов^{*/}.

Характеристика исходной информации и блок-схема решения геологической задачи

Основная масса дифференцированных трапзовых интрузий, взятых для исследования, сосредоточена в северо-западной части Сибирской платформы, где их пространственное положение контролируется зонами крупных тектонических нарушений. Здесь же располагаются наиболее крупные и хорошо изученные месторождения данного типа. Другая часть дифференцированных трапзовых интрузий находится в северном и восточном бортах Тунгусской синеклизы.

Степень изученности массивов далеко не одинакова и зависит от целого ряда причин, прежде всего от потенциальной рудоносности. В процессе поисков, разведки и эксплуатации наиболее разнообразная и детальная информация накапливается по промышленным месторождениям, последовательно убывая для объектов, рудоносность которых считается менее перспективной или не представляет практического интереса.

Геологическая задача, подлежащая решению, сформулирована следующим образом: на основе полученной информации средствами математической обработки оценить перспективность дифференцированных трапзовых интрузий севера Сибирской платформы на сульфидное медно-никелевое оруденение по совокупности косвенных признаков и выявить среди них признаки рудоносных интрузий.

Для решения этой задачи было взято несколько десятков дифференцированных трапзовых интрузий, по которым удалось собрать информацию одного уровня изученности. Каждая интрузия была охарактеризована четырьмя группами признаков (табл. 1).

После характеристики каждого объекта вся информация была приведена к табличному виду* (табл. 2), где указаны названия и

^{*/} См. статью А.А. Бишаева в настоящем сборнике.

Т а б л и ц а I

Список признаков, характеризующих дифференцированные трапловые интрузии

I. ГЕОЛОГО-ТЕКТОНИЧЕСКАЯ ОБСТАНОВКА

1. Связь интрузий с глубинными разломами
2. Наличие метасоматитов в экзо- и эндоконтактных зонах
3. Наличие скарнов
4. Присутствие дифференцированных серий лав во вмещающих толщах
5. Наличие пикритовых базальтов во вмещающих вулканогенных образованиях
6. Наличие субщелочных трапов во вмещающих вулканогенных образованиях
7. Туфогенно-лавовая толща и интрузивные породы базальтоидного состава в нижнем экзоконтакте
8. Конгломераты, песчаники, сланцы в нижнем экзоконтакте
9. Карбонатно-глинистые породы, ангидрит, гипс в нижнем экзоконтакте
10. Туфогенно-лавовая толща и интрузивные породы базальтоидного состава в верхнем экзоконтакте
11. Конгломераты, песчаники, сланцы в верхнем экзоконтакте
12. Карбонатно-глинистые породы, ангидрит, гипс в верхнем экзоконтакте.

II. ВНУТРЕННЕЕ СТРОЕНИЕ ИНТРУЗИВНЫХ ТЕЛ

13. Характер дифференцированности (трех- и двухчленное строение)
14. Наличие нижних такситовых габбро-долеритов
15. Наличие верхних такситовых габбро-долеритов
16. Наличие пикритовых габбро-долеритов
17. Присутствие кислых гибридных пород в кровле интрузий
18. Форма интрузий (хонолит или другая)
19. Наличие "эруптивных брекчий" в периферических частях интрузий

III. МИНЕРАЛОГИЧЕСКИЕ ПРИЗНАКИ

20. Железистость оливинов нижних горизонтов
21. Железистость оливинов верхних горизонтов

22. Присутствие ортопироксена в нижних дифференциатах интрузий
23. Основность плагиоклаза нижних горизонтов
24. Наличие биотита
25. Присутствие неправильных выделений титаномagnetита
26. Наличие сульфидной вкрапленности в нижних горизонтах
27. Наличие сульфидной вкрапленности в верхних горизонтах
28. Проявление жильного оруденения
29. Наличие скаполита

IV . ПЕТРОХИМИЧЕСКИЕ ПРИЗНАКИ

30. Магнезиальность предполагаемого исходного расплава (по средневзвешенным или средним величинам в вес. %)
31. Железистость предполагаемого исходного расплава.
(молекул.%)
$$\frac{(FeO + Fe_2O_3) \times 100}{FeO + Fe_2O_3 + MgO}$$
32. Содержание щелочей в предполагаемом исходном расплаве (вес. %)
33. Калиевость предполагаемого исходного расплава (мол.%)
$$\left(\frac{K_2O}{K_2O + Na_2O} \times 100 \right)$$
34. Кремнекислотность (SiO_2 - в атомных количествах) предполагаемого исходного расплава
35. Щелочность-кислотность предполагаемого исходного расплава (в ат.%)
$$\left(\frac{SiO_2 + Al_2O_3}{Na_2O + K_2O + SiO_2 + Al_2O_3} \right)$$
36. Условный ионный потенциал для предполагаемого исходного расплава
37. Магнезиальность наиболее богатых оливином горизонтов (вес.%)
38. Степень магнезиальности наиболее богатых оливином горизонтов (вес.%)
$$\left(\frac{MgO}{FeO + Fe_2O_3 + MgO} \times 100 \right)$$
39. Содержание щелочей в наиболее богатых оливином горизонтах (вес.%)
40. Степень дифференциации интрузий:

$$\Delta K\Phi = \left(\frac{FeO + Fe_2O_3}{FeO + Fe_2O_3 + MgO} \times 100 \right) - \text{(верх. дифференциат)}$$

$$- \left(\frac{FeO + Fe_2O_3}{FeO + Fe_2O_3 + MgO} \times 100 \right) \text{(нижн. дифференциат)}$$

номера объектов, значения и номера признаков (номера признаков соответствуют таковым в таблице I). Затем было проведено целевое разбиение всех исходных интрузий на группы в соответствии с масштабом проявления сульфидного медно-никелевого оруденения и уровнями детальности (подробностью) этого разбиения. Первому уровню детальности соответствует целевое разделение всех интрузий с проявленным медно-никелевым оруденением на группы с промышленным ("месторождения") и с непромышленным ("рудопроявления") масштабами оруденения, и группу "неизвестных" объектов (проб) относительно этого разделения. Второму уровню детальности соответствует целевое разделение всех месторождений на группы: "более значимые" и "менее значимые", и группу "неизвестных" относительно этого разделения объектов (рис. I).

На следующем шаге было осуществлено преобразование геологически поставленной задачи в формализованную. В процессе формализации учитываются во взаимосвязи следующие основные моменты: содержание геологической задачи, целеуказание - целевое требование (сравнительное изучение объектов, классов объектов по принципу сходства, различия), вид целевого признака, вид исходной информации и особенности используемого математического метода [3]. Выделенные при формировании геологической задачи группы интрузий в соответствии с масштабом медно-никелевого оруденения (заданного в виде целевого признака) в формализованной постановке представлены как три последовательных класса эталонных объектов - рудопроявления, менее значимые и более значимые месторождения, и класс проб-неизвестные объекты (см. табл. 2). Блок-схема решения этой задачи показана на рис.2

Алгоритмическое (формальное) решение задачи проведено методом "Цикл" по программам, составленным для ЭВМ М-222.

Способ и результаты решения задачи

В формализованной постановке материал "обучения" представлен восемнадцатью объектами (табл. 2) - пять из класса более значимые месторождения (1,2,3,4,5), четыре - из класса менее значимые месторождения (7,8,9,10), девять - из класса рудопроявления (21,23,25,28,32,34,35,36,37). Остальные объекты (тридцать

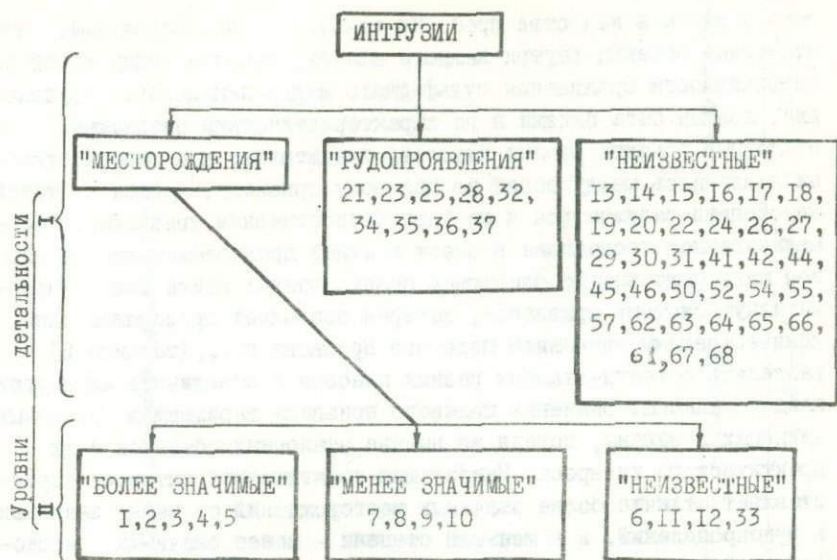


Рис. 1. Априорное разделение объектов на группы и уровни детальности.

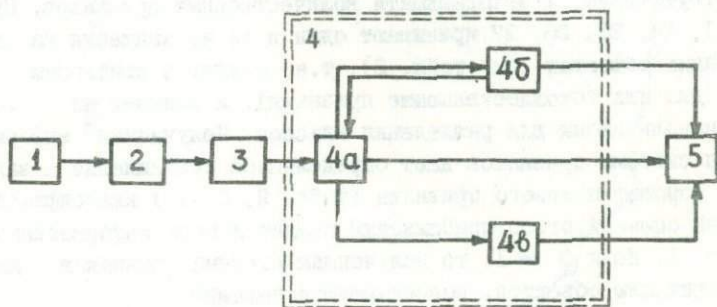


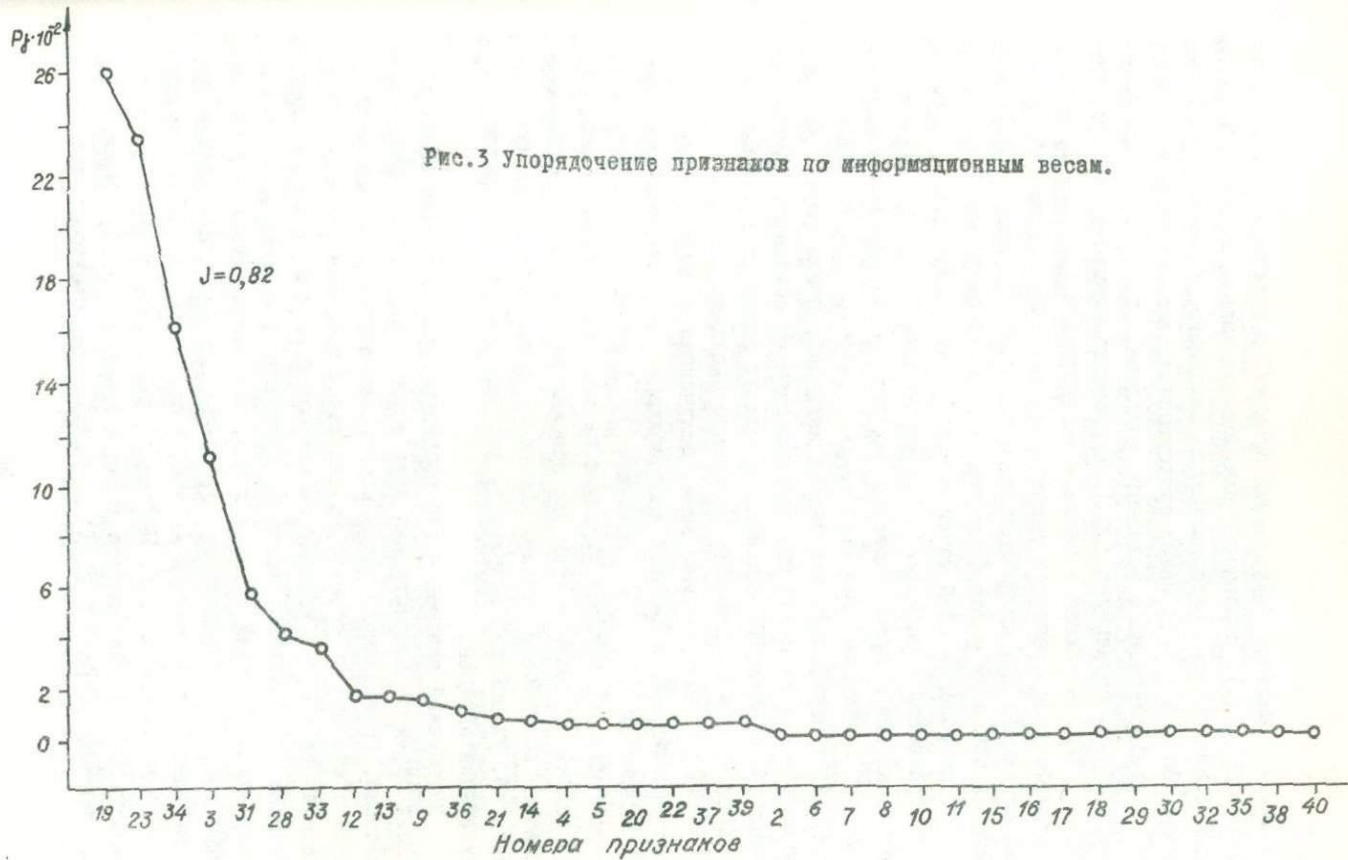
Рис. 2. Блок-схема решения задачи.

1 - формирование геологической задачи; 2 - постановка задачи в формализованном виде; 3 - подготовка информации к решению; 4 - алгоритмическое решение; 4а - нахождение информативной системы признаков, 4б - "внутренний" экзамен, 4в - распознавание проб; 5 - интерпретация результатов решения и выводы.

семь) взяты в качестве проб. Мы исходим из предположений, что эталонные объекты внутри каждого класса, близкие между собой по интенсивности проявления сульфидного медно-никелевого оруденения, должны быть близки и по характеристическим признакам, а эталонные объекты разных классов, значительно и в разной степени различаясь между собой по целевому признаку, должны в такой же степени различаться и по характеристическим признакам. В формализованной постановке в связи с этими предположениями на первом шаге было задано следующее целеуказание: найти такую информативную систему признаков, которая позволяет пропорционально количественным значениям целевого признака τ_{n+1} (таблица 2) разделять объекты-эталонные разных классов и объединять их внутри каждого класса. Значения целевого признака выражены в условных единицах и заданы, исходя из знания эталонных объектов и их практического интереса. Наибольший практический интерес представляет отличие более значимых месторождений от менее значимых и рудопроявлений, и в меньшей степени - менее значимых месторождений от рудопроявлений (соответственно разница между значениями целевого признака).

Информативная система признаков находилась из двадцати одного качественного признака, которые принимают значения: "1" - признак выполнен, "0" - признак не выполнен, прочерк - информация отсутствует, и четырнадцати количественных признаков. Признаки 1, 24, 25, 26, 27 принимают одни и те же значения на всех эталонных объектах (см. табл. 2), т.е. являются критерием общности для них (отождествляющие признаки), и поэтому не могут быть использованы для разделения классов. Получаемая информативная система признаков дает определенное приближение к заданной с помощью целевого признака (табл. 2, τ_{n+1}) классификации. Средней оценкой этого приближения является мера информативности $0 \leq J \leq 1$. Если $J = 1$, то полученная система признаков дает классификацию объектов, совпадающую с целевой.

В нашем случае полученная система признаков дает довольно хорошее приближение к целевой классификации, что выражается соответствующим значением меры информативности $J = 0,82$. В эту систему вошли девятнадцать признаков, из них наиболее существенные (имеющие наибольшие информационные веса P_j , см. рис. 3)



следующие: наличие "эруптивных брекчий" в периферических частях интрузий (x_{19}), основность плагиоклаза нижних горизонтов (x_{23}), кремнекислотность предполагаемого исходного расплава (x_{34}), наличие скарнов (x_3); менее существенные: железистость и калиевость предполагаемого исходного расплава (x_{31} , x_{33}), проявление жильного оруденения (x_{28}), карбонатно-глинистые породы, ангидрид, гипс в нижнем и верхнем экзоконтакте (x_9 , x_{12}), характер дифференцированности интрузивных тел (x_{13}), условный ионный потенциал для предполагаемого исходного расплава (x_{36}); остальные признаки малоинформативные и неинформативные.

С петрологической точки зрения значительная роль целого ряда признаков в разделении "более значимых", "менее значимых" месторождений и рудопроявлений находит вполне определенное объяснение. К таковым относятся наличие скарнов и "эруптивных брекчий" в периферических частях интрузий, пониженная кремнекислотность и железистость предполагаемого исходного расплава, повышение основности плагиоклаза нижних горизонтов и повышенная калиевость предполагаемого исходного расплава.

Как уже было сказано выше, полученная система признаков дает приближение к целевой классификации со значением меры информативности $J = 0,82$. Величина $J < I$ означает, что есть эталонные объекты, которые ухудшают результаты решения поставленной задачи. В связи с этим для дополнительной оценки информативности полученной системы признаков и выявления эталонных объектов, ухудшающих результаты решения задачи, проводился внутренний экзамен.

Внутренний экзамен — это процедура распознавания самих эталонных объектов относительно друг друга. Результаты внутреннего экзамена, проведенного путем последовательного распознавания каждого эталонного объекта по значениям коэффициента его принадлежности (V) к каждому из классов и по коэффициентам предпочтения (W) одного класса по сравнению с другим, приведены в таблице 3. Из этой таблицы видно, что максимальные величины коэффициентов принадлежности эталонные объекты имеют каждый на своем классе, т.е. каждый эталонный объект больше всего похож на объекты своего класса. Например, для объектов первого класса коэффициенты принадлежности лежат в диапазоне $0,878 < V_1 < 0,896$, для объектов второго $0,227 < V_{II} < 0,756$, для объектов третьего

$0.382 \leq V_{III} \leq 0.855$. Заметим, что из всех эталонных объектов только девятый (Норильск-II) и двадцать восьмой (Куломбинская) имеют небольшие по величине коэффициенты принадлежности ($V_{II} = 0.227$, $V_{III} = 0.382$). Окончательное решение об отнесении эталонных объектов к классу принимается по коэффициентам предпочтения (W), т.е. по разности величин принадлежности этих объектов к сравниваемым классам.

Для всех эталонных объектов коэффициенты принадлежности и предпочтения отражают целевые требования, т.е. эталонные объекты внутри своего класса близки между собой, в большей степени различаются с объектами максимально удаленного класса и в меньшей степени различаются с объектами ближайшего класса (в соответствии со средними величинами сходства (\bar{g}_{n+i}) и различия (\bar{d}_{n+i}) по целевому признаку). Например, эталонные объекты первого класса близки между собой ($0.878 \leq V_I \leq 0.896$), в большей степени различаются с объектами третьего класса ($0.585 \leq W_{I/III} \leq 0.642$) и в меньшей степени с объектами второго класса ($0.378 \leq W_{I/II} \leq 0.439$).

Более определенное принятие решений на внутреннем экзамене проводится с использованием порогов значимости (см. стр. 83-85).

1) $\xi_I = \xi_{II} = \xi_{III} = 0.5$ - порог для уверенного принятия решения по коэффициентам принадлежности ($V \geq 0.5$) к каждому из классов, где $I/2 \xi_I = I/2 \xi_{II} = I/2 \xi_{III} = 0.25$ - порог для приближенного принятия решения ($V \geq 0.25$). Если коэффициент принадлежности объекта к одному из классов меньше 0.25, то решение о принадлежности объекта к этому классу не принимается, т.е. неопределенное.

2) $\xi_I = \xi_{II} = \xi_{III} = -0.5$ - порог уверенного исключения объекта из каждого класса по коэффициентам принадлежности ($V \leq -0.5$), где $I/2 \xi_I = I/2 \xi_{II} = I/2 \xi_{III} = -0.25$ - порог приближенного исключения ($V \leq -0.25$). Если коэффициент принадлежности объекта к одному из трех классов больше -0.25, то решение об исключении объекта из этого класса не принимается.

3) $\eta_{I/III} = 0.5$, $\eta_{I/II} = 0.35$, $\eta_{II/III} = 0.15$ - пороги уверенного принятия решения по коэффициентам предпочтения ($|W_{I/III}| \geq 0.5$, $|W_{I/II}| \geq 0.35$, $|W_{II/III}| \geq 0.15$), где $I/2 \eta_{I/III} = 0.25$, $I/2 \eta_{I/II} = 0.175$, $I/2 \eta_{II/III} = 0.075$ - пороги для приближенного принятия решения. Если коэффициент предпочтения (W) объекта для пары сравниваемых классов меньше по абсолютной величине соответствующего из

Т а б л и ц а 3
Результаты внутреннего экзамена

клас- сы объек- тов	По полученной системе признаков				По исходной системе признаков			
	№ об- екта	V_I	V_{II}	V_{III}	№ об- екта	V_I	V_{II}	V_{III}
первый	1	0,878	0,021	-0,395	1	0,744	0,570	0,002
	2	0,896	0,140	-0,276	2	0,760	0,548	0,032
	3	0,895	0,102	-0,310	3	0,839	0,485	0,104
	4	0,878	0,000	-0,406	4	0,830	0,467	0,115
	5	0,892	0,113	-0,278	5	0,651	0,471	0,150
второй	7	0,004	0,756	0,445	7	0,558	0,700	0,232
	8	0,000	0,745	0,437	8	0,528	0,714	0,271
	9	0,186	0,227	0,197	9	0,481	0,389	0,240
	10	-0,064	0,670	0,500	10	0,545	0,719	0,232
третий	21	-0,333	0,403	0,730	21	0,205	0,261	0,422
	23	-0,541	0,368	0,791	23	0,029	0,212	0,426
	25	-0,435	0,412	0,786	25	0,044	0,380	0,432
	28	-0,145	0,157	0,382	28	0,181	0,037	0,181
	32	-0,302	0,516	0,691	32	0,195	0,338	0,272
	34	-0,441	0,478	0,718	34	-0,023	0,187	0,268
	35	-0,578	0,331	0,773	35	-0,346	-0,219	0,328
	36	-0,548	0,355	0,855	36	-0,063	0,078	0,621
	37	-0,592	0,320	0,854	37	-0,002	0,150	0,631

V_I, V_{II}, V_{III} - коэффициенты принадлежности эталонных объектов к первому, второму и третьему классам.

порогов $|w_{I/II}| < 0.25$, $|w_{I/III}| < 0.175$, $|w_{II/III}| < 0.075$, то решение о предпочтении одного из классов для этого объекта не принимается. При этом, если коэффициент принадлежности объекта $V \geq 0.25$ к одному из двух классов, то считается, что объект занимает промежуточное положение между ними.

Для тринадцати эталонных объектов выполнены все пороги для уверенного принятия решений. Уверенное решение о принадлежности к своему классу и о предпочтении своего класса в сравнении с

первым выполняется для трех объектов (10, 32, 34), но решение о предпочтении своего класса в сравнении с третьим для десятого объекта и со вторым для тридцать второго и тридцать четвертого объектов принимается приближенно. Для двадцать восьмого объекта (Куломбинская) выполнены пороги приближенного принятия решений как по принадлежности, так и по предпочтительности. Исключение составляет девятый объект (Норильск-П), для которого не выполнены пороги для приближенного принятия решений, можно лишь говорить о тенденции распознавания его в свой класс. В соответствии с этим, ошибка распознавания объекта Норильск-П на внутреннем экзамене равна половине меры различия между вторым и первым классами по целевому признаку ($1/2 \bar{a}_{n+1} (П-I) = 0.35$).

Результаты разделения эталонных объектов на внутреннем экзамене по информативной системе признаков можно сравнить с начальным разделением (табл. 3), полученным на основании исходной системы признаков, т.е. когда информационные веса всех признаков равны между собой. При самом общем рассмотрении видно, что результаты начального разделения значительно хуже: все величины коэффициентов принадлежности эталонных объектов каждого к своему классу меньше по абсолютной величине; три объекта Норильск-П, Куломбинская, Моронго (номера объектов 9, 28, 32) имеют максимальные коэффициенты принадлежности не на своих классах, т.е. дают ошибку в распознавании; неопределено решение о предпочтительности своего класса для трех объектов первого класса (1, 2, 5), трех объектов второго класса (7, 8, 10) и для четырех объектов третьего класса (21, 23, 25, 34). Общая ошибка на внутреннем экзамене по исходной системе признаков равна 5,45, т.е. в несколько раз больше, чем по полученной системе признаков.

Определенный интерес представляет в данном случае вопрос поведения коэффициента корреляции, рассчитанного для результатов внутреннего экзамена по исходной и полученной системам признаков. Корреляция считалась между коэффициентами принадлежности (V_0), полученными в соответствии со значениями целевого признака - "идеальный" внутренний экзамен, и коэффициентами принадлежности из таблицы 3 (V_1 - коэффициент принадлежности по исходной системе признаков, V_2 - по полученной системе признаков). Сравнение двух коэффициентов корреляции $r_{V_0 V_1} = 0,56$, $r_{V_0 V_2} = 0,95$ по χ^2 -критерию, после соответствующего Z преобразова-

ния Фишера показало, что эти коэффициенты значительно различаются, при величине второго близкого к единице. Таким образом, результаты внутреннего экзамена по полученной системе признаков значительно лучше соответствуют "идеальному" внутреннему экзамену. При более детальном исследовании этого вопроса, характеристики такого типа могут оказаться полезными, в дополнение к мере информативности J , для оценки системы признаков сверху по результатам внутреннего экзамена.

Учитывая вышесказанное, полученная система признаков считается вполне информативной для требуемого целевого объединения и разделения эталонных объектов и используется затем для распознавания объектов-проб к одному из трех эталонных классов. Результаты распознавания проб приведены в табл. 4.

Уверенное решение о принадлежности к первому классу принимается для трех интрузий: Ханарская ($v_I = 0.627$), Имангда ($v_I = 0.556$), Пясинско-Вологочанская ($v_I = 0.539$). Из них Ханарская и Пясинско-Вологочанская распознаются приблизительно по коэффициентам предпочтения ($|w_{I/II}| \geq 0.25$, $|w_{I/III}| \geq 0.175$) в первый класс, Имангда распознается приблизительно по коэффициенту предпочтения ($|w_{I/III}| = 0.314$) в первый класс по сравнению с третьим и неопределенно по сравнению со вторым ($|w_{I/II}| < 0.175$), т.е. она занимает промежуточное положение между вторым и первым классами, тяготея к классу более значимых месторождений. По другой группе интрузий Верхне-Кунтыкахская, Накохоз и Средне-Чангадинская принимается приближенное решение о принадлежности их к первому классу ($0.25 \leq v_I < 0.5$), решение о предпочтительности первого класса для этих интрузий не принимается (решение неопределенное). Последние интрузии занимают (менее уверенно, чем Имангда) промежуточное положение между первым и вторым классами.

Уверенное решение о принадлежности ко второму классу (v_{II}) принимается для четырех интрузий: Нижне-Норильская ($v_{II} = 0.711$), Хупери ($v_{II} = 0.642$), Арылахская ($v_{II} = 0.594$), Горстроевская ($v_{II} = 0.553$); и приближенное для трех: Зеленая Грива ($v_{II} = 0.479$), Ручей Магнитный ($v_{II} = 0.473$) и Борговая ($v_{II} = 0.425$). Из них только Нижне-Норильская уверенно по коэффициентам предпочтения распознается в свой класс ($|w_{II/I}| > 0.35$, $|w_{II/III}| > 0.15$). Интрузии Хупери, Арылахская, Горстроевская, Зеленая Грива и Ру-

чей Магнитный значимо отличаются от первого класса ($|w_{п/г}| > 0.175$), решение о предпочтительности второго класса по сравнению с третьим для этих интрузий неопределенное ($|w_{п/ш}| < 0.075$), т.е. можно предположить промежуточное положение этих интрузий между вторым и третьим классами. Определенный интерес вызывает результат распознавания для интрузии Бортовая, которая уверенно отличаясь от третьего класса ($|w_{п/ш}| > 0.15$), занимает промежуточное положение между вторым и первым классами, тяготея к классу менее значимых месторождений. Эта интрузия является краевой частью Талнахской интрузии (эталонный объект класса более значимые месторождения) и отнесение ее к классу месторождений указывает на то, что полученная информативная система признаков учитывает те особенности интрузии, которые свидетельствуют о связи всей интрузии в целом с богатым сульфидным медно-никелевым оруденением.

По девятнадцати интрузиям (I3, I4, I5, I6, I9, 24, 26, 27, 30, 3I, 54, 4I, 42, 45, 46, 50, 52, 57, 67) принимается уверенное решение о принадлежности их к третьему классу. Шестнадцать из них распознаются в третий класс по предпочтению с первым уверенно и три (I4, 26 и 54) — приближенно. Шестнадцать интрузий уверенно по коэффициенту предпочтения отличаются от второго класса, одна интрузия (45) — приближенно и две (26 и 54) занимают положение между третьим и вторым классами. Приближенное решение о принадлежности к третьему классу принимается для трех интрузий (6, 22 и 65), из них две (22, 65) приближенно отличаются от первого и второго класса, а Мантуровская (6) занимает промежуточное положение между вторым и третьим классами.

По двум интрузиям (I7 и 44) решение о принадлежности их к тому или иному эталонному классу не принимается. При этом для интрузии Макус (44) решение вообще неопределено, а для интрузии Мирдзинская (I7) принимается приближенное решение об отличии ее от всех трех эталонных классов.

Результаты решения поставленной геологической задачи позволяют сделать следующие основные выводы:

I). Методом "Целевой итерационной классификации" получена информативная система признаков, которая позволила уверенно разделить эталонные классы дифференцированных трапсовых интрузий в соответствии с заданным целевым признаком. Полученная система

Т а б л и ц а 4

Результаты распознавания проб

реше- ние по V	№№ объ- екта	V_I	V_{II}	V_{III}	реше- ние по V	№№ объ- екта	V_I	V_{II}	V_{III}				
к первому классу	62	0,627	0,246	-0,052	к третьему классу	16	-0,523	0,231	0,737				
	63	0,539	0,193	-0,191		19	-0,482	0,032	0,562				
	II	0,556	0,377	-0,072		24	-0,491	0,452	0,760				
	20	0,436	0,306	0,019		27	-0,648	0,262	0,699				
	33	0,388	0,341	-0,022		30	-0,566	0,382	0,758				
	55	0,310	0,196	0,060		31	-0,527	0,418	0,768				
ко второму классу	61	0,397	0,425	-0,054		41	-0,800	0,144	0,612				
	68	-0,101	0,711	0,402		42	-0,780	0,163	0,681				
	I2	-0,301	0,553	0,534		46	-0,734	0,206	0,639				
	18	0,199	0,594	0,482		50	-0,636	0,308	0,772				
	29	-0,172	0,642	0,545		52	-0,751	0,192	0,645				
	64	0,064	0,479	0,415		57	-0,812	0,133	0,634				
	66	-0,174	0,473	0,454		67	-0,542	0,369	0,733				
к третьему классу	6	-0,043	0,375	0,429		22	-0,448	0,054	0,499				
	26	-0,323	0,551	0,599		65	-0,172	0,284	0,475				
	54	-0,244	0,627	0,643	не принимается	I7	-0,412	-0,406	-0,362				
	45	-0,375	0,399	0,627						44	0,213	0,200	0,014
	I3	-0,651	0,098	0,558									
	I4	-0,450	0,020	0,519									
	I5	-0,589	0,343	0,702									

V_I , V_{II} , V_{III} - коэффициенты принадлежности эталонных объектов к первому, второму и третьему классам.

признаков может быть рекомендована как совокупность косвенных поисковых признаков рудоносных интрузий.

2). Произведена разбраковка дифференцированных траптовых интрузий из класса проб на группы объектов, преимущественно отнесенных к первому, второму и третьему классам, и на группы объектов, занимающих промежуточное положение между первым и вторым, между вторым и третьим классами.

3). Выявлены объекты первоочередной важности. К классу наиболее значимых месторождений отнесена интрузия Ханарская. Ука-

занный объект может быть рекомендован для детальных разведочных работ. Заслуживают внимания и дальнейшего изучения объекты, идущие по важности вслед за первоочередным, а именно: интрузии Пясино-Вологочанская и Имангдинская.

Л и т е р а т у р а

1. ДМИТРИЕВ А.Н., ЗОЛОТУХИН В.В., ВАСИЛЬЕВ Ю.Р. Опыт применения дискретной математической обработки информации по дифференцированным рудоносным трапповым интрузиям северо-запада Сибирской платформы. - "Советская геология", 1968, № 12, с. 98-108.
2. ВАСИЛЬЕВ Ю.Р., ДМИТРИЕВ А.Н., ЗОЛОТУХИН В.В. Распознавание и оценка никеленосных трапповых интрузий севера Сибирской платформы. - "Геология и геофизика", 1973, № 1, с. 13-23.
3. БИШАЕВ А.А., КАРБЫШЕВ В.Д. Методика решения задач классификации геологических объектов с помощью ЭВМ (на примере дифференцированных трапповых интрузий) - В кн.: Тезисы докладов на семинаре "Применение математических методов и ЭВМ в геологии", Алма-Ата, 1974, с. 192-194.

В.А. Каштанов, А.Д. Соколов

ПРОГНОЗ ПРОДУКТИВНОСТИ ЛОКАЛЬНЫХ ПОДНЯТИЙ ДО ИХ ВВОДА В БУРЕНИЕ (НА ПРИМЕРЕ МЕЗОЗОЙСКО-КАЙНОЗОЙСКИХ ОТЛОЖЕНИЙ СЕВЕРА ЗАПАДНО-СИБИРСКОЙ ПЛИТЫ)

В пределах осадочного чехла Западно-Сибирской плиты широко распространены положительные пликативные дислокации III порядка - локальные поднятия. Такие структуры с достаточной степенью достоверности выявляются при помощи детально разработанных методов сейсморазведки, которыми в пределах исследуемой территории выделено несколько региональных отражающих сейсмических поверхностей

(горизонтов). Основными являются следующие пять горизонтов: Ф - поверхность доюрского фундамента; Па - кровля отложений заводоуловской серии ($J_1 - J_2$); Ш - кровля отложений альмской свиты и ее аналогов (K_1); LV - кровля покурской свиты и ее аналогов (K_2); У - кровля отложений ганькинской свиты и ее аналогов (граница K_2 и P_9). В северной части Западно-Сибирской плиты, преимущественно в отложениях сеномана, открыты крупнейшие и уникальные залежи газа и газоконденсата, приуроченные к локальным поднятиям. Нижние горизонты этих структур в силу больших глубин залегания за редким исключением не были опробованы. В последнее время бурением вскрыты средние и крупные по масштабам залежи нефти и газоконденсата в юрских и нижнемеловых отложениях на Арктической, Новопортовской, Уренгойской, Пурпейской, Северо-Варь-Еганской, Ярайнерской и других площадях. Все это предопределяет большие перспективы севера Западно-Сибирской плиты на возможное обнаружение здесь залежей нефти и газа в нижних горизонтах. В связи с этим решение задач по прогнозу продуктивности локальных поднятий на добуровой стадии изученности имеет большое практическое значение.

Решаемая задача формулируется следующим образом: на основе изучения эталонных объектов - продуктивных и "пустых" структур произвести оценку перспективности менее изученных в этом отношении объектов с использованием логико-математических методов обработки информации. Оценивать ожидаемые запасы для проб, отнесенных в разряд перспективных, не требуется. Следовательно, решаемая задача по виду целевого задания соответствует задаче классифицирования с двухклассовой постановкой [1, 2].

