Московская финансово-промышленная академия

Фомин Я.А.

«Диагностика кризисного состояния предприятия»

Москва 2004

УДК 65.016.7
ББК 65.2965.290-2
Ф 762762

Фомин Я.А. Диагностика кризисного состояния предприятия - М. Московская финансово-промышленная академия. 2004. – 61 с.

Рекомендовано Учебно-методическим объединением по образованию в области антикризисного управления в качестве учебного пособия для студентов высших учебных заведений, обучающихся по специальности 351000 «Антикризисное управление» и другим экономическим специальностям.

В учебном пособие рассмотрены основные методы и алгоритмы распознавания, позволяющие с гарантированной достоверностью определять кризисные (или некризисные) состояния фирмы в условиях риска.

Рецензