

ТЕОРИЯ РАСПОЗНАВАНИЯ

ОБ АЛГЕБРАИЧЕСКОМ ПОДХОДЕ К РЕШЕНИЮ ЗАДАЧ РАСПОЗНАВАНИЯ ИЛИ КЛАССИФИКАЦИИ

Ю. И. ЖУРАВЛЕВ

(МОСКВА)

Введение

В работе рассматривается широко известная задача отнесения объекта к некоторым классам из заданного списка классов.

Постановка задачи состоит в следующем. Дано множество M объектов, относительно которых проводится классификация. Известно, что множество M представимо в виде суммы подмножеств K_1, \dots, K_l , называемых обычно классами.

Задана информация I о классах K_1, \dots, K_l , описание множества M и описание $I(S)$ объекта S , о котором, вообще говоря, неизвестно — к каким из классов K_1, \dots, K_l он принадлежит. Требуется по информации I , описанию $I(S)$ установить для каждого j значение свойства $S \in K_j$, $j = 1, 2, \dots, l$.

Сформулированная выше задача обычно называется задачей распознавания образов или задачей классификации. В дальнейшем мы будем называть ее задачей Z или задачей распознавания.

Задачи типа Z рассматривались сначала как чисто прикладные. Большое внимание уделялось, например, задаче автоматического чтения текстов. При этом каждой букве сопоставляется класс K ее различных изображений.

Требовалось по предъявленному изображению (считалось, что это априори изображение некоторой буквы) установить, к какому из классов K оно принадлежит. Другими словами, требовалось идентифицировать букву по ее изображению. В дальнейшем возникли аналогичные задачи идентификации речи и т. п.

Несколько позднее внимание исследователей стали привлекать проблемы классификации более сложно описываемых объектов. Появились задачи медицинской диагностики, геологического прогнозирования, оценки экономических и политических ситуаций, прогнозирования свойств химических соединений и т. п.

Нет необходимости рассматривать подробно эти проблемы. По каждой из них существует большая научная литература, а главное, с точки зрения математика, существенного различия между этими задачами нет. Отметим только, что проблемы распознавания являются весьма общими и к ним приводится множество разнообразных практических вопросов. Так, при некоторой идеализации, к схеме распознавания приводится любая задача принятия решений, если только процесс принятия решения базируется в основном на изучении ранее накопленного опыта. Действительно, допустим, что имеется совокупность ранее исследованных ситуаций S_1, \dots

..., S_m и в каком-то виде даны их описания $I(S_1), \dots, I(S_m)$. Для каждой из ситуаций S_i известно, какое решение является наилучшим в данной ситуации, и это решение $R(S_i)$ также описано. Предположим также, что может быть дано описание множества $\{R\}$ всевозможных решений. Допустим, что в пространстве $\{R\}$ существует понятие «близость», т. е. решения могут быть разбиты на классы K_1, \dots, K_l таким образом, что в один класс попадают «близкие» решения (решения одного типа). Решения разных классов не являются близкими.

Дано описание новой ситуации $I(S)$. Требуется по набору описаний ранее исследованных ситуаций $I(S_1), \dots, I(S_m)$ и найденных для них решений $R(S_1), \dots, R(S_m)$ найти — к какому из классов решений должно быть отнесено наилучшее решение.

Таким образом, задача распознавания является в частном случае дискретным аналогом проблемы поиска оптимальных решений. К задачам распознавания сводятся не только задачи синтеза наилучших решений, но и другие «важные» классы прикладных проблем.

Итак, первая причина, по которой внимание большого числа исследователей было привлечено в последние десятилетия к решению задач классификации — это обилие прикладных вопросов, исследование которых сводится к решению задач этого типа. Вторая причина, весьма важная для математиков, состоит в том, что решение таких задач ввело в обиход большое число некорректных, или, как их обычно называют, эвристических алгоритмов. Дело в том, что подавляющее большинство применений теории распознавания связано с плохо формализованными областями науки и практики, такими, как медицина, геология, социология, химия и т. п.

В этих областях трудно строить формальные теории и применять стандартные математические методы. В лучшем случае удается дать математическое оформление некоторым интуитивным принципам и затем применить построенные «эмпирические» формализмы для решения специальных типов проблем. Это обстоятельство определило тот факт, что на первом этапе развития теории и практики распознавания образов возникло большое число различных методов и алгоритмов, применявшихся без какого-либо серьезного обоснования для решения практических задач. Такие методы, как это принято во всех экспериментальных науках, обосновывались непосредственной проверкой — успехом решения реальных задач. Многие из них выдержали такую проверку и применяются, несмотря на отсутствие математических обоснований.

Следует, очевидно, признать, что существование некорректных алгоритмов уже давно стало фактом. По-видимому, появление каждого из таких алгоритмов можно рассматривать как эксперимент, а со всем множеством таких экспериментов и их результатов — работать как с новым для математики множеством объектов. Итак, задача состоит в том, чтобы, признав как реальность существование и пользу для практики некорректных процедур решения плохо формализованных задач, изучить с помощью строгих математических методов само множество таких процедур. Антитезой такой идеологии могло бы быть построение формальных моделей в областях, ныне не поддающихся формализации. Однако этот последний путь вряд ли может привести к успеху, например, в описательной биологии или геологии.

Рассмотрим совокупность некорректных алгоритмов, предназначенных для решения задач распознавания. По мере их накопления описывались не только индивидуальные алгоритмы, но и принципы их формирования. Эти принципы, действующие уже над подмножествами алгоритмов и высказываемые поначалу также в плохо формализованном виде, в дальнейшем получали (или могли получить) точные математические описания.

На этом этапе эвристикой является уже выбор принципа, а алгоритмы, порождаемые при помощи данного принципа, могут получаться стандартным образом. Таким образом, формализация различных принципов приводит к появлению моделей распознающих алгоритмов.

Рассмотрим простой пример. Интуитивная формулировка принципа: во многих задачах, где описания объектов задаются наборами значений числовых признаков (объекты суть точки n -мерного пространства), такие описания, принадлежащие разным классам, могут быть разделены поверхностью достаточно простого вида. Назовем этот принцип принципом разделения.

Одна из возможных формализаций. Рассмотрим простейший класс разделяющих поверхностей — гиперплоскости

$$\sum_{i=1}^n a_i x_i + a_{n+1} = 0.$$

Пусть множество допустимых объектов разделено на два класса: $K_1, K_2, K_1 \cap K_2 = \emptyset$. Пусть также известно, что объекты S_1, \dots, S_m принадлежат K_1 , объекты $S_{m+1}, \dots, S_q \in K_2$. Эти объекты, вообще говоря, неравнозначны. Поэтому введем их числовые характеристики $\gamma(S_i) = \gamma_i$ — вес объекта S_i , $i = 1, 2, \dots, m, m+1, \dots, q$. Таким образом, множество алгоритмов характеризуется заданием параметров a_1, \dots, a_{n+1} — коэффициентов в уравнении гиперплоскости и $\gamma_1, \dots, \gamma_q$ — весов объектов, классификация которых проведена ранее. Процесс распознавания для $I(S) = (\alpha_1, \dots, \alpha_n)$ производится следующим образом.

Пусть

$$f(x_1, \dots, x_n) = \sum_{i=1}^n a_i x_i + a_{n+1}.$$

Разделим объекты S_1, \dots, S_m на множества K_1^+, K_1^- : $S_i \in K_1^+$, если $f(I(S_i)) \geq 0$; $S \in K_1^-$, если $f(I(S_i)) < 0$. Аналогично объекты S_{m+1}, \dots, S_q разделим на множества K_2^+, K_2^- . Рассмотрим величины

$$\gamma(K_1^+) = \sum_{S_i \in K_1^+} \gamma(S_i), \quad \gamma(K_1^-) = \sum_{S_j \in K_1^-} \gamma(S_j)$$

и аналогичные им величины $\gamma(K_2^+), \gamma(K_2^-)$.

Вычисляем $f(I(S))$. Сопоставим S два числа: $\Gamma_1(S), \Gamma_2(S)$ — соответственно, значение функции принадлежности S классам K_1, K_2 . Если $f(I(S)) \geq 0$, то

$$\Gamma_1(S) = \frac{\gamma(K_1^+) + \gamma(K_2^-)}{\gamma(K_1^-) + \gamma(K_2^+)}, \quad \Gamma_2(S) = \frac{\gamma(K_2^+) + \gamma(K_1^-)}{\gamma(K_1^+) + \gamma(K_2^+)}.$$

При $f(I(S)) < 0$: $\Gamma_1(S) = \frac{\gamma(K_1^-) + \gamma(K_2^+)}{\gamma(K_1^+) + \gamma(K_2^-)}$ и аналогично $\Gamma_2(S)$.

По числам $\Gamma_1(S), \Gamma_2(S)$ принимается решение о зачислении S в K_1 или K_2 . Эта процедура задается решающим правилом. Рассмотрим класс решающих правил, определяемых параметром $\delta \geq 0$:

если $\Gamma_1(S) - \Gamma_2(S) > \delta$, то $S \in K_1$;

если $\Gamma_2(S) - \Gamma_1(S) > \delta$, то $S \in K_2$;

если $|\Gamma_1(S) - \Gamma_2(S)| \leq \delta$, то решение не принимается, алгоритм отказывается от классификации S .

Мы построили одну из возможных моделей, основанных на принципе разделения. Эта модель основывается на гипотезах: а) элементы классов

K_1 и K_2 разделяются гиперплоскостью (по крайней мере значительная часть элементов, классификация которых представляет интерес); б) элементы классов не равнозначны по важности; меру этой важности можно выразить числом.

Реализация гипотез проведена при построении модели

$$\mathfrak{M}(a_1, \dots, a_{n+1}, \gamma_1, \dots, \gamma_q, \delta), \quad -\infty < \gamma_i, \quad a_i < +\infty, \quad \delta \geq 0.$$

Задание значений всех параметров модели определяет ее элемент — конкретный алгоритм распознавания.

Второй этап развития теории распознавания характеризуется прежде всего переходом от отдельных алгоритмов к рассмотрению моделей — семейств алгоритмов для решения классификационных задач. К настоящему времени построено и изучено несколько типов моделей. Выделим среди них несколько моделей, наиболее часто применяемых при решении различных прикладных задач.

1. Модели, построенные с использованием принципа разделения (R -модели) [55, 56, 61, 76]. Важно рассмотрен пример такой модели. R -модели различаются главным образом заданием класса поверхностей, среди которых выбирается поверхность (или набор поверхностей), разделяющая элементы разных классов.

2. Статистические модели. Формируются на базе принципов математической статистики. Используются в основном в тех случаях, когда известны (или могут быть просто определены) вероятностные характеристики классов K_1, \dots, K_i , например, соответствующие функции распределения. Так как в большинстве задач начальная информация не дает возможности определить эти характеристики с достаточным уровнем достоверности, то естественно эти модели также считать моделью, формируемыми на базе эвристических принципов.

3. Модели, построенные на принципе потенциалов (Π -модели) [4].

В основе формирования — аналогия с известным физическим принципом — сила притяжения прямо пропорциональна произведению масс и обратно пропорциональна расстоянию.

Если рассмотреть объекты, принадлежность которых классу K_j , ранее установлена, то можно различными способами ввести понятие массы этого множества, расстояние от множества до распознаваемого объекта S и выбрать в качестве значения функции принадлежности S классу K_j величину, являющуюся монотонно возрастающей функцией массы и монотонно убывающей функцией расстояния. Построение Π -моделей может быть проведено различными способами.

4. Модели вычисления оценок (голосования) (Γ -модели) [27, 28, 31]. Основаны на принципе частичной прецедентности. Анализируется «близость» между частями описаний ранее классифицированных объектов и объекта, подлежащего распознаванию. Наличие близости является частичным прецедентом и оценивается по некоторому заданному правилу.

По набору оценок вырабатывается общая оценка объекта по классу, которая и является значением функции принадлежности объекта классу.

При формировании модели по одному из принципов могут использоваться идеи и других принципов. Так, в построенной выше R -модели использован также и принцип потенциалов. В методе потенциалов [4] прямо сочетается использование принципов потенциалов и разделения.

Метод комитетов [55, 56, 61, 76] использует идеи принципов разделения классов гиперплоскостями и голосования.

Если на уровне создания отдельных фиксированных алгоритмов основная проблематика была связана с построением эффективных вычислительных схем и проведением экспериментов — решением прикладных задач —

то на уровне моделей возникает множество новых задач, интересных для математика.

Среди них особо следует выделить проблемы синтеза алгоритмов, экстремальных по качеству распознавания в рамках данной модели. Функционал качества алгоритма может определяться различными способами. Обычно его определение базируется на следующем принципе. Задается способ построения объектов каждого из классов. Оценивается для фиксированного алгоритма из данной модели — какую долю объектов он классифицирует правильно, т. е. отнесет к данному классу. Полученная величина усредняется по классам и называется функционалом качества алгоритма. Задача состоит в том, чтобы в рамках модели найти алгоритм с максимальным значением функционала качества. Например, может быть задан следующий закон порождения классов K_1, K_2 . Пусть описания $I(S)$ объектов S являются наборами $(a_1(S), \dots, a_n(S))$ числовых признаков $-\infty < a_i(S) < +\infty, i = 1, 2, \dots, n$.

В n -мерном пространстве заданы два нормальных распределения соответственно с математическими ожиданиями m_1, m_2 и дисперсиями σ_1^2, σ_2^2 . Производится случайный выбор точек (описаний объектов) и разыгрывается по заданным законам класс, в который они зачисляются. После этого объект S , занесенный, например, в K_1 , с вероятностью p причисляется к обучающей выборке, с вероятностью $1 - p$ — к контрольной. То же самое делается с объектами из K_2 . Пусть таким образом сформированы обучающая и контрольная выборки. В первую из них зачислены из K_1 объекты S_{11}, \dots, S_{1m} , из K_2 — S_{21}, \dots, S_{2t} ; во вторую из K_1 — объекты S_{31}, \dots, S_{3v} , из K_2 — S_{41}, \dots, S_{4u} . В модели строится алгоритм A , который по описаниям $I(S_{11}), \dots, I(S_{1m}), I(S_{21}), \dots, I(S_{2t})$ дает максимальное значение функционала качества $\varphi(A) = \frac{q'}{q''}$, где q' — число объектов из контрольной выборки, правильно классифицированных алгоритмом A ; $q'' = v + u$ — число объектов в контрольной выборке.

Значение $\varphi(A)$ есть случайная величина и ее характеристики (моменты) дают представление о точности модели на определенном типе задач распознавания. Вычисление этих характеристик является обычно далеко не тривиальным. Результаты в таких задачах удается получить лишь для сравнительно простых моделей и законов образования классов (см., например, [67, 69]).

Более стандартным является подход, когда при фиксированной начальной информации I_0 и модели требуется найти в модели алгоритм, позволяющий максимально точно классифицировать данную совокупность $S_i, i = 1, 2, \dots, m$, контрольных объектов, принадлежность или непринадлежность которых классам K_1, \dots, K_l известна.

Естественно, что информация типа $S_i \notin K_j, S_i \in K_j$ не вводится в алгоритм. Построение экстремальных, в модели на заданной контрольной выборке, алгоритмов приводит к решению и исследованию новых типов экстремальных задач. Таким исследованиям посвящено большое число работ, особенно в R -моделях и Γ -моделях. Приведем один пример. Пусть даны описания объектов $I(S_1), \dots, I(S_m)$ из $K_1, I(S_{m+1}), \dots, I(S_q)$ из $K_2, I(S_i) = (\alpha_{i1}, \dots, \alpha_{in})$. Начальной информации нет. Строится

R -модель, разбиение проводится гиперплоскостью $f(\vec{x}) = \sum_{i=1}^n a_i x_i + a_{n+1}$.

Параметрами модели являются коэффициенты гиперплоскости a_1, \dots, a_{n+1} .

Решающее правило. Если $f(I(S_i)) \geq 0$, то $S_i \in K_1$; при $f(I(S_i)) < 0$ объект S_i заносится в $K_2, i = 1, 2, \dots, q$.

Для каждого S_i нетрудно написать условие правильной классификации. Написав эти условия последовательно для $S_1, \dots, S_m, S_{m+1}, \dots, S_q$, получим систему линейных неравенств с неизвестными a_1, \dots, a_{n+1} :

$$\begin{aligned}
 & a_1\alpha_{11} + \dots + a_n\alpha_{1n} + a_{n+1} \geq 0, \\
 & \vdots \quad \vdots \\
 & a_1\alpha_{m1} + \dots + a_n\alpha_{mn} + a_{n+1} \geq 0, \\
 a_1\alpha_{m+1,1} + \dots + a_n\alpha_{m+1,n} + a_{n+1} < 0, \\
 & \vdots \quad \vdots \\
 & a_1\alpha_{q1} + \dots + a_n\alpha_{qn} + a_{n+1} < 0.
 \end{aligned} \tag{1}$$

Система (1), вообще говоря, несовместна. Для построения искомого алгоритма требуется найти максимальную совместную подсистему в (1). Решив ее, получим значения a_1, \dots, a_{n+1} , а следовательно, экстремальный на выборке S_1, \dots, S_a алгоритм.

Выделение максимальной совместной подсистемы даже для линейных систем является трудной задачей, и ее решение требует создания специальных методов. В более сложных моделях построение экстремальных на заданной выборке алгоритмов также приводится к отысканию максимальных совместных подсистем, но неравенства в аналогах (1) не являются линейными.

Наконец, во многих случаях интересно само построение удобного описания модели и решение внутренних для модели задач, не связанных прямо с задачами синтеза и исследования экстремальных алгоритмов. Так, в моделях вычисления оценок $\Gamma_j(S)$, отыскание их величин, основанное на прямом переборе по «близким» частям описаний объектов, часто реально невыполнимо. Поэтому потребовалось построить специальный «аналитический» аппарат для эффективного вычисления величин $\Gamma_j(S)$.

Параллельно процессу перехода от отдельных алгоритмов к моделям развивалась другая ветвь, направленная главным образом на облегчение подготовки обучающей информации для решения задач распознавания. В рамках этого направления предприняты успешные попытки расширить класс обучающих информаций, годных для дальнейшей обработки распознавающими алгоритмами. Это направление получило название структурного распознавания [78].

Ранее для каждого класса K_j , $j = 1, 2, \dots, l$, требовалось подготовить список объектов, заведомо входящих и заведомо не входящих в K_j , выбрать набор характеристик (признаков) и задать описания подготовленных объектов в виде наборов признаков. Построение достаточного числа эталонных объектов и их описаний требовало много труда и времени. Фактически, требовалось задать классы частичным перечислением их элементов. Вместе с тем, в дискретной математике (алгебре, математической логике, теории алгоритмов и т. д.) уже давно применяется другой способ задания множеств. Выбирается система базисных элементов $\{B\}$ и система операций $\{O\}$. Последовательным применением операций из $\{O\}$ к базисным элементам из $\{B\}$, т. е. применением стандартной процедуры, последовательно порождается множество. Для задач распознавания класс K_j задается набором базисных элементов $\{B\}$, и набором правил порождения (операций) $\{O\}$.

Набор признаков (свойств) строится аналогично. Задается множество базисных свойств (элементарных предикатов), не обязательно двузначных. Из них индуктивно, например, при помощи логических операций, кванторов и т. п. строится множество предикатов, которые и являются призна-

ками, описывающими объекты. Совокупность таких предикатов по аналогии с формулами формальной теории обозначим через $\{\Phi\}_j$. Значения $\Phi(S)$ вычисляются на конструкциях S и порожденных элементах из K_j (фактически в ограниченное время порождается конечный подкласс из K_j).

Выделяется набор элементарных свойств, описывающих отношение объекта S к классу K_j . Например, $(S \in K_j)$; более вероятно вхождение S в K_u , чем в K_t , функция принадлежности (по Заде) [79] S классу K_j не больше ε и т. д. Над набором элементарных свойств по аналогии с $\{\Phi\}_j$ строится класс формул $\{\Psi\}_j$.

Наконец, рассматриваются формулы $\Phi \rightarrow \Psi$, $\Psi \in \{\Psi\}_j$, $\Phi \in \{\Phi\}_j$, т. е. выводы в терминах $\{\Psi\}_j$ из выполненных Φ (можно рассматривать необходимые условия: $\Psi \rightarrow \Phi$).

Набор формул $\Phi \rightarrow \Psi$ с выполненной Φ получает числовую оценку, на основе которой строится числовая величина $\Gamma_j(S)$ (явно или неявно) — функция принадлежности S классу K_j , $j = 1, 2, \dots, l$. По вектору $(\Gamma_1(S), \dots, \Gamma_l(S))$ или матрице $\{\Gamma_j(S_i)\}$, $i = 1, 2, \dots, q$, принимается решение о том, в какие из классов входит S или каждый из S_1, \dots, S_q .

Структурный подход имеет свою внутреннюю проблематику и проблемы, связанные с решением прикладных задач.

В нашей работе мы не будем останавливаться подробно на этих проблемах. Приведем только некоторые частные результаты, показывающие, что принципиально структурный подход в распознавании не отличается от подхода, связанного с заданием классов перечислением описаний конечного числа объектов.

Суммируя изложенное, следует отметить, что теория и практика распознавания накопили большой экспериментальный материал (различные алгоритмы), некоторый опыт построения и исследования моделей — описаний классов распознающих алгоритмов. На базе указанной предыстории можно поставить задачу построения общей теории распознающих алгоритмов. В этом направлении следует указать интересную работу У. Гренандера [77]. В настоящей статье развивается другой подход к построению такой общей теории.

Предлагается вариант определения алгоритма распознавания, в рамках которого укладываются все существующие типы алгоритмов. Термины, в которых даются определения, позволяют использовать последние как для теоретических исследований, так и при решении прикладных задач.

Далее, для эффективного исследования класса алгоритмов распознавания и их конструктивного описания предлагается подход, аналогичный изложенному при описании структурных методов. Выделяются базисные алгоритмы и модели и вводятся операции над ними, позволяющие последовательно порождать новые алгоритмы и модели. При таком подходе удается не только эффективно описать многие новые классы распознающих алгоритмов, но установить некоторые их важные свойства. Так, например, получены условия, при которых данное семейство алгоритмов относительно введенных операций является базисным. На примере двух моделей показана проверка этих условий. Выяснено, какими свойствами должна обладать модель, чтобы для произвольной конечной выборки в модели нашелся алгоритм, правильно классифицирующий все объекты этой выборки; для различных моделей даны методы построения таких алгоритмов.

Краткое содержание работы по главам.

В первой главе вводятся основные определения, постановка задач, причем особое внимание уделяется понятию обучающей начальной информации.

Во второй главе дается краткое описание R - и P -моделей и более подробное — G -модели (вычисление оценок).

На примере последней подробно показан переход от конкретных алгоритмов к интуитивному формированию принципа и затем к формальному описанию модели.

Дано построение аналитического аппарата, позволяющего эффективно вычислять оценки $\Gamma_j(S)$ для классифицируемых объектов S . В этой главе рассматриваются модели, созданные на базе структурного подхода.

В третьей главе на основе анализа конкретных моделей предлагается определение алгоритма распознавания. Такой алгоритм A рассматривается как последовательное выполнение двух операторов R_A и r_A , $A = R_A \cdot r_A$. Оператор R_A , называемый в дальнейшем распознающим оператором, применяется к начальной (обучающей) информации I_0 и описаниям распознаваемых объектов $I(S'_1), \dots, I(S'_q)$. Он переводит их в числовую матрицу $\{a_{ij}\}_{q \times l}$, где a_{ij} есть значение функции принадлежности S_i классу K_j , $j = 1, 2, \dots, l$, $i = 1, 2, \dots, q$. Термин «функция принадлежности» взят из теории размытых множеств [79].

Во многих случаях вместо него мы будем употреблять более короткий термин «оценка», который к тому же, на наш взгляд, лучше отражает содержательный смысл величины a_{ij} .

Оператор r_A — решающее правило алгоритма A — переводит матрицу $\{a_{ij}\}_{q \times l}$ в информационную матрицу $\{\alpha_{ij}\}_{q \times l}$, $\alpha_{ij} \in \{0, 1, \Delta\}$. При $\alpha_{ij} \in \{0, 1\}$ алгоритм A утверждает, что $P_j(S_i) = \alpha_{ij}$, $\alpha_{ij} = \Delta$ означает, что алгоритм A не вычислил значение свойства $P_j(S'_i)$.

Для распознающих операторов R_A (а следовательно, для алгоритмов с фиксированным r_A) вводятся операции сложения и умножения на скаляр. По этим операциям они образуют линейное векторное пространство. При ограниченном сверху объеме контрольной выборки это пространство имеет конечный базис при достаточно широких предположениях о начальной информации и возможных описаниях контрольных объектов. Вводится также произведение (последовательное применение распознающих операторов). В этой же главе доказывается теорема, устанавливающая свойства моделей, позволяющих безошибочно классифицировать любую конечную контрольную выборку. Доказанные теоремы применяются для исследования линейных замыканий R - и Γ -моделей.

В четвертой главе рассматриваются операции над алгоритмами на уровне информационных матриц. Описаны подгруппы по таким операциям. Введены также более общие типы операций, которые интерпретируются как многоместные функции трехзначной логики. Показано применение этих операций к синтезу новых моделей распознающих алгоритмов.

В пятой главе описаны основные типы задач, возникающих при синтезе экстремальных алгоритмов и изложены принципы их решения.

ГЛАВА I

ПОСТАНОВКА ЗАДАЧИ. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ. ОБУЧАЮЩАЯ ИНФОРМАЦИЯ. ФУНКЦИОНАЛ КАЧЕСТВА

§ 1. Постановка задачи

Пусть дано множество M объектов, называемых в дальнейшем допустимыми объектами. Множество допустимых объектов покрыто конечным числом подмножеств K_1, \dots, K_l : $M = \bigcup_{i=1}^l K_i$. Подмножества K_j , $j = 1, 2, \dots, l$, называются классами. Разбиение M определено не полностью. Задана лишь некоторая информация I_0 о классах K_1, \dots, K_l . Аналогично допустимый объект S определен значениями некоторых харак-

теристик. Совокупность заданных значений определяет описание $I(S)$ объекта S .

Основная задача (задача Z) состоит в том, чтобы по информации I_0 о классах K_1, \dots, K_l , $I_0(K_1, \dots, K_l)$ и описанию $I(S)$ допустимого объекта S вычислить значения предикатов $P_j(S)$ — « $S \in K_j$ », $j = 1, 2, \dots, l$. Информацию I_0 принято называть обучающей информацией, предикаты $P_j(S)$ — элементарными предикатами.

Перефразировка Z состоит в следующем. Пусть заданы два множества $\{I_0\}$, $\{I(S)\}$. Первое из них является множеством допустимых обучающих информаций $I_0(K_1, \dots, K_l)$, второе — множеством описаний $I(S)$ допустимых $S \in M$.

Требуется построить алгоритм A такой, что

$$A(I_0(K_1, \dots, K_l), I(S)) = (\alpha_1^A(S) \dots \alpha_l^A(S)), \text{ где } \alpha_i^A(S) = P_j(S), \quad (2)$$

$$j = 1, 2, \dots, l,$$

или, в применении к совокупности допустимых объектов,

$$A(I_0(K_1, \dots, K_l), I(S'_1), \dots, I(S'_q)) = \{\alpha_j^A(S'_i)\}_{q \times l}, \quad \alpha_j^A(S'_i) = P_j(S'_i), \quad (3)$$

$$i = 1, 2, \dots, q, \quad j = 1, 2, \dots, l.$$

При решении задачи Z рассматриваются также некорректные алгоритмы A , такие, что $\alpha_j^A(S'_i) \in \{0, 1\}$, но не обязательно $\alpha_j^A(S'_i) = P_j(S'_i)$. Наконец, рассматриваются алгоритмы, которые для некоторых допустимых S'_i отказываются вычислять значение $P_j(S'_i)$, $1 \leq j \leq l$. Этот факт символически записывается так: $\alpha_j^A(S'_i) = \Delta$.

При решении задачи Z в классе некорректных алгоритмов на алгоритм A обычно накладываются дополнительные условия.

Пусть задано множество $\{A\}$ алгоритмов таких, что

$$A(I_0(K_1, \dots, K_l), I(S)) = (\beta_1^A(S), \dots, \beta_l^A(S)), \quad \beta_j^A(S) \in \{0, 1, \Delta\}. \quad (4)$$

На множестве $\{A\}$ таких алгоритмов задан числовой функционал $\varphi(A)$ — функционал качества алгоритма A .

Уточненная основная задача (задача \tilde{Z}): среди алгоритмов, определенных (4), найти алгоритм A^* такой, что:

$$\varphi(A^*) = \sup_{A \in \{A\}} \varphi(A). \quad (5)$$

Для полной формулировки задачи необходимо уточнить понятия:

- 1) обучающей информации $I_0(K_1, \dots, K_l)$;
- 2) описания $I(S)$ допустимого объекта S ;
- 3) функционала качества $\varphi(A)$ алгоритма A .

§ 2. Обучающая информация $I_0(K_1, \dots, K_l)$; описания допустимых объектов

I. Стандартный тип обучающей информации и описаний

1. Пусть задана совокупность признаков $\{1, 2, \dots, n\}$. Каждый из признаков i имеет множество допустимых значений M_i , $i = 1, 2, \dots, n$. Обычно рассматриваются признаки, имеющие следующие множества значений:

1°. $M_i^1 = \{0, 1\}$ — признак выполнен или не выполнен на объекте;

2°. $M_i^k = \{0, 1, 2, \dots, k - 1\}$ — признак имеет несколько градаций; $k > 2$;

3°. $\tilde{M}_i^k = \{a_1, \dots, a_k\}$ — признак принимает конечное число значений, элементы из \tilde{M}_i^k , вообще говоря, не являются числами; $k > 2$;

4°. $M_i = [a, b], (a, b], [a, b), (a, b), a, b$ — произвольные числа или символы $-\infty, +\infty$;

5°. M_i — некоторое более сложно устроенное подмножество множества действительных чисел;

6°. M_i^f — значениями признака i являются функции из некоторого класса функций.

7°. M_i^μ — значениями признака являются функции распределения некоторой случайной величины.

Приведенные здесь множества 1° — 7° не исчерпывают, очевидно, многообразия признаков, встречающихся в задачах распознавания.

В дальнейшем будем считать, что множества M_i значений признака могут быть пополнены элементом Δ , означающим, что значение i -го признака неизвестно. Множество $M_i \cup \Delta$ будем обозначать $M_i(\Delta)$.

Пусть даны наборы $(b_1, \dots, b_n) = \tilde{b}, (c_1, \dots, c_n) = \tilde{c}, b_i, c_i \in M_i(\Delta)$. Наборы \tilde{b}, \tilde{c} называются различными, если существует хотя бы один номер i такой, что $b_i \neq \Delta, c_i \neq \Delta, b_i \neq c_i, 1 \leq i \leq n$.

Определение 1. Описание $I(S) = (a_1(S), \dots, a_n(S))$, $a_i(S) \in M_i(\Delta)$ допустимого объекта S ($S \in M$) называется стандартным описанием S . Описание $I(S)$ называется полным, если $a_i(S) \neq \Delta, i = 1, 2, \dots, n$.

По традиции существует несколько специально выделяемых классов признаков и наименований для них. Так, признаки i с множеством значений M_i^2 называются бинарными, с множеством (M_i , 4°) — простыми числовыми, с множеством (M_i , 5°) — сложными числовыми, с множеством M_i^f — функциональными, с множеством M_i^μ — вероятностными.

Особое значение в задачах распознавания имеют признаки со специальными дополнительными условиями на множество M_i .

Определение 2. Признак i , $1 \leq i \leq n$, такой, что его M_i является метрическим пространством, называется метрическим.

Если метрика в M_i обозначена через ρ_i , то признак i может в дальнейшем обозначаться как (M_i, ρ_i) . В некоторых случаях функция ρ_i удовлетворяет всем аксиомам расстояния, кроме аксиомы треугольника. Тогда ρ_i называется полуметрикой, а признак (M_i, ρ_i) — полуметрическим.

В реальных задачах стандартные описания допустимых объектов обычно включают признаки разных типов.

2. Информационные векторы. Задача распознавания (классификации) состоит в том, чтобы для данного S и набора классов K_1, \dots, K_l вычислить элементарные свойства $P_j(S)$ — « S принадлежит классу K_j ».

Рассматриваются три вида информации о $P_j(S)$: установлено, что $S \in K_j$, установлено, что $S \notin K_j$, неизвестно — принадлежит S классу K_j или нет. Эти факты кодируются, соответственно, символами 1, 0, Δ . Информация о вхождении S в классы K_1, \dots, K_l кодируется вектором $(\alpha_1 \alpha_2 \dots \alpha_l)$, $\alpha_i \in \{0, 1, \Delta\}$, $i = 1, \dots, l$. Полученная информация может быть правильной и неправильной, полной или неполной. Различные возникающие здесь ситуации дают основания для введения следующих определений.

Определение 3. Вектор $\tilde{\alpha} = (\alpha_1 \dots \alpha_l)$ называется информационным, если $\alpha_i \in \{0, 1, \Delta\}$. Информационный вектор $\tilde{\alpha}$, сопоставленный допустимому объекту S , обозначается $\tilde{\alpha}(S) = (\alpha_1(S), \dots, \alpha_l(S))$.

Вектор $\tilde{\alpha}(S)$ называется полным, если $\alpha_i(S) \neq \Delta$, $i = 1, 2, \dots, l$. Вектор $\tilde{\alpha}(S)$ называется корректным для S , если из условия $\alpha_j(S) \neq \Delta$ следует, что $P_j(S) = \alpha_j(S)$, $j = 1, 2, \dots, l$. Корректный для S полный вектор $\tilde{\alpha}(S)$ называется истинным для S [32].

3. Пусть задана совокупность S_1, \dots, S_m допустимых объектов и их стандартные описания $I(S_1), \dots, I(S_m)$. Пусть также каждому из объектов S_t сопоставлен корректный информационный вектор $\tilde{\alpha}(S_t)$, отличный от $(\Delta \Delta \dots \Delta \Delta)$.

Определение 4. Стандартной информацией $I_0(K_1, \dots, K_l)$ называется совокупность множеств $\mathfrak{M}_1 = (I(S_1), \dots, I(S_m))$, $\mathfrak{M}_2 = (\tilde{\alpha}(S_1), \dots, \tilde{\alpha}(S_m))$. Стандартная информация $I_0(K_1, \dots, K_l)$ будет в дальнейшем записываться в виде

$$I_0(K_1, \dots, K_l) = \mathfrak{M}_1 \cup \mathfrak{M}_2 \quad (6)$$

или

$$I_0(K_1, \dots, K_l) = (I(S_1), \tilde{\alpha}(S_1), \dots, I(S_m), \tilde{\alpha}(S_m)). \quad (7)$$

Стандартная информация называется корректной (истинной), если векторы $\tilde{\alpha}(S_i)$ являются корректными (истинными) для S_i .

В дальнейшем мы будем рассматривать истинные стандартные информации. При рассмотрении корректных стандартных информаций в дальнейшем предполагается, что среди элементов \mathfrak{M}_2 нет вектора $(\Delta \Delta \dots \Delta \Delta)$.

Стандартная информация, содержащая описания объектов S_1, \dots, S_m по признакам $1, 2, \dots, n$ и информационные векторы длины l , обозначается в дальнейшем

$$I(S_1, \dots, S_m, K_1, \dots, K_l).$$

Если априори понятно, о каких объектах, признаках или классах идет речь или точное перечисление соответствующих компонент не нужно, будут употребляться обозначения

$$I_0(l), I_0(m, l) \quad \text{и т. д.}$$

4. Обучающая информация в задаче с непересекающимися классами K_1, \dots, K_l часто задается в специальной форме — в виде таблицы обучения T_{nm}^0 . Строками таблицы являются стандартные описания объектов S_1, \dots, S_m (таблица 1).

Таблица 1

	1	2	...	i	...	$n-1$	n	
S_1	1	a_{12}	...	a_{1i}	...	a_{1n-1}	a_{1n}	
	
S_{m_1}	m_{11}	a_{m_12}	...	a_{m_1i}	...	a_{m_1n-1}	a_{m_1n}	клас K_1
	
$S_{m_{l-1}+1}$	$a_{m_{l-1}+1,1}$	$a_{m_{l-1}+1,2}$...	$a_{m_{l-1}+1,i}$...	$a_{m_{l-1}+1,n-1}$	$a_{m_{l-1}+1,n}$	клас K_l
	
S_m	a_{m1}	...	a_{mi}	a_{mn}		

В дальнейшем будем считать, что объекты S_1, \dots, S_{m_1} принадлежат классу K_1, \dots , объекты $S_{m_{j-1}+1}, \dots, S_{m_j}$ принадлежат классу K_j, \dots , объекты $S_{m_{l-1}+1}, \dots, S_m$ — классу K_l .

5. Части стандартной обучающей информации. Пусть в множестве S_1, \dots, S_m выделено подмножество S_{i_1}, \dots, S_{i_k} , в системе признаков $1, 2, \dots, n$ — подмножество признаков u_1, \dots, u_t .

Оставим в $I_0(K_1, \dots, K_l)$ только описания $I(S_{i_1}), \dots, I(S_{i_k})$, а в этих описаниях — только координаты с номерами u_1, \dots, u_t . Информационные векторы остаются без изменений. Полученная таким образом информация называется частью стандартной информации.

Очевидно, часть стандартной информации является стандартной информацией.

II. Структурная обучающая информация и описание

Пусть заданы конечные алфавиты $\mathfrak{A} = \{a_1, \dots, a_n\}$, $\Lambda = \{\Lambda_1, \dots, \Lambda_q\}$. Из них образуем язык Σ следующим образом:

1°. Все конечные слова с буквами из \mathfrak{A} принадлежат Σ ;

2°. Если $S', S'' \in \Sigma$, то $S'\Lambda_i S'' \in \Sigma$; если S', S'' — непустые слова. Слова из Σ называются в дальнейшем фразами.

Определение 5. Описание $I(S)$ допустимого объекта называется структурным, если оно является элементом из Σ (фразой).

Очевидно, всякое стандартное описание является структурным. Часть фразы от первой буквы (включительно) до первой буквы из Λ (исключительно), заключенная между последовательными вхождениями букв из Λ , от последнего вхождения буквы из Λ (исключительно) до последней буквы фразы (включительно) называется правильной частью фразы или словом.

Остальные части фразы называются подфразами.

Пусть для каждого класса K_j задан набор фраз $B_1^j, \dots, B_{k(j)}^j$ и операции $O_1^j, \dots, O_{q(j)}^j$, применение которых к фразам образует фразу. Пусть также класс K_j состоит только и только из тех фраз, которые получаются из фраз B_u^j применением к ним операций O_t^j , $u = 1, 2, \dots, k(j)$, $t = 1, 2, \dots, q(j)$.

Определение 6. Описание $I(K_j) = \{B_1^j, \dots, B_{k(j)}^j, O_1^j, \dots, O_{q(j)}^j\}$ класса K_j называется структурным или алгебраическим, набор $B^j = \{B_1^j, \dots, B_{k(j)}^j\}$ — порождающим набором в $I(K_j)$ (базисом).

Набор $I(K_1), \dots, I(K_l)$ является обобщением набора описаний $I(S_1), \dots, I(S_m)$ в стандартной обучающей информации.

Пусть задан набор свойств $P_1^j(\Phi), \dots, P_{n(j)}^j(\Phi)$, определенных на фразах Φ из Σ , являющихся описаниями допустимых объектов из K_j . Построим над $P_1^j, \dots, P_{n(j)}^j$ класс формул $\{\phi\}_j$, используя, например, логические операции $\wedge, \vee, \neg, \rightarrow$ и кванторы \exists, \forall . Заметим, что предикаты $P_1^j, \dots, P_{n(j)}^j$ являются, вообще говоря, многоместными и Φ — это лишь один из аргументов в P_i^j , $i = 1, 2, \dots, n$.

Например, $P(\Phi, \Phi')$ — « Φ' есть слово в Φ ». Выделим набор элементарных отношений, описывающих вхождения объектов S в классы K_j . Среди них особо отметим: $Q_j(S)(Q_j(\Phi))$ — объект S принадлежит классу K_j , $j = 1, 2, \dots, l$, $Q_{ut}^{p1}(S)(Q_{ut}^{p1}(\Phi))$ — « $S \in K_u$ с большей вероятностью, чем $S \in K_t$ »; $Q_{ut}^{p0}(S)(Q_{ut}^{p0}(\Phi))$ — события $S \in K_u, S \in K_t$ равновероятны. $Q_j^{p, a, b}(S)(Q_j^{p, a, b}(\Phi))$ — событие $S \in K_j$ имеет вероятность p , заключенную в пределах $[a, b]$, т. е. $a \leq p \leq b$.

Аналогично вводятся $Q_{ut}^{f, 1}, Q_{ut}^{f, 0}, Q_j^{f, a, b}$, где вместо вероятности p рассматриваются значения функций принадлежности f объекта S классам K_u, K_t, K_j .

В набор элементарных отношений (предикатов) могут включаться и другие элементарные отношения.

Над элементарными отношениями Q^1, \dots, Q^r строим систему формул $\{\psi\}_j$, при помощи логических операций $\wedge, \vee, \neg, \rightarrow$.

Наконец, задаем набор формул $\{\chi\}_j$, где формулы имеют вид: а) $\phi \rightarrow \psi$, б) $\psi \rightarrow \phi$, $\phi \in \{\varphi\}_j$, $\psi \in \{\psi\}_j$.

Знак следования в $A \rightarrow B$ определен здесь лишь для случая $A = 1$ стандартным образом: если $A = 1$, то $B = 1$.

Определение 7. Набор формул $\{\chi\}_j$ называется информационным множеством K_j в структурной обучающей информации.

Определение 8. Множество $\tilde{I}_0(K_1, \dots, K_l) = \{I(K_1), \{\chi(K_1)\}, \dots, I(K_l), \{\chi(K_l)\}\}$ называется структурной обучающей информацией по классам K_1, \dots, K_l .

Пример 1. Пусть $\Phi(S)$ — описание объекта S и для класса K_j указана совокупность слов $\Phi^1, \dots, \Phi^{r(j)}$.

Элементарные предикаты: $P_i(\Phi(S), \Phi^i)$ — слово Φ^i является правильной частью фразы $\Phi(S)$, $i = 1, 2, \dots, r(j)$.

Для построения $\{\varphi\}_j$ применяем к $\{P_i\}$ операции \wedge, \vee, \neg . Множество $\{\varphi\}_j$ состоит из всевозможных д. н. ф. $\bigvee P_{i_1}^{\sigma_{i_1}} \dots P_{i_k}^{\sigma_{i_k}}$.

Элементарные предикаты для системы $\{\psi\}_j$: $Q_{ut}^{p,1}(\Phi)$: $S \in K_u$ с большей вероятностью, чем $S \in K_t$, $u = 1, 2, \dots, l$, $t = 1, 2, \dots, u-1, u+1, \dots, l$. В данном примере предикат $Q_{ut}^{p,1}(\Phi)$ будет обозначать $K_u(\Phi) > K_t(\Phi)$.

Систему $\{\psi\}_j$ образуем, применяя к элементарным предикатам операции \neg, \wedge, \vee .

Набор формул $\{\chi\}_j$: формулы типа $\psi \rightarrow \phi$ отсутствуют. Каждому элементарному предикату $P_i(\Phi(S), \Phi^i)$ сопоставляем единственную формулу из ψ вида $\prod_t (K_j(\Phi) > K_t(\Phi))$, произведение берется по индексам, входящим в подмножество $\{1, 2, \dots, l\} \setminus j$. Включаем в $\{\chi\}_j$ формулу:

$$P_i(\Phi(S), \Phi^i) \rightarrow \prod_t (K_j(\Phi) > K_t(\Phi)), \quad i = 1, 2, \dots, r(j).$$

Формулам из $\{\varphi\}_j$, содержащим хотя бы один P_i , не сопоставляем никаких формул из $\{\psi\}_j$. Если $A, B \in \{\varphi\}_j$, в $\{\chi\}_j$ включены формулы

$$A \rightarrow \prod_u (K_j(\Phi) > K_u(\Phi)), \quad B \rightarrow \prod_v (K_j(\Phi) > K_v(\Phi)),$$

причем индексы u и v пробегают соответственно подмножества M_u, M_v , то в $\{\chi\}_j$ включаются формулы:

$$A \cdot B \rightarrow \prod_w (K_j(\Phi) > K_w(\Phi)), \quad w \in M_u \cup M_v,$$

$$A \vee B \rightarrow \prod_{w \in (M_u \cap M_v)} (K_j(\Phi) > K_w(\Phi)), \quad \text{если } M_u \cap M_v \text{ непусто.}$$

Тем самым полностью построена система формул $\{\chi\}_j$, $j = 1, 2, \dots, l$.

Очевидно, можно также рассмотреть понятие части структурной обучающей информации:

$\tilde{I}_\alpha(K_1, \dots, K_l)$ есть часть $\tilde{I}(K_1, \dots, K_l)$,
если

$$\tilde{I}_\alpha(K_1, \dots, K_l) = \{I^\alpha(K_1), \{\chi_1\}^\alpha, \dots, I^\alpha(K_l), \{\chi_l\}^\alpha\},$$

$$\tilde{I}(K_1, \dots, K_l) = \{I(K_1), \{\chi_1\}_1, \dots, I(K_l), \{\chi_l\}_l\}$$

$$\{\chi_j\}^\alpha \subseteq \{\chi\}_j, \quad I^\alpha(K_j) \subseteq I(K_j).$$

Наконец, можно объединять произвольную часть стандартной информации $I(K_1, \dots, K_l)$ и произвольную часть структурной информации $\tilde{I}(K_1, \dots, K_l)$.

Другие виды обучающей информации нами рассматриваться не будут.

§ 3. Функционал качества $\varphi(A)$ алгоритма A

В работе рассматриваются следующие типы функционалов $\varphi(A)$. Пусть S'_1, \dots, S'_q — совокупность допустимых объектов, называемая в дальнейшем контрольным множеством; $\tilde{\alpha}_i = (\alpha_{i1} \dots \alpha_{il})$ — истинные информационные векторы объектов S'_i ; $\tilde{\alpha}_i^A = (\alpha_{i1}^A \dots \alpha_{il}^A)$ — информационные векторы, построенные для S'_i алгоритмом A , $i = 1, 2, \dots, q$. Обозначим через $\rho(\tilde{\alpha}, \tilde{\alpha}')$ произвольную полуметрику в пространстве информационных векторов.

Рассмотрим последовательность функций $f_1(x), f_2(x_1, x_2), \dots, f_t(x_1, \dots, x_t)$ удовлетворяющую следующим условиям:

1° все f_i определены при $x_i \geq 0$, $i = 1, 2, \dots, t$;

2° все f_i не возрастают по каждой переменной;

3° функция f_t достигает абсолютного максимума в точке $(0, 0, \dots, 0)$, и этот максимум равен единице.

Последнее условие всегда можно удовлетворить выбором нормирующего множителя.

Определение 9. Функционалом качества $\varphi(A)$ по контрольному множеству S'_1, \dots, S'_q называется величина

$$f_q(\rho(\tilde{\alpha}_1, \tilde{\alpha}_1^A), \dots, \rho(\tilde{\alpha}_q, \tilde{\alpha}_q^A)).$$

В дальнейшем мы будем иногда опускать слова «по контрольному множеству».

Определение 9 является весьма общим. При конструктивном построении оптимальных алгоритмов используется специальное понятие функционала качества.

Пусть дана пара (S'_i, K_j) — объект из контрольной выборки, класс из числа K_1, \dots, K_l , и для каждой пары задана таблица 2.

Таблица 2

		α_{ij}^A	0	1	Δ
		$P_j(S'_i)$	0	1	Δ
$P_j(S'_i)$	0	Φ_{ij}^{00}	Φ_{ij}^{01}	$\Phi_{ij}^{0\Delta}$	
	1	Φ_{ij}^{10}	Φ_{ij}^{11}	$\Phi_{ij}^{1\Delta}$	

Таблица 3

		α_{ij}^A	0	1	Δ
		$P_j(S'_i)$	0	1	Δ
$P_j(S'_i)$	0	0	1	0	0
	1	1	0	1	0

Здесь α_{ij}^A — значение предиката $P_j(S'_i)$ — « $S'_i \in K_j$ », вычисленное в алгоритме A , $P_j(S'_i)$ — истинное значение этого предиката, $\Phi_{ij}^{\alpha, \beta}$ — числовая оценка (поощрение или штраф) события: $P_j(S'_i) = \alpha$, $\alpha_{ij}^A = \beta$, $\alpha \in \{0, 1\}$, $\beta \in \{0, 1, \Delta\}$. Для контрольной выборки S'_1, \dots, S'_q по набору истинных информационных векторов $\tilde{\alpha}_i$ и векторов $\tilde{\alpha}_i^A$ строится матрица $\{\gamma_{ij}\}_{q \times l}$, где $\gamma_{ij} = \Phi_{ij}^{\alpha_{ij}, \alpha_{ij}^A}$.

Определение 10. Линейным функционалом качества называется функционал $\varphi(A) = \frac{1}{q \cdot l} \sum_{i=1}^q \sum_{j=1}^l \gamma_{ij}$ [32].

Важным частным случаем линейного функционала является функционал $\varphi^*(A)$, у которого таблицы 2 одинаковы для всех (i, j) , $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, и имеют вид, указанный в таблице 3.

Функционал $\varphi^*(A)$ называется долей правильных прогнозов.

ГЛАВА II
НЕКОТОРЫЕ МОДЕЛИ В ТЕОРИИ РАСПОЗНАЮЩИХ АЛГОРИТМОВ

Раздел 1. МОДЕЛИ АЛГОРИТМОВ ВЫЧИСЛЕНИЯ ОЦЕНОК

§ 1. Общие замечания

Алгоритмы, построенные на принципе вычисления оценок, являются формализацией идеи прецедентности или частичной прецедентности. Допустим, что заданы стандартные описания объектов $\{\bar{S}\}$, $\bar{S} \in K_j$, и $\{S'\}$, $S' \in \bar{K}_j$. Предъявлен для распознавания по j -му классу объект S , $j = 1, 2, \dots, l$. Допустим, что задан способ определения величины близости для некоторых частей описания S и соответствующих частей описаний $\{I(\bar{S})\}$, $\{I(S')\}$. Вычислив близости между частями описаний $I(S)$ и $I(\bar{S})$, соответственно между $I(S)$ и $I(S')$, можно построить обобщенную близость между S и множествами объектов $\{\bar{S}\}$, $\{S'\}$. В простейшем случае обобщенная близость равна сумме близостей между частями описаний.

Обозначив обобщенные близости через Γ_j^+ , Γ_j^- , построим характеристику типа $\Gamma_j(S) = \Gamma_j^+ - \Gamma_j^-$, которую естественно считать значением функции принадлежности объекта S классу K_j . Если $S = S'_i$, то в дальнейшем $\Gamma_j(S'_i)$ обозначается через Γ_{ij} . Величина Γ_{ij} называется оценкой S'_i по классу K_j . Очевидно, предъявленные для распознавания описания объектов S'_i переводятся алгоритмом вычисления оценок в числовую матрицу $\{\Gamma_{ij}\}_{q \times l}$, называемую далее матрицей оценок. По этой матрице применением решающего правила строится матрица $\{\alpha_{ij}\}_{q \times l}$ информационных векторов объектов S'_1, \dots, S'_q .

§ 2. Некоторые алгоритмы типа вычисления оценок

Первые алгоритмы вычисления оценок появились в 60-е годы [18]. Они хорошо зарекомендовали себя при решении практических задач, таких, как задачи геологического прогнозирования, медицинской диагностики и т. п. На базе этих алгоритмов были в дальнейшем [28] созданы модели алгоритмов вычисления оценок. В большом цикле работ эти модели были подробно изучены, а также использованы для решения разнообразных прикладных задач.

Особую роль в формировании модели вычисления оценок сыграл тестовый алгоритм. Этот алгоритм базируется на понятии теста, введенном в 1956 г. С. В. Яблонским. Описано большое число модификаций тестового алгоритма, их машинных реализаций и применений [19, 22, 45–50, 66, 75, 3].

В настоящей работе описывается модификация тестового алгоритма, удобная для иллюстрации формирования моделей вычисления оценок. Тестовый алгоритм описывается для задач с непересекающимися классами K_1, \dots, K_l . При этом информация I_0 представляется в виде таблицы обучения T_{nm}^0 (§ 2, глава I, таблица 1), описания объектов S_1, \dots, S_m являются бинарными векторами.

Определение 11. Тестом таблицы T_{nm}^0 называется совокупность столбцов i_1, \dots, i_k таких, что после удаления из T_{nm}^0 всех столбцов, за исключением имеющих номера i_1, \dots, i_k , в полученной таблице $T_{n-k, m}$ все пары строк, принадлежащих разным классам, различны. Тест $\{i_1, \dots, i_k\}$ называется туниковым, если никакая его истинная часть не является тестом [71].

Рассмотрим множество $\{T\}$ тупиковых тестов таблицы T_{nm}^0 . Пусть $T = \{i_1, \dots, i_k\} \in \{T\}$. Выделим в описании $I(S'_i)$ распознаваемого объекта S'_i подописание $(a_{i_1}, \dots, a_{i_k})$ по признакам i_1, \dots, i_k . Сравним $(a_{i_1}, \dots, a_{i_k})$ со всеми подописаниями $(a_{ui_1}, \dots, a_{ui_k})$ объектов S_1, \dots, S_m таблицы T_{nm}^0 .

Число совпадений $(a_{i_1}, \dots, a_{i_k})$ с $(a_{ui_1}, \dots, a_{ui_k})$, $u = m_{j-1} + 1, \dots, m_j$, $j = 1, 2, \dots, l$, $m_0 = 0$, $m_l = m$, (по описаниям объектов j -го класса) обозначим через $\Gamma_{ij}(T)$. Величина

$$\Gamma_{ij} = \Gamma_j(S'_i) = \frac{1}{m_j - m_{j-1}} \sum_{T \in \{T\}} \Gamma_{ij}(T)$$

называется оценкой S'_i по классу K_j .

Имея оценки $\Gamma_1(S'_1), \dots, \Gamma_l(S'_l)$, легко при помощи простых решающих правил классифицировать S'_i . Например, если в наборе оценок по крайней мере два максимума — объект не классифицируется. При одном максимуме объект заносится в тот класс, по которому он получил максимальную оценку.

Несколько позднее были применены алгоритмы, в которых в качестве множества $\{(i_1 i_2 \dots i_k)\}$ рассматривались всевозможные непустые подмножества множества $\{1, 2, \dots, n\}$ или всевозможные подмножества одинаковой мощности k , $1 \leq k \leq n$.

Для решения задач геологического прогнозирования применялись алгоритмы типа «Кора» [11]. В них в качестве подмножеств $(i_1 \dots i_k)$ рассматривается подкласс специально отбираемых по T_{nm}^0 подмножеств длины 3.

В работе [75] была высказана идея о том, что не все эталоны в T_{nm}^0 следует считать равноправными. Их следует снабдить числовыми весами, которые необходимо учитывать при формировании величин $\Gamma_{ij}(T)$, $\Gamma_j(S'_i)$. Ранее на базе анализа множества тупиковых тестов [18] были введены веса $p(i)$ признаков i , $i = 1, 2, \dots, n$, которые также учитываются при вычислении $\Gamma_{ij}(T)$, $\Gamma_j(S'_i)$.

В тестовом алгоритме $p(i)$ вычисляется следующим образом. Пусть $\tau(n, m)$ — число тупиковых тестов таблицы T_{nm}^0 , $\tau_i(n, m)$ — число таких тестов, содержащих столбец i . Тогда

$$p(i) = \frac{\tau_i(n, m)}{\tau(n, m)}.$$

Чем больше вес $p(i)$, тем важнее признак i для описания допустимых объектов из M . Последнее утверждение обосновывается при помощи следующих правдоподобных рассуждений. Таблица T_{nm}^0 представляет собой некоторое описание множества M . Она интересна тем, что сообщает некоторые сведения о разбиении M на классы K_1, \dots, K_l .

Тупиковый тест — это не сжимаемое далее описание, которое еще содержит все сведения о разделении M на классы. Чем в большее число таких неизбыточных описаний входит признак i , тем он важнее.

Если определены веса признаков $p(1), \dots, p(n)$ и веса $\gamma(S_1), \dots, \gamma(S_m)$ объектов, описания которых составляют таблицу обучения, то при совпадении $(a_{i_1}, \dots, a_{i_k})$ из распознаваемого S' с подстрокой $(a_{ui_1}, \dots, a_{ui_k})$ объекта S_u , $S_u \in K_j$, такое совпадение поопределяется величиной $\gamma(S_u) \cdot (p(i_1) + \dots + p(i_k)) = \Gamma_T(S_u, S')$.

Далее

$$\Gamma_j(S') = \frac{1}{m_j - m_{j-1}} \sum_{T \in \{S\}} \sum_{u=m_{j-1}+1}^{m_j} \Gamma_T(S_u, S'). \quad (8)$$

Исследованию свойств туниковых тестов бинарных таблиц посвящена значительная литература [58, 63—65].

§ 3. Модели алгоритмов вычисления оценок для стандартной обучающей информации

В настоящем параграфе мы рассмотрим лишь некоторые модели вычисления оценок. В дальнейшем будем считать, что признаки $1, 2, \dots, n$ являются метрическими (определение 2), т. е. множества M_i являются метрическими пространствами с метриками ρ_i , $i = 1, 2, \dots, n$.

I. Первым шагом задания алгоритма A в модели является указание системы Ω_A подмножеств множества $\{1, 2, \dots, n\}$. Элементы Ω_A называются *опорными множествами* алгоритма, а система Ω_A — системой опорных подмножеств алгоритма A . Примерами систем Ω_A являются множества туниковых тестов (в задаче с непересекающимися классами и бинарной таблицей T_{nm}^0); система всех подмножеств мощности k (в этом случае Ω_A задается параметром k), $1 \leq k \leq n - 1$, система всех непустых подмножеств множества $\{1, 2, \dots, n\}$.

Очевидно, каждому подмножеству $\Omega = \{i_1, i_2, \dots, i_k\}$ может быть взаимно однозначно сопоставлен *характеристический* булевский *вектор* $\tilde{\omega} = (\alpha_1 \dots \alpha_n)$, где $\alpha_{i_1} = \dots = \alpha_{i_k} = 1$, и остальные координаты равны 0.

Определение 12. Характеристической функцией $f_\Omega^A(x_1, \dots, x_n)$ системы Ω_A (алгоритма A) называется булевская функция, определяемая соотношением:

$$f_\Omega^A(\tilde{\omega}) = \begin{cases} 1, & \text{если } \tilde{\omega} \text{ является характеристическим} \\ & \text{вектором элемента } \Omega_A, \\ 0, & \text{в противном случае.} \end{cases} \quad (9)$$

Во многих ситуациях задание системы Ω_A через $f_\Omega^A(x_1, \dots, x_n)$ удобно.

Пример 2. Если Ω_A — система всех непустых подмножеств множества $\{1, 2, \dots, n\}$, то $f_\Omega^A = 0$ только на наборе $(00 \dots 00)$. Тогда $f_\Omega^A = x_1 \vee \dots \vee x_n$. Если Ω_A — система всех подмножеств мощности k , то f_Ω^A — симметрическая булева функция, равная 1 на и только на k -м слое вершин единичного n -мерного куба.

II. Вторым шагом задания алгоритма A является определение функции близости $B_{\tilde{\omega}}(S_u, S_t)$, $\tilde{\omega}$ соответствует $\Omega \in \Omega_A$, S_u, S_t — произвольные допустимые объекты.

Введем понятие $\tilde{\omega}$ -части описания $I(S)$ (объекта S). Пусть в $\tilde{\omega}$ координаты с номерами i_1, \dots, i_k равны 1 и остальные координаты — 0.

Определение 13. $\tilde{\omega}$ -частью $I(S)$ — обозначение $\tilde{\omega} I(S)$ — называется описание $(a_{i_1}, \dots, a_{i_k})$. Иногда для краткости объект S отождествляется со своим описанием. Мы будем, если это не приводит к путанице, пользоваться также обозначением: $\tilde{\omega} S$.

Соответственно $\tilde{\omega}$ -частью множества описаний $I(S_1), \dots, I(S_m)$ матрицы T_{nm} называется множество $\{\tilde{\omega} I(S_1), \dots, \tilde{\omega} I(S_m)\}$ (матрица

$T_{km} = \tilde{\omega} T_{nm}$, из которой удалены все столбцы, за исключением столбцов с номерами i_1, \dots, i_k .

Функцию близости $B_{\tilde{\omega}}(S_u, S_t)$ будем в дальнейшем также обозначать $B(\tilde{\omega}S_u, \tilde{\omega}S_t)$.

В различных моделях рассматривались следующие функции близости [7—9, 12, 13, 24, 25, 29, 31—37, 44, 53, 67—69, 72, 73, 80].

1. $B_e^{\vec{e}}(\tilde{\omega}S_u, \tilde{\omega}S_t)$. Пусть $e_i \geq 0$, $i = 1, 2, \dots, n$, $\vec{e} = (e_1 \dots e_n)$, e — целое число, $e \geq 0$, $\tilde{\omega}S_u = (a_{u_1}, \dots, a_{u_k})$, $\tilde{\omega}S_t = (a_{t_1}, \dots, a_{t_k})$. Напишем систему неравенств

$$\rho_i(a_{u_i}, b_{t_i}) \leq e_i, \quad i = 1, 2, \dots, k. \quad (10)$$

Здесь ρ_i — метрика в множестве M_i значений i -го признака, $i = 1, 2, \dots, n$. Тогда

$$B_e^{\vec{e}}(\tilde{\omega}S_u, \tilde{\omega}S_t) = \begin{cases} 1, & \text{если число невыполненных неравенств} \\ & \text{в системе (10) не больше } e; \\ 0 & \text{в противном случае.} \end{cases}$$

Важными частными видами функции $B_e^{\vec{e}}$ являются: $B_e^{(0\dots 0)}$; эта функция равна 1 в том и только том случае, когда в $\tilde{\omega}S_u$, $\tilde{\omega}S_t$ не равны не более e координат; $B_e^{\vec{e}} = 1$, если выполнены все неравенства системы (10), $B_e^{(0\dots 0)} = 1$, если $\tilde{\omega}S_u = \tilde{\omega}S_t$.

$$2. B_e(\tilde{\omega}S_u, \tilde{\omega}S_t) = \begin{cases} 1, & \text{если } \sum_{j=1}^k \rho_j(a_{u_j}, b_{t_j}) \leq e, \\ 0 & \text{в противном случае.} \end{cases}$$

Для случая числовых признаков, когда $\rho_i(x, y) = |x - y|$, функция $B_e(\tilde{\omega}S_u, \tilde{\omega}S_t)$ подробно изучалась в [57].

В дальнейшем будут рассматриваться в основном модели с функцией $B_e^{\vec{e}}$, а также с функциями B , существенно зависящими от неравенств, входящих в систему (10), но более общего, чем $B_e^{\vec{e}}$, вида.

III. Оценка $\Gamma_{\omega}(I(S), I(S_i))$, $\Gamma_{\omega}(S, S_i)$. Пусть S_i входит в стандартную обучающую информацию $I_0(K_1, \dots, K_l)$, S — допустимый объект, $\tilde{\omega}$ — характеристический вектор Ω , $\Omega \in \Omega_A$.

Пусть также объекту S_i сопоставлены числовые параметры $(\gamma_1(S_i), \dots, \gamma_q(S_i)) = \tilde{\gamma}(S_i)$; множеству Ω (вектору $\tilde{\omega}$) — числовые параметры $(p_1(\tilde{\omega}), \dots, p_r(\tilde{\omega})) = \tilde{p}(\tilde{\omega})$. Тогда

$$\Gamma_{\tilde{\omega}}(S, S_i) = f(B_{\omega}(S, S_i), \tilde{\gamma}(S_i), \tilde{p}(\tilde{\omega})) \quad (11)$$

называется оценкой S по объекту S_i и множеству Ω .

В приложениях чаще других рассматривался следующий вид функции $\Gamma_{\tilde{\omega}}$:

$$\Gamma_{\tilde{\omega}}(S, S_i) = (\tilde{\gamma}(S_i) \cdot \tilde{p}(\tilde{\omega})) \cdot B_{\tilde{\omega}}(S, S_i). \quad (12)$$

Во многих случаях, если множество Ω , которому сопоставлен вектор $\tilde{\omega}$, состоит из элементов i_1, \dots, i_k , то $p(\tilde{\omega}) = p_{i_1} + \dots + p_{i_k}$, $q = 1$ и тогда

$$\Gamma_{\tilde{\omega}}(S, S_i) = \gamma_1(S_i) (p_{i_1} + \dots + p_{i_k}) \cdot B_{\tilde{\omega}}(S, S_i). \quad (13)$$

В общем случае, если за $B_{\tilde{\omega}}^e(S, S_i)$ принять функцию $B_e^{\vec{e}}(\tilde{\omega}S_i, \tilde{\omega}S)$, величины $\Gamma_{\tilde{\omega}}$ (11) задаются наборами значений параметров e_1, \dots, e_n , $\gamma_1(S_1) = \gamma_1, \dots, \gamma_1(S_m) = \gamma_m, p_1, \dots, p_n$. Если опорные подмножества есть совокупность всех подмножеств множества $\{1, 2, \dots, n\}$ из k элементов, то пункты I — III задаются набором значений параметров $(k, e, \vec{e}, \vec{p}, \vec{\gamma})$, т. е. набором значений $2n + m + 2$ параметров. Параметры e_i обычно называют точностями измерения признаков, параметры p_i — весами признаков, параметры $\gamma(S_u) = \gamma_u$ — весами объектов. Тем самым они получают простое физическое истолкование. Набор параметров $(k, e, \vec{e}, \vec{p}, \vec{\gamma})$ является важнейшим для определения наиболее часто применяемой модели, содержащей подкласс алгоритмов вычисления оценок.

IV. Оценка $\Gamma_{\tilde{\omega}}^j(S)$ объекта S по опорному множеству Ω (или его характеристическому вектору $\tilde{\omega}$) по классу K_j :

$$\Gamma_{\tilde{\omega}}^j(S) = \varphi(\Gamma_{\tilde{\omega}}(S, S_{u_1}^j), \dots, \Gamma_{\tilde{\omega}}(S, S_{u_t}^j)), \quad (14)$$

$\{S_{u_i}^j\}$ — подмножество объектов из S_1, \dots, S_m , принадлежащих K_j , $i = 1, 2, \dots, t$. Наиболее часто в (14) функция $\Gamma_{\tilde{\omega}}^j(S)$ задается в виде

$$\Gamma_{\tilde{\omega}}^j(S) = \frac{1}{\mu(W_j^1)} \sum_{S_{u_i}^j \in W_j^1} \Gamma_{\tilde{\omega}}(S, S_{u_i}^j), \quad \mu(W_j^1) \quad (15)$$

— число элементов в W_j^1 , $W_j^1 = K_j \cap \{S_1, \dots, S_m\}$.

Тогда в модели $(k, e, \vec{e}, \vec{p}, \vec{\gamma})$:

$$\Gamma_{\tilde{\omega}}^j(S) = \frac{1}{\mu(W_j^1)} \sum_{S_{u_i}^j \in W_j^1} \gamma(S_{u_i}^j) \cdot (p_{i_1} + \dots + p_{i_k(\tilde{\omega})}) \cdot B_{\tilde{\omega}}^e(S, S_{u_i}^j). \quad (16)$$

В задачах с непересекающимися классами

$$\Gamma_{\tilde{\omega}}^j(S) = \frac{1}{m_j - m_{j-1}} \sum_{i=m_{j-1}+1}^{m_j} \gamma(S_i) (p_{i_1} + \dots + p_{i_k(\tilde{\omega})}) \cdot B_e^{\vec{e}}(S, S_i). \quad (17)$$

V. Оценка $\Gamma_j(S)$ по классу K_j . $\Gamma_j(S)$ определяется как функция от оценок $\Gamma_{\tilde{\omega}}^j(S)$ по различным подмножествам Ω , $\tilde{\omega} \leftrightarrow \Omega$. Обычно

$$\Gamma_j(S) = \sum_{\tilde{\omega} \leftrightarrow \Omega \in \Omega_A} \Gamma_{\tilde{\omega}}^j(S) \quad \text{или} \quad \Gamma_j(S) = \frac{1}{N} \sum_{\tilde{\omega} \leftrightarrow \Omega \in \Omega_A} \Gamma_{\tilde{\omega}}^j(S), \quad (18)$$

здесь N — нормирующий множитель. С учетом (15)

$$\Gamma_j(S) = \frac{1}{N} \sum_{\tilde{\omega} \leftrightarrow \Omega \in \Omega_A} \sum_{S_{u_i}^j \in W_j^1} \Gamma_{\tilde{\omega}}(S, S_{u_i}^j). \quad (19)$$

С учетом (16), (19) в модели $(k, e, \vec{e}, \vec{p}, \vec{\gamma})$ в общем случае

$$\Gamma_j(S) = \frac{1}{N} \cdot \frac{1}{\mu(W_j^1)} \sum_{\{\tilde{\omega}^k\}} \sum_{S_{u_i}^j} (p_{i_1} + \dots + p_{i_k}) B_e^{\vec{e}}(\tilde{\omega}S, \tilde{\omega}S_{u_i}^j). \quad (20)$$

Здесь первое суммирование производится по векторам $\tilde{\omega}^k$, имеющим среди координат ровно k единиц, второе суммирование — по элементам $S_{u_i}^j$, входящим в $W_j^1 = K_j \cap \{S_1, \dots, S_m\}$.

В задаче с непересекающимися классами

$$\Gamma_j(S) = \frac{1}{N} \cdot \frac{1}{m_j - m_{j-1}} \sum_{\{\tilde{\omega}^k\}} \sum_{i=m_{j-1}+1}^{m_j} \gamma(S_i) \cdot (p_{i_1} + \dots + p_{i_k}) \cdot \vec{B}_e(\tilde{\omega}S, \tilde{\omega}S_i). \quad (21)$$

Если в качестве S рассматривать последовательно объекты контрольной выборки S'_1, \dots, S'_q , то начальная информация

$$(I(S_1), \tilde{\alpha}(S_1), \dots, I(S_m), \tilde{\alpha}(S_m), I(S'_1), \dots, I(S'_q)) = (I_0(l), I_S(q))$$

переводится при помощи вычислений по I — V в матрицу $\{\Gamma_{ij}\}_{q \times l}$. Здесь $\Gamma_{ij} = \Gamma_j(S'_i)$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$. Матрица $\{\Gamma_{ij}\}$ называется матрицей оценок. Оператор R_A такой, что

$$R_A(I_0(l), I_S(q)) = \{\Gamma_{ij}\}_{q \times l}, \quad (22)$$

называется оператором вычисления оценок. В тех случаях, когда величины Γ_{ij} определяются набором значений параметров π_1, \dots, π_t , оператор R_A иногда будет обозначаться

$$R_A(I_0(l), I_S(q)) = \{\Gamma_{ij}\}_{q \times l} \text{ или } R_A(\pi_1, \dots, \pi_t). \quad (23)$$

Так, для модели с параметрами $(k, \varepsilon, \tilde{\varepsilon}, \vec{p}, \vec{\gamma})$:

$$R_A = R_A(k, \varepsilon, \tilde{\varepsilon}, \vec{p}, \vec{\gamma}).$$

VI. Решающие правила в алгоритмах вычисления оценок. Такое правило r по строке $(\Gamma_{i1}, \dots, \Gamma_{il})$ матрицы оценок для каждого класса K_j , $j = 1, 2, \dots, l$, вычисляет значение предиката $P_j(S'_i) = \langle S'_i \in K_j \rangle$ или отказывается от вычисления $P_j(S'_i)$. В последнем случае производится запись $P_j(S'_i) = \Delta$.

Таким образом, $r(\Gamma_{i1}, \dots, \Gamma_{il}) = (\alpha_{i1}^A \dots \alpha_{il}^A) = \tilde{\alpha}_A(S'_i)$, где $\tilde{\alpha}_A(S'_i)$ — информационный вектор S'_i в алгоритме A , определяемом оператором R_A и решающим правилом r . Очевидно,

$$r(\{\Gamma_{ij}\}_{q \times l}) = \{\alpha_{ij}^A\}_{q \times l}.$$

Здесь $\{\alpha_{ij}^A\}_{q \times l}$ — матрица информационных векторов элементов S'_1, \dots, S'_q , построенных алгоритмом A .

В различных моделях вычисления оценок рассматриваются различные решающие правила. Так, в задачах с непересекающимися классами употреблялось следующее правило:

среди элементов строки $(\Gamma_{i1}, \dots, \Gamma_{il})$ находится максимальный элемент Γ_{ij} ; если $\Gamma_{ij} - \Gamma_{iu} > \delta_1$, $u \neq j$ и $\frac{\Gamma_{ij}}{l} > \delta_2$, то $S'_i \in K_j$ и, следова-

$$\sum_{u \neq j} \Gamma_{iu}$$

тельно, $S'_i \notin K_u$, $u \neq j$; если хотя бы одно из перечисленных условий не выполнено, то в строке $(\alpha_{i1}^A, \dots, \alpha_{il}^A)$ матрицы $\{\Gamma_{ij}\}$ все элементы равны Δ . Модель, определенная оператором $R(k, \varepsilon, \tilde{\varepsilon}, \vec{p}, \vec{\gamma})$ и этим решающим правилом с параметрами δ_1, δ_2 , обозначается через

$$\mathfrak{M}(k, \varepsilon, \tilde{\varepsilon}, \vec{p}, \vec{\gamma}, \delta_1, \delta_2). \quad (24)$$

Эта модель и ее подмодели наиболее часто употреблялись при решении практических задач алгоритмами вычисления оценок [7, 13, 34, 36, 72].

Более общими видами решающих правил являются следующие:

1°. Для каждого класса K_j задана линейная форма

$$b_1^j \cdot x_1 + \dots + b_l^j x_l + b_{l+1}^j = L_j(\tilde{b}^j, \tilde{x})$$

и пороговые константы c_{1j}, c_{2j} , $l = 1, 2, \dots, l$. Пусть $\tilde{\Gamma}_i = (\Gamma_{i1}, \dots, \Gamma_{il})$.

Линейное решающее правило:

если $L_j(\tilde{b}^j, \tilde{\Gamma}_i) > c_{2j}$, то $S'_i \in K_j$ или $\alpha_{ij}^A = 1$,

если $L_j(\tilde{b}^j, \tilde{\Gamma}_i) < c_{1j}$, то $S'_i \notin K_j$ или $\alpha_{ij}^A = 0$,

при $c_{2j} \geq L_j(\tilde{b}^j, \tilde{\Gamma}_i) \geq c_{1j}$ вхождение S'_i в K_j алгоритмом A не определяется, $\alpha_{ij}^A = \Delta$.

2°. Аналогично 1° может быть задано нелинейное решающее правило. В этом случае вместо линейных форм $L_j(\tilde{b}^j, \tilde{x})$ задается система непрерывных функций $f_j(x_1, \dots, x_l)$, $j = 1, 2, \dots, l$.

Важным частным случаем правил 1°, 2° являются:

$$L_j(\tilde{b}^j, \tilde{x}) \equiv L(\tilde{b}, \tilde{x}), \quad j = 1, 2, \dots, l, \quad (25)$$

$$f_j(x_1, \dots, x_l) \equiv f(x_1, \dots, x_l), \quad j = 1, 2, \dots, l. \quad (26)$$

Варьируя в моделях вычисления оценок правила определения характеристик I — VI, можно получать различные модели распознающих алгоритмов типа вычисления оценок.

§ 4. Формулы для вычисления оценок $\Gamma_j(S)$

Вычисление элементов матрицы оценок непосредственно по их определениям из § 3 обычно практически невозможно.

Действительно, в модели $\mathfrak{M}(k, \varepsilon, \vec{\varepsilon}, \vec{p}, \vec{\gamma})$ (24) число опорных подмножеств равно C_n^k , и число слагаемых в формуле (19) для $\Gamma_j(S)$ равно $\mu(W_j) \cdot C_n^k$, n — число признаков в описании объектов, $\mu(W_j)$ — число описаний объектов из K_j в обучающей информации.

Оказывается, однако, что во многих случаях можно получить практически эффективные формулы для вычисления $\Gamma_j(S)$.

В настоящей работе не ставится цель перечисления всех таких случаев. Будут описаны два достаточно общих метода получения простых формул для $\Gamma_j(S)$ и показано несколько их применений.

I. Первый метод базируется на использовании специального свойства оценок, получаемых при помощи (19), когда

$$\Gamma_{\tilde{\omega}}^j(S) = \frac{1}{\mu(W_j)} \cdot \sum_{S_u^i \in W_j^1} \gamma(S_u^i) \cdot (p_{i_1} + \dots + p_{i_{k(\tilde{\omega})}}) \cdot B_{\tilde{\omega}}(S, S_i), \quad (27)$$

где $i_1, \dots, i_{k(\tilde{\omega})}$ — совокупность всех единичных координат вектора $\tilde{\omega}$, функция близости $B_{\tilde{\omega}}$ принимает только значения 0 и 1.

Величина $\Gamma_j(S)$ получается из $\Gamma_{\tilde{\omega}}^j(S)$, $\tilde{\omega} \leftrightarrow \Omega \in \Omega_A$ по (18). Зафиксируем признак t , $1 \leq t \leq n$. Обозначим через $V_t(S, S_i)$ число подмножеств Ω из Ω_A (число соответствующих им характеристических векторов $\tilde{\omega}$), содержащих признак t (имеющих единичную координату с номером t) и таких, что

$$B(\tilde{\omega}S, \tilde{\omega}S_i) = 1. \quad (28)$$

Множество $K_j \cap \{S_1, \dots, S_m\}$ обозначим через W_j^1 .

Теорема 1.

$$\Gamma_j(S) = \frac{1}{\mu(W_j^1)} \cdot \frac{1}{N} \sum_{S_i \in W_j^1} \gamma(S_i) \cdot \sum_{t=1}^n p_t V_t(S, S_i).$$

Доказательство. В самом деле,

$$\begin{aligned} \sum_{\tilde{\omega} \in \Omega_A} \Gamma_{\tilde{\omega}}^j(S, S_i) &= \gamma(S_i) \cdot \sum_{\Omega \in \Omega_A} B_{\tilde{\omega}}(S, S_i) \cdot \sum_{t \in \Omega} p_t = \\ &= \gamma(S_i) \sum_{\Omega : B_{\tilde{\omega}}(S, S_i) = 1} \sum_{t \in \Omega} p_t = \gamma(S_i) \sum_{t=1}^n p_t \cdot V_t(S, S_i). \end{aligned}$$

Подставляя полученное выражение в (18), получим утверждение теоремы.

Следствие. Обозначим через $\Gamma(S, S_i)$ величину $\sum_{\Omega \in \Omega_A} \Gamma(\tilde{\omega}S, \tilde{\omega}S_i)$. Тогда

$$\Gamma(S, S_i) = \gamma(S_i) \sum_{t=1}^n p_t \cdot V_t(S, S_i) \quad (29)$$

и

$$\Gamma_j(S) = \frac{1}{N} \cdot \frac{1}{\mu(W_j^1)} \sum_{S_i \in W_j^1} \Gamma(S, S_i). \quad (30)$$

Заметим, что если в (29) число различных значений $V_t(S, S_i)$ невелико, то формула для $\Gamma(S, S_i)$ упрощается, а следовательно, упрощается формула для $\Gamma_j(S)$. В этом случае в (19) для $\Gamma_j(S)$ практически исчезает суммирование по системе опорных множеств и остается только суммирование по элементам S_i из W_j^1 .

Опишем два вида функций близости, для которых число различных значений $V_t(S, S_i)$ мало при всех S_i из W_j^1 .

Рассмотрим модель \mathfrak{M} , в которой функция близости для $(\tilde{\omega}I(S_i), \tilde{\omega}I(S))$, $I(S) = (a_1, \dots, a_n)$, $I(S_i) = (b_1, \dots, b_n)$, определяется системой неравенств $\rho_1(a_1, b_1) \leq \varepsilon_1, \dots, \rho_n(a_n, b_n) \leq \varepsilon_n$ следующим образом. Если для пар $(\tilde{\omega}S_i, \tilde{\omega}S)$, $(\tilde{\omega}S_u, \tilde{\omega}S')$ неравенства с одинаковыми номерами одновременно выполнены или невыполнены, то

$$B(\tilde{\omega}I(S_i), \tilde{\omega}I(S)) = B(\tilde{\omega}I(S_u), \tilde{\omega}I(S')). \quad (31)$$

Определение 14. Функции B , удовлетворяющие (31) и принимающие значения 0, 1, называются пороговыми.

Множество $\{B\}$ пороговых функций обозначается в дальнейшем через $\{B^e\}$.

Сопоставим паре $(I(S) = (a_1, \dots, a_n), I(S') = (b_1, \dots, b_n))$ допустимых объектов характеристический вектор $\vec{\delta} = (\delta_1 \dots \delta_n) = \vec{\delta}(S, S')$:

если $\rho_i(a_i, b_i) \leq \varepsilon_i$, то $\delta_i = 1$,

если $\rho(a_i, b_i) > \varepsilon_i$, то $\delta_i = 0$, $i = 1, \dots, n$.

Очевидно, если $\vec{\delta}(S_i, S) = \vec{\delta}(S_u, S')$, то для любого $\tilde{\omega}$ пороговая функция B^e удовлетворяет равенству

$$B(\tilde{\omega}S_i, \tilde{\omega}S) = B(\tilde{\omega}S_u, \tilde{\omega}S'). \quad (32)$$

Пусть даны векторы $\tilde{\delta}(S, S_i)$ и характеристические векторы $\tilde{\omega}^1, \tilde{\omega}^2$ опорных подмножеств Ω_1, Ω_2 .

Определение 15. Пороговая функция $B^{\vec{e}}$ называется симметрической, если из условия: векторы $\tilde{\delta}(S, S_i) \cdot \tilde{\omega}^1, \tilde{\delta}(S, S_i) \cdot \tilde{\omega}^2$ (здесь под умножением векторов понимается операция их покоординатного умножения) имеют одинаковое число единичных и одинаковое число нулевых координат следует:

$$B^{\vec{e}}(\tilde{\omega}^1 I(S), \tilde{\omega}^1 I(S_i)) = B^{\vec{e}}(\tilde{\omega}^2 I(S), \tilde{\omega}^2 I(S_i)).$$

Если в $\tilde{\omega}$ единичные координаты суть i_1, \dots, i_k , то любая пороговая функция $B^{\vec{e}}$, значение которой на $(\tilde{\omega} I(S), \tilde{\omega} I(S_i))$ определяется только числом выполненных и числом невыполненных неравенств в системе (10), является симметрической.

Определение 17. Система опорных подмножеств Ω_A называется правильной, если из условия $\Omega \in \Omega_A$ и Ω имеет мощность k , $1 \leq k \leq n - 1$, следует, что все подмножества мощности k принадлежат Ω_A .

Теорема 2. Если функция близости $B_{\tilde{\omega}}^{\vec{e}}$ является симметрической, и система опорных подмножеств — правильной, то для каждой пары (S, S_i) величины $V_t(S, S_i)$, $t = 1, 2, \dots, n$, принимают не более двух различных значений.

Доказательство. Выделим в одну группу признаки j_1, \dots, j_q такие, что соответствующие им координаты вектора $\tilde{\delta}(S, S_i)$ равны 1. Возьмем произвольную пару признаков j_u, j_v из этой группы. Рассмотрим произвольный $\tilde{\omega}$, такой, что

$$j_u \in \Omega (\tilde{\omega} \leftrightarrow \Omega), \quad j_v \notin \Omega \text{ и } B_{\tilde{\omega}}^{\vec{e}}(S, S_i) = 1.$$

Тогда для $\tilde{\omega}^1$, соответствующего множеству $(\Omega \cup j_v) \setminus j_u$, в силу симметричности функции близости

$$B^{\vec{e}}(\tilde{\omega}^1 I(S), \tilde{\omega}^1 I(S_i)) = 1. \quad (33)$$

В силу (33) и того, что соответствие $\Omega \leftrightarrow (\Omega \cup j_v) \setminus j_u$ взаимно однозначно для совокупностей M_{10}, M_{01} , элементами которых являются опорные множества, содержащие j_u и не содержащие j_v (M_{10}), соответственно не содержащие j_u и содержащие j_v (M_{01}), имеем: $V_{j_u}(S, S_i) = V_{j_v}(S, S_i)$. Мы использовали также правильность системы опорных подмножеств.

Аналогично доказывается такое же равенство для признаков, которым соответствуют нулевые координаты вектора $\tilde{\delta}(S, S_i)$.

Теорема доказана.

Мы доказали, что все признаки, которым соответствуют в $\tilde{\delta}(S, S_i)$ нулевые, соответственно единичные координаты, входят в одинаковое число V^0 , соответственно V^1 опорных Ω , по которым функция близости между $I(S), I(S_i)$, равна 1.

Формулировка теоремы может быть дана в следующем виде.

Если $(\tilde{v} \cdot \tilde{w})$ — скалярное произведение векторов \tilde{v} и \tilde{w} , то

$$\Gamma(S, S_i) = (\tilde{\delta}(S, S_i) \cdot \tilde{p}) \cdot V^1 + (\tilde{\delta}(S, S_i) \cdot \tilde{p}) \cdot V^0, \quad (34)$$

где $\tilde{p} = (p_1 \dots p_n)$.

Приведем несколько примеров применения теоремы. Пусть $B^{\vec{e}}(\tilde{\omega}S, \tilde{\omega}S_i) = B^{\vec{e}}_i$ (см. § 3, II, 4°),

$$\Gamma(\tilde{\omega}S, \tilde{\omega}S_i) = \gamma(S_i) \cdot B^{\vec{e}}_i(\tilde{\omega}S, \tilde{\omega}S_i) \cdot (p_{i_1} + \dots + p_{i_k}).$$

Тогда (теорема 1)

$$\Gamma_j(S) = \frac{1}{N \cdot \mu(W_j^1)} \sum_{S_i \in W_j^1} \Gamma(S, S_i) \quad \text{и} \quad \Gamma(S, S_i) = V_1 \cdot \sum_u p_u + V_0 \cdot \sum_v p_v,$$

суммирование производится по единичным (u) и нулевым (v) координатам вектора $\tilde{\delta}(S, S_i)$. Рассмотрим систему Ω_A опорных множеств, состоящую из всех подмножеств мощности k множества $\{1, 2, \dots, n\}$.

Пусть t — признак, которому соответствует единичная координата вектора $\tilde{\delta}(S, S_i)$. Тогда $V_t = V^1$. Величина V^1 равна числу подмножеств мощности k , содержащих t , и не более e признаков, которым соответствуют нулевые координаты в $\tilde{\delta}(S, S_i)$. Пусть $q(S, S_i)$ — число единичных координат в $\tilde{\delta}(S, S_i)$.

Легко подсчитать, что

$$V^1(S, S_i) = \sum_{u=0}^e C_{n-q(S, S_i)}^u \cdot C_{q(S, S_i)-1}^{k-u-1}.$$

Аналогично,

$$V^0(S, S_i) = \sum_{u=1}^e C_{n-q(S, S_i)-1}^{u-1} \cdot C_{q(S, S_i)}^{k-u}.$$

Подставляя в (34), получаем

$$\begin{aligned} \Gamma_j(S) = & \frac{1}{N \cdot \mu(W_j^1)} \sum_{S_i \in W_j^1} \gamma(S_i) \left\{ (\tilde{\delta}(S, S_i) \cdot \vec{p}) \cdot \sum_{u=0}^e C_{n-q}^u \cdot C_q^{k-u-1} + \right. \\ & \left. + (\tilde{\delta}(S, S_i) \cdot \vec{p}) \cdot \sum_{u=1}^e C_{n-q-1}^{u-1} \cdot C_q^{k-u} \right\}. \end{aligned} \quad (35)$$

Пусть теперь в качестве системы Ω_A выбрана совокупность всех подмножеств множества $\{1, 2, \dots, n\}$. Функция $B(\tilde{\omega}S, \tilde{\omega}S_i)$ зависит от параметров $\vec{e}, \tilde{\varepsilon}_1, \tilde{\varepsilon}_2$.

$$B(\tilde{\omega}S, \tilde{\omega}S_i) = \begin{cases} 1, & \text{если среди координат вектора } \tilde{\delta}(S, S_i) \\ & \text{с номерами из } \Omega \text{ (} \tilde{\omega} \leftrightarrow \Omega \text{) число единич-} \\ & \text{ных не меньше } \tilde{\varepsilon}_1, \text{ число нулей — не} \\ & \text{больше } \tilde{\varepsilon}_2, \\ 0 & \text{во всех остальных случаях.} \end{cases} \quad (36)$$

Остальные этапы вычисления $\Gamma_j(S)$ определены как в предыдущем случае.

Нетрудно видеть, что функция $B^{\vec{e}}$, определенная в (36), является симметрической. Если опять число единичных координат в $\tilde{\delta}(S, S_i)$ равно $q(S, S_i)$, то величины V^1 и V^0 определяются как в первом примере.

Тогда

$$V^1 = \sum_{u=0}^{q(S, S_i) - \tilde{e}_1} C_{q(S, S_i) - 1}^{\tilde{e}_1 + u - 1} \cdot \sum_{u=0}^{\tilde{e}_2} C_{n-q(S, S_i)}^u, \quad (37)$$

$$V^0 = \sum_{u=0}^{q(S, S_i) - \tilde{e}_1} C_{q(S, S_i)}^{u + \tilde{e}_1} \cdot \sum_{u=0}^{\tilde{e}_2} C_{n-q(S, S_i) - 1}^{u - 1}. \quad (38)$$

Имея (37), (38) и используя теорему, нетрудно получить формулу для $\Gamma_j(S)$, $j = 1, 2, \dots, l$.

Для иллюстрации докажем (37).

Кроме одного фиксированного признака t , в опорное подмножество должно войти либо $\tilde{e}_1 - 1$, либо \tilde{e}_1, \dots , либо $q(S, S_i)$ элементов, которым в $\tilde{\delta}(S, S_i)$ соответствуют единичные координаты.

Выбор делается из $q(S, S_i) - 1$ элементов, элемент t фиксирован, число единичных координат $\tilde{\delta}(S, S_i)$ равно $q(S, S_i)$. Всего имеем возможностей выбора

$$\sum_{u=0}^{q(S, S_i) - \tilde{e}_1} C_{q(S, S_i) - 1}^{\tilde{e}_1 + u - 1}. \quad (39)$$

К каждому из выбранных подмножеств может быть присоединено не более \tilde{e}_2 элементов из числа $n - q(S, S_i)$ элементов, которым в $\tilde{\delta}(S, S_i)$ соответствуют нулевые координаты. Число возможных присоединений равно

$$\sum_{u=0}^{\tilde{e}_2} C_{n-q(S, S_i)}^u. \quad (40)$$

Из (39) и (40) без труда получается (37). Формула (38) доказывается аналогично.

Имея формулы (37) и (38) для V^1 и V^0 , без труда получаем формулу для $\Gamma_j(S)$.

Обобщением понятия симметрической пороговой функции является понятие пороговой функции, симметрической по разбиению.

Пусть $\{1, 2, \dots, n\} = \bigcup_{i=1}^v N_i$ — разбиение R множества признаков на подмножества N_1, \dots, N_v . Пусть даны два подмножества Ω^1, Ω^2 с характеристическими векторами $\tilde{\omega}^1, \tilde{\omega}^2$. Характеристические векторы подмножеств N_i обозначим через $\tilde{\omega}_i$, $i = 1, 2, \dots, v$. Подмножества Ω_1, Ω_2 назовем эквивалентными по разбиению R и описаниям $I(S), I(S_u)$, если векторы $(\tilde{\omega}^1 \cdot \tilde{\omega}_i)$ и $(\tilde{\omega}^2 \cdot \tilde{\omega}_i)$ имеют одинаковое число единичных координат, $i = 1, 2, \dots, v$.

Определение 18. Функция близости $B(\tilde{\omega}^1 S, \tilde{\omega}^2 S')$ называется симметрической по разбиению R , если

$$B(\tilde{\omega}^1 S, \tilde{\omega}^1 S_u) = B(\tilde{\omega}^2 S, \tilde{\omega}^2 S_u).$$

Разобьем множество признаков $\{1, 2, \dots, n\}$ на классы эквивалентности по разбиению R : признаки t' и t'' являются R -эквивалентными, если они одновременно входят (или не входят) в каждое подмножество N_i , $i = 1, 2, \dots, v$. Число классов R -эквивалентных признаков обозначим через $r(R)$.

Теорема 3. Если B — пороговая функция, симметрическая по разбиению R , и система опорных подмножеств является правильной, то число различных значений $V_u(S, S_i)$ не превосходит $2r(R)$.

Доказательство является незначительным усложнением доказательства теоремы 2. Пусть признаки t' , t'' являются R -эквивалентными и соответствующие им координаты вектора $\tilde{\delta}(S, S_i)$ равны между собой.

Рассмотрим произвольное опорное множество Ω' , $t' \in \Omega'$, $t'' \notin \Omega'$. В силу правильности системы опорных множеств $\Omega'' = (\Omega' \cup t'') \setminus t'$ также является опорным. Установим взаимно однозначное соответствие между опорными множествами, содержащими хотя бы один из элементов t' , t'' . Если $\{t', t''\} \in \Omega$, то $\Omega \leftrightarrow \Omega$; если $t' \in \Omega'$ и $t'' \notin \Omega'$, то $\Omega' \leftrightarrow \Omega''$. Соответствие указывает, что t' и t'' входят в одинаковое число опорных подмножеств. Так как t' и t'' являются R -эквивалентными, то они одновременно входят или не входят в каждое из N_i , $i = 1, 2, \dots, v$.

Так как соответствующие t' и t'' координаты $\tilde{\delta}(S, S_i)$ равны между собой, то элементам множеств $\Omega' \cap N_i$, $\Omega'' \cap N_i$ соответствует одинаковое число единичных и нулевых координат в $\tilde{\delta}(S, S_i)$. Если $\Omega' \leftrightarrow \tilde{\omega}'$ и $\Omega'' \leftrightarrow \tilde{\omega}''$, то в силу симметричности по R функции близости B получаем

$$B(\tilde{\omega}'S, \tilde{\omega}'S') = B(\tilde{\omega}''S, \tilde{\omega}''S'). \quad (41)$$

Из (41) легко следует, что $V_{t'}(S, S_i) = V_{t''}(S, S_i)$. Из последнего равенства, условия, что число классов R -эквивалентных признаков равно $r(R)$ и того факта, что координаты в $\tilde{\delta}(S, S_i)$ принимают не более двух значений, легко следует утверждение теоремы.

Следствие. Если разбиение R состоит из v непересекающихся подмножеств, то число различных значений $V_t(S, S_i)$ не превосходит $2v$.

Подсчет величин $V_t(S, S_i)$ для правильных опорных подмножеств и пороговых функций, симметрических для N_1, \dots, N_v , обычно связан лишь с преодолением технических трудностей. Подсчитаем эти величины для двух случаев.

Пусть системой опорных подмножеств Ω_A^k является совокупность всех подмножеств из k элементов, функция $B(\tilde{\omega}S, \tilde{\omega}S_i)$ — симметрическая по системе непересекающихся подмножеств и определяется при помощи параметров $\varepsilon_1, \dots, \varepsilon_n$ (они определяют вектор $\tilde{\delta}(S, S_i)$) и параметров ε_j^1 , $j = 1, 2, \dots, v$ следующим образом:

$$B(\tilde{\omega}S, \tilde{\omega}S_i) = \begin{cases} 1, & \text{если в } \tilde{\delta}(S, S_i) \text{ среди координат с номерами из } N_j \cap \Omega \text{ содержится не более } \varepsilon_j^1 \text{ нулевых координат,} \\ & j = 1, 2, \dots, v, \\ 0, & \text{во всех остальных случаях.} \end{cases} \quad (42)$$

Из следствия теоремы 3 следует, что величина $V_t(S, S_i)$ принимает не более $2v$ значений. Все признаки t из N_j с одинаковой координатой α вектора $\tilde{\delta}(S, S_i)$ имеют одинаковое значение величины V_u . Будем обозначать это значение через V_j^α , $\alpha = 0, 1$; $j = 1, 2, \dots, v$.

Переходим к выводу формул для V_j^0 , V_j^1 .

Пусть N_j содержит Δ_j элементов, из них Δ_j^1 , Δ_j^0 , $\Delta_j^1 + \Delta_j^0 = \Delta_j$, таковы, что координата в $\tilde{\delta}(S, S_i)$ с номером, равным элементу из Δ_j^1 , Δ_j^0 , есть соответственно 1, 0. Аналогично, число k_j элементов в $N_j \cap \Omega$ представим в виде $k_j^0 + k_j^1 = k_j$ (по значениям соответствующих координат в $\tilde{\delta}$), $j = 1, 2, \dots, v$.

Будем говорить, что множество $N_j \cap \Omega$ удовлетворяет условию Q_j , если число нулевых координат в $\tilde{\delta}(S, S_i)$ с номерами из $N_j \cap \Omega$ не превосходит ε_j^1 , $1 \leq j \leq v$.

Подсчитаем число $B(k_j, \Delta_j^1, \Delta_j^0, \varepsilon_j^1)$ таких множеств. Это делается аналогично подсчетам, проведенным в примере применения теоремы 2. Заметим, что если $\Delta_j^0 \leq \varepsilon_j^1$, то

$$B(k_j, \Delta_j^1, \Delta_j^0, \varepsilon_j^1) = C_{\Delta_j}^{k_j}. \quad (43)$$

При $\Delta_j^0 > \varepsilon_j^1$

$$B(k_j, \Delta_j^1, \Delta_j^0, \varepsilon_j^1) = \sum_{t=0}^{\varepsilon_j^1} C_{\Delta_j^1}^{k_j-t} \cdot C_{\Delta_j^0}^t. \quad (44)$$

Пусть теперь $B_\alpha(k_j, \Delta_j^0, \Delta_j^1, \varepsilon_j^1)$ — число множеств $N_j \cap \Omega$, удовлетворяющих условию Q_j и содержащих признак α такой, что α -я координата вектора $\tilde{\delta}(S, S_i)$ равна α , $\alpha = 0, 1$.

Нетрудно показать, что

$$B_0(k_j, \Delta_j^0, \Delta_j^1, \varepsilon_j^1) = \sum_{t=0}^{\varepsilon_j^1} C_{\Delta_j^1}^{k_j-t} \cdot C_{\Delta_j^0-1}^{t-1}, \quad (45)$$

$$B_1(k_j, \Delta_j^0, \Delta_j^1, \varepsilon_j^1) = \sum_{t=0}^{\varepsilon_j^1} C_{\Delta_j^1-1}^{k_j-t-1} \cdot C_{\Delta_j^0}^t. \quad (46)$$

Очевидно также, что $k = k_1 + \dots + k_v$.
С учетом (43) — (46), получаем

$$V_j^\alpha = \sum_{\substack{(k_1, \dots, k_v): \\ k_1 + \dots + k_v = k \\ \Delta_u \leq \Delta_u \\ \Delta_u^0 + \Delta_u^1 = \Delta_u}} \prod_{i \neq j} B(k_i, \Delta_i^0, \Delta_i^1, \varepsilon_i^1) \cdot B_\alpha(k_j, \Delta_j^0, \Delta_j^1, \varepsilon_j^1).$$

Аналогично можно получить формулы для V_j^α в случае, когда система Ω_A опорных подмножеств алгоритма A состоит из всех непустых подмножеств множества $\{1, 2, \dots, n\}$, функция $B(\tilde{\omega}S, \tilde{\omega}S_i)$ определяется параметрами $\varepsilon_1, \dots, \varepsilon_n, \varepsilon_1^1, \varepsilon_1^2, \dots, \varepsilon_v^1, \varepsilon_v^2$ и является симметрической по системе непересекающихся подмножеств N, \dots, N_v . Пусть

$$B(\tilde{\omega}S, \tilde{\omega}S_i) = \begin{cases} 1, & \text{если в векторе } \tilde{\delta}(S, S_i) \text{ среди координат с номерами из } \Omega_A \cap N_j \text{ число единичных координат не меньше } \varepsilon_j^1, \text{ число нулевых — не больше } \varepsilon_j^2, j = 1, 2, \dots, v, \\ 0 & \text{во всех остальных случаях.} \end{cases} \quad (47)$$

Как и ранее, через $\Delta_j, \Delta_j^0, \Delta_j^1$ обозначим число элементов в N_j , число элементов в N_j с нулевыми и единичными координатами в векторе $\tilde{\delta}(S, S_i)$, $j = 1, 2, \dots, v$.

Теорема 4.

$$V_j^0 = N \cdot \frac{\sum_{k=e_j^1}^{\Delta_j^1} C_{\Delta_j^1}^k \cdot \sum_{h=0}^{e_j^2-1} C_{\Delta_j^0}^h}{\sum_{k=e_j^1}^{\Delta_j^1} C_{\Delta_j^1}^k \cdot \sum_{h=0}^{e_j^2} C_{\Delta_j^0}^h}, \quad V_j^1 = N \cdot \frac{\sum_{k=e_j^1-1}^{\Delta_j^1-1} C_{\Delta_j^1-1}^k \cdot \sum_{h=0}^{-e_j^2} C_{\Delta_j^0}^h}{\sum_{k=e_j^1}^{\Delta_j^1} C_{\Delta_j^1}^k \cdot \sum_{h=0}^{e_j^2} C_{\Delta_j^0}^h},$$

$$N = \prod_{l=1}^v \left(\sum_{h=e_l^1}^{\Delta_l^1} C_{\Delta_l^1}^h \cdot \sum_{h=0}^{e_l^2} C_{\Delta_l^0}^h \right).$$

Доказательство. Заметим, что если хотя бы для одного j $\Delta_j^1 < e_j^1$, $j = 1, 2, \dots, v$, то не существует ни одного спорного Ω такого, что $B(\tilde{\omega}S, \tilde{\omega}S_i) \neq 0$. В этом случае $V_j^\alpha = 0$, $\alpha = 0, 1$, $j = 1, 2, \dots, v$.

Поэтому в дальнейшем будем считать, что $\Delta_j^1 \geq e_j^1$, $j = 1, 2, \dots, v$. Будем говорить, что $\Omega \cap N_j$ удовлетворяет условию P_j , если в векторе $\tilde{\delta}(S, S_i)$ не менее e_j^1 координат из $\Omega \cap N_j$ равны 1 и не более e_j^2 таких координат равны 0. Число множеств, удовлетворяющих P_j , обозначим через $D(e_j^1, e_j^2, \Delta_j^1, \Delta_j^0)$.

Здесь через Δ_j^1 , Δ_j^0 обозначено число единичных, соответственно нулевых координат в $\tilde{\delta}(S, S_i)$ с номерами из N_j , $j = 1, 2, \dots, v$.

Зафиксируем признак u в $\Omega \cap N_j$, сформируем множество $(\Omega \cap N_j) \setminus u = \Omega(j, u)$ и введем на нем свойства P_j^0 , P_j^1 .

$P_j^0(\Omega(j, u))$ — число единичных координат $\tilde{\delta}(S, S_i)$ с номерами из $\Omega(j, u)$ не меньше e_j^1 , число таких нулевых координат не больше $e_j^2 - 1$.

$P_j^1(\Omega(j, u))$ — число единичных координат $\tilde{\delta}(S, S_i)$ с номерами из $\Omega(j, u)$ не меньше $e_j^1 - 1$, число нулевых — не больше e_j^2 .

Число множеств Ω , $\Omega(j, u)$, удовлетворяющих соответственно условию P_j (для Ω), условиям P_j^0 , P_j^1 (для $\Omega(j, u)$) обозначим через $D(e_j^1, e_j^2, \Delta_j^0, \Delta_j^1)$, $D_j^0(e_j^1, e_j^2, \Delta_j^0, \Delta_j^1)$, $D_j^1(e_j^1, e_j^2, \Delta_j^0, \Delta_j^1)$. Аналогично через D_j^{0t} , D_j^{1t} обозначим число множеств $\Omega(j, u)$ при условии выполнения соответственно свойств P_j^0 , P_j^1 и дополнительного условия: u -я координата $\tilde{\delta}(S, S_i)$ равна t , $t \in \{0, 1\}$.

Нетрудно показать, что все величины D_j^{0t} (аналогично D_j^{1t}) равны для всех таких признаков u .

Имеет место равенство:

$$V_j^t = \prod_{l \in \{1, 2, \dots, v\} \setminus j} D(e_l^1, e_l^2, \Delta_l^1, \Delta_l^0) \cdot D_l^{tt}(e_j^1, e_j^2, \Delta_j^1, \Delta_j^0) = \\ = \prod_{l=1}^v D(e_l^1, e_l^2, \Delta_l^1, \Delta_l^2) \cdot \frac{D_j^{tt}(e_j^1, e_j^2, \Delta_j^1, \Delta_j^0)}{D(e_j^1, e_j^2, \Delta_j^1, \Delta_j^0)} \quad (48)$$

для всех признаков u из N_j , таких, что u -я координата вектора $\tilde{\delta}(S, S_i)$ равна t , $j = 1, 2, \dots, v$, $t = 0, 1$.

Нетрудно непосредственно подсчитать, что

$$\begin{aligned} D(e_i^1, e_i^2, \Delta_i^1, \Delta_i^2) &= (C_{\Delta_j^1}^{e_j^1} + C_{\Delta_j^1}^{e_j^1+1} + \dots + C_{\Delta_j^1}^{\Delta_j^1}) (C_{\Delta_j^2}^{e_j^2} + C_{\Delta_j^2}^{e_j^2-1} + \dots + C_{\Delta_j^2}^{\Delta_j^2}) = \\ &= \sum_{h=e_j^1}^{\Delta_j^1} C_{\Delta_j^1}^h \cdot \sum_{k=0}^{e_j^2} C_{\Delta_j^2}^k. \end{aligned} \quad (49)$$

Также прямо подсчитывается, что

$$D_j^{00} = \sum_{h=e_j^1}^{\Delta_j^1} C_{\Delta_j^1}^h \cdot \sum_{k=0}^{e_j^2-1} C_{\Delta_j^2}^k. \quad (50)$$

D_j^{11} подсчитывается аналогично.

Подставляя выражения D_j^{00}, D_j^{11}, D в (48), доказываем теорему.

В простых случаях формулы для величин $\Gamma_j(S)$ выглядят несложно. Так, если через $\|\tilde{\delta}(S, S_i)\|$ обозначить число единиц $\tilde{\delta}(S, S_i)$, рассмотреть функцию $B(\tilde{\omega}S, \tilde{\omega}S_i)$, которая равна 1 в том и только том случае, когда единичные координаты ω в $\tilde{\omega}(S, S_i)$ содержатся среди единичных координат $\tilde{\delta}(S, S_i)$ (все неравенства $\rho_\omega \leq \varepsilon_\omega$ выполнены), то

1° если система опорных подмножеств состоит из всех k -элементных подмножеств множества $\{1, 2, \dots, n\}$:

$$\Gamma_j(S) = \frac{1}{\mu(W_j^1)} \sum_{S_i \in W_j^1} \gamma(S_i) \cdot C_{\|\tilde{\delta}(S, S_i)\|}^k,$$

2° если система опорных подмножеств состоит из всех непустых подмножеств множества $\{1, 2, \dots, n\}$, то

$$\Gamma_j(S) = \frac{1}{\mu(W_j^1)} \sum_{S_i \in W_j^1} \gamma(S_i) \cdot (2^{\|\tilde{\delta}(S, S_i)\|} - 1).$$

Здесь $W_j^1 = K_j \cap \{S_1, \dots, S_m\}$, $I(S_u) \in I_0$, $u = 1, 2, \dots, m$.

II. Второй принцип построения формул для $\Gamma_j(S)$ основан на возможности специального представления характеристической функции $f_\Omega^A(\tilde{\omega})$ системы Ω_A опорных множеств алгоритма A . Известно, что всякую булевскую функцию f , $f \neq 0$, (а следовательно, и $f_\Omega^A(\tilde{\omega})$) можно представить в виде д. н. ф., составленной из ортогональных элементарных конъюнкций [74], т. е.

$$f_\Omega^A(\tilde{\omega}) = K_1 \vee \dots \vee K_r, \quad K_i \cdot K_j = 0 \quad \text{при } i \neq j, \quad (51)$$

$$K_i = x_{i_1}^{\sigma_{i_1}} \dots x_{i_q}^{\sigma_{i_q}}, \quad x^\sigma = \begin{cases} x, & \sigma = 1, \\ \bar{x}, & \sigma = 0. \end{cases}$$

В дальнейшем $f_\Omega^A(\tilde{\omega})$ будем обозначать как $f(x_1 \dots x_n)$ или просто f . Известно, что всякому представлению (51) соответствует разбиение множества N_f единиц функции на систему непересекающихся интервалов N_{K_1}, \dots, N_{K_r} : (N_{K_i} — множество единиц конъюнкции K_i), т. е.

$$N_f = N_{K_1} \cup \dots \cup N_{K_r}, \quad N_{K_i} \cap N_{K_j} = \emptyset \quad \text{при } i \neq j. \quad (52)$$

Пусть $\tilde{\omega}_A$ — совокупность характеристических векторов системы Ω_A опорных множеств алгоритма A . Задано разбиение $\tilde{\omega}_A = \bigcup_{i=1}^r \{\tilde{\omega}\}_i$ (следовательно, $\Omega_A = \bigcup_{i=1}^r \{\Omega\}_i$) на совокупность непересекающихся подмножеств $\{\tilde{\omega}\}_i$ (соответственно $\{\Omega\}_i$), $i = 1, 2, \dots, r$.

Рассмотрим совокупность алгоритмов A_1, \dots, A_r типа вычисления оценок с системами опорных множеств $\{\Omega\}_1 = \Omega_{A_1}, \dots, \{\Omega\}_r = \Omega_{A_r}$, у которых все остальные этапы II — VI определены одинаково так, что

$$\Gamma_j^{A_u}(S) = \frac{1}{N} \cdot \frac{1}{\mu(W_j^1)} \sum_{s_i \in W_j^1} \sum_{\tilde{\omega} \in \{\tilde{\omega}\}_i} \varphi(B(\tilde{\omega}S, \tilde{\omega}s_i, \pi_1, \dots, \pi_t)) \quad (53)$$

(здесь $\varphi(B, \pi_1, \dots, \pi_t)$ — произвольная числовая функция, определяющая оценку $\Gamma(\tilde{\omega}S, \tilde{\omega}s_i)$ по значению функции близости и фиксированным значениям числовых параметров π_1, \dots, π_t).

Так, ранее рассматривалась в качестве φ : $\gamma(S_i) \cdot (\vec{p}, \vec{\omega}) \cdot B(\tilde{\omega}S, \tilde{\omega}s_i)$, $(\vec{p}, \vec{\omega})$ — скалярное произведение числового вектора $(p_1 \dots p_n)$ на $\tilde{\omega}$, B — пороговая функция близости. Имеет место

Теорема 5 [14].

$$\Gamma_j^A(S) = \sum_{u=1}^r \Gamma_j^{A_u}(S).$$

Доказательство. Нетрудно видеть, что

$$\begin{aligned} \Gamma_j^A(S) &= \frac{1}{N} \cdot \frac{1}{\mu(W_j^1)} \cdot \sum_{s_i \in W_j^1} \sum_{\Omega \in \Omega_A} \varphi(B(\tilde{\omega}S, \tilde{\omega}s_i), \pi_1, \dots, \pi_t) = \\ &= \sum_{u=1}^r \left(\frac{1}{N \cdot \mu(W_j^1)} \cdot \sum_{s_i \in W_j^1} \sum_{\Omega \in \{\Omega\}_i} \varphi(B(\tilde{\omega}S, \tilde{\omega}s_i), \pi_1, \dots, \pi_t) \right) = \sum_{u=1}^r \Gamma_j^{A_u}(S). \end{aligned}$$

Теорема доказана.

Пусть алгоритмы A_1, A_2, A, B определены системами опорных подмножеств $\Omega_1, \Omega_2, \Omega_1 \cup \Omega_2, \Omega_{12} = \Omega_1 \cap \Omega_2$ и величины $\Gamma_j(S)$ вычисляются по формуле (53), где вместо $\{\tilde{\omega}\}_i$ подставлены соответственно $\Omega_1, \Omega_2, \Omega_1 \cup \Omega_2, \Omega_{12}$. Легко доказывается (доказательство не приводится)

Теорема 6 [14].

$$\Gamma_j^A(S) = \Gamma_j^{A_1}(S) + \Gamma_j^{A_2}(S) - \Gamma_j^B(S).$$

Из теорем 5, 6 следует, если для некоторых базисных подмножеств характеристических векторов (соответствующих им систем опорных множеств) вычисление оценок $\Gamma_j(S)$ производится по полученным формулам то можно получать формулы, представляя произвольные системы опорных множеств в виде сумм базисных подмножеств.

Оказывается, что если совокупность $\{\tilde{\omega}\}_A$ характеристических векторов образует интервал [74] N_K , $K = x_{i_1}^{\sigma_1} \dots x_t^{\sigma_t}$, а функция близости B

является пороговой, то тоже можно дать общий метод синтеза формул для величин $\Gamma_j(S)$. Далее с использованием представлений (51), (52) и теорем 5, 6 можно строить формулы для алгоритмов с достаточно разнообразными системами опорных множеств. Построение формулы для случая $\{\tilde{\omega}\}_A = N_K$, $K = x_{i_1}^{\sigma_1} \cdot \dots \cdot x_t^{\sigma_t}$ приводится ниже.

Конъюнкция K обращается в единицу на любых $\tilde{\omega} = (\omega_1 \dots \omega_n)$ таких, что $\omega_{i_1} = \sigma_1, \dots, \omega_{i_t} = \sigma_t$, остальные координаты — произвольные. Для простоты записи, не ограничивая общности, можно считать, что $K = x_1 \cdot \dots \cdot x_r \cdot x_{r+1} \cdot \dots \cdot x_t$ (такой записи всегда можно добиться перестановкой переменных). Очевидно, что в систему $\Omega_A = N_K$ войдут все подмножества, не содержащие $\{1, 2, \dots, r\}$ и включающие в себя $\{r+1, \dots, t\}$. Вычислим отдельно

$$\Gamma_j^0(S) = \frac{1}{\mu(W_j^1)} \sum_{S_i \in W_j^1} \Gamma(\tilde{\omega}S, \tilde{\omega}S_i), \quad \Omega = \{r+1, \dots, t\}.$$

Рассмотрим изученную ранее функцию $B^{\vec{e}}(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2)$ (36). Для нее в (37), (38) получены формулы для величин V^1 и V^0 , через которые с помощью теоремы 1 легко записывается формула для $\Gamma_j(S)$.

Обозначим число единичных, соответственно нулевых, координат вектора $\tilde{\delta}(S, S_i)$ с номерами из $\{r+1, \dots, t\}$ через $\Delta_{t-r}^1, \Delta_{t-r}^0$. Тогда $B^{\vec{e}}(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2) = 1$ на $(\tilde{\omega}S, \tilde{\omega}S_i)$ если и только если: 1° число единиц среди координат вектора $\tilde{\delta}(S, S_i)$ с номерами из $\{t+1, \dots, n\}$ не меньше $\tilde{\varepsilon}_1 - \Delta_{t-r}^1$ и 2° число нулей среди тех же координат не больше $\tilde{\varepsilon}_2 - \Delta_{t-r}^0$.

Таким образом, задача приведена к ранее решенной заменой: $n \rightarrow n-t, \tilde{\varepsilon}_1 \rightarrow \tilde{\varepsilon}_1 - \Delta_{t-r}^1, \tilde{\varepsilon}_2 \rightarrow \tilde{\varepsilon}_2 - \Delta_{t-r}^0$. Если выполнить соответствующую подстановку в (37) и (38), учесть слагаемое $\Gamma_j^0(S)$ и применить теорему 1, то получится

Теорема 7.

$$\begin{aligned} \Gamma_j(S) = & \frac{1}{N \cdot \mu(W_j^1)} \left(\Gamma_j^0(S) + \sum_{S_i \in W_j^1} \gamma(S_i) \cdot \left((\tilde{\delta}(S, S_i) \cdot \vec{p}) \times \right. \right. \\ & \times \sum_{u=0}^{\tilde{q}(S, S_i) - \tilde{\varepsilon}_1 + \Delta_{t-r}^1} C_{\tilde{q}(S, S_i)}^{\tilde{\varepsilon}_1 - \Delta_{t-r}^1} \cdot \sum_{u=0}^{\tilde{\varepsilon}_2 - \Delta_{t-r}^0} C_{n - \tilde{q}(S, S_i)}^u + \\ & + (\tilde{\delta}(S, S_i) \cdot \vec{p}) \cdot \sum_{u=0}^{\tilde{q}(S, S_i) - \tilde{\varepsilon}_1 + \Delta_{t-r}^1} C_{\tilde{q}(S, S_i)}^{\tilde{\varepsilon}_1 - \Delta_{t-r}^1 + u} \cdot \sum_{u=0}^{\tilde{\varepsilon}_2 - \Delta_{t-r}^0} C_{n - \tilde{q}(S, S_i)}^{u-1} \left. \right) \right), \end{aligned}$$

$\tilde{q}(S, S_i)$ — число единиц среди координат $\tilde{\delta}(S, S_i)$ с номерами из $\{t+1, \dots, n\}$.

Вообще, если получена формула для $\Gamma_j(S)$ в алгоритме с системой опорных подмножеств Ω_A , совпадающих со всеми подмножествами $\{1, 2, \dots, n\}$, то обычно приемом, подобным описанному при доказательстве теоремы 1, получается формула для случая $\{\tilde{\omega}\}_A = N_K$.

Раздел II. ДРУГИЕ МОДЕЛИ РАСПОЗНАЮЩИХ АЛГОРИТМОВ

§ 5. Модели алгоритмов, основанные на принципе разделения (R -модели) [10]

Алгоритмы, основанные на построении поверхностей, отделяющих классы друг от друга, появились ранее других алгоритмов. Они применяются главным образом для решения задач с обучающей информацией $I_0(K_1, \dots, K_l)$, заданной в стандартной форме. При этом рассматриваются обычно числовые признаки.

Рассмотрим сначала задачу с двумя непересекающимися классами K_1, K_2 . В этом случае в явной или неявной форме постулируется, что существует поверхность R достаточно простого вида, разделяющие элементы классов K_1 и K_2 , например, таким образом, что,

- если $S \in K_1$ и $I(S) = (a_1, \dots, a_n) = \tilde{a}$, то $R(\tilde{a}) > 0$;
- если $S' \in K_2$ и $I(S') = (b_1, \dots, b_n) = \tilde{b}$, то $R(\tilde{b}) \leq 0$.

Различные алгоритмы отличаются множеством поверхностей $\{R\}$, из которого выделяется разделяющая поверхность, а также способом выбора поверхности. Чаще всего в качестве $\{R\}$ рассматривается класс гиперплоскостей $\alpha_1x_1 + \dots + \alpha_nx_n + \alpha_{n+1} = 0$, кусочно-линейных поверхностей, поверхностей второго порядка.

Если все признаки в описаниях S являются бинарными, то совокупность элементов первого, соответственно второго класса отождествляется с множеством единиц, соответственно нулей не всюду определенной булевой функции $F(x_1, \dots, x_n)$ (область ее определения совпадает с множеством описаний $I(S_1), \dots, I(S_m)$ в стандартной информации) [5, 6]. Произвольная логическая формула $\Phi(x_1, \dots, x_n)$, реализующая F , является разделяющей.

В этом случае объект S относится к K_1 , соответственно к K_2 , если $F(I(S)) = 1$, соответственно $F(I(S)) = 0$.

Аналогично строится класс разделяющих поверхностей, если признаки $1, 2, \dots, n$ принимают каждый не более k градаций, и множество допустимых объектов разбито не более чем на k непересекающихся классов. В этом случае аналогично булевскому случаю строится не всюду определенная функция $F(x_1, \dots, x_n)$ k -значной логики, принимающая значение $j - 1$ на описаниях $I(S)$ объектов, принадлежащих классу K_j , $j = 1, 2, \dots, l$; $l \leq k$. Тогда разделятелем является произвольная формула k -значной логики, реализующая $F(x_1, \dots, x_n)$. В качестве соответствующей формулы естественно выбирать дизъюнктивную нормальную форму (случай двух классов) или ее аналог в k -значной логике.

Идея построения разделяющей поверхности методом последовательных приближений составляет основу метода потенциальных функций [4]. В этом методе в качестве класса возможных разделяющих поверхностей рассматривается весьма широкий класс поверхностей.

После того как класс $\{R\}$ разделяющих поверхностей зафиксирован, выбор одной из них обычно осуществляется решением некоторой экстремальной задачи. Например, может выбираться поверхность, которая на заданном контрольном множестве объектов дает максимальную точность распознавания.

Главный вывод, который можно сделать, изучив различные R -модели, состоит в следующем. Выполнение каждого из таких алгоритмов сводится к выполнению двух этапов. Пусть задача распознавания решается для допустимых объектов S'_1, \dots, S'_q .

1) По обучающей информации строится числовая матрица $\{R_{ij}\}_{q \times l}$. Элемент R_{ij} может быть интерпретирован как значение функции принадлежности S'_i классу K_j , $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$.

2) По матрице $\{R_{ij}\}_{q \times l}$ при помощи решающего правила алгоритма строится матрица $\{\alpha_{ij}\}_{q \times l}$, $\alpha_{ij} \in \{0, 1, \Delta\}$, информационных векторов элементов S'_1, \dots, S'_q . Если переход от I_0 к $\{R_{ij}\}_{q \times l}$ рассматривать как выполнение оператора R_A , а переход от $\{R_{ij}\}_{q \times l}$ к $\{\alpha_{ij}\}_{q \times l}$ как выполнение другого оператора r_A , то произвольный алгоритм из R -модели представим в виде $A = (R_A(I_0(l), I_S(q)) \cdot r_A \{R_{ij}\}_{q \times l})$. Под произведением операторов здесь понимается их последовательное применение.

Основные R -модели

I. $\mathfrak{M}(a_1, \dots, a_n, a_{n+1})$. Разделение производится гиперплоскостью $a_1x_1 + \dots + a_nx_n + a_{n+1} = 0$. Модель определяется заданием параметров a_1, \dots, a_{n+1} и решающего правила.

II. \mathfrak{M}_L^k — разделение производится кусочно-линейной поверхностью.

Любая разделяющая поверхность R делит пространство описаний $I(S)$ на два подмножества: $M^+(R) = \{I(S): R(I(S)) > 0\}$, $M^-(R) = \{I(S): R(I(S)) \leq 0\}$ запись $R(I(S)) \leq 0$ условна.

III. Модели $\mathfrak{M}^1(l)$ отличаются от моделей I — II тем, что разделение производится поверхностями R_1, \dots, R_l , причем поверхность R_j разделяет объекты, принадлежащие K_j и объекты, не принадлежащие K_j , $j = 1, 2, \dots, l$. Операторы R_A вводятся аналогично 1°, 2°.

IV. Модель $\mathfrak{M}^1(\gamma)$.

Пусть задана стандартная обучающая информация

$$I_0 = (I(S_1), \tilde{\alpha}(S_1), \dots, I(S_m), \tilde{\alpha}(S_m))$$

и описания $I(S'_1), \dots, I(S'_q)$ допустимых объектов. Каждому S_i сопоставлен параметр $\gamma(S_i) = \gamma_i > 0$. В классе $\{R\}$ выбрана разделяющая поверхность R .

Рассмотрим предикат $P_R(S_i)$:

$$P_R(S_i) = \begin{cases} 1, & \text{если } R(I(S_i)) > 0, \\ 0, & \text{если } R(I(S_i)) \leq 0. \end{cases}$$

Сопоставим каждому S_i пару (α, β) , где $\alpha = P_R(S_i)$, $\beta = \alpha_{ij} — j$ -я координата информационного вектора $\tilde{\alpha}(S_i)$.

Множество S_1, \dots, S_m разбивается на четыре непересекающихся подмножества $S_j^{00}, S_j^{01}, S_j^{10}, S_j^{11}$ элементов S_i , которым сопоставлены пары (α, β) , равные соответственно $(0, 0), (0, 1), (1, 0), (1, 1)$. Пусть

$$Q_{11}^j = \sum_{S_t \in S_j^{11}} \gamma_t, \quad Q_{10}^j = \sum_{S_t \in S_j^{10}} \gamma_t, \quad Q_{01}^j = \sum_{S_t \in S_j^{01}} \gamma_t, \quad Q_{00}^j = \sum_{S_t \in S_j^{00}} \gamma_t.$$

Сопоставим S'_i числовой вектор (r_{i1}, \dots, r_{il}) следующим образом.

$$\text{Если } R(I(S'_i)) > 0, \text{ то } R_{ij} = \frac{Q_{11}^j + Q_{00}^j}{Q_{10}^j + Q_{01}^j + 1}, \quad (54)$$

$$\text{если } R(I(S'_i)) \leq 0, \text{ то } R_{ij} = \frac{Q_{01}^j + Q_{10}^j}{Q_{11}^j + Q_{00}^j + 1}.$$

Условия (54) задают оператор R_A , переводящий I_0 в матрицу $\{R_{ij}\}_{q \times l}$.