Приуроченность основной массы месторождений нефти и газа к локальным поднятиям, достоверность и надежность их выявления геофизическими методами, предопределили выбор объектов исследования, положенных в основу этой работы. Исходная геологическая информация представлена описаниями объектов обучения, объектов внешнего экзамена и проб. Обучающая выборка включает в себя два класса: I класс - месторождения (20 объектов), II класс - "пустые" структуры (24 объекта). Месторождениями считаются структуры, имеющие залежь с промышленными запасами полезного

компонента. К "пустым" структурам относятся объекты без промышленных притоков нефти и газа. Контрольная выборка, состоящая из объектов с известной принадлежностью к классам, предназначена для выяснения качества распознавания применяемого алгоритма на заданном материале обучения.

При выборе объектов обучения и внешнего экзамена помимо близкой пространственной группировки месторождений и "пустых" структур были учтены и данные по количественному распределению залежей в районе, поэтому предпочтение было отдано газовым залежам сеноманского возраста. В качестве объектов внешнего экзамена были взяты: Арктическое, Нидинское, Уренгойское, Губкинское, Северо-Варь-Еганское, Сузунское и Вахтинское месторождения, и Большелайдинская, Нань-Яхская, Звездная, Октябрьская, Кутоп-Еганская, Асомкинская, Вершинная-I и Пойменная "пустые" структуры. Объекты прогноза (пробы) представлены I -Харамшурским, II - Часельским, III -Толькинским, IV - Каралькинским, V - Верхнеартюгинским, VI - Русскореченским, VII - Советским, VIII - Вадинским, IX - Ванкорским, X - Пайяхским и XII - Верхне-Кубинским локальными поднятиями.

Если ранее (см. например [3-4]) локальные поднятия характеризовались только по одному из сейсмических горизонтов, соответствующему основному продуктивному горизонту того или иного региона, то в решаемой задаче локальное поднятие рассматривалось как геологическое тело, прослеживаемое по всем сейсмическим горизонтам, начиная с горизонта "Ф", а информация об объектах исследования снималась с одномасштабных структурных карт одного и того же авторского коллектива.

В основу признакового пространства, состоящего из 134 признаков, объединенных в 4 группы, были положены структурно-тектонические построения, как наиболее четко отражающие наличие или отсутствие нефтегазоносности осадочного чехла Западно-Сибирской плиты (табл. I). Признаки группы А введены для выяснения влияния унаследованности структурного плана на нефтегазоносность. Признаки группы Б характеризуют: 1) гипсометрическое положение локального поднятия в пределах структурно-тектонических зон; 2) влияние региональных впадин на нефтегазоносность; 3) возможность катагенетических преобразований углеводородов в зависимости от глубины погружения осадочных толщ; 4) гидродинамический

Таблица I
Объединенный набор структурно-тектонических признаков

Описание группы признаков	Номер признака	Содержание	Градация	Значение	Тип ¹
I	2	3	4	5	6
А. Общая характеристика структурно-тектонического плана региона в пределах контура локальных поднятий по основным сейсмическим горизонтам (Ф, Па, Ю, Юв, Юа)	1-5	Наличие структурной ловушки на региональных структурных картах	1	Есть	I
	6-9	Унаследованность структурного плана по смежным сравниваемым горизонтам	0	Нет	I
			1	Есть	
	10-13	Вид унаследованности структурного плана по смежным сравниваемым горизонтам	0	Нет	2
			1	Поднятия по обоим горизонтам	
			2	Поднятие по нижнему, моноклинали по верхнему горизонтам	
			3	Поднятие по нижнему, впадина по верхнему горизонтам	
			4	Моноклинали по нижнему, поднятие по верхнему горизонтам	
			5	Впадина по нижнему, моноклинали по верхнему горизонтам	
	14-19	Унаследованность структурного плана по горизонтам, удаленным друг от друга более чем на один	1	Есть	I
			0	Нет	
	20-25	Вид унаследованности структурного плана по горизонтам, удаленным друг от друга более чем на один	1	Поднятие по обоим горизонтам	2
			2	Поднятие по нижнему, моноклинали по верхнему горизонтам	
3			Поднятие по нижнему, впадина по верхнему горизонтам		
4			Моноклинали по нижнему, поднятие по верхнему горизонтам		
5			Впадина по нижнему, моноклинали по верхнему горизонтам		
6			Моноклинали по обоим сравниваемым горизонтам		
26-30	Расстояние от центра л.п. или его проекции на соответствующий горизонт до ближайшей зоны инклинивания (км)			3	
Б. Гипсометрическое положение локального поднятия в пределах структурно-тектонических зон	31-35	Положение л.п. или его проекции на соответствующий горизонт в пределах региональной зоны	1 ^{жж}	Внутренняя мегасинеклиза	2
			2	Внешняя мегасинеклиза	
			3	Сложная моноклинали	
			4	Простая моноклинали	

1	2	3	4	5	6
	36-45	Абсолютные отметки региональных изогипс (min, max), околтуривающих структурно-тектоническую зону, в пределах которой находится л.п. или его проекция на соответствующий горизонт (км)			3
	46-55	Расстояние от центра л.п. или его проекции на соответствующий горизонт до региональных изогипс (min, max), околтуривающих структурно-тектоническую зону (км)			3
	56-65	Превышение вершины л.п. или его проекции на соответствующий горизонт над региональными изогипсами (min, max), околтуривающими структурно-тектоническую зону (км)			3
	66-70	Амплитуда л.п. (только по изогипсам, околтуривающим л.п.) (км)			3
	71-75	Абсолютные отметки нижней замкнутой изогипсы, околтуривающей л.п. или его проекции на соответствующий горизонт (км)			3
	76-80	Абсолютные отметки вершины л.п. или его проекции на соответствующий горизонт (км)			3
	81	Расстояние от центра л.п. до наиболее погруженной части ближайшей впадины по основному нефтепроизводящему горизонту (км)			3
	82-86	Превышение вершины л.п. или его проекции на соответствующий горизонт над наиболее погруженной частью ближайшей впадины (км)			3
	87-91	Максимальная глубина ближайшей впадины (км)			3
В. Динамические характеристики структурного плана локальных поднятий	92-101	Прирост глубины ближайшей впадины за раннеюрский, поздне-юрско-неокомский, аптосенноманский, турон-палеоценовый и другие этапы тектонического развития региона (км)			3
	102-103	Прирост амплитуды л.п. за этапы (см. признак 92-101) тектонического развития региона (км)			3
	112-121	Расстояние между подошвами л.п. по сравняваемым горизонтам (км)			3
	122-131	Расстояние между вершинами л.п. по сравняваемым горизонтам (км)			3
Г. Отождествляющие признаки	I32	Тип ловушек	0	Неструктурный	2
	I33	Возраст нефтегазоносных и водоносных горизонтов	I	Структурный	2
	I34	Географическая приуроченность локальных поднятий	I	Мезозой-кайнозойский	2
			2	Палеозойский	
			I	Север Запдно-Сибирской плиты	2
			0	Юг Запдно-Сибирской плиты	

*/ I - логические, 2 - в виде наименований, 3 - количественные

км/ I - наиболее погруженная часть региона по данному горизонту, околтуривающаяся замкнутыми региональными изогипсами; 2 - менее погруженная, но также околтуривающаяся региональными изогипсами часть региона; 3 - приборговая часть региона, осложненная структурными миссами, заливами и т.д., не околтуривающаяся замкнутыми изогипсами; 4 - бортовая часть региона, не несущая никаких осложнений структурного плана.

(пьезометрический) режим подземных потоков, способствующих переносу углеводородов; 5) гипсометрическое положение гидрозамка залежей и 6) возможность латеральной миграции углеводородов из областей газонефтеобразования в область газонефтенакопления. Признаки группы В дают возможность исследовать изменение структурного плана региона во времени и пространстве (палеотектоника). Отождествляющие признаки группы Г характеризуют общность объектов исследования обоих классов, выбранных для данной задачи. Учитывая большой объем признакового пространства, включающего 18 логических, 101 количественных и 15 имеющих шкалу наименований признаков, для обработки исходной информации был применен алгоритм "Каскад" [5, 2]. Кодирование и подсчет информативности количественных признаков производились по методу "скользящей границы", а остальных — методом "перебора" [6]. В итоге были получены оптимальные варианты кодирования (табл. 2), в соответствии с которыми вся исходная информация была приведена к бинарному виду (табл. 3). Подсчет строчечных нагрузок $JL P_j^c$ для эталонов на полном признаковом пространстве показал, что один из объектов I класса (№ 3) обладает большим сходством с "идеальным" объектом II класса [2]. Для повышения "контрастности" заданных классов была применена процедура минимизации по способу упорядочения признаков по информативности [6]. В результате минимизации был получен набор из 46 признаков, позволивший получить максимальную растяжку классов ($p^e = 2,95$).

Для проверки проб на допустимость к распознаванию [7] была составлена схема блокового анализа [2] (табл. 4), в которой выделяются блоки А, С₁ и С₂ не являются тесторами, а блок В отсутствует, логический анализ сводился, по существу, к оценке выполнения пробами критерия общности [7] (выполнения блока А). Несмотря на незначительную длину блока А, он допускает к распознаванию лишь объекты, имеющие структурный тип залежи и мезозойский возраст вмещающих толщ, локализующихся в северной части Западно-Сибирской плиты. В результате анализа выяснилось, что все пробы, а также объекты внешнего экзамена полностью выполняют блок А и допускаются к распознаванию. Для оценки эффективности работы системы распознавания был проведен внешний экзамен как по полному, так и по минимизированному пространству признаков (рис. 1).

На полном пространстве признаков качество распознавания сос-

Таблица 2

Таблица кодов логических (I) со шкалой наименований (II)
и количественных (III) признаков

I		
Номер призна.	Значен. "1"	Значен. "0"
I	2	3
I-9	I	0
I4-I9	I	0
I32-I34	I	0

II		
Номер призна.	Значен. "1"	Значен. "0"
I	2	3
I0-II	I	2,3,4,5,6
20,24,25	I	2,3,4,5,6
21,23	I,4	2,3,5,6
31,32,34	I,2	3,4
35	I	2,3,4
33	I,2,3	4

III			III			III		
I	2	3	I	2	3	I	2	3
26	>0	≤0	65	≤0	>0	99	>0,17	≤0,17
27	>70	≤70	66	>0,2	≤0,2	I00	>0,17	≤0,17
28	>I65	≤I65	67	>0,07	≤0,07	I01	>0,04	≤0,04
29	>235	≤235	68	>0,07	≤0,07	I02	<0	≥0
30	>80	≤80	69	>0	≤0	I03	≤-0,1	>-0,1
36	>3,5	≤3,5	70	>0,03	≤0,03	I04	>0,01	≤0,01
37	>2,5	≤2,5	71	>3,7	≤3,7	I05	≤0,05	>0,05
38	>2,0	≤2,0	72	>2,45	≤2,45	I06	<-0,35	>-0,35
39	>0,2	≤0,2	73	>1,47	≤1,47	I07	≤-0,3	>-0,3
40	>I,5	≤I,5	74	>0,67	≤0,67	I08	≥0,05	<0,05
41	>0,2	≤0,2	75	>0,5	≤0,5	I09	<-0,05	≥-0,05
42	>0,95	≤0,95	76	>2,95	≤2,95	I10	≤-0,15	>-0,15
43	>0,65	≤0,65	77	>2,27	≤2,27	I11	>0	≤0
44	>0,25	≤0,25	78	>I,47	≤I,47	I12	>0,8	≤0,8
45	>0,07	≤0,07	79	>0,28	≤0,28	I13	>0,7	≤0,7
46	>8	≤8	80	>0,44	≤0,44	I14	>0,8	≤0,8
47	>52	≤52	81	≤80	>80	I15	>0,4	≤0,4
48	≤42	>42	82	≤4,5	>4,5	I16	>2,2	≤2,2
49	≤I00	>I00	83	≤I	>I	I17	>2,4	≤2,4
50	≤22	>22	84	≤0,9	>0,9	I18	>2,65	≤2,65
51	≤I00	>I00	85	≤0,5	>0,5	I19	>I,65	≤I,65
52	>5I	≤5I	86	≤0,4	>0,4	I20	>2,15	≤2,15
53	>27	≤27	87	>0,25	≤0,25	I21	>I,12	≤I,12
54	<27	≥27	88	>0,25	≤0,25	I22	>0,6	≤0,6
55	≥45	<45	89	>0,05	≤0,05	I23	>0,97	≤0,97
56	>I,62	≤I,62	90	>0,03	≤0,03	I24	>0,75	≤0,75
57	>2,05	≤2,05	91	>0,03	≤0,03	I25	>0,27	≤0,27
58	>I,05	≤I,05	92	>0,2	≤0,2	I26	>I,32	≤I,32
59	≤0,72	>0,72	93	≤-0,35	>-0,35	I27	>2,10	≤2,10
60	≤0,5	>0,5	94	>0,02	≤0,02	I28	>2,40	≤2,40
61	≤0,4	>0,4	95	>0	≤0	I29	>I,67	≤I,67
62	>0,32	≤0,32	96	>0,3	≤0,3	I30	>I,97	≤I,97
63	≤0	>0	97	>0,22	≤0,22	I31	>I,12	≤I,12
64	<0,24	≥0,24	98	>0,24	≤0,24			

Таблица 3

РАБОЧАЯ ИНФОРМАЦИОННАЯ ТАБЛИЦА

№ признаков	76	128	120	127	72	118	36	37	71	131	31	119	130	121	117	73	77	122	112
Названия объектов	76	128	120	127	72	118	36	37	71	131	31	119	130	121	117	73	77	122	112
1. Новопортовское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
2. Надымское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
3. С. Казымское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
4. Сикторское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
5. Варь-Еганское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
6. Вэнгапурское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
7. Ярайнерское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
8. Таркосалинское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
9. Русско-Часельское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
10. Ю. Русское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
11. Русское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
12. Заполярное	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
13. Тазовское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
14. Соленинское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
15. Мессо-Яхское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
16. Зимнее	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
17. Пеляткинское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
18. Казанцевское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
19. Озерное	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
20. Джангодское	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
21. Тундровая	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
22. Суходудинская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
23. Точинская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
24. Долганская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
25. Щучинская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
26. Ермаковская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
27. Фарковская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
28. Кыксинская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
29. Чимулякская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
30. Эмторская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
31. Лык-Пайская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
32. Казымская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
33. Амнинская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
34. Кислорская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
35. Радомская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
36. Проточинская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
37. В. Сартыньинская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
38. Шоганская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
39. В. Куноватская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
40. Ярудейская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
41. Мшистая	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
42. Таноичинская	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
43. Медведевская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
44. Яр-Салинская	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
P_j^v	0,88	0,88	0,85	0,85	0,82	0,82	0,81	0,81	0,81	0,81	0,81	0,78	0,78	0,78	0,77	0,75	0,75	0,75	0,75

Таблица 3 (продолжение)

	1.	I	129
	2.	I	116
	3.	I	I
	4.	I	I
	5.	I	I
	6.	I	I
	7.	I	I
	8.	I	I
	9.	I	I
	10.	I	I
	11.	I	I
	12.	I	I
	13.	I	I
	14.	I	I
	15.	I	I
	16.	I	I
	17.	I	I
	18.	I	I
	19.	I	I
	20.	I	I
	21.	I	I
	22.	I	I
	23.	I	I
	24.	I	I
	25.	I	I
	26.	I	I
	27.	I	I
	28.	I	I
	29.	I	I
	30.	I	I
	31.	I	I
	32.	I	I
	33.	I	I
	34.	I	I
	35.	I	I
	36.	I	I
	37.	I	I
	38.	I	I
	39.	I	I
	40.	I	I
	41.	I	I
	42.	I	I
	43.	I	I
	44.	I	I
p_j^v	0,73	I	125
	0,71	I	79
	0,70	I	27
	0,70	I	78
	0,69	I	115
	0,68	I	80
	0,67	I	70
	0,67	I	38
	0,66	I	39
	0,66	I	34
	0,66	I	42
	0,63	I	43
	0,63	I	40
	0,62	I	12
	0,62	I	41
	0,61	I	33
	0,61	I	32
	0,61	I	67
	0,60	I	23
	0,60	I	4
	0,58	I	13
	0,58	I	21
	0,58	I	69
	0,58	I	113

Таблица 3 (продолжение)

1.	I	21.	0	0,58	123
2.	I	22.	0	0,58	124
3.	I	23.	I	0,56	62
4.	I	24.	I	0,55	50
5.	I	25.	0	0,54	5
6.	I	26.	0	0,54	25
7.	I	27.	0	0,54	25
8.	I	28.	0	0,54	59
9.	I	29.	0	0,54	89
10.	II	30.	I	0,53	24
11.	I	31.	0	0,53	56
12.	I	32.	0	0,50	55
13.	I	33.	0	0,50	81
14.	I	34.	0	0,48	26
15.	I	35.	0	0,46	64
16.	I	36.	0	0,46	68
17.	I	37.	0	0,45	114
18.	I	38.	0	0,44	83
19.	I	39.	0	0,42	28
20.	I	40.	0	0,42	54
	I	41.	0	0,41	85
	I	42.	I	0,40	20
	I	43.	0	0,40	53
	I	44.	I	0,39	87
	I		I	0,39	94
	I		I	0,39	96
	I		I	0,39	97
	I		I	0,39	98

Таблица 3 (продолжение)