V. Модель $\mathfrak{M}^1(\tilde{\gamma}, l)$, описывается аналогично модели IV, и отличается от нее тем, что для каждого класса K_j вычисление R_{ij} производится относительно поверхности R_j , $j = 1, 2, \dots, l$.

§ 6. Статистические модели и модели типа потенциалов

I. Статистические модели. Мы не будем подробно изучать эти модели. Для нас важно, что алгоритмы, принадлежащие к ним, переводят обучающую информацию I_0 и описания допустимых объектов $I(S'_1 \dots S'_q)$, т. е. пару $(I_0(l), I_S(q))$ в числовую матрицу $\{p_{ij}\}_{q \times l}$.

Затем по элементам этой матрицы формируются информационные векторы $\tilde{\alpha}_i^A(S) = (\alpha_{i1} \dots \alpha_{il})$ для объектов S'_i . Элементами p_{ij} могут быть например, вероятности вхождения S'_i в K_j , $j = 1, 2, \dots, l$, $i = 1, 2, \dots, q$; или другие статистические характеристики. Констатируем только факт, что в статистических алгоритмах, как и в других моделях, выделяются два этапа:

1° действует оператор R_A , переводящий $(I_0(l), I_S(q))$ в числовую матрицу $\{p_{ij}\}_{q \times l}$;

2° действует решающее правило r_A , переводящее $\{p_{ij}\}_{q \times l}$ в матрицу информационных векторов $\{\alpha_{ij}^A\}$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$.

II. Алгоритмы типа потенциальных функций. Рассмотрим случай числовых признаков и будем предполагать, что исходная информация задана в стандартной форме.

Выделим среди S_1, \dots, S_m множество W_j^1 объектов, принадлежащих классу K_j . Алгоритмы задаются следующим образом.

Рассматривается совокупность функций $\{F\}$. Потенциалом S'_i относительно K_j называется величина $F(I(W_j^1), I(S'_i))$, где $I(W_j^1)$ — совокупность описаний элементов из W_j^1 .

В качестве $\{F\}$ можно рассматривать, например, следующее множество функций. Пусть $\rho(I(S), I(S'_i))$ — расстояние в множестве описаний допустимых объектов, $\gamma(S_i) = \gamma_i$ — масса (или вес) объекта S_i . Положим

$$\rho(I(S'_i), I(W_j^1)) = \sum_{S_t \in W_j^1} \gamma_t \cdot \rho(I(S_t), I(S'_i)).$$

Тогда $\{F\} = \{e^{-\alpha \cdot \rho(x, y)}\}$ или $\{F\} = \left\{ \frac{c_1}{c_2 \cdot \rho(x, y) + c_3} \right\}$. (Возможны и другие способы задания класса $\{F\}$.) Потенциал S'_i относительно K_j равен соответственно

$$e^{-\alpha \cdot \sum_{S_t \in W_j^1} \gamma_t \cdot \rho(I(S_t), I(S'_i))}, \quad \frac{c_1}{c_2 \cdot \sum_{S_t \in W_j^1} \rho(I(S_t), I(S'_i)) + c_3}.$$

В общем случае величину потенциала объекта S'_i относительно K_j (множества W_j^1) будем обозначать через Π_{ij} .

Алгоритмы типа потенциалов также выполняются в два этапа.

1. На первом этапе по предъявленной контрольной выборке S'_1, \dots, S'_q строится матрица $\{\Pi_{ij}\}_{q \times l}$ значений потенциалов, т. е. работает оператор R_A , переводящий $(I_0(l), I_S(q))$ в числовую матрицу $\{\Pi_{ij}\}_{q \times l}$.

2. По набору потенциалов $\Pi_{i1}, \dots, \Pi_{il}$ (или по всей матрице $\{\Pi_{ij}\}$) делается вывод о вхождении S'_i в K_j , $j = 1, \dots, l$. Другими словами, по матрице потенциалов $\{\Pi_{ij}\}_{q \times l}$ строится матрица информационных векторов $\{\alpha_{ij}^A\}_{q \times l}$, т. е. применяется решающее правило r_A и $r_A(\Pi_{ij}) = \{\alpha_{ij}^A\}_{q \times l}$.

Статистические алгоритмы, алгоритмы типа потенциалов и их различные подклассы могут описываться при помощи различных моделей. Принципы построения таких моделей родственны тем, которые применялись при описании Г-моделей (вычисления оценок) или R -моделей.

§ 7. Структурные алгоритмы распознавания [78, 33]

Во введении была описана принципиальная схема структурных методов распознавания. Поэтому мы ограничимся разбором одного примера.

Пусть задан конечный алфавит $\mathcal{A} = \{\Lambda, a_1, \dots, a_n\}$, и описаниями объектов, подлежащих распознаванию, являются фразы из Σ (определение 5).

Класс $\{\Phi\}_j$ строится из элементарных предикатов, $Q_{jt}^{f,1}(S)$ — функция принадлежности (оценка) объекта классу K_j , больше, чем функция принадлежности (оценка) S классу K_t .

Вместе с $Q_{jt}^{f,1}(S)$ в $\{\Phi\}_j$ включаются также предикаты $\bar{Q}_{jt}^{f,1}(S)$ — оценка S для K_j не больше оценки S для K_t .

Вместе с указанными выше предикатами $Q_{jt}^{f,1}, \bar{Q}_{jt}^{f,1}$ включаются также их всевозможные логические произведения.

Множество формул $\{\Phi\}_j$ составляется из элементарных предикатов $P(r, a_i, S)$ — во фразе S r -я буква есть a_i , $r = 1, 2, \dots, t, \dots, i = 1, 2, \dots, n$.

Пусть также известно, что классы K_j не пересекаются, $j = 1, 2, \dots, l$.

Из базисных элементов $\{B\}^j$ при помощи операций $\{O\}^j$ построено множество описаний M_j объектов из K_j . Введены числовые параметры $\delta_{ut}^1, \delta_{ut}^2, u = 1, 2, \dots, l, t \neq u$.

Рассматриваем описание S опознаваемого объекта, $S \in M_j, j = 1, 2, \dots, l$.

Пусть r -я буква в S есть a_i , число слов в M_j , содержащих r -ю букву a_i обозначим через $Q_j(r, a_i)$, аналогичную величину для M_t — через $Q_t(r, a_i)$.

Если $Q_j(r, a_i) - Q_t(r, a_i) > \delta_{jt}^1$, включаем в $\{\chi\}_j$ формулу $P(r, a_i, S) \rightarrow Q_{jt}^{f,1}$.

Если $Q_t(r, a_i) - Q_j(r, a_i) > \delta_{jt}^2$, включаем в $\{\chi\}_j$ формулу $P(r, a_i, S) \rightarrow \bar{Q}_{jt}^{f,1}$.

Если в $\{\chi\}_j$ включены формулы $\Phi_1 = (P(k, a_u, S) \rightarrow \bar{Q}_{jt}^{f,1}), \Phi_2 = (P(k, a_u, S) \rightarrow \bar{Q}_{ju}^{f,1})$, то в $\{\chi\}_j$ включается также формула $P(k, a_u, S) \rightarrow \bar{Q}_{jt}^{f,1} \cdot \bar{Q}_{ju}^{f,1}$.

Знак \tilde{Q} означает, что $\tilde{Q} = Q$ либо $\tilde{Q} = \bar{Q}$.

Введем параметры $x_{tj}^1(r, a_i), x_{tj}^2(r, a_i), j = 1, 2, \dots, l, t = 1, 2, \dots, j-1, j+1, \dots, l$.

Пусть $S = (b_1 b_2 \dots b_q)$. Сопоставим $b_i, i = 1, 2, \dots, q$, число

$$\Gamma_j(b_i, S) = \begin{cases} 0, & \text{если } b_i = \Lambda \text{ или в } \{\chi\}_j \text{ нет формул,} \\ & \text{содержащих символ } b_i, \\ \sum_t x_{tj}^\alpha & \text{в остальных случаях.} \end{cases}$$

Суммирование производится по номерам t , для которых в $\{\chi\}_j$ включена формула $P(i, b_i, S) \rightarrow \bar{Q}_{jt}^{f,1}$. Если $\bar{Q}_{jt}^{f,1} = Q_{jt}^{f,1}$, то $\alpha = 1$. Если $\bar{Q}_{jt}^{f,1} = \bar{Q}_{jt}^{f,1}$, то $\alpha = 2$.

Объекту S сопоставляется оценка

$$\Gamma_j(S) = \frac{1}{q} \sum_{i=1}^q \Gamma_j(b_i, S) \cdot \gamma_i.$$

Величины $\gamma_1, \dots, \gamma_q, \dots$ являются числовыми параметрами.

Мы описали модель структурных распознающих алгоритмов с параметрами $\delta_{it}^1, \delta_{it}^2, x_{ij}^1(r, a_i), x_{ij}^2(r, a_i), \gamma_i$.

Вычисления по этой модели переводят структурную обучающую информацию и описания распознаваемых объектов S_1, \dots, S_q в числовую матрицу оценок $\{\Gamma_{ij}\}_{q \times l}$. Параметры модели могут выбираться из априорных соображений или отыскиваться как решение экстремальной задачи.

§ 8. Решающие правила в алгоритмах распознавания

Решающим правилом r_A алгоритма распознавания A называется оператор, переводящий матрицу оценок $\{a_{ij}\}_{q \times l}$ в матрицу информационных векторов $\{\alpha_{ij}^A\}_{q \times l}$ распознаваемых элементов. Ниже будут рассмотрены три основных типа решающих правил.

I. Функциональные r_A порядка 1. Вычисление α_{ij}^A происходит следующим образом. Вводится числовая функция $f(x)$. В множестве M_f значений $f(x)$ выделяются три непересекающихся множества M_f^Δ, M_f^1, M_f^0 :

$$\alpha_{ij}^A(a_{ij}) = \begin{cases} \Delta, & \text{если } f(a_{ij}) \in M_f^\Delta, \\ 1, & \text{если } f(a_{ij}) \in M_f^1, \\ 0, & \text{если } f(a_{ij}) \in M_f^0. \end{cases} \quad (55)$$

Функция $f(x)$ называется базой решающего правила r_A .

II. Функциональные правила порядка l . Задаются l функций $f_1(x_1, \dots, x_l), \dots, f_l(x_1, \dots, x_l)$ — база решающего правила, функции $f_j, j = 1, 2, \dots, l$, необязательно различны.

Как и в I, множество M_{f_j} значений f_j разбивается на три подмножества: $M_{f_j}^\Delta, M_{f_j}^1, M_{f_j}^0$,

$$\alpha_{ij}^A = \begin{cases} \Delta, & \text{если } f_j(a_{i1}, \dots, a_{il}) \in M_{f_j}^\Delta, \\ 0, & \text{если } f_j(a_{i1}, \dots, a_{il}) \in M_{f_j}^0, \\ 1, & \text{если } f_j(a_{i1}, \dots, a_{il}) \in M_{f_j}^1. \end{cases} \quad (5.6)$$

Другой класс решающих правил порядка l задается базой из одной функции $f(x_1, \dots, x_l)$. Пусть $\{\tilde{\alpha}\}$ — множество информационных векторов длины l и в M_f выделены подмножества $M_f^{\tilde{\alpha}}, \tilde{\alpha} \in \{\tilde{\alpha}\}$,

$$f(a_{i1}, \dots, a_{il}) = (\alpha_{i1}^A, \dots, \alpha_{il}^A), \quad \text{если } f(a_{i1}, \dots, a_{il}) \in M_f^{\tilde{\alpha}}. \quad (57)$$

Среди функциональных решающих правил порядка l особенно часто применяются линейные правила. Базой в этих правилах являются линейные функции: $b_{j1}x_1 + \dots + b_{jl}x_l + b_{j,l+1}, j = 1, 2, \dots, l$, или функция $a_1x_1 + \dots + a_lx_l + a_{l+1}$.

Разбиение в линейных правилах производится с помощью констант $c_{1j}, c_{2j}, c_{1j} \leq c_{2j}, j = 1, 2, \dots, l$.

III. Решающие правила счетного порядка. Базой является последовательность функций $f_1^j(x_1), \dots, f_h^j(x_1, \dots, x_k), \dots$

Выбирается функция $f_{q,l}(x_1, \dots, x_{q,l})$. Пусть $(\alpha_{ij})_{q \times l} = N_{q \times l}$ — совокупность информационных матриц размерности $q \times l$.

В $M_{f_{q,l}}$ выбираются для каждой $N \in N_{q \times l}$ подмножества $M_{f_{q,l}}(N)$

$$r_A(\{\alpha_{ij}\}_{q \times l}) = N, \text{ если } f_{q,l}(a_{11}, \dots, a_{ij}, \dots, a_{q,l}) \in M_{f_{q,l}}(N). \quad (58)$$

ГЛАВА III

ОБЩАЯ ТЕОРИЯ РАСПОЗНАЮЩИХ (КЛАССИФИЦИРУЮЩИХ) АЛГОРИТМОВ

§ 1. Определение класса распознающих алгоритмов [26]

На основании изучения моделей распознавания может быть предложено следующее определение распознающего алгоритма.

Пусть задано $I = \{I_0(K_1, \dots, K_l)\}$ и для каждого S из M определен класс допустимых описаний $I_S = \{I(S)\}$.

Определение 19. Алгоритм A называется распознающим, если он переводит обучающую информацию $I_0(K_1, \dots, K_l)$ и описания произвольного конечного числа q допустимых объектов $I(S'_1), \dots, I(S'_q)$ в матрицу $\{\alpha_{ij}^A\}_{q \times l}$, составленную из элементов $\{1, 0, \Delta\}$.

В определении предполагается, что $I(S'_u) \in I_{S'_u}, u = 1, 2, \dots, q, I_0 \in I$.

Набор $(I(S'_1), \dots, I(S'_q))$ будем обозначать в дальнейшем через $I(S'_1, \dots, S'_q)$. Тогда

$$A(I_0(K_1, \dots, K_l), I(S'_1, \dots, S'_q)) = \{\alpha_{ij}^A\}_{q \times l} \quad (59)$$

является символической записью распознающего алгоритма.

Информацию $I_0(K_1, \dots, K_l)$ будем иногда обозначать через $I_0(l)$, набор $I(S'_1, \dots, S'_q)$ — через $I_S(q)$. Тогда вместо (59) можно написать

$$A(I_0(l), I_S(q)) = \{\alpha_{ij}^A\}_{q \times l}. \quad (60)$$

В дальнейшем для элементов α_{ij}^A принятая стандартная интерпретация:

$\alpha_{ij}^A = 1$ — элемент S'_i входит в класс K_j ;

$\alpha_{ij}^A = 0$ — элемент S'_i не входит в класс K_j ;

$\alpha_{ij}^A = \Delta$ — алгоритм A не установил, входит или нет элемент S'_i в класс $K_j, i = 1, 2, \dots, q; j = 1, 2, \dots, l$.

В конкретных моделях алгоритмов распознавания алгоритм A выполнялся в два этапа. На первом этапе к набору $(I_0(l), I_S(q))$ применялся оператор, переводящий $(I_0(l), I_S(q))$ в числовую матрицу $\{a_{ij}\}_{q \times l}$. На втором этапе применением решающего правила матрица $\{a_{ij}\}_{q \times l}$ переводилась в информационную матрицу $\{\alpha_{ij}^A\}_{q \times l}$.

Введем подкласс распознающих алгоритмов, выполняемых в два этапа.

Определение 20. Оператор R_A называется распознающим оператором, если он перерабатывает $(I_0(K_1, \dots, K_l), I(S'_1, \dots, S'_q), I_0(K_1, \dots, K_l) \in I, S'_i$ — произвольные допустимые объекты, $i = 1, 2, \dots, q, q$ — произвольное целое число, $I(S'_i) \in I_{S'_i}$ — в числовую матрицу $\{a_{ij}\}_{q \times l} = M_{q \times l}$.

Очевидно, $M_{q \times l} = M_{q \times l}(I_0(l), I_S(q))$. Символическая запись распознающего оператора:

$$R_A(I_0(K_1, \dots, K_l), I(S'_1, \dots, S'_q)) = \{a_{ij}\}_{q \times l}$$

или, сокращенно,

$$R_A(I_0(l), I_S(q)) = \{a_{ij}\}_{q \times l}. \quad (61)$$

Определение 21. Оператор r_A называется решающим правилом, если он переводит произвольную числовую матрицу $\{a_{ij}\}_{q \times l} = M_{q \times l}$ в информационную матрицу $\{\alpha_{ij}^A\}_{q \times l} = I_{q \times l}$, т. е. матрицу, составленную из элементов $\{1, 0, \Delta\}$.

Символическая запись решающего правила:

$$r_A(\{a_{ij}\})_{q \times l} = \{\alpha_{ij}^A\}_{q \times l}, \quad r_A(M_{q \times l}) = I_{q \times l}. \quad (62)$$

Очевидно определена операция произведения операторов R_A и r_A — обозначение: $R_A \cdot r_A$, которая сводится к последовательному выполнению R_A , r_A .

Определение 22. Оператор $R_A \cdot r_A = \tilde{A}$ называется стандартным распознающим алгоритмом (р. с. а.).

В дальнейшем мы будем рассматривать в основном р. с. а. Символическая запись р. с. а. \tilde{A} та же, что для произвольного распознающего алгоритма:

$$\tilde{A}(I_0(l), I_S(q)) = \{\alpha_{ij}^{\tilde{A}}\}_{q \times l} = I_{q \times l}. \quad (63)$$

Очевидно, что стандартный распознающий алгоритм является распознающим алгоритмом, т. е. $\{\tilde{A}\} \subseteq \{A\}$.

Нетрудно показать, что всякий распознающий алгоритм A можно превратить в стандартный \tilde{A} . Действительно, пусть $A(I_0(l), I_S(q)) = \{\beta_{ij}\}_{q \times l}$. Рассмотрим алгоритм $O_N \{\beta_{ij}\}_{q \times l} = \{a_{ij}\}_{q \times l}$, где $\{a_{ij}\}$ — числовая матрица, определенная следующими соотношениями:

- если $\beta_{ij} \in \{0, 1\}$, то $a_{ij} = N \cdot \beta_{ij}$, $N > 1$;
- если $\beta_{ij} = \Delta$, то $a_{ij} = 1$.

Применим к $\{a_{ij}\}_{q \times l}$ решающее правило \tilde{r}_A :

если $a_{ij} \geq N$, то $S'_i \in K_j$, т. е. $\beta_{ij} = 1$; при $a_{ij} \leq 0$: $S'_i \notin K_j$, т. е. $\beta_{ij} = 0$;
если $0 < a_{ij} < N$, то $\beta_{ij} = \Delta$. Из определений O_N и \tilde{r}_A следует, что $A \cdot O_N \cdot \tilde{r}_A = A$, и алгоритм $(A \cdot O_N) \tilde{r}_A$ является стандартным распознающим алгоритмом. Нами доказана

Теорема 8. Для каждого распознающего алгоритма A существует стандартный распознающий алгоритм \tilde{A} такой, что для каждой пары $(I_0(l), I_S(q))$ выполнено равенство:

$$A(I_0(l), I_S(q)) = \tilde{A}(I_0(l), I_S(q)).$$

В дальнейшем мы будем рассматривать лишь решающие правила \tilde{r}_A , удовлетворяющие дополнительному условию.

Пусть S'_1, \dots, S'_q — произвольная конечная совокупность допустимых объектов, $P_j(S'_i)$ — предикат: $S'_i \in K_j$, $j = 1, 2, \dots, l$. Напомним, что информационная матрица $\{\alpha_{ij}^A\}_{q \times l}$, $\alpha_{ij}^A \in \{0, 1\}$ называется истинной для S'_1, \dots, S'_q , если $\alpha_{ij}^A = P_j(S'_i)$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$.

Определение 23. Решающее правило \tilde{r}_A называется корректным, если для всякой конечной совокупности допустимых объектов S'_1, \dots, S'_q существует числовая матрица $\{a_{ij}\}_{q \times l}$ такая, что \tilde{r}_A переводит $\{a_{ij}\}_{q \times l}$ в матрицу, истинную для S'_1, \dots, S'_q .

В дальнейшем мы будем рассматривать только корректные решаютые правила.

§ 2. Линейное пространство распознающих операторов

В множестве распознающих операторов $\{R_A\}$ естественным образом вводятся операции сложения и умножения на скаляр.

Пусть $R_A^1(I_0(l), I_S(q)) = \{a_{ij}\}_{q \times l}$, $R_A^{2l}(I_0(l), I_S(q)) = \{b_{ij}\}_{q \times l}$.

Определение 24. Оператор $R_A = R_{A_1} + R_{A_2}$ называется суммой операторов R_{A_1}, R_{A_2} , если $R_A(I_0(l), I_S(q)) = \{c_{ij}\}_{q \times l} = \{a_{ij} + b_{ij}\}_{q \times l}$.

Распознающий оператор $R_A = c \cdot R_A^1$, c — скаляр, называется произведением скаляра c на оператор R_A^1 , если

$$R_A(I_0(l), I_S(q)) = \{c \cdot a_{ij}\}_{q \times l}.$$

Теорема 9. По операциям сложения и умножения на скаляр множество распознающих операторов образует линейное векторное пространство. Умножение на скаляр коммутативно.

Доказательство легко следует из того, что множество числовых матриц одной размерности образует коммутативную группу по сложению. Из этого факта и из определения распознающего оператора легко следует, что множество распознающих операторов образует абелеву группу по сложению. Аксиомы, связанные с умножением на скаляр и дистрибутивностью умножения на скаляр легко проверяются. Очевидна также коммутативность умножения на скаляр, т. е. $a \cdot b \cdot R_A = b \cdot a \cdot R_A$. Теорема доказана.

Таким образом, по заданной серии р. с. а. $A_1 = (R_A, r_A), \dots, A_t = (R_{A_t}, r_A)$ можно порождать новые алгоритмы $A = (\sum_{i=1}^t a_i R_{A_i}, r_A)$, a_i — константы. Алгоритм A называется в этом случае линейным по A_1, \dots, A_t . Набор A_1, \dots, A_t называется базой A . Пусть задана совокупность $\{A\}$ р. с. а. с фиксированным решающим правилом r_A .

Определение 25. Линейным замыканием $L\{A\}$ множества (модели) $\{A\}$ называется совокупность линейных по всевозможным конечным базам из $\{A\}$ распознающих алгоритмов с фиксированным решающим правилом r .

Очевидно, каждый алгоритм B из $L\{A\}$ представим в виде $A = (\sum_{A_i \in \{A\}} a_i R_{A_i}, r)$. Определение позволяет строить линейное замыкание для любой модели $\mathfrak{M}(A)$ распознающих алгоритмов.

Если в определении 25 вместо р. с. а. рассматривать распознающие операторы, то получается определение линейного замыкания множества $\{R_A\}$ распознающих операторов $L\{R_A\}$.

Пусть, наконец, задано множество (модель) р. с. а. $\mathfrak{M} = \{(R_A, r_A)\}$ с различными решающими правилами. Представим модель \mathfrak{M} в виде $\bigcup_{r_A} \mathfrak{M}(r_A)$. Здесь $\mathfrak{M}(r_A)$ — совокупность всех алгоритмов модели \mathfrak{M} с фиксированным решающим правилом r_A .

Определение 26. Множество $L(\mathfrak{M}) = \bigcup_{r_A \in \{r_A\}} L(\mathfrak{M}(r_A))$ называется линейным замыканием модели \mathfrak{M} .

Пусть задано конечное множество моделей $\mathfrak{M}_1, \dots, \mathfrak{M}_t$,

$$\mathfrak{M}_i = \{(R_A^i r_A^i)\}, R_A^i \in \{R_A^i\}, r_A^i \in \{r_A^i\}, i = 1, 2, \dots, t.$$

Определение 27. Линейным замыканием моделей $\mathfrak{M}_1, \dots, \mathfrak{M}_t$ называется множество $L(\bigcup_{i=1}^t L(\mathfrak{M}_i)) = L(\mathfrak{M}_1, \dots, \mathfrak{M}_t)$.

Всякий алгоритм A из $L(\mathfrak{M}_1, \dots, \mathfrak{M}_t)$ может быть построен следующим образом. Из множества $\bigcup_{i=1}^t \mathfrak{M}_i$ выбирается база A'_u, \dots, A_u^p (т. е. произвольный набор алгоритмов с одинаковым решающим правилом r).
 $A = (R_A r_A), r_A = r, R_A = \sum_{i=1}^p c_i R_{A'_u}$.

Доказательство последнего утверждения легко следует из определений линейного замыкания модели и линейного замыкания совокупности моделей. Достаточно показать, что

$$L(\bigcup_{i=1}^t \mathfrak{M}_i) = L(\bigcup_{i=1}^t L(\mathfrak{M}_i)). \quad (64)$$

Распознающие операторы R_A могут применяться, вообще говоря, к частям информации I_0 . В дальнейшем мы будем рассматривать линейные замыкания моделей, где в качестве различных слагаемых фигурируют одинаковые операторы, применяемые к различным частям $I(K_1, \dots, K_l)$. В этих случаях будет указываться часть I_0 — область действия оператора R_A .

§ 2. Экстремальные алгоритмы в линейных замыканиях моделей. Полные модели

Мы будем рассматривать общий вид функционала качества

$$f_q(\rho(\tilde{\alpha}(S'_1), \tilde{\alpha}^A(S'_1)), \dots, \rho(\tilde{\alpha}(S'_q), \tilde{\alpha}^A(S'_q))),$$

где $\tilde{\alpha}(S'_i)$ — истинный информационный вектор S'_i , $\tilde{\alpha}^A(S'_i)$ — информационный вектор для S'_i , сформированный алгоритмом A .

Напомним, что $f_q(0, \dots, 0)$ есть абсолютный максимум по всей области определения f_q , т. е. f_q достигает абсолютного максимума, если $\tilde{\alpha}(S'_i) = \tilde{\alpha}^A(S'_i)$, $i = 1, 2, \dots, l$.

Пусть задана обучающая информация I_0 и произвольная конечная выборка S'_1, \dots, S'_q из M , $I_0 \in \{I_0\}$ — множества допустимых обучающих информаций.

Определение 28. Модель \mathfrak{M} называется корректной, если для любых $I_0, q, S'_1, \dots, S'_q$ в \mathfrak{M} существует алгоритм A , такой, что на A достигается абсолютный максимум функционала качества f_q .

Пусть модель \mathfrak{M} составлена из $A = R_A r_A$. Обозначим через $\mathfrak{M}\{R_A\}$ совокупность распознающих операторов модели \mathfrak{M} .

Определение 29. Модель \mathfrak{M} называется полной, если для любых $I_0, q, S'_1, \dots, S'_q$ множество матриц размерности $q \times l: R_A \{I_0(K_1, \dots, K_l), I(S'_1, \dots, S'_q)\} = M(R_A)$, $R_A \in \mathfrak{M}\{R_A\}$ содержит базис в пространстве числовых матриц $q \times l$. Совокупность $\mathfrak{M}\{R_A\}$ операторов полной модели называется полной.

Теорема 10. Если модель \mathfrak{M} является полной и алгоритмы из \mathfrak{M} имеют только корректные решающие правила, то линейное замыкание $L(\mathfrak{M})$ модели \mathfrak{M} является корректным.

Доказательство. Пусть задана произвольная конечная выборка S'_1, \dots, S'_q , начальная информация I_0 и множество описаний $\{I(S'_1), \dots, I(S'_q)\} = I_S(q)$. Базис в $M_{q \times l}$ состоит из $q \cdot l$ матриц $M^1, \dots, M^{q \cdot l}$.

Рассмотрим решающее правило r . Так как мы рассматриваем только корректные решающие правила, то существует числовая матрица $M_q(S) = \{a_{ij}\}_{q \times l}$, такая, что $r(M_q(S)) = \{P_j(S'_i)\}_{q \times l}$, т. е. информационная матрица $r(M_q(S))$ есть матрица истинных информационных векторов для объектов S'_1, \dots, S'_q . (напомним, что через $P_j(S'_i)$ обозначалось значение предиката «объект S'_i входит в класс K_j »).

Теперь достаточно построить матрицу $M_q(S)$ по $(I_0(l), I_S(q))$. Так как множество матриц $M\{R_A\}$ содержит базис, и этот базис является конечным (состоит из $q \cdot l$ элементов), то можно выделить $q \cdot l$ операторов $R_{A_1}, \dots, R_{A_{q \cdot l}}$, которые в применении к $(I_0(l), I_S(q))$ порождают базисные матрицы $M^1, \dots, M^{q \cdot l}$.

Тогда (по определению базиса в линейном пространстве) существуют константы $a_1, \dots, a_{q \cdot l}$, такие, что

$$M_q(S) = \sum_{i=1}^{q \cdot l} a_i \cdot M^i.$$

Распознающий оператор $R_A = \sum_{i=1}^{q \cdot l} a_i \cdot R_{A_i}$ входит в линейное замыкание системы операторов, содержащей операторы $R_{A_1}, \dots, R_{A_{q \cdot l}}$.

Рассмотрим алгоритм $(R_A r) = A$, где $R_A = \sum_{i=1}^{q \cdot l} a_i \cdot R_{A_i}$. Очевидно, $A(I_0(l), I_S(q)) = \{\alpha_{ij}^A\}_{q \times l} = \{P_j(S'_i)\}_{q \times l}$.

Тогда $\rho(\tilde{\alpha}_A(S'_1), \tilde{\alpha}(S'_1)) = \rho(\tilde{x}, \tilde{x}) = 0$ — функция ρ удовлетворяет аксиоме расстояния: $\rho(a, a) = 0$. Поэтому значение функционала качества

$$f_q(\rho(\tilde{\alpha}(S'_1), \alpha_A(S'_1)), \dots, \rho(\tilde{\alpha}_q(S'_q), \alpha_A(S'_q))) = f(0, 0, \dots, 0, 0).$$

По определению абсолютный максимум функционала f_q достигается на наборе $(0, 0, \dots, 0, 0)$. Поэтому линейное замыкание $L(\mathfrak{M})$ модели \mathfrak{M} является корректным. Теорема доказана.

Следствие. Если модель \mathfrak{M} является полной (и, следовательно, корректной), то модель \mathfrak{M}' с одним фиксированным решающим правилом r и совокупностью распознающих операторов $\mathfrak{M}(R_A)$ является полной, и, следовательно, корректной.

Рассмотрим одно применение теоремы 10.

Мы будем рассматривать случай, когда информация $I_0(K_1, \dots, K_l) = I_0(l)$ и описания $\{I(S'_1), \dots, I(S'_q)\} = I_S(q)$ объектов, подлежащих распознаванию, заданы в стандартной форме.

Рассмотрим R -алгоритмы с кусочно-линейными разделяющими поверхностями и параметрами $\gamma_1, \dots, \gamma_m$ (модель $\mathfrak{M}^1(\tilde{\gamma}, l)$, где $R_1 = \dots = R_l$). Описания объектов S_1, \dots, S_m , входящие в I_0 , и описания $I(S'_1), \dots, I(S'_q)$ являются точками n -мерного евклидова пространства. Разделение производится одной поверхностью R для всех классов. Алгоритм определяется заданием поверхности R , и параметров $\gamma_1, \dots, \gamma_m$ — весов объектов S_1, \dots, S_m .

В дальнейшем стандартную информацию $I_0 = (I(S_1), \tilde{\alpha}(S_1), \dots, I(S_m), \tilde{\alpha}(S_m))$ подчиним следующим дополнительным условиям:

1) векторы $\tilde{\alpha}(S_i)$ не содержат знаков Δ и являются истинными для S_i , т.е. если $\tilde{\alpha}(S_i) = (\alpha_{i1} \dots \alpha_{il})$, то α_{ij} есть значение предиката $P_j(S_i)$ — « S_i входит в класс K_j »;

2) представим множество $\{S_1, \dots, S_m\} = W_1^1 \cup W_1^0$; подмножества W_1^1, W_1^0 — непересекающиеся и составленные из элементов S_i , для которых $\alpha_{ij} = 1$, соответственно $\alpha_{ij} = 0$, $i = 1, 2, \dots, m$; $j = 1, 2, \dots, l$.

Условие: $W_u^1 \not\subseteq W_v^1$, если $v \neq u$, $u = 1, 2, \dots, l$. Условие 2) означает, что в обучающей выборке для каждой пары классов K_u, K_v существует хотя бы один объект, входящий в K_u , и не входящий в K_v .

Кроме того будем считать, что $S_t \in \{S_1, \dots, S_m\}$, $t = 1, 2, \dots, q$.

Теорема 11*). Класс распознающих операторов R_A , определяемых кусочно-линейной разделяющей поверхностью R и параметрами $\{\gamma_1, \dots, \gamma_m\} = \{\gamma(S_1), \dots, \gamma(S_m)\}$, является полным ($\gamma_i \geq c > 0$).

Доказательство. Зафиксируем произвольную контрольную выборку S'_1, \dots, S'_q . Покажем, что множество матриц $M_{q \times l}$, порождаемых операторами R_A по стандартной начальной информации $(I_0(l), I_S(q))$, содержит базис в линейном пространстве числовых матриц размерности $q \times l$.

Доказательство состоит в прямом построении операторов $R_A^{i,j}$ из $L(R_A)$ (линейного замыкания класса R_A), переводящих $(I_0(l), I_S(q))$ в числовую матрицу $\{b_{uv}\}_{q \times l}^{i,j}$, такую, что $b_{ij} = 1$ и остальные $b_{uv} = 0$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$. Очевидно, что матрицы $\{b_{uv}\}_{q \times l}^{i,j}$ образуют базис в множестве $M_{q \times l}$ числовых матриц размерности $q \times l$.

I. Проведем произвольную кусочно-линейную поверхность R_1 , с выполнением следующих условий.

$$1^\circ R_1(S'_i) > 0;$$

$$2^\circ R(I(S_u)) > 0 \text{ для всех } S_u \text{ из } W_1^1;$$

$$3^\circ R_1(I(S_v)) \leq 0, v = 1, 2, \dots, i-1, i+1, \dots, q;$$

$$4^\circ R_1(I(S_u)) \leq 0 \text{ для всех } S_u \text{ из } W_1^0.$$

Пусть $\gamma_1 = \dots = \gamma_m = N^k$, $N > 1$, k — натуральное число. Вычислим матрицу $\{r_{uv}^1\}_{q \times l}$, в которую оператор R_1 переводит начальную информацию. Напомним, что для объектов S'_u таких, что

$$R(S'_u) > 0: r_{uv}^1 = \frac{Q_v^{11} + Q_v^{00}}{Q_v^{01} + Q_v^{10} + 1};$$

для объектов S'_u таких, что

$$R(S'_u) \leq 0: r_{uv}^1 = \frac{Q_v^{01} + Q_v^{10}}{Q_v^{00} + Q_v^{11} + 1}.$$

($Q_v^{\alpha_1 \alpha_2}$ — сумма весов γ_t элементов S_t , входящих в $S_v^{\alpha_1 \alpha_2}$, $\alpha_1, \alpha_2 \in \{0, 1\}$; α_2 — значение v -й координаты информационного вектора $\tilde{\alpha}(S_t)$; α_1 равно 1, если в $S_v^{\alpha_1 \alpha_2}$ включаются объекты S_t , для которых выполнено неравенство $R(S_t) > 0$, и $\alpha_1 = 0$ в противном случае).

По построению R_1 (условие 1°): $R_1(I(S'_i)) > 0$. Множество S_v^{11} для него состоит из объектов, входящих в W_1^1 (условие 2° при построении R_1), множество S_v^{00} — из объектов, входящих в W_1^0 (условие 4° построения R_1).

* См. примечание при корректуре на стр. 68.

Поскольку $\{S_1, \dots, S_m\} = W_j^0 \cup W_j^1$, то множества S_v^{01}, S_v^{10} для S_i' пусты. Учитывая, что $\gamma_1 = \dots = \gamma_m = N^h$, получаем: $r_{iv}^1 = m \cdot N^h$.

Оценим r_{iv}^1 при $v = 1, 2, \dots, j - 1, j + 1, \dots, l$. Покажем, что в этом случае $r_{iv}^1 \leq m - 1$. Действительно, по дополнительному условию 2) на начальную информацию $W_j^0 \neq W_j^1$. Поэтому существует хотя бы один элемент S_t из W_j^0 , не входящий в W_j^1 или $S_t \in W_j^1$ и $R(S_t) \leq 0$.

Рассмотрим множества $S_v^{\alpha_1 \alpha_2}$ для S_i' . Имеем $R(S_i') \geq 0$, $R(S_t) > 0$, ($S_t \in W_j^1$), v -я координата информационного вектора $\tilde{\alpha}(S_t)$ равна 0. Поэтому S_t входит для S_i' в множество S_v^{10} .

Следовательно, в выражении для r_{iv} в числителе будет отсутствовать слагаемое $\gamma_t = \gamma(S_t) = N^h$, в знаменателе появится такое слагаемое. Имеем:

$$r_{iv}^1 \leq \frac{m \cdot N^h - N^h}{N^h + 1} < m - 1. \quad (65)$$

Нетрудно показать, что при $u = 1, 2, \dots, i - 1, i + 1, \dots, l$ величины r_{ui}^1 равны 0. Действительно, по построению $R_1: R_1(S_u) < 0$, для S_u множество $S_u^{00} = W_j^0$, $S_u^j = W_j^1$, множества S_u^{01}, S_u^{10} пусты.

II. Проведем опять произвольную кусочно-линейную поверхность R_2 , удовлетворяющую условиям $1^\circ - 4^\circ$. Здесь условия $2^\circ - 4^\circ$ те же, что в I, а условие 1° изменено: $R_2(S_i') < 0$. Опять все $\gamma_i = N^h$.

Аналогично I исследуем матрицу $\{r_{uv}^2\}_{q \times l}$, в которую переводит $(I_0(l), I_s(q))$ оператор, определяемый заданным набором $R_2, \gamma_1, \dots, \gamma_m$.

Заметим, что для объектов $S_1', \dots, S_{i-1}', S_{i+1}', \dots, S_q'$ ничего не изменилось по сравнению с I. Поэтому

$$r_{uv}^2 = r_{uv}^1 \text{ при } u \neq i, \quad v = 1, 2, \dots, l. \quad (66)$$

Нетрудно показать, что

$$r_{ij}^2 = 0. \quad (67)$$

Действительно, относительно R_2 и класса K_j для S_i' множества Q_{01} и Q_{10} пусты.

Аналогично I доказывается неравенство

$$r_{iv}^2 \leq m - 1, \quad v = 1, 2, \dots, j - 1, j + 1, \dots, l. \quad (68)$$

Из (66)–(68) и аналогичных результатов в I вытекает следующее утверждение:

Пусть $R_A^{1,h}, R_A^{2,h}$ — операторы, определяемые соответственно наборами $(R_1, \gamma_1 = \dots = \gamma_m = N^h)$, $(R_2, \gamma_1 = \dots = \gamma_m = N^h)$. Тогда оператор $R_A^h = \frac{1}{m \cdot N^h} (R_A^{1,h} - R_A^{2,h})$ входит в линейное замыкание $L(R_A)$ класса операторов, определяемых кусочно-линейной разделяющей поверхностью R , и набором параметров $\gamma_1 \dots \gamma_m$, и

$$M_{q \times l}^h = R_A^h (I_0(l), I_s(q)) = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ a_{i1} & a_{i2} & \dots & a_{ij-1} & 1 & a_{ij+1} & \dots & a_{il-1} & a_{il} \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Здесь $|a_{iu}| \leq \frac{2(m-1)}{m \cdot N^k}$, $u = 1, 2, \dots, j-1, j+1, \dots, l$.

Очевидно, последовательность $M_{q \times l}^h$ равномерно сходится к матрице $\{b_{uv}\}_{q \times l}^{i,j}$, у которой $b_{ij} = 1$ и остальные элементы равны 0.

Совокупность числовых матриц $M_{q \times l}$, порождаемых операторами, входящими в линейное замыкание $L(R_A)$, образует линейное векторное пространство конечной размерности. Оно содержит свои предельные точки с ограниченной нормой. Следовательно, $\{b_{uv}\}_{q \times l}^{i,j} \in M_{q \times l}$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$. Теорема доказана.

Следствие. Класс распознающих алгоритмов (R_A, r_A) , определяемый набором разделяющих кусочно-линейных поверхностей R_1, \dots, R_l (элементы каждого класса разделяются своей поверхностью) и набором параметров $\gamma_1, \dots, \gamma_m$ (модель $\mathfrak{M}_k^L(\vec{\gamma}, l)$), является полным.

Доказательство следует из того, что полон подкласс

$$R_1 = \dots = R_l = R.$$

§ 3. Экстремальные алгоритмы в линейных замыканиях моделей. Слабая полнота моделей *)

Определения корректной и полной модели, введенные в предыдущем параграфе, позволяют сформулировать два условия, при которых линейное замыкание модели содержит алгоритм, абсолютно точный на каждой контрольной выборке фиксированной длины.

При этом накладываются слабые условия на класс решающих правил и довольно сильные — на класс распознающих операторов модели \mathfrak{M} . Введем другое условие корректности модели, в котором накладываются более жесткие ограничения на класс решающих правил и более слабые — на класс распознающих операторов модели \mathfrak{M} .

Определение 30. Модель \mathfrak{M} (набор распознающих операторов $\mathfrak{M}(R_A)$) называется полной (полным) в слабом смысле, если для любых $I_0, q, S'_1, \dots, S'_q$ множество матриц $M(R_A) = \{R_A(I_0(l), I_S(q))\}$ размерности $q \times l$ содержит в себе подмножество матриц $\{M_{q \times l}^{i,j}\}$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, таких, что:

- 1° максимальный по модулю элемент в $M_{q \times l}^{i,j} = \{c_{ij}\}_{q \times l}$ есть c_{ij} ;
- 2° $c_{ij} > c_{ut}$ при $(u, t) \neq (i, j)$.

В дальнейшем будем считать, что элементы S_1, \dots, S_m , описания которых составляют I_0 , удовлетворяют условию:

$$\emptyset \neq W_j = K_j \cap \{S_1 \dots S_m\} \neq W_u = K_u \cap \{S_1 \dots S_m\}, \\ j = 1, 2, \dots, l, u \neq j.$$

Пусть в обучающую информацию I_0 входят описания $I(S_u) = (a_{u1}, \dots, a_{un})$, в контрольное множество — описания $I(S'_i) = (b_{i1}, \dots, b_{in})$, $u = 1, 2, \dots, m$, $i = 1, 2, \dots, q$.

Определение 31. Множество $I = \{I(S_1), \dots, I(S_m), I(S'_1), \dots, I(S'_q)\}$ называется правильным, если для каждой пары (S'_v, S'_w) и каждого K_j найдется $S_u \in K_j$ и признак $r(u = u(v, w), r = r(u, w))$ такой, что $\rho_r(a_{ur}, b_{vr}) < \rho_r(a_{ur}, b_{wr})$, $1 \leq v, w \leq n$, $v \neq w$, $1 \leq r \leq n$. Стандартная информация, содержащая правильное множество I , называется правильной.

Рассмотрим модель $\mathfrak{M}(\vec{\varepsilon}, \vec{p}, \vec{\gamma})$ вычисления оценок, определяемую следующим образом.

*) См. примечание при корректуре на стр. 68.

1. Система опорных подмножеств состоит из подмножеств длины 1 множества $\{1, 2, \dots, n\}$, т. е. в обозначениях раздела I главы II параметр k равен единице.

2. Функция близости $B_0^{\vec{\varepsilon}}$ является пороговой, в ней $\varepsilon = 0$, она зависит от параметров $\varepsilon_1, \dots, \varepsilon_n$. Части $\tilde{\omega}S$ в определяемом подклассе состоят из одного элемента. Пусть $\tilde{\omega}S_u = a_{ui}$, $\tilde{\omega}S'_t = b_{ti}$; $B_0^{\vec{\varepsilon}}(\tilde{\omega}S_u, \tilde{\omega}S'_t) = 1 = B(a_{ui}, b_{ti})$ в том и только том случае, когда $\rho_i(a_{ui}, b_{ui}) \leq \varepsilon_i$, $1 \leq i \leq n$.

$$3. \Gamma_j(S'_t) = \tilde{Q} \cdot \sum_{S_u \in W_j^1} \gamma(S_u) \cdot \sum_{i=1}^n p_i B(a_{ui}, b_{ti}).$$

Оценки $\Gamma_j(S'_t)$, $t = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, являются функциями параметров $\vec{\varepsilon} = (\varepsilon_1 \dots \varepsilon_n)$, $\vec{p} = (p_1 \dots p_n)$, $\vec{\gamma} = (\gamma_1 \dots \gamma_m)$. Множество операторов, вычисляющих по начальной информации оценки $\Gamma_j(S'_t)$, обозначим через $R_A(\vec{\varepsilon}, \vec{p}, \vec{\gamma})$.

Пусть $\{I_0, I(S'_1), \dots, I(S'_q)\} = \mathfrak{M}(I_0, q)$ — совокупность множеств, для которых $\{I(S_1), \dots, I(S_m), I(S'_1), \dots, I(S'_q)\}$ является правильным.

Теорема 12. *Множество $R_A(\vec{\varepsilon}, \vec{p}, \vec{\gamma})$ является полным в слабом смысле на $\mathfrak{M}(I_0, q)$ ($\tilde{Q}=1$).*

Доказательство. Требуется показать, что в линейном замыкании $L(R_A(\vec{\varepsilon}, \vec{p}, \vec{\gamma}))$ для всех пар (i, j) , $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, найдется оператор $R_A^{i,j}$, переводящий начальную информацию из $\mathfrak{M}(I_0, q)$ в матрицу $\{\Gamma_{ut}\}_{q \times l}$, в которой элемент Γ_{ij} является единственным максимальным элементом. Построение $R_A^{i,j}$ проводится в несколько шагов.

1. Фиксируется объект S'_i и класс K_j . В силу определения 31 для любого S'_w найдется признак r и элемент $S_u \in W_j^1$ такой, что $\rho_r(a_{ur}, b_{ir}) < \rho_r(a_{ur}, b_{wr})$. Вводится оператор R_{ij} . При его определении все p_i , кроме p_r , полагаются равными нулю, $p_r = P > 0$; все параметры γ , кроме γ_u , полагаются равными 0; $\gamma_u = 1$.

Параметр ε_r выбирается так, чтобы выполнялось неравенство

$$\rho_r(a_{ur}, b_{ir}) < \varepsilon_r < \rho(a_{ur}, b_{wr}). \quad (69)$$

Остальные ε_i выбираются произвольно.

Из (69) следует, что при применении R_{ij}^w к $(I_0(l), I_S(q))$ получается матрица $\{\Gamma_{ij}^w\}$, в которой

$$\Gamma_{ij} = P, \quad \Gamma_{wj} = 0, \quad \Gamma_{vs} \leq P, \quad (v, s) \neq (w, j). \quad (70)$$

Оператор $\sum_{w \neq i} R_{ij}^w = R_1(S'_i)$ входит, очевидно, в линейное замыкание $L(R_A(\vec{\varepsilon}, \vec{p}, \vec{\gamma}))$. Из определения $R_1(S'_i)$ и (70) следует, что в матрице

$$R_1(I_0(l), I_S(q)) = \{\Gamma_{uv}^1\}_{q \times l}$$

$$1^\circ \quad \Gamma_{ij}^1 = (q-1) \cdot P;$$

$$2^\circ \quad \Gamma_{uj}^1 \leq (q-2) \cdot P, \quad u \neq i;$$

$$3^\circ \quad \Gamma_{vt}^1 \leq (q-1) \cdot P, \quad v = 1, 2, \dots, q, \quad t = 1, 2, \dots, j-1, j+1, \dots, l.$$

2. Фиксируется класс K_j . Так как $W_j^1 \not\subseteq W_r^1$, в множестве S_1, \dots, S_m могут быть выбраны элементы S_{i_r} , $r = 1, 2, \dots, j-1, j+1, \dots, l$,

принадлежащие K_r , и не принадлежащие соответственно K_r . Строятся операторы R_r^j , в которых параметры определяются следующим образом: $\gamma(S_{i_r}) = \gamma_{i_r} = 1$, остальные $\gamma_i = 0$; $p_1 = \dots = p_n = P$ (величина P здесь та же, что и при построении операторов R_{ij}^w); если $I(S_i) = (a_{i1} \dots a_{in})$, $I(S'_u) = (b_{u1} \dots b_{un})$, то $\varepsilon_t > \rho_t (a_{it}, b_{ut})$.

Оператор R_r^j переводит исходную информацию в числовую матрицу $\{\Gamma_{vw}^r(r)\}$, в которой

$$\begin{aligned} 4^\circ \quad \Gamma_{ir}(r) &= 0; \\ 5^\circ \quad \Gamma_{ij}(r) &= n \cdot P; \\ 6^\circ \quad \text{остальные элементы } \Gamma_{vw}(r) &\leq n \cdot P. \end{aligned}$$

Оператор $R_2^j = \sum_{r \neq j} R_r^j$ входит, очевидно, в $L\{R_A(\vec{e}, \vec{p}, \vec{v})\}$. Из его определения и свойств $4^\circ - 6^\circ$ следует, что в матрице $R_2^j(I_0(l), I_S(q)) = \{\Gamma_{uv}^2\}_{q \times l}$.

$$\begin{aligned} 7^\circ \quad \Gamma_{ij}^2 &= (l-1) \cdot n \cdot P; \\ 8^\circ \quad \Gamma_{wj}^2 &\leq \Gamma_{ij}^2, \quad w \neq i; \\ 9^\circ \quad \Gamma_{ww}^2 &\leq (l-2) \cdot n \cdot P < (l-1) \cdot n \cdot P = \Gamma_{ij}^2. \end{aligned}$$

3. Оператор $R_{ij} = R_1(S'_i) + R_2^j$ переводит исходную информацию в матрицу $\{\Gamma_{vw}\}_{q \times l}$, причем

$$\begin{aligned} 10^\circ \quad \Gamma_{ij} &= P((l-1) \cdot n + q - 1)); \\ 11^\circ \quad \Gamma_{ij} - \Gamma_{rt} &\geq P, \quad (rt) \neq (i, j). \end{aligned}$$

Положим $R_A^{i,j} = R_{ij}$. Теорема доказана.

Следствие. В $L\{R_A(\vec{e}, \vec{p}, \vec{v})\}$ для каждой пары (i, j) , $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, и каждой правильной стандартной информации существует оператор $R_N(S'_i, K_j)$, N — произвольное положительное число, переводящий правильную стандартную информацию в числовую матрицу $\{\Gamma_{uv}\}_{q \times l}$, причем:

$$\Gamma_{ij} - \Gamma_{uv} \geq N, \quad (uv) \neq (ij).$$

Доказательство следует из 11° в доказательстве теоремы, если положить $P = N$.

Вообще говоря, модель $\mathfrak{M}(R_A)$ операторов, полная в слабом смысле, не является полной. Но, добавив в нее еще один оператор, ее можно сделать полной. Рассмотрим оператор R в пространстве числовых матриц:

$$R(\{a_{ij}\}_{q \times l}) = \{b_{ij}\}_{q \times l}, \quad \text{где } b_{uv} = 1,$$

если $a_{uv} = \max a_{ij}$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, в остальных случаях $b_{uv} = 0$.

Определение 32. Модель $\mathfrak{M}(R_A \cdot R)$ называется элементарным расширением модели $\mathfrak{M}\{R_A\}$ распознающих операторов.

Теорема 13. Элементарное расширение $\mathfrak{M}\{R_A \cdot R\}$ модели $\mathfrak{M}\{R_A\}$, полной в слабом смысле, является полной моделью.

Доказательство. В модели $\mathfrak{M}\{R_A\}$, полной в слабом смысле, по $I_0, I(S'_1, \dots, S'_q)$ при помощи операторов из $\mathfrak{M}\{R_A\}$ могут быть построены $q \cdot l$ матриц $\{\Gamma_{vw}^u\}_{q \times l}$, $u = 1, 2, \dots, q \cdot l$ таких, что единственным максимальным элементом $\{\Gamma_{vw}^u\}$ является элемент Γ_{ij} . Применяя к этим матрицам оператор R , получим матрицы, образующие базис в пространстве числовых матриц $\{a_{uv}\}_{q \times l}$. Теорема доказана.

Пусть $\mathfrak{M}\{A\}$ — модель р.с.а., $\mathfrak{M} = \{R_A r_A\}$ с корректными решающими правилами r_A , причем $\mathfrak{M}\{R_A\}$ полно в слабом смысле. Пусть также R — оператор, выделяющий максимум.

Из предыдущего легко доказывается

Теорема 14. Модель $\tilde{\mathfrak{M}}(A) = \{R_A \cdot R \cdot r_A\}$ является полной и, следовательно, корректной.

Модель $\tilde{\mathfrak{M}}(A)$ называется элементарным расширением модели $\mathfrak{M}(A)$.

Во многих случаях доказательство полноты модели технически существенно более сложно, чем доказательство полноты в слабом смысле. Вместе с тем элементарное пополнение множества распознающих операторов практически не усложняет модель. Новая модель является полной на том же множестве $\{I_0\}$ и, следовательно, корректной.

Для достаточно широких классов решающих правил модели, в которых совокупность распознающих операторов является слабо полной, обладают свойством корректности.

Пусть множество $\mathfrak{M}\{R_A\}$ на $\{I_0\}$ обладает свойством слабой полноты и, кроме того, дополнительным свойством. Для всяких положительных Q и N , $N < Q$, существуют операторы $R_{ij}^{Q,N}$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, такие, что $R_{ij}^{Q,N}(I_0) = \{\Gamma_{uv}\}_{q \times l}$, $\max \Gamma_{uv} = \Gamma_{ij} = Q$, $\Gamma_{ij} - \Gamma_{uv} \geq N$, $(uv) \neq (i,j)$.

Слабо полные модели с дополнительным свойством будем называть регулярными моделями.

Теорема 15. Модель $\tilde{\mathfrak{M}} = \{R(\vec{e}, \vec{p}, \vec{v})\}$ является регулярной.

Доказательство. 1) Из доказательства теоремы 12 и следствия из нее видно, что в $L(\mathfrak{M})$ существует оператор R_{ij} (см. 10° в теореме 12), который является оператором $R_{ij}^{Q,N'}$ с произвольным Q' и $N' = \frac{Q'}{(l-1) \cdot n + q - 1}$.

Тогда нетрудно указать y такое, что $y \cdot R_{ij}^{Q,N'} = R_{ij}^{Q' \cdot y \cdot N' \cdot y}$, причем $N' \cdot y = N$, N — произвольное положительное число, меньшее $Q' \cdot y$.

2) В модели $L\{R(\vec{e}, \vec{p}, \vec{v})\}$ для любой константы c существует оператор R_c , переводящий I_0 в матрицу $\{\Gamma_{ij}\}_{q \times l}$, где все $\Gamma_{ij} = c$.

Очевидно, достаточно доказать утверждение для $c = 1$. Если $I(S_u) = (a_{u1} \dots a_{un})$, $u = 1, 2, \dots, m$; $I(S'_t) = (b_{t1} \dots b_{tn})$, то $e_i > \max p_i (a_{ui}, b_{ti})$, $i = 1, 2, \dots, n$, $t = 1, 2, \dots, q$.

При выбранных таким образом параметрах e_1, \dots, e_n , $\tilde{Q} = \frac{1}{l \mu(W^j)}$.

$$\Gamma_j(S'_t) = \frac{1}{l} \cdot \frac{1}{\mu(W^j)} \sum_{S_i \in W^j} \gamma_i \cdot \sum_{i=1}^n p_i.$$

Положим $\frac{1}{l} \cdot \frac{1}{\mu(W^j)} = N$. Тогда при $p_i = \frac{1}{n}$, $i = 1, 2, \dots, n$, $\gamma_i = \frac{1}{N \cdot |W^j|}$ имеем: $\Gamma_j(S'_t) = 1$ для $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$.

3) Очевидно, $R_{ij}^{Q' \cdot y \cdot N'} + R_{Q-Q' \cdot y} = R_{ij}^{Q,N'}$.

Теорема доказана.

Рассмотрим сначала простейшие функциональные решающие правила порядка 1 с константами $0 < c_{1j} < c_{2j}$, $j = 1, 2, \dots, l$: $S'_i \in K_j$, если $a_{ij} > c_{2j}$; $S'_i \notin K_j$, если $a_{ij} < c_{1j}$, при $c_{1j} \leq a_{ij} \leq c_{2j}$ решающее правило вырабатывает $\alpha_{ij}^A = \Delta$.

Теорема 16. Модель $\tilde{\mathfrak{M}}\{R_A \cdot r_A\}$ с регулярным множеством $\mathfrak{M}\{R_A\}$ и решающими правилами порядка 1 с константами c_{1j} , c_{2j} , $j = 1, 2, \dots, l$, является корректной.

Доказательство. Рассмотрим произвольную конечную выборку допустимых объектов S'_1, \dots, S'_q и их истинные информационные векторы $\tilde{\alpha}(S'_1) = \tilde{\alpha}_1 = (\alpha_{11} \dots \alpha_{1l}), \dots, \tilde{\alpha}(S'_q) = \tilde{\alpha}_q = (\alpha_{q1} \dots \alpha_{ql})$. Разобьем номера (i, j) , $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, на подмножества $M_0 : (i, j) \in M_0$, если $\alpha_{ij} = 0$, и $M_1 : (i, j) \in M_1$, если $\alpha_{ij} = 1$.

Пусть $\max_j c_{2j} = b_2$, $\min_j c_{1j} = b_1$, $b_{12} = b_2 - b_1$. Для каждой пары (i, j) из M_1 выберем в $L\{\mathfrak{M}(R_A)\}$ оператор $R_{ij}^{Q, N}$ (см. определение регулярной модели), причем

$$Q = b_2 + \varepsilon > b_2 \geq c_{2j}, \quad Q > N > Q - \frac{b_2 - b_{12}}{\mu(M_1)}, \quad (71)$$

где $\mu(M_1)$ — мощность множества M_1 .

Применением оператора $\sum_{(i, j) \in M_1} R_{ij}^{Q, N}$ к I_0 получаем матрицу $\{\Gamma_{ij}\}_{q \times l}$, в которой

1° если $(i, j) \in M_1$, то $\Gamma_{ij} \geq b_2 + \varepsilon > c_{2j}$,

2° если $(i, j) \in M_0$, то $\Gamma_{ij} \leq c_{1j}$.

Первое утверждение тривиально следует из (71). Докажем 2°. В $R_{ij}^{Q, N}(I_0) = \{\Gamma_{ij}^{Q, N}\}$ величина $\Gamma_{uv}^{QN} \leq Q - N < \frac{b_2 - b_{12}}{\mu(M_1)}$. Тогда $|\Gamma_{uv}| \leq \mu(M_1)(Q - N) < b_2 - b_{12} = \min_j c_{1j}$.

Применяя к $\{\Gamma_{ij}\}_{q \times l}$ решающее правило, получаем истинные информационные векторы $\alpha(S_i)$, $i = 1, 2, \dots, q$.

Указанное выше построение возможно, так как $\min_i c_{1j} > 0$ и поэтому $N < Q$. Теорема доказана.

Следствие 1. Если модель $\mathfrak{M} = \{R_A \cdot r_A\}$ с регулярным $\mathfrak{M}(R_A)$ для каждой I_0 содержит $R_A^{I_0}$ такой, что $R_A^{I_0}(I_0) = \{\Gamma_{ij}\}_{q \times l}$ и $\Gamma_{ij} = \text{const}$, то модель $\mathfrak{M}(R_A \cdot r_A)$ является корректной при решающих правилах с произвольными c_{1j} , c_{2j} , $c_{1j} < c_{2j}$.

Доказательство. Достаточно рассмотреть случай, когда $\max_j c_{2j} \geq 0$, $\min_j c_{1j} \leq 0$. Рассмотрим сначала новые решающие правила, где $\tilde{c}_{2j} = c_{2j} + \max_j |c_{1j}| + \delta$, $\tilde{c}_{1j} = c_{1j} + \max_j |c_{1j}| + \delta''$, $\delta'' > 0$. Для этих новых решающих правил верна теорема 16. Рассмотрев затем оператор $\sum_{(i, j) \in M_1} R_{ij}^{Q, N} - R \max_j |c_{1j}| + \delta''$, докажем следствие 1.

Следствие 2. Модель $\mathfrak{M}(\vec{\varepsilon}, \vec{p}, \vec{\gamma})$, где $\mathfrak{M}(R_A) = \{R(\vec{\varepsilon}, \vec{p}, \vec{\gamma})\}$ и решающие правила порядка 1 с произвольными константами c_{1j} , c_{2j} , $c_{1j} < c_{2j}$, $j = 1, 2, \dots, l$, является корректной.

Действительно, при доказательстве теоремы мы построили оператор $R(\vec{\varepsilon}, \vec{p}, \vec{\gamma})$, который для произвольной правильной I_0 строит матрицу $\{\Gamma_{ij}\}_{q \times l}$, $\Gamma_{ij} = 1$.

Если рассмотреть произвольное решающее правило порядка 1 с функциями $f_1(x), \dots, f_l(x)$, монотонно возрастающими, определенными на всей числовой оси, неограниченными сверху и снизу, то, очевидно, теорема остается справедливой, как и следствия из нее.

Для пороговых решающих правил порядка f с константами c_{1j} , c_{2j} теорема о корректности регулярных моделей остается справедливой при дополнительных ограничениях на базу решающего правила. Пусть $f_l^j(x_1, \dots, x_l)$ удовлетворяет условиям:

1° $f_l^j(x_1, \dots, x_l) = \psi_l^j(x_j - x_1, \dots, x_j - x_{j-1}, x_j, x_{j+1} - x_j, \dots, x_l - x_l)$;

2° ψ_l^j монотонно возрастает по каждому из аргументов $x_j - x_1, \dots, x_j - x_{j-1}, \dots, x_j, \dots, x_l - x_l$;

$$3^\circ \lim_{x_j \rightarrow \infty} \psi_l^j(c_1, \dots, c_{j-1}, x_j, c_{j+1}, \dots, c_l) = \infty;$$

$$4^\circ \lim_{x_j \rightarrow -\infty} \psi_l^j(c_1, \dots, c_{j-1}, x_j, c_{j+1}, \dots, c_l) = -\infty.$$

Свойства 3° и 4° выполняются для любых постоянных $c_1, \dots, c_{j-1}, c_{j+1}, \dots, c_l$.