1.	I	101
2.	I	101
3.	I	101
4.	I	101
5.	I	101
6.	I	101
7.	I	101
8.	I	101
9.	I	101
10.	I	101
11.	I	101
12.	I	101
13.	I	101
14.	I	101
15.	I	101
16.	I	101
17.	I	101
18.	I	101
19.	I	101
20.	I	101
21.	0	101
22.	I	101
23.	0	101
24.	I	101
25.	I	101
26.	0	101
27.	0	101
28.	0	101
29.	I	101
30.	I	101
31.	I	101
32.	0	101
33.	I	101
34.	0	101
35.	I	101
36.	I	101
37.	0	101
38.	I	101
39.	I	101
40.	I	101
41.	0	101
42.	0	101
43.	I	101
44.	I	101
45.	I	101
46.	I	101
47.	I	101
48.	I	101
49.	I	101
50.	I	101
51.	I	101
52.	I	101
53.	I	101
54.	I	101
55.	I	101
56.	I	101
57.	I	101
58.	I	101
59.	I	101
60.	I	101
61.	I	101
62.	I	101
63.	I	101
64.	I	101
65.	I	101
66.	I	101
67.	I	101
68.	I	101
69.	I	101
70.	I	101
71.	I	101
72.	I	101
73.	I	101
74.	I	101
75.	I	101
76.	I	101
77.	I	101
78.	I	101
79.	I	101
80.	I	101
81.	I	101
82.	I	101
83.	I	101
84.	I	101
85.	I	101
86.	I	101
87.	I	101
88.	I	101
89.	I	101
90.	I	101
91.	I	101
92.	I	101
93.	I	101
94.	I	101
95.	I	101
96.	I	101
97.	I	101
98.	I	101
99.	I	101
100.	I	101
101.	I	101
102.	I	101
103.	I	101
104.	I	101
105.	I	101
106.	I	101
107.	I	101
108.	I	101
109.	I	101
110.	I	101
111.	I	101
112.	I	101
113.	I	101
114.	I	101
115.	I	101
116.	I	101
117.	I	101
118.	I	101
119.	I	101
120.	I	101
121.	I	101
122.	I	101
123.	I	101
124.	I	101
125.	I	101
126.	I	101
127.	I	101
128.	I	101
129.	I	101
130.	I	101
131.	I	101
132.	I	101
133.	I	101
134.	I	101
135.	I	101
136.	I	101
137.	I	101
138.	I	101
139.	I	101
140.	I	101
141.	I	101
142.	I	101
143.	I	101
144.	I	101
145.	I	101
146.	I	101
147.	I	101
148.	I	101
149.	I	101
150.	I	101

Таблица 3 (окончание)

	III	109	19	16	44	45	58	80	14	82	35	65	2	49	51	52	93	108	48	57	106	107	46	15	6	9
1.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
2.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
6.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
8.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
10.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
11.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
12.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p_d^v	0,25	0,24	0,23	0,22	0,22	0,22	0,22	0,22	0,21	0,21	0,20	0,20	0,18	0,18	0,18	0,18	0,18	0,16	0,16	0,15	0,15	0,15	0,13	0,12	0,08	0,03

№ п/п	№ гр.																						
	131	132	133	76	128	127	27	38	39	113	59	82	46	120	32	114	95	111	57	106	107	35	
1.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
2.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
3.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
4.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
5.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
6.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
7.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
8.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
9.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
10.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
11.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
12.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
13.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
14.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
15.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
16.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
17.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
18.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
19.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
20.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
21.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
21.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
22.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
23.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
24.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
25.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
26.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
27.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
28.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
29.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
30.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
31.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
32.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
33.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
34.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
35.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
36.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
37.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
38.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
39.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
40.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
41.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
42.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
43.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
44.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
БЛОКИ	A			C ₁										C ₂									

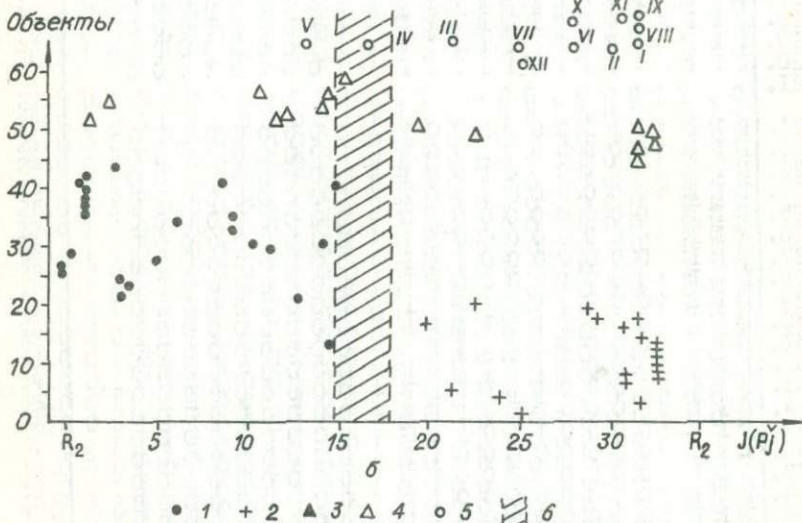
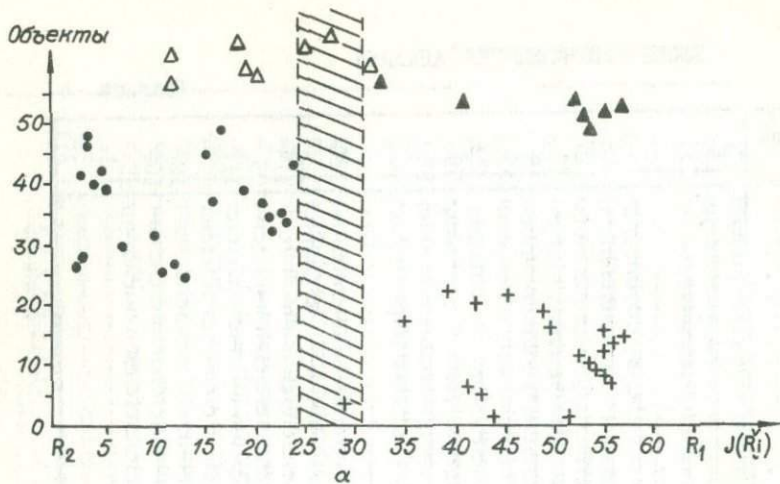


Рис. 1. Внешний экзамен на полном (а) и минимизированном (б) признаков пространства. Распознавание проб выполнено в минимизированном пространстве признаков. I - объекты обучения I класса; 2 - объекты обучения II класса; 3 - объекты экзамена I класса; 4 - объекты экзамена II класса; 5 - пробы; 6 - зона неопределенности.

тавляет 86,7%, на минимизированном пространстве признаков качество распознавания повысилось до 96,7%. Распознавание проб проводилось по минимизированному набору признаков (рис. 1). Из 12 объектов, представленных к распознаванию, 10 отнеслись к I классу (зона "а"), 1 - к зоне неопределенности [2] (зона "в") и 1 - ко II классу (зона "с").

Как уже отмечалось, минимизированное пространство, обеспечивающее наилучшее разделение классов, содержит 46 признаков, распределяющихся по группам следующим образом (табл. 5).

Т а б л и ц а 5
Средняя информативность группы признаков

Индекс группы признаков	Номера признаков	(средняя) (P_j^y)
А	$X_4, X_{12}, X_{13}, X_{21}, X_{27}, X_{30}$	0,61
Б	$X_{31}-X_{34}, X_{36}-X_{43}, X_{67}, X_{70}-X_{74},$ $X_{76}-X_{79}, X_{90}$	0,69
В	$X_{112}, X_{115}-X_{123}, X_{126}-X_{131}$	0,77

Из приведенной таблицы следует, что:

1. Максимальной средней информативностью для разделения эталонных классов обладают признаки группы В, характеризующие динамику структурного плана локальных поднятий (палеотектонику).

2. В следующей по средней информативности группе признаков, характеризующих гипсометрическое положение локального поднятия в пределах структурно-тектонических зон (группа Б), наиболее значимыми являются признаки, определяющие абсолютные отметки нижней замкнутой изогипсы локального поднятия по всем горизонтам, кроме верхнего.

3. Наименьшей средней информативностью обладают признаки, определяющие общий структурно-тектонический план региона (группа А). В этой группе наиболее значимыми являются признаки, характеризующие расстояние от центра локального поднятия до зоны выклинивания отложений по горизонтам Π^a и У.

Признаки, характеризующие отдельные горизонты (статические признаки), а также их взаимосвязи с выше- и нижележащими гори-

зонтами (динамические), имеют следующую информативность (табл.6).

Т а б л и ц а 6
Информативность признаков по горизонтам

Гори- зонты	Статистические признаки (гр. Б)		Динамические признаки (гр. В)	
	Количество	P_j^y (средняя)	Количество	P_j^y (средняя)
Ф	5	0,82	9	0,75
Па	7	0,67	6	0,74
Ш	5	0,65	5	0,72
IV	7	0,64	10	0,68
У	2	0,66	9	0,76

Как видно из таблицы, наибольшей способностью различать продуктивные и "пустые" структуры обладают статические признаки, характеризующие горизонт Ф, а по группе динамических признаков — характеристики горизонтов Ф, Па и V. При совместном рассмотрении всех групп признаков наиболее часто встречаются признаки, характеризующие горизонты Ф и IV.