Решающее правило, у которого все функции базы удовлетворяют условиям 1°—4°, называется монотонным.

Будем говорить, что $L(\mathfrak{M}(R_A))$ содержит константы, если для любой начальной информации $(I_0(l), I_S(q))$ в $L(\mathfrak{M}(R_A))$ существует R_A такой, что $R_A(I_0(l), I_S(q)) = \{a_{ij}\}_{q \times l}$, $a_{ij} \equiv 1$.

Теорема 17. Регулярная модель, линейное замыкание которой содержит константы и решающие правила суть монотонные функциональные пороговые правила порядка l , является корректной.

Доказательство полностью аналогично доказательству теоремы т16. Поэтому приведем только схему, не объясняя подробно отдельные этапы.

Пусть $c_1^* = \min_j c_{1j}$, $c_2^* = \max_j c_{2j}$.

Разбиваем пары (i, j) на множества M_0 , M_1 по матрице информационных векторов $\{\tilde{\alpha}(S_i)\}_{q \times l}$.

Используя свойство регулярности и конструкции теоремы, строим по $(I_0(l), I_S(q))$ матрицу $\{a_{ij}\}_{q \times l}$, такую, что ее элементы a_{ij} , $(i, j) \in M_1$ столь велики, что $\psi_l^j(a_{ij} - a_{i1}, \dots, a_{ij}, \dots, a_{ij} - a_{il}) \gg c_2^*$, и значения a_{ij} , $(i, j) \in M_0$ существенно больше значений a_{ij} при $(i, j) \in M_0$.

Разность между a_{ij} , $(i, j) \in M_1$ и a_{uv} , $(u, v) \in M_0$ выбирается так, чтобы был возможен сдвиг элементов матрицы $\{a_{ij}\}$ (вычитание константы N из элементов матрицы, при котором

$$\begin{aligned} \psi_l^j(a_{ij} - a_{i1}, \dots, a_{ij} - N, \dots, a_{ij} - a_{il}) &> c_2^*, \quad (i, j) \in M_1, \\ \psi_l^j(a_{ij} - a_{i1}, \dots, a_{ij} - N, \dots, a_{ij} - a_{il}) &< c_1^*, \quad (i, j) \in M_0. \end{aligned} \quad (72)$$

Любой сдвиг на N может быть произведен, так как модель содержит константы.

Существование N , при котором выполнены условия (72), гарантируется регулярностью модели и свойствами 1°—4° базы решающего правила.

ГЛАВА IV ОПЕРАЦИИ НАД РАСПОЗНАЮЩИМИ АЛГОРИТМАМИ

Подобно тому как было введено в главе III сложение распознающих операторов и умножение их на скаляр, можно рассматривать операции над р.а. как операции над информационными матрицами

$$\{\alpha_{ij}^A\}_{q \times l} = A(I_0(K_1, \dots, K_l), I(S'_1, \dots, S'_q)).$$

Однако, если множество распознающих операторов образует линейное векторное пространство, т. е. алгебру «с хорошими» свойствами, то в множествах информационных матриц таких хороших операций нет. Операции с более бедными свойствами могут быть введены и изучены. Применяя их к простым моделям р. а., можно получить и изучить более сложные модели.

§ 1. Поэлементные операции над информационными матрицами

Пусть сложение $A + B$ распознающих алгоритмов определено как поэлементное сложение соответствующих информационных матриц, т. е., если $A(I_0) = \{\alpha_{ij}^A\}$, $B(I_0) = \{\alpha_{ij}^B\}$, то $(A + B)(I_0) = \{\alpha_{ij}^A + \alpha_{ij}^B\}$. Так как элементы информационных матриц принимают лишь значения $\{0, 1, \Delta\}$, то сложение алгоритмов в этом случае определяется заданием операции

Таблица 4

посредственno. Другие операции (их всего 2⁴) не обладают свойством ассоциативности. Пусть, например $0 + 1 = \Delta$. Положим $0 + \Delta = a_{0\Delta}$, $1 + \Delta = a_\Delta$. Тогда

1. $a_{0\Delta} = 0$, $a_{1\Delta} = 1$, тогда $(0 + 0) + 1 = \Delta \neq 0 + (0 + 1) = 0$,
 2. $a_{0\Delta} = 0$, $a_{1\Delta} = \Delta$, тогда $(0 + 0) + 1 \neq 0 + (0 + 1)$,
 3. $a_{0\Delta} = \Delta$, $a_{1\Delta} = \Delta$ — операция 1, отсутствует 0 операции,
 4. $a_{0\Delta} = \Delta$, $a_{1\Delta} = 1$, тогда $(1 + 1) + 0 \neq 1 + (1 + 0)$.

Остальные случаи разбираются аналогично. Теорема доказана.

Следствие. Из семи полугрупповых операций таблицы ни одна не является групповой. Две из них имеют нулевой элемент операции.

Все введенные в таблице операции не интересны.

Пусть алгоритмы A_1, A_2, \dots, A_n при вычислении свойства $S'_i \in K_i$, дали соответственно ответы $\alpha'_{ij}, \alpha''_{ij}, \dots, \alpha'''_{ij}$. Тогда $\alpha'_{ij} \oplus \dots \oplus \alpha'''_{ij} = \Delta$, если среди α'_{ij} есть хотя бы одно Δ или $\alpha''_{ij} \neq 1$ или $\alpha'''_{ij} \neq 0$ — при 1-й операции, $q = 1, 2, \dots, n$. Остальные операции сводятся к заданию порядка в $\{0, 1, \Delta\}$.

- 2-я: $\Delta > 1 > 0$;
 3-я: $1 > \Delta > 0$;
 4-я: $1 > 0 > \Delta$;
 5-я: $\Delta > 0 > 1$;
 6-я: $0 > \Delta > 1$;
 7-я: $0 > 1 > \Delta$.

Применение этих операций к $\alpha_{ij}^1 \oplus \dots \oplus \alpha_{ij}^n$ дает результат, равный максимальному по соответствующему порядку элементу из множества $\{\alpha_{ij}^1, \dots, \alpha_{ij}^n\}$. Содержательно это означает, например, при операции 4, что один положительный ответ $S'_i \in K_j$ уничтожает любое число отрицательных ответов $S_i \in K_j$ и ответов «не знаю».

§ 2. Многоместные операции над информационными матрицами

Результаты § 1 показывают, что операции над информационными матрицами, и, следовательно, над алгоритмами распознавания должны существенно зависеть от числа алгоритмов, участвующих в операции и, может быть, от порядка применения алгоритмов к начальной информации. Такие операции применялись для формирования новых моделей. Приведем два примера.

I. Рассматриваются R -алгоритмы A_1, \dots, A_n с линейными разделяющими поверхностями R_1, \dots, R_n . Решается задача классификации с двумя пересекающимися классами K_1, K_2 . Относительно объекта S_i каждый из алгоритмов A_u дает ответ: $S_i \in K_1 (R_u(S_i) \geq 0, \alpha_{ii}^u = 1)$ или $S_i \notin K_1 (R_u(S_i) < 0, \alpha_{ii}^u = 0)$. К полученному вектору $(\alpha_{i1}^1 \dots \alpha_{in}^n)$ применяется булева функция $l(x_1, \dots, x_n)$:

$$f(\alpha_{i1}^1, \dots, \alpha_{in}^n) = \begin{cases} 1, & \text{если число единиц среди } (\alpha_{i1}^1 \dots, \alpha_{in}^n) \\ & \text{больше или равно } \left[\frac{n}{2} \right] + 1, \\ 0 & \text{во всех остальных случаях.} \end{cases}$$

Таким образом описывается метод комитетов [61].

II. Описания объектов задаются наборами бинарных признаков. Известно, что объекты S_1, \dots, S_m принадлежат K_1 , объекты S_{m+1}, \dots, S_t — к классу K_2 . Классы не пересекаются.

Рассматриваются не всюду определенные функции:

$F_1(I(S_i)) = 1, i = 1, 2, \dots, m; F_1(I(S_i)) = 0, i = m+1, \dots, t$, на остальных наборах F_1 не определена;

$F_1(I(S_i)) = 0, i = 1, 2, \dots, m; F_2(I(S_i)) = 1, i = m+1, \dots, t$, на остальных наборах F_2 не определена.

Рассматриваются сокращенные или туниковые д.н.ф. Φ_1, Φ_2 реализующие функции F_1, F_2 соответственно [5, 6].

Пусть S — распознаваемый объект, $I(S)$ — его описание системой бинарных признаков.

Алгоритм A_1 : $\alpha_1^1(S) = 1 (S \in K_1)$, если $F_1(I(S)) = 1$,
 $\alpha_1^1(S) = 0 (S \notin K_1)$, если $F_1(I(S)) = 0$.

Алгоритм A_2 : $\alpha_1^2(S) = 1 (S \in K_1)$, если $F_2(I(S)) = 0$,
 $\alpha_1^2(S) = 0 (S \notin K_1)$, если $F_2(I(S)) = 0$.

Алгоритм A_3 . Алгоритм вычисления оценок с обучающей информацией $I(S_1), \tilde{\alpha}(S_1), \dots, I(S_t), \tilde{\alpha}(S_t)$. Алгоритм выбирается так, что $\alpha_1^3(S) \in \{0, 1\}$.

Над набором $\alpha_1^1(S), \alpha_1^2(S), \alpha_1^3(S)$ выполняется операция, задаваемая булевой функцией $f(x, y, z)$: если $x = y$, то $f(x, y, z) = x$, если $x \neq y$, то $f(x, y, z) = z$. Алгоритм, сводящийся к выполнению f над $\alpha_1^1(S), \alpha_1^2(S), \alpha_1^3(S)$, описан в [6, 5]. Функция f допускает простую запись: $f = x \cdot (y \vee z) \vee y \cdot z$. Она реализует метод комитетов в применении к алгоритмам A_1, A_2, A_3 .

Рассмотренные примеры дают основание для введения следующих определений.

Пусть задана модель \mathfrak{M} распознающих алгоритмов и всевозможные конечные упорядоченные наборы алгоритмов A_1, \dots, A_n из \mathfrak{M} . Пусть результаты вычисления предиката $P_j(S_i) = (S_i \in K_j)$ алгоритмами A_1, \dots, A_n соответственно равны $\alpha_{ij}^1, \dots, \alpha_{ij}^n, \alpha_{ij}^t \in \{0, 1, \Delta\}$.

Определение 33. В модели \mathfrak{M} задана операция $F = (f_1(x), \dots, f_n(x_1, \dots, x_n), \dots)$, если определена последовательность f_1, \dots, f_n, \dots функций 3-значной логики.

Аргументы функций f_i принимают значения из множества $\{0, 1, \Delta\}$, область определения f_i есть также $\{0, 1, \Delta\}$.

Будем пользоваться символической записью $f_n(A_1, \dots, A_n)$ для нового алгоритма (A_1, \dots, A_n, f_n) :

$$(A_1, \dots, A_n, f_n)(\alpha_{ij}^1, \dots, \alpha_{ij}^n) = f_n(\alpha_{ij}^1, \dots, \alpha_{ij}^n).$$

Совокупность алгоритмов (A_1, \dots, A_n, f_n) , $n = 1, 2, \dots, r, \dots$ (здесь рассматриваются всевозможные упорядоченные наборы алгоритмов A_1, \dots, A_n из модели \mathfrak{M}), будем обозначать через $F(\mathfrak{M})$.

Определение 34. Множество $F(\mathfrak{M})$ называется функциональным замыканием \mathfrak{M} по последовательности F .

Так как существует 3^{3^n} различных функций 3-значной логики от переменных x_1, \dots, x_n , то функция f_n в F может выбираться 3^{3^n} различными способами. Поэтому из модели \mathfrak{M} при помощи различных F может быть построено весьма большое число различных функциональных замыканий $F(\mathfrak{M})$.

Следует отметить, что функции трехзначной логики детально изучены в работе С. В. Яблонского [74]. Основываясь на этой работе, а также на более поздних исследованиях [1, 2, 59], можно предложить аппарат для записи формул, реализующих функции трехзначной логики как суперпозиции элементарных функций. Удобной системой таких элементарных функций является:

$$f_1(x_1) = 0, \quad f_2(x) = 1, \quad f_3(x) = \Delta, \quad j_i(x) = \begin{cases} \Delta & \text{при } x = i, \\ 0 & \text{при } x \neq i, \end{cases}$$

$i = 0, 1, \Delta, \min(x, y), \max(x, y)$. Последние две функции определены относительно порядка $0 < 1 < \Delta$ [74].

Представление функции трехзначной логики суперпозицией указанных элементарных функций аналогично представлению булевой функции в виде д.п.ф. Действительно,

$$\begin{aligned} f(x_1, \dots, x_{n-1}, x_n) = \max \{ & \min j_0(x_n), f(x_1, \dots, x_{n-1}, 0), \\ & \min [j_1(x_n), f(x_1, \dots, x_{n-1}, 1)], \min [j_\Delta(x_n), f(x_1, \dots, x_{n-1}, \Delta)] \}. \end{aligned}$$

Таким образом, для функций из F могут быть построены формулы и вычисление (A_1, \dots, A_n, f_n) проводится эффективно.

Из содержательных соображений в F не могут включаться произвольные f_n .

Рассмотрим классы функций трехзначной логики, из которых могут выбираться функции $f_n(x_1, \dots, x_n)$ в F .

Первая серия условий:

$$1^\circ f_n(a, a, \dots, a, a) = a, \quad a \in \{0, 1, \Delta\};$$

2° $f_n(a_1, \dots, a_n) \in \{a_1, \dots, a_n\}$, если $\{a_1, \dots, a_n\}$ есть $\{0, \Delta\}$ или $\{1, \Delta\}$.

Действительно, если ни один из алгоритмов A_1, \dots, A_n не произвел вычисления ($S \in K_j = 1$), то не имеет смысла, объединяя результаты этих вычислений, считать, что $(S \in K_j) = 1$. Аналогично для ($S \in K_j = 0$).

Множество f , удовлетворяющих условиям 1°—2°, образует замкнутый класс σ_1 функций, сохраняющих множества $\{0\}$, $\{1\}$, $\{\Delta\}$, $\{0, \Delta\}$, $\{\Delta, 1\}$. Этот класс имеет конечный базис *).

*). То, что класс σ_1 имеет конечный базис, сообщил автору С. В. Яблонский.

Вторая серия условий: Пусть $f_n(\alpha_1, \alpha_2, \dots, \alpha_n) = \alpha$, $\alpha \in \{0, 1, \Delta\}$. Заменим любые из значений $\alpha_1, \alpha_2, \dots, \alpha_n$ на α . Получим новый набор $(\alpha'_1, \dots, \alpha'_n)$:

$$3^{\circ} f_n(\alpha'_1, \dots, \alpha'_n) = \alpha.$$

Система функций f , удовлетворяющих условиям 3° , образует класс σ_2 .

Естественно в дальнейшем рассматривать функции в F из $\sigma_1 \cap \sigma_2$. Третья серия условий: Пусть x, y, z — число 0, 1, Δ в наборе x_1, \dots, x_n .

$$4^{\circ} f_n(x_1, \dots, x_n) = \varphi(x, y, z).$$

Функции, удовлетворяющие 4° , не образуют замкнутого класса.

5^о Пусть $\varphi(x, y, z) = f_n(x_1, \dots, x_n) = \alpha$. Тогда, если в наборе (x'_1, \dots, x'_n) число единиц x' , нулей — y' , символов Δ — z' удовлетворяет условиям:

а) $x' \geqslant x, y' \leqslant y, z' \leqslant z,$

б) $x' \leqslant x, y' \geqslant y, z' \leqslant z,$

в) $x' \leqslant x, y' \leqslant y, z' \leqslant z$, и если $\alpha \in \{0, 1, \Delta\}$ и выполнены соответственно условия а), б), в), то

$$f_n(x'_1, \dots, x'_n) = \varphi(x', y', z') = \alpha.$$

Функции, удовлетворяющие условиям 4° — 5° , естественно называть функциями комитетов. Класс таких функций будем обозначать в дальнейшем через K .

Естественным обобщением моделей $F(\mathfrak{M})$ являются модели (F_1, \dots, F_l) (\mathfrak{M}), где вычисление свойства $S \in K_j$ проводится для каждого K_j по функции F_j , $j = 1, 2, \dots, l$.

§ 3. Другие операции над распознающими алгоритмами

1. Произведение распознающих операторов. Пусть операторы R_A из модели \mathfrak{M} включают в область определения обучающие информации, в которых описания объектов даны наборами числовых признаков.

Пусть $R_A(I_0(l), I_S(q)) = \{a_{ij}\}_{q \times l}$,

$$R_A(I(S_1), \tilde{\alpha}(S_1), \dots, I(S_m), \tilde{\alpha}(S_m); I(S_1, \dots, S_m)) =$$

$$= \{b_{ij}\}_{m \times l} = R_A(I_0(l), I_S(m)).$$

Пусть $I^A(S_i) = (b_{i1}, \dots, b_{il})$, $I^A(S_u) = (a_{u1}, \dots, a_{ul})$, $i = 1, 2, \dots, m$, $u = 1, 2, \dots, q$. Тогда

$$R_{A'}(I^A(S_1), \tilde{\alpha}(S_1), \dots, I^A(S_m), \tilde{\alpha}(S_m), \\ I^A(S'_1 \dots S'_q)) = \{a'_{ij}\}_{q \times l} = R_A \cdot R_{A'}(I_0(l), I_S(q)).$$

Последовательное применение $R_A, R_{A'}$ ($I_0(l), I_S(q)$) называется произведением операторов $R_A, R_{A'}$ (обозначение $R_A \cdot R_{A'}$).

Рассматривая над распознающими операторами операции сложения, умножения и умножения на скаляр, легко ввести многочлены от распознающих операторов.

2. Пополнение признакового пространства. Вычисленные распознающими операторами для S'_i оценки (a_{i1}, \dots, a_{il}) и аналогичные оценки для S_1, \dots, S_m могут быть введены как значения новых признаков в описания $I(S_i), I(S_u)$, $i = 1, 2, \dots, q$, $u = 1, 2, \dots, m$. Новый признак интерпретируется следующим образом: алгоритмом A (оператором R_A)

установлено, что оценка $S'_i (S_u)$ за класс K_j есть a_{ij} (a_{uj}). Аналогично в качестве новых признаков могут быть взяты: алгоритм A , $A \in \mathfrak{M}$, установил, что значение свойства $S'_i \in K_j$ равно $a_{ij}^A \in \{0, 1, \Delta\}$.

С помощью операций настоящего параграфа можно образовывать новые модели из ранее построенных.

ГЛАВА V

МЕТОДЫ ПОСТРОЕНИЯ РАСПОЗНАЮЩИХ СТАНДАРТНЫХ АЛГОРИТМОВ, ОПТИМАЛЬНЫХ ПО ФУНКЦИОНАЛУ КАЧЕСТВА

§ 1. Оптимизация в рамках одной параметрической модели

Рассмотрим модель $\mathfrak{M}\{A\} = \{R_A r_A\}$ распознающих алгоритмов, представленных в виде $(R_A r_A); R_A(I_0(l), I(S'_1, \dots, S'_q)) = \{a_{ij}^A\}_{q \times l}, r_A(a_{ij})_{q \times l} = \{\alpha_{ij}^A\}$, a_{ij}^A — числа, $\alpha_{ij}^A \in \{0, 1, \Delta\}$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$.

Будем считать, что матрица $\{\alpha_{ij}^A\}$ определяется обучающей информацией, набором описаний объектов, представляемых для распознавания и набором числовых параметров π_1^1, \dots, π_t^1 , так что

$$a_{ij}^A = a_{ij}^A(\pi_1^1, \dots, \pi_t^1).$$

Матрица $\{\alpha_{ij}^A\}$ определяется матрицей $\{a_{ij}^A\}$ и набором числовых параметров π_1^2, \dots, π_v^2 . В этом случае модель \mathfrak{M} называется параметрической.

Мы будем рассматривать функциональные решающие правила с пороговыми константами c_{1j}, c_{2j} .

В качестве базы решающих правил обычно будут выбираться линейные формы L_j , $j = 1, 2, \dots, l$, от всех или некоторых элементов матрицы $\{a_{ij}\}_{q \times l}$.

В дальнейшем рассматриваются только линейные функционалы качества (определение 10 главы I).

Задача выбора по $(I_0(l), I_S(q))$ экстремального в модели \mathfrak{M} алгоритма A состоит в нахождении набора параметров $\hat{\pi}_1^1, \dots, \hat{\pi}_t^1$ распознающего оператора и $\hat{\pi}_1^2, \dots, \hat{\pi}_v^2$ — решающего правила, определяющего алгоритм $\hat{A} \in \mathfrak{M}$ такой, что

$$\begin{aligned} \Phi(\hat{A}) = \Phi(\hat{\pi}_1^1, \dots, \hat{\pi}_t^1, \hat{\pi}_1^2, \dots, \hat{\pi}_v^2) &= \max_{A \in \mathfrak{M}} \Phi(A) = \\ &= \max_{A \in A(\pi_1^1, \dots, \pi_t^1, \pi_1^2, \dots, \pi_v^2)} \Phi(\pi_1^1, \dots, \pi_t^1, \pi_1^2, \dots, \pi_v^2). \end{aligned}$$

Вторая задача состоит в выделении в \mathfrak{M} множества $\hat{\mathfrak{M}}$, которые содержит в себе алгоритмы, экстремальные по произвольному $(I_0(l), I_S(q))$, где $I_0 \in \{I_0\}$ — множество допустимых обучающих информаций, $I(S'_1), \dots, I(S'_q)$ — допустимые описания произвольной конечной совокупности допустимых объектов.

Рассмотрим сначала первую задачу. Пусть дано функциональное решающее правило $f_j(x_{11}, \dots, x_{q \cdot l})$ с порогами c_{1j}, c_{2j} . Тогда значение $f_j(a_{11}, \dots, a_{q \cdot l})$ на элементах матрицы $\{a_{ij}\}$ в зависимости от того, какое неравенство выполнено: $b_j > c_{2j}$, $f_j < c_{1j}$, $c_{1j} \leq f_j \leq c_{2j}$, определяет значение, равное соответственно 1, 0, Δ . По полученному значению однозначно определяется величина поощрения γ_{ij} , $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$. Напомним, что f_j может зависеть лишь от части элемента $a_{11}, \dots, a_{q \cdot l}$ ($f_j = f(a_{ij}), f_j = f_j(a_{11}, \dots, a_{il})$).

Имея способ вычислять функционал качества при любом фиксированном наборе значений параметров, можно искать экстремальный алгоритм методами локального спуска, случайного поиска, направленного перебора или комбинацией этих методов. Большое число таких работ выполнено для алгоритмов вычисления оценок [15—17, 23, 38, 39, 42], а также и для других моделей алгоритмов.

Здесь следует выделить исследования [40, 41, 70], позволяющие найти хорошее начальное приближение для параметров модели.

Другая идея [62, 20, 21] состоит в том, что функционал качества заменяется на более простой, и для него находятся в явной форме или простыми вычислительными процедурами оптимальные значения параметров. Разнообразие методов оптимизации здесь весьма велико.

В данном направлении существует один большой подкласс задач, который может быть описан и решен единообразно [35].

Рассмотрим функционал $\varphi(A)$, равный доле правильных прогнозов (определение 10). При пороговых функциональных решающих правилах условие правильного вычисления $P_j(S'_i)$ легко записывается в виде неравенства. Если $P_j(S'_i) = 0$, то

$$f_j(a_{i1}, \dots, a_{il}) < c_{1j} \quad (73)$$

или

$$f_j(a_{i1}(\pi_1^1, \dots, \pi_t^1), \dots, a_{il}(\pi_1^1, \dots, \pi_t^1)) < c_{1j}.$$

При $P_j(S'_i) = 1$

$$f_j(a_{i1}(\pi_1^1, \dots, \pi_t^1), \dots, a_{il}(\pi_1^1, \dots, \pi_t^1)) < c_{2j} \quad (74)$$

Напишем $q \cdot l$ неравенств вида (73), (74) для каждой из пар (K_j, S'_i) . Получим систему

$$f_j(a_{i1}, \dots, a_{il}) \leq c_{uj}, u \in \{1, 2\}, i = 1, 2, \dots, q, j = 1, 2, \dots, l. \quad (75)$$

Оптимальному значению функционала качества, определяемого таблицей 3, соответствует такой набор параметров π_1^1, \dots, π_t^1 , параметров, определяющих f_j , и порогов c_{1j}, c_{2j} , при котором в системе (75) выполнено максимальное число неравенств. Для нахождения оптимального по функционалу качества алгоритма следует выделить в (75) максимальную совместную подсистему и решить ее.

Найденные таким образом значения параметров определяют оптимальный алгоритм.

Задача выделения максимальной совместной подсистемы в системе неравенств при любом разнообразии функций f_j может быть сведена к стандартной дискретной экстремальной задаче.

Сопоставим неравенствам системы (75) булевые переменные $y_1, \dots, y_{q \cdot l}$. Введем булеву функцию $\chi(y_1, \dots, y_{q \cdot l})$ следующим образом. Пусть $\tilde{\alpha} = (\alpha_1, \dots, \alpha_{q \cdot l})$ — булев вектор, $\alpha_i \in \{0, 1\}$. Выделим в $\tilde{\alpha}$ все единичные координаты. Пусть их номера суть i_1, i_2, \dots, i_k .

Выделим в (75) подсистему неравенств, составленную из неравенств с номерами i_1, i_2, \dots, i_k . Обозначим эту подсистему через $\sigma(i_1, i_2, \dots, i_k)$ или через $\sigma(\tilde{\alpha})$.

Положим

$$\chi(\alpha_1, \dots, \alpha_{q \cdot l}) = \begin{cases} 1, & \text{если подсистема } \sigma(i_1, \dots, i_k) \text{ несовместна,} \\ 0, & \text{если подсистема } \sigma(i_1, \dots, i_k) \text{ совместна.} \end{cases}$$

Напомним определение монотонной булевой функции f . Пусть на множестве бинарных наборов $\tilde{\alpha}$ одинаковой длины задано отношение

частичного порядка: $\tilde{\alpha} = (\alpha_1 \dots \alpha_n) \leqslant \tilde{\beta} = (\beta_1 \dots \beta_n)$, если $\alpha_i \leqslant \beta_i$, $i = 1, 2, \dots, n$. Функция f называется монотонной, если из соотношения $\tilde{\alpha} \leqslant \tilde{\beta}$ следует: $f(\tilde{\alpha}) \leqslant f(\tilde{\beta})$. Верхним нулем (нижней единицей) монотонной f называется точка $\tilde{\gamma}$ такая, что $f(\tilde{\gamma}) = 0$ ($f(\tilde{\gamma}) = 1$), и для всех $\tilde{\delta} \neq \tilde{\gamma}$ таких, что $\tilde{\delta} \geqslant \tilde{\gamma}$: $f(\tilde{\delta}) = 1$ (для $\tilde{\delta} \leqslant \tilde{\gamma}$: $f(\tilde{\delta}) = 0$).

Верхний нуль $\tilde{\alpha}$ называется максимальным (нижняя единица $\tilde{\alpha}$ — минимальной), если $\|\tilde{\alpha}\|$ максимальна (минимальна) по множеству верхних нулей (нижних единиц). Под нормой бинарного вектора $\|\alpha\|$ понимается, как обычно, число единичных координат в $\tilde{\alpha}$.

Совместная подсистема неравенств называется туниковой, если добавление к ней любого одного неравенства делает ее несовместной.

Монотонная функция однозначно определяется заданием множества ее нижних единиц или верхних нулей. Относительно функции χ имеет место

Теорема 19. *Функция $\chi(x_1, \dots, x_{q.i})$ является монотонной булевой функцией. Множество совместных подсистем (75) взаимно однозначно соответствует множеству нулей функции χ . Множеству туниковых подсистем взаимно однозначно соответствует множество верхних нулей χ . Множеству верхних нулей с максимальной нормой взаимно однозначно соответствует множество максимальных совместных подсистем.*

Доказательство. По построению χ множество ее нулей взаимно однозначно соответствует множеству совместных подсистем системы (75).

Монотонность χ следует из двух очевидных утверждений:

а) если к несовместной подсистеме добавить неравенство, то новая подсистема останется несовместной;

б) если из совместной подсистемы удалить неравенство, то новая подсистема останется совместной.

Если подсистеме соответствует не максимальный верхний нуль $\tilde{\alpha}$, то найдется $\tilde{\beta}$ такое, что $\chi(\tilde{\beta}) = 0$, $\tilde{\beta} \geqslant \tilde{\alpha}$; все единичные координаты $\tilde{\alpha}$ остаются единичными и добавляется хотя бы одна новая единичная координата. Номер этой новой координаты обозначим через i . По построению χ система $\sigma(\tilde{\alpha}) \cup i$ является совместной и, следовательно, система $\sigma(\tilde{\alpha})$ не является туниковой. Соответствие максимальных верхних нулей максимальным совместным подсистемам очевидно. Теорема доказана.

Задача нахождения максимальной совместной подсистемы в системе неравенств сведена к задаче поиска максимального верхнего нуля монотонной булевой функции.

Эта последняя задача хорошо известна, и ее решению посвящены различные работы. Некоторые из этих работ выполнены в терминах теории монотонных булевых функций [51, 52], другие — в терминах теории неравенств. Они также могут быть переписаны в терминах теории монотонных булевых функций. Стандартной при поиске максимального верхнего нуля является следующая постановка. Пусть задан элементарный оператор B , который по предъявленной точке $\tilde{\alpha}$ вычисляет $\chi(\tilde{\alpha})$, доопределяет функцию χ по монотонности $\tilde{\alpha}$ и проверяет условие остановки. При остановке в $\tilde{\alpha}$ гарантируется, что $\tilde{\alpha}$ является максимальным верхним нулем функции χ .

Требуется построить алгоритм, который минимальным при данной χ числом применений оператора B указывает некоторый верхний нуль для χ . Реально выполнение B не всегда эффективно, так как требуется при реализации B проверять совместность некоторой подсистемы. В настоящей работе мы не будем останавливаться на этой проблеме.

Задача минимизации числа применений B решена в шенноновской постановке [43].

Обозначим минимальное число применений B для поиска максимального верхнего нуля функции $f(x_1, \dots, x_n)$ через $\mu_f(n)$ и $L(n) = \max \mu_f(n)$ по монотонным функциям от n переменных. Требуется вычислить величину $L(n)$ и построить алгоритм, гарантирующий для любой монотонной $f(x_1, \dots, x_n)$ нахождение максимального верхнего нуля не более, чем за $L(n)$ шагов. Такой алгоритм построен в [43], где и показано, что

$$L(n) = C_n^{\left[\frac{n}{2}\right] + 1}.$$

Область реальных применений оптимального по Шенону алгоритма невелика ($L(n)$ растет сростом n , как $c \cdot \frac{2^n}{\sqrt{n}}$). Реально применяемые алгоритмы либо 1°: ограничиваются построением некоторого количества верхних нулей и выбором среди них элемента с максимальной нормой [51, 52], либо 2°: отправляясь от некоторой начальной точки $\tilde{\alpha}$, в которой вычисляется $\chi(\tilde{\alpha})$, проводят направленный выбор последующих точек для применения B с учетом информации, накопленной о функции χ . Процесс может быть оборван на любом шаге, при этом выдается 0 с максимальной из найденных нулей нормой. Рассмотрим один пример такой процедуры [35].