Сравнительный анализ результатов математической обработки исходной информации показал, что для всех месторождений обучающей выборки характерны следующие условия:

1. Абсолютные отметки вершин продуктивных структур, выраженных по поверхности домезозойского фундамента, превышают 2,95 км.

2. Абсолютные отметки вершин л.п. по поверхности домезозойского фундамента к концу формирования отложений сеномана превышают 2,1 км.

3. Расстояние от центра л.п. до зоны выклинивания отложений заводоуковской серии превышает 70 км.

4. Абсолютные отметки нижней и верхней региональных изогипс, оконтуривающих структурно-тектоническую зону, в пределах которой находится л.п. по поверхности отложений заводоуковской серии превышают 2,0 км и 0,2 км соответственно.

Все условия, необходимые для отнесения л.п. к продуктивным структурам, выполнены на Тарко-Салинском, Русско-Часельском, Юж-

но-Русском, Русском, Заполярном и Тазовском газовых месторождениях.

Как видно из таблицы (см. табл. 3), чаще всего на месторождениях обучающей выборки (кроме вышепоименованных) не выполняется условие, характеризующее положение л.п. в пределах внутренней или внешней мегасинеклизы региональной тектонической зоны, выделяемой по поверхности отложений заводоуковской серии. Это обстоятельство связано с тем, что наиболее прогнутая часть территории находится на севере и крайнем северо-востоке региона, в силу чего часть месторождений (8 из 20) располагается вне его. Необходимо отметить также и то, что по поверхности отложений сеномана условие, характеризующее положение л.п. в пределах региональной зоны, не выполняется в пяти случаях. По-видимому, эти невыполнения условий, свойственных основной массе месторождений, вызваны интенсивной миграцией углеводородов из областей нефтегазообразования в области нефтегазонакопления во время формирования основных продуктивных горизонтов (средняя + верхняя юра; сеноман) или позднее (сеноман), в силу чего локальные поднятия, расположенные вне мегасинеклиз (в пределах моноклиналей), заполнены нефтью и газом.

По величине строчечных нагрузок объекты исследования располагаются в определенной последовательности, группируясь в подклассы. Минимальным значением строчечной нагрузки обладает объект № 3 (Северо-Казымское месторождение). Это можно объяснить тем, что это мелкое по масштабу месторождение имеет залежи в отложениях нижнеюрской продуктивной толщи, что нетипично для севера Западно-Сибирской плиты. Несколько большее значение $J[P_j^y]$ получено по Вахскому месторождению, на котором известны мелкая и крупная залежи нефти в отложениях верхней юры, и средняя по масштабу залежь нефти в отложениях валанжина, что также нетипично для исследуемого района. Зимнее, Варь-Еганское, Джангодское, Сикторское, Новопортовское и Северо-Варь-Еганское месторождения также отличаются от характерных для района месторождений либо величиной залежей, либо их многопластовостью, либо качеством углеводородов (нефть, газоконденсат). Более высокие значения $J[P_j^y]$ получены для Казанцевского и Озерного месторождений со средними по масштабам газовыми залежами в отложениях валанжина. Ближе к "идеальному" объекту по величине строчечных

нагрузок стоят Вэнга-Пурское, Ярайнерское и Мессо-Яхское место – рождения, что находит свое подтверждение и в геологической их близости. Надымское, Соленинское и Пеляткинское месторождения имеют весьма высокие значения $J[P_j^v]$, однако Соленинское и Пеляткинское месторождения отличаются более древним возрастом продуктивных отложений. Группа из II месторождений (Тарко-Салинское, Русско-Часельское, Южно-Русское, Русское, Заполярное, Тазовское, Арктическое, Надынское, Уренгойское, Губкинское, Сузунское) представлена наиболее типичными для района объектами, что нашло свое отражение в приобретении ими максимальных $J[P_j^v]$.

Таким образом, проведенные исследования оценки продуктивности локальных поднятий севера Западно-Сибирской плиты показали, что на основе описаний объектов, охарактеризованных только структурно-тектоническими параметрами, используя алгоритм "Каскад", можно разделять структуры на "пустые" и продуктивные. По близости значений $J[P_j^v]$ объектов прогноза с $J[P_j^v]$ объектов обучения можно с определенной долей уверенности относить пробы к похожим изученным структурам (табл. 7).

Т а б л и ц а 7
Близость проб к объектам обучения

№ проб	Значение $J(P_j^v)$	Степень сходства	Название близких или аналогичных объектов обучения
I	3I,34	2	Мессо-Яхское месторождение
II	30,54	3	Вэнга-Пурское и Ярайнерское месторожден.
III	2I,85	3	Джангодское месторождение
IV	16,99	1	Северо-Казымское месторождение
V	13,53	2	Эмторская водоносная структура
VI	28,26	1	Казанцевское и Озерное месторождения
VII	25,28	2	Новопортовское месторождение
VIII	3I,84	3	Соленинское и Пеляткинское месторожден.
IX	3I,10	3	Мессо-Яхское месторождение
X	28,26	1	Озерное месторождение
XI	3I,12	3	Мессо-Яхское месторождение
XII	25,4I	2	Новопортовское месторождение

Примечание: Степень сходства: I-очень слабая, 2-слабая, 3-полная.

На основании проведенных исследований уже сейчас выделяются направления, которые могут быть развиты на базе этой задачи:

1. Прогноз продуктивности локальных поднятий по величине залежей.
2. Прогноз продуктивности л.п. по качеству углеводородов (нефть, газ, газоконденсат).
3. Прогноз продуктивности л.п. по возрасту основных продуктивных залежей.
4. Построение моделей типичных месторождений и "пустых" структур.

Л и т е р а т у р а

1. Распознавание образов гигантских нефтяных месторождений.- В кн.: Проблема нефтеносности Сибири. Новосибирск, "Наука", 1971, с.34-50. -Авт.: Трофимук А.А., Вышемирский В.С., Дмитриев А.Н. и др.
2. БАБИЧ В.В., ФЕДОСЕЕВ Г.С. Метод целевого классифицирования и упорядочения объектов ("Каскад-1"). См. настоящий сборник. с.42-70.
3. Использование программ распознавания для разделения подготовленных сейсморазведкой локальных поднятий на нефтегазоносные и "пустые". - В кн.: Математические методы в геологии и геофизике Новосибирск, 1968, с.62-67 (Тр. СНИИГГТИМС, вып.79).- Авт.:Бенько В.И., Еханин Е.В., Жаднова В.П., Миталев И.А.
4. НЕСТЕРОВ И.И. Критерии прогноза нефтегазоносности.М."Недра", 1969. 335с. (Тр. ЗапСибНИГНИ, вып.15).
5. БАБИЧ В.В., КРАСАВЧИКОВ В.О., ФЕДОСЕЕВ Г.С. Программы метода суммарного учета мер приуроченности и согласования. - В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.122-126.
6. СОКОЛОВ А.Д. Программа П5 "Оптимальное бинарное кодирование признаков, подсчет строчечных нагрузок и минимизация.- В кн.: Логико-математическая обработка геологической информации. Новосибирск, 1975, с.151-158.
7. Организация и обработка геологической информации с помощью ЭВМ на основе построения тупиковых тестов. - В кн.: Логико-информационные решения геологических задач. М. "Наука", 1975, с. 83-128. -Авт.: Дмитриев А.Н., Кренделев Ф.П., Бишаев А.А. и др.

СО Д Е Р Ж А Н И Е

I. ТЕОРЕТИЧЕСКИЕ ВОПРОСЫ И МЕТОДЫ ЛОГИКО - МАТЕМАТИЧЕСКИХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ

А.Н.Дмитриев. Вопросы формализованных постановок геологических задач прогнозно-поискового профиля	3
Г.С.Федосеев. Предварительный анализ данных в логико - математических исследованиях	22
В.В.Бабич, Г.С.Федосеев. Метод целевого классифицирования и упорядочения объектов ("Каскад"-I").	42
А.А.Бишаев. Метод "Целевая итерационная классификация" ("Цикл")	70
С.В.Макаров, Е.А.Смертин. Центрированная качельная процедура для нахождения согласованной системы информационных весов. .	92

II. ВОПРОСЫ ПРАКТИЧЕСКОГО ПРИМЕНЕНИЯ МАТЕМАТИЧЕСКИХ ПРОЦЕДУР

В.В.Бабич, Г.С.Федосеев. Прогнозная оценка железорудных объектов Кондомского района Горной Шории	101
В.Д.Карбышев, А.А.Бишаев. Прогнозирование продуктивности дифференцированных трапповых интрузий севера Сибирской платформы	117
В.А. Каштанов, А.Д.Соколов. Прогноз продуктивности локальных поднятий до их ввода в бурение (на примере мезозойско - кайнозойских отложений севера Западно-Сибирской плиты)	133

Технический редактор *Л. А. Панина*

Подписано к печати 25.Х.1976г. МН 03011
Бумага 60x84/16. Печ.л. 9,5. Уч.-изд.л. 9,0.
Тираж 600. Заказ 260. Цена 63 коп.

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

Цена 63 коп.

1927