Предварительно случайно выбирается несколько точек $\tilde{\alpha}_1, \dots, \tilde{\alpha}_t$, к ним применяется B и вычисляются $\chi(\tilde{\alpha}_1), \dots, \chi(\tilde{\alpha}_t)$. Процесс ведется до тех пор, пока среди вычисленных значений не будет по крайней мере одного нуля и одной единицы. Среди $\chi(\tilde{\alpha}_i), i = 1, 2, \dots, t$, выбираем 0 с максимальной нормой N и полагаем $\chi = 0$ во всех точках, норма которых не больше N .

Число таких точек обозначим через Q_0 , а множество их — через M_0 . Подсчитываем число точек $\tilde{\beta}_t$, расположенных в множестве $E_n \setminus M_0$, не меньших, чем найденные единицы функции χ (множество M_1). Число таких точек обозначаем через Q_1 . Полагаем, что $\chi = 0$ на M_0 и $\chi = 1$ на M_1 .

Описанный выше процесс назовем процессом восстановления функции, имеющей хотя бы один верхний 0 с максимальной нормой, общий с χ или, коротко, χ -восстановлением.

В множестве $\chi(\tilde{\alpha}_1), \dots, \chi(\tilde{\alpha}_t)$ фиксируем хотя бы один 0 с максимальной нормой. В случае остановки после вычислений в $\tilde{\alpha}_1, \dots, \tilde{\alpha}_t$ выдается этот нуль.

Выбор следующей точки производится применением процедуры следующего вида.

Пусть

$$p_0(\chi) = \frac{Q_0}{Q_0 + Q_1}, \quad p_1(\chi) = \frac{Q_1}{Q_0 + Q_1},$$

$\tilde{\beta}$ — произвольная точка, в которой значение χ неизвестно в процессе χ -восстановления. Сопоставим $\tilde{\beta}$ величину

$$Q(\tilde{\beta}) = p_0 \Delta_0(\tilde{\beta}) + p_1 \cdot \Delta_1(\tilde{\beta}), \quad (75')$$

где $\Delta_0(\tilde{\beta})$, $\Delta_1(\tilde{\beta})$ — число точек, в которых значение χ неизвестно и в которых оно будет известно при условии, что $\chi(\tilde{\beta}) = 0$, соответственно $\chi(\tilde{\beta}) = 1$ после χ -восстановления, по точке $\tilde{\beta}$ функции χ , построенной ранее. Для следующего применения B выбирается $\tilde{\beta}$, на которой достигается $\max_{\tilde{\beta}} \chi(\tilde{\beta})$.

$\tilde{\beta}$

В реальных вычислительных схемах вид (75') может быть изменен, может вычисляться не точное значение $Q(\tilde{\beta})$, а верхняя или нижняя оценка Q , наконец, может выбираться $\tilde{\beta}$, доставляющая локальный экстремум для $Q(\tilde{\beta})$.

Если $\chi(\tilde{\beta}) = 0$, то фиксируется точка $\tilde{\beta}$, если $\chi(\tilde{\beta}) = 1$, то фиксированным остается 0, выбранный на предыдущем шаге. При остановке после выполнения шага выдается фиксированный 0.

Решение второй задачи — указание множества, в котором заведомо содержатся все алгоритмы, экстремальные на произвольных конечных выборках допустимых объектов, — сводится к следующему.

В модели $\mathfrak{M}(A)$ рассматривается подмодель $\mathfrak{M}'(A)$ и устанавливается полнота или слабая полнота модели $\mathfrak{M}'(R_A)$.

Если $\mathfrak{M}'(R_A)$ — полная модель и r — произвольное корректное решающее правило, то модель $L\{\mathfrak{M}'(R_A) \cdot r\}$ заведомо удовлетворяет условию второй задачи.

При слабой полноте в качестве r может быть взято произвольное функциональное решающее правило с базой из монотонно возрастающих функций. В некоторых случаях еще требуется строгая положительность пороговых констант.

Так, в главе III было показано, что в модели вычисления оценок с параметрами $k, \varepsilon, \vec{\varepsilon}, p, \gamma$ достаточно положить $k = 1, \varepsilon = 0$ и искать экстремум лишь по операторам, представляемым линейными комбинациями операторов из $L\{\mathfrak{M}(\varepsilon, p, \gamma)\}$.

§ 2. Оптимизация в линейных замыканиях

Пусть заданы параметрические модели $\mathfrak{M}_1 = \{(R_{A_1}, r_{A_1})\}, \dots, \mathfrak{M}_k = \{(R_{A_k}, r_{A_k})\}$. Сформируем линейное замыкание $L(\mathfrak{M}_1, \dots, \mathfrak{M}_k)$, т.е. совокупность алгоритмов $L^t = \{c_1 R_A^{1t} + \dots + c_k R_A^{kt}, |r\}$, $r \equiv \bigcup_{i=1}^k r_A^i$, $R_A^{it} \in R_A^i, i = 1, 2, \dots, k$. $L(\mathfrak{M}_1, \dots, \mathfrak{M}_k)$ снова является параметрической моделью: объединены параметры моделей $\mathfrak{M}_1, \dots, \mathfrak{M}_k$ и добавлены параметры c_1, \dots, c_k . В новой модели можно снова решать задачу оптимизации. Но при больших k или в тех случаях, когда некоторые \mathfrak{M}_i сами определяются заданием большого числа параметров, реальная оптимизация в L связана с большими трудностями. Приближенная оптимизация может быть проведена в два этапа:

1°. Выбираются экстремальные алгоритмы A_i^* в $\mathfrak{M}_i, i = 1, 2, \dots, k$ (или алгоритмы — представители в каждой из моделей). Пусть $A_i^* = (R_{A_i}^*, r_{A_i})$.

2°. Рассматриваются алгоритмы $(\sum c_i R_{A_i}^*, r_{A_i}), r_{A_i} \in L$. Из этого множества выбирается алгоритм, экстремальный по параметрам c_1, \dots, c_k и параметрам решающего правила r_{A_i} .

Пусть $R_{A_i}^*(I_0(l), I_S(q)) = \{a_{ut}^i\}, i = 1, 2, \dots, k$. Тогда, если $A = (R_A \cdot r)$,

где $R_A = \sum_{i=1}^k c_i R_{A_i}^*$, то

$$R_A(I_0(l), I_S(q)) = \left\{ \sum_{i=1}^k c_i a_{ut}^i \right\} = \{d_{ut}\}_{q \times l}.$$

Если решающие правила определяются линейными формами $b_1^j x_1 + \dots + b_l^j x_l, c_{1j} = c_{2j} = 0$, то система (75) состоит из неравенств

$$b_1^j \left(\sum_{i=1}^k c_i a_{uj}^i \right) + \dots + b_l^j \left(\sum_{i=1}^k c_i a_{uj}^i \right) > 0 \quad (76)$$

для пар (S'_u, K_j) таких, что $S'_u \in K_j$,

$$b_1^j \left(\sum_{i=1}^k a_{tj}^i \cdot c_i \right) + \dots + b_l^j \left(\sum_{i=1}^k a_{tj}^i \cdot c_i \right) < 0 \quad (77)$$

для пар (S'_t, K_j) , таких, что $S'_t \notin K_j$.

Объединяя (76) и (77), получаем систему $q \cdot l$ (в случае непересекающихся классов — l) неравенств, билинейных по $b_1^j, \dots, b_l^j, c_1, \dots, c_k$ от $k \cdot l$ неизвестных. В этом случае выбор максимальной совместной подсистемы может быть выполнен одним из известных методов (например [60]). Таким образом, если в каждой модели выбраны алгоритмы, — коэффициенты линейной комбинации алгоритмов и коэффициенты линейных решающих правил могут быть найдены решением сравнительно простой экстремальной задачи.

Следует заметить, что линейные замыкания могут охватывать операторы из одной модели, но работающие на различных частях начальной информации.

§ 3. Оптимизация в функциональных замыканиях

Аналогично § 2 пусть $\mathfrak{M} = \mathfrak{M}_1 \cup \dots \cup \mathfrak{M}_n$, т. е. \mathfrak{M} представляет собой систему моделей $\mathfrak{M}_1, \dots, \mathfrak{M}_n$. Пусть в каждой из \mathfrak{M}_i построен оптимальный по контрольной выборке S'_1, \dots, S'_q алгоритм A_i^* , $i = 1, 2, \dots, n$. Задача оптимизации состоит в подборе функции f_n трехзначной логики такой, что $j_n(A_1^*, \dots, A_n^*)$ является алгоритмом, оптимальным по функционалу качества $\phi(A)$.

Мы будем рассматривать линейные функционалы $\phi(A)$. Пусть результаты вычислений свойства $S'_i \in K_j$ алгоритмом A_i^* есть α_{ij}^u , $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, $u = 1, 2, \dots, n$. Пусть в действительности предикат $S'_i \in K_j$ равен α_{ij} . Тогда для правильного вычисления $S'_i \in K_j$ необходимо и достаточно выполнение условия

$$f_n(\alpha_{1j}^1, \dots, \alpha_{1j}^n) = \alpha_{1j}, \quad \alpha_{ij}^u \in \{0, 1, \Delta\}, \quad \alpha_{ij} \in \{0, 1\}. \quad (78)$$

Аналогичные условия записываются для всех (i, j) .

Получается система $q \cdot l$ уравнений для определения f_n . Система, вообще говоря, несовместна.

Пусть при определении функции f_n набору $\alpha_{1j}^1, \dots, \alpha_{1j}^n$ приписано значение $\tilde{\alpha}_{1j} \in \{0, 1, \Delta\}$. В линейном функционале качества паре $(\alpha_{1j}, \tilde{\alpha}_{1j})$ приписывается поощрение $\gamma_{1j} = \gamma(\alpha_{1j}, \tilde{\alpha}_{1j})$. Значение функционала равно $\frac{1}{q \cdot l} \sum_{i,j} \gamma_{ij}$. Задача состоит в выборе определения f_n , доставляющего максимум величине $\frac{1}{q \cdot l} \sum_{i,j} \gamma_{ij}$.

В пространстве наборов $(\alpha_{1j}^1, \dots, \alpha_{1j}^n) = \vec{\alpha}_{1j}$, $i = 1, 2, \dots, q$, $j = 1, 2, \dots, l$, выделим все подмножества $\{M\}$ одинаковых наборов.

Рассмотрим такое произвольное подмножество $M = \{\vec{\alpha}_{i_1 j_1}, \dots, \vec{\alpha}_{i_t j_t}\}$. Если оно состоит из одного элемента $\vec{\alpha}_{ij}$, то по согласованию с (78)

$$f_n(\vec{\alpha}_{ij}) = \alpha_{ij}. \quad (79)$$

Соответствующее уравнение в (78) в этом случае выполнено. Если в подмножестве $M = \{\vec{\alpha}_{i_1 j_1}, \dots, \vec{\alpha}_{i_t j_t}\}$ более одного элемента, то, учитывая, что $\vec{\alpha}_{i_1 j_1} = \dots = \vec{\alpha}_{i_t j_t}$, имеем $f_n(\vec{\alpha}_{i_1 j_1}) = \dots = f_n(\vec{\alpha}_{i_t j_t}) = \alpha_{ij}^t$. Величина α_{ij}^t может принимать одно из трех значений 0, 1, Δ . Пусть $\alpha \in \{0, 1, \Delta\}$. Положим

$$\Phi_M^\alpha(A) = \frac{1}{q \cdot l} \sum_{(i, j) \in \{(i_1, j_1), \dots, (i_t, j_t)\}} \gamma(\alpha, \alpha_{ij}). \quad (80)$$

Выбирается значение α^* , при котором достигается $\max_\alpha \Phi_M^\alpha(A)$.

Полагается $\alpha_{ij}^t = f_n(\vec{\alpha}_{i_1 j_1}) = \dots = f_n(\vec{\alpha}_{i_t j_t}) = \alpha^*$.

Перебирая подмножества M , аналогично определяем f_n на всех наборах $(\alpha_{ij}^1, \dots, \alpha_{ij}^n)$.

Теорема 20. Определенная описанным выше алгоритмом функция f_n дает алгоритм $f_n(A_1^*, \dots, A_n^*)$, оптимальный по линейному функционалу.

Доказательство. В силу линейности: $\Phi(A) = \sum_{M \in \{M\}} \Phi_M^\alpha(A)$.

Достаточно показать, что на выбранной f_n максимальны все $\Phi_M^\alpha(A)$. Для M , составленных более чем из одного элемента, максимальность $\Phi_M^\alpha(A)$ следует из построения f_n на наборах из M . Для одноЭлементных M оптимальность $\Phi_M^\alpha(A)$ следует из условия: $\gamma(\alpha_{ij}, \alpha_{ij}) \geq \gamma(\alpha_{ij}, \tilde{\alpha}_{ij})$ при $\tilde{\alpha}_{ij} \neq \alpha_{ij}$. Последнее неравенство легко вытекает из определения функционала качества, как последовательности функций, причем каждая функция $f(\rho_1, \dots, \rho_m)$ достигает абсолютного максимума на наборе $(0, \dots, 0)$. Так как $\gamma(x, y) = \gamma(\rho(x, y))$, то $\gamma(\rho(\alpha_{ij}, \alpha_{ij})) = \gamma(0) \geq \gamma(\rho(\alpha_{ij}, \tilde{\alpha}_{ij}))$. Теорема доказана.

Заметим, что построение оптимальной f_n требует не более $3 \cdot \left[\frac{q \cdot l}{2} \right]$ вычислений значений функционалов $\Phi_M^\alpha(A)$, $\alpha \in \{0, 1, \Delta\}$.

Доопределение f_n на наборы, не входящие в множество $\{\alpha_{ij}^1, \dots, \alpha_{ij}^n\}$, может быть проведено различными способами. Например, можно выбрать доопределение, при котором f_n реализуется простейшей формулой в выбранном базисе. Проще всего рассмотреть систему функций 0, 1, Δ , $j_0(x)$, $j_1(x)$, $j_\Delta(x)$, $\max(x, y)$, $\min(x, y)$ относительно порядка $0 < 1 < \Delta$.

Для нее легко вводится аналог д. н. ф. в трехзначной логике. Выбирается доопределение, при котором этот аналог д. н. ф. имеет минимальную сложность или неупрощаем в стандартной системе тождественных преобразований (аналог тупиковой д. н. ф.).

Если f_n выбирается в классе функций, удовлетворяющих 1-й серии условий (глава IV, § 2), то процесс построения оптимальной f_n усложняется незначительно. Тогда $f_n(0, \dots, 0) = 0$, $f_n(1, \dots, 1) = 1$, $f_n(\Delta, \dots, \Delta) = \Delta$. На M , составленных из наборов, содержащих только $\{0, \Delta\}$, соответственно только $(1, \Delta)$, независимо от числа наборов в M , проверяются значения 0, Δ , 1, Δ , соответственно.

Если потребовать одновременного выполнения для f_n первой серии условий и условия $f_n(x_1, \dots, x_n) \equiv \phi(x, y, z)$ из третьей серии, то проблема выбора оптимальной f_n также остается несложной. В этом случае в один класс $M(x, y, z)$ объединяются все классы M , у которых в наборах совпадает число единиц и число нулей. На наборах каждого из классов $M(x, y, z)$ функция f_n должна принимать одно значение. Поэтому для каждого $M(x, y, z)$ производится выбор α по значению $\phi_{M(x, y, z)}^\alpha(A)$ из множеств $\{0, \Delta\}$ или $\{1, \Delta\}$ или $\{0, 1, \Delta\}$.

ЛИТЕРАТУРА

1. Абдугалиев У. А., О нормальных формах k -значной логики. Кибернетика, № 1, 1967, 16—20.
 2. Абдугалиев У. А., К вопросу представления функций k -значной логики нормальными формами. I. Сб. «Дискретный анализ», вып. 15, Новосибирск, ИМ СО АН СССР, 1969, 3—24.
 3. Айзенберг Н. Н., Циткин А. И., Вопросы применения простых тестов. Кибернетика, № 1, 1974, 135—141.
 4. Айзerman M. A., Braverman E. M., Rosenoer L. I., Метод потенциальных функций в теории обучения машин. М., «Наука», 1970.
 5. Асланян Л. А., Об одном методе распознавания, основанном на разделении классов дизъюнктивными нормальными формами. Кибернетика, № 5, 1975, 103—110.
 6. Асланян Л. А., Алгоритмы распознавания с логическими отделителями. Сборник работ по математической кибернетике, вып. 1, М., ВЦ АН СССР, 1976, 116—131.
 7. Алиев Э. М., Камилов М. М., Выбор существенных параметров одного технологического процесса методом вычисления оценок. Труды Международного симпозиума по практическим применением методов распознавания образов. М., ВЦ АН СССР, 1973, 219—225.
 8. Багровская З. В., О некоторых свойствах алгоритмов I -голосования. Кибернетика, № 6, 1975, 96—104.
 9. Бак Хынг Кханг, О геометрических задачах, связанных с оценкой точности одного класса распознающих алгоритмов. Журнал выч. математики и матем. физики 14, 6, 1974, 1571—1580.
 10. Вапник В. Н., Червоненкис А. Я., Теория распознавания образов. М., «Наука», 1974.
 11. Губерман Ш. А., Марипов Т. М., Оценка перспективности месторождений гидротермальных руд с применением ЭВМ. Труды МИНХ и ГП им. Губкина, вып. 62, М., Изд-во «Недра», 1966.
 12. Гулямов Д. Х., Журавлев Ю. И., Оценка качества экспертов, формирующих таблицу обучения в задаче распознавания образов. Журнал выч. матем. и матем. физики 12, 4, 1972, 1066—1070.
 13. Гулямов Д. Х., Опыт применения алгоритмов голосования для прогнозирования двойных неорганических соединений. Труды Международного симпозиума по практическим применением методов распознавания образов. М., ВЦ АН СССР, 1973, 231—236.
 14. Гуревич И. Б., Журавлев Ю. И., Минимизация булевых функций и эффективные алгоритмы распознавания. Кибернетика, № 3, 1974, 16—20.
 15. Дадашев Т. М., Об алгоритмах распознавания, экстремальных по порогам осторожности. Журнал выч. математики и матем. физики 13, 1973, 737—749.
 16. Дадашев Т. М., О логических орбитах распознающих систем. Кибернетика, № 6, 1974, 121—133.
 17. Дадашев Т. М., Локально-экстремальные алгоритмы распознавания, основанные на вычислении оценок. Сборник работ по математической кибернетике, вып. 1, М., ВЦ АН СССР, 1976, 132—153.
 18. Дмитриев А. Н., Журавлев Ю. И., Крендель Ф. П., О математических принципах классификации предметов и явлений. Сб. «Дискретный анализ», вып. 7, Новосибирск, ИМ СО АН СССР, 1966, 3—11.
 19. Дмитриев А. Н., Журавлев Ю. И., Крендель Ф. П., Об одном принципе классификации и прогноза геологических объектов и явлений. Известия Сиб. отд. АН СССР, Геология и геофизика, 5, 1968, 50—64.
 20. Докторович А. Б., Задача выбора оптимального алгоритма распознавания в одном классе алгоритмов голосования. Кибернетика, № 4, 1974, 135—140.
- 5 Проблемы кибернетики, вып. 33

- 21. Докторович А. Б., Исследование критерия эффективности алгоритмов вычисления оценок. Сб. работ по матем. кибернетике, вып. 1, ВЦ АН СССР, М., 1976, 154—166.
- 22. Дюкова Е. В., Об одном алгоритме построения тупиковых тестов. Сб. работ по матем. кибернетике, вып. 1, М., ВЦ АН СССР, 1976, 167—185.
- 23. Енич В., Алгоритм оптимизации параметров одного класса распознающих алгоритмов. Труды Международного симпозиума по практическим применениям методов распознавания образов. М., ВЦ АН СССР, 1973, 237—245.
- 24. Журавлев Ю. И., Экстремальные задачи, возникающие при обосновании эвристических процедур. Сб. «Проблемы прикладной математики и механики». М., «Наука», 1971.
- 25. Журавлев Ю. И., Алгоритмы распознавания, основанные на вычислении оценок. Содержательный смысл параметров, задающих алгоритм. Труды Международного симпозиума по практическим применениям методов распознавания образов. М., ВЦ АН СССР, 1973, 205—218.
- 26. Журавлев Ю. И., Экстремальные алгоритмы в математических моделях для задач распознавания и классификации. ДАН СССР 231, 3, 1976, 532—535.
- 27. Журавлев Ю. И., Тулягин Ш. Е., Измерение важности признака. Сб. «Вопросы кибернетики», вып. 38. Ташкент, ИК с ВЦ АН Уз. ССР, 1970, 29—42.
- 28. Журавлев Ю. И., Никифоров В. В., Алгоритмы распознавания, основанные на вычислении оценок. Кибернетика, № 3, 1971, 1—11.
- 29. Журавлев Ю. И., Юнусов Р., Об одном способе уточнения алгоритма таксономии при помощи распознающих методов типа голосования. Журнал выч. матем. и матем. физики 11, 5, 1971, 1344—1347.
- 30. Журавлев Ю. И., Тулягин Ш. Е., Меры важности объектов в сложных системах. Журнал выч. матем. и матем. физики 12, 1, 1972, 170—184.
- 31. Журавлев Ю. И., Камилов М. М., Тулягин Ш. Е., Алгоритмы вычисления оценок и их применение. «ФАН», Ташкент, 1974.
- 32. Журавлев Ю. И., Михалевич М., Алгоритмы распознавания, основанные на вычислении оценок, для задач с пересекающимися классами. Труды ВЦ Польской АН, вып. 145, Варшава, 1974.
- 33. Журавлев Ю. И., Карчевский Е., Михалевич М., Алгоритмы распознавания для задач, задаваемых наборами качественных признаков. Труды ВЦ Польской АН, вып. 146, Варшава, 1974.
- 34. Журавлев Ю. И., Зенкин А. И., Скоробогатов И. Б., Узаков Я. А., Опыт применения методов теории распознавания образов к задаче оценки деловых качеств работника. Вопросы радиоэлектроники. Серия общетехническая, вып. 3, М., 1974, 28—44.
- 35. Журавлев Ю. И., Мирошник С. Н., Швартина С. М., Об одном подходе к оптимизации в классе параметрических алгоритмов распознавания. Журнал выч. матем. и матем. физики 16, 1, 1976, 209—218.
- 36. Журавлев Ю. И., Зенкин А. И., Рязанов В. В., Алгоритм прогнозирования состояний производственных процессов. Вопросы радиоэлектроники, серия АСУ, вып. 1, М., 1976, 32—42.
- 37. Ищук В. И., О поиске оптимальных весовых коэффициентов для одного класса алгоритмов вычисления оценок. Сб. работ по матем. кибернетике, вып. 1, М., ВЦ АН СССР, 1976, 186—194.
- 38. Камилов М. М., О программном распознающем комплексе ПРАСК-1. Сб. «Вопросы кибернетики», вып. 51, Ташкент, ИК с ВЦ АН Уз. ССР, 1972, 63—65.
- 39. Камилов М. М., Об оптимизации и некоторых применениях алгоритмов вычисления оценок. Труды Международного симпозиума по практическим применениям методов распознавания образов. М., ВЦ АН СССР, 1973, 246—258.
- 40. Камилов М. М., Алиев Э. М., Выбор длины голосующих наборов в алгоритмах вычисления оценок. Сб. «Вопросы кибернетики», вып. 44, Ташкент, ИК с ВЦ АН Уз. ССР, 1971, 162—165.
- 41. Камилов М. М., Алиев Э. М., Ким А. Н., О вычислении ε -порогов при распознавании объектов алгоритмами голосования. Сб. «Вопросы кибернетики», вып. 43, Ташкент, ИК с ВЦ АН Уз. ССР, 1971.
- 42. Камилов М. М., Ким А. Н., Распознающий комплекс, основанный на алгоритмах вычисления оценок. Сб. «Вопросы кибернетики», вып. 52, Ташкент, ИК с ВЦ АН Уз. ССР, 1972, 68—74.
- 43. Катериничкина Н. Н., Поиск максимального верхнего нуля монотонной функции алгебры логики. ДАН СССР 224, 3, 1975, 557—560.
- 44. Колыцов Н. П., Задача распознавания при ограничениях во времени. Журнал выч. матем. и матем. физики 14, 5, 1974, 1292—1308.
- 45. Константинов Р. М., Королева З. Е., Применение тестовых алгоритмов к задачам геологического прогнозирования. Труды Международного симпозиума по практическим применениям методов распознавания образов. М., ВЦ АН СССР, 1973, 194—204.

46. Константинов Р. М., Королева З. Е., Кудрявцев В. Б., Комбинаторнологический подход к задачам прогнозарудоносности. Сб. «Проблемы кибернетики», вып. 31, М., «Наука», 1976, 5—33.
47. Королева З. Е., О сравнении тестовых алгоритмов распознавания. Журнал выч. матем. и матем. физики 15, 3, 1975, 749—756.
48. Ренделев Ф. П., Дмитриев А. Н., Журавлев Ю. И., Сравнение геологического строения зарубежных месторождений докембрийских конгломераторов с помощью дискретной математики. ДАН СССР 173, 5, 1967, 1149—1152.
49. Кудрявцев В. Б., Кудрявцев В. Б. О тестовом подходе к задаче о перспективности населенных пунктов. Сб. «Исследование операций», М., ВЦ АН СССР, 1972.
50. Кузнецов В. Е., Об одном стохастическом алгоритме вычисления информационных характеристик таблиц по методу тестов. Сб. «Дискретный анализ», вып. 23, Новосибирск, ИМ СО АН СССР, 1973, 8—23.
51. Кульянов Е. Г., Об оптимизации одного класса алгоритмов распознавания. Журнал выч. матем. и матем. физики 14, 3, 1974, 756—767.
52. Кульянов Е. Г., Алгоритм поиска максимального верхнего нуля произвольной монотонной функции алгебры логики. Журнал выч. матем. и матем. физики 15, 4, 1975, 1080—1082.
53. Леонович А. С., Построение алгоритма распознавания причин брака в стальеплавильном производстве. Сб. работ по матем. кибернетике, вып. 1, М., ВЦ АН СССР, 1976, 212—222.
54. Мадатян Х. А., Построение оптимального алгоритма в одном классе задач распознавания. Сб. работ по матем. кибернетике, вып. 4, М., ВЦ АН СССР, 1976, 223—232.
55. Мазуров В. Д., О комитете системы выпуклых неравенств. Труды ICM-1966, М., МГУ, 1966.
56. Мазуров В. Д., Комитеты системы неравенств и задача распознавания. Кибернетика, № 3, 1971.
57. Миросник С. Н., Алгоритм голосования с непрерывной метрикой. Кибернетика, № 2, 1972, 54—63.
58. Носков В. Н. О тупиковых и минимальных тестах для одного класса таблиц. Сб. «Дискретный анализ», вып. 12, Новосибирск, ИМ СО АН СССР, 1968, 27—49.
59. Нурыбаев А., О нормальных формах k -значной логики. Сборник работ по математической кибернетике, вып. 1, ВЦ АН СССР, М., 1976, 56—68.
60. Рязанов В. В., Оптимизация алгоритмов вычисления оценок по параметрам, характеризующим представительность эталонных строк. Журнал выч. матем. и матем. физики 16, 6, 1976, 1559—1570.
61. Сборник. Метод комитетов в распознавании образов. АН СССР, Уральский научный центр, Свердловск, 1974.
62. Себастьян Г. С., Процессы принятия решений при распознавании образов. М., Изд-во «Техника», 1965.
63. Слепян В. А., Вероятностные характеристики распределения тупиковых тестов. Сб. «Дискретный анализ», вып. 12, Новосибирск, ИМ СО АН СССР, 1968, 50—74.
64. Слепян В. А., Параметры распределения тупиковых тестов и информационные веса столбцов в бинарных таблицах. Сб. «Дискретный анализ», вып. 14, Новосибирск, ИМ СО АН СССР, 1969, 28—43.
65. Слепян В. А., Длина минимального теста для некоторого класса таблиц. Сб. «Дискретный анализ», вып. 23, Новосибирск, ИМ СО АН СССР, 1973, 59—71.
66. Слуцкая Т. Л., Алгоритм вычисления информационных весов признаков. Сб. «Дискретный анализ», вып. 12, Новосибирск, ИМ СО АН СССР, 1968, 75—90.
67. Слуцкая Т. Л., Об эффективности алгоритмов голосования для одного класса бинарных таблиц. Кибернетика, № 2, 1973, 80—86.
68. Теренков В. Н., О точности алгоритмов вычисления оценок для таблиц, порождаемых монотонными булевыми функциями. Журнал выч. матем. и матем. физики 13, 6, 1973, 1620—1625.
69. Теренков В. Н., О точности алгоритмов вычисления оценок для одного класса таблиц. Кибернетика, № 3, 1974, 27—31.
70. Узаков Я. А. Об одном определении меры значимости признаков, участвующих в описании объекта. Известия АН Уз. ССР, серия технических наук, № 1, 1976, 3—6.
71. Чегис И. А., Яблонский С. В., Логические способы контроля электрических схем. Труды Матем. ин-та им. В. А. Стеклова АН СССР, 51, 1958, 270—360.
72. Шаукова Л. З., Оценка информативности функциональных показателей и их совокупностей в динамике лечения. Журнал выч. матем. и матем. физики 15, 3, 1975, 803—806.

73. Шауцукова Л. З., Решение одной задачи классификации методом вычисления оценок. Журнал выч. матем. и матем. физики 15, 4, 1975, 1020—1030.
74. Яблонский С. В., Функциональные построения в k -значной логике. Труды Матем. ин-та им. В. А. Стеклова АН СССР 51, 1958, 5—142.
75. Яблонский С. В., Демидова Н. Г., Константинов Р. М., Королева З. Е., Кудрявцев В. Б., Сиротинская С. В., Тестовый подход к количественной оценке геолого-структурных факторов и масштабов оруденения (на примере ртутных месторождений). Геология ртутных месторождений 13, 2, 1971, 30—42.
76. Ablow C. M., Kaylog D. J., A committee solution of the pattern recognition problem. IEEE Trans. IT-11, 3, 1965.
77. Grenander U., Foundations of pattern analysis. Quarterly of Applied Mathematics, XXVII, 1, 1969, 1—55.
78. Fu K. S., Syntactic Approaches to pattern recognition. Acad. Press. New York, 1974.
79. Zadeh L. A., Fuzzy sets. Inform. Contr. 8, 1965, 338—353.
80. Zhuravlev Y. I., Algorithms for assessing the quality of expert data. International Seminar on Trends in Math. Modelling. 80, Springer-Verlag. Berlin — Heidelberg — New-York, 1973, 238—242.

Поступило в редакцию 17 VI 1976.

Приложение при корректуре. После сдачи работы в печать автором были получены более общие результаты: см. Ю. И. Журавлев. Корректные алгебры над множеством некорректных (эвристических) алгоритмов. I, II. Кибернетика, № 4, № 6, 1977.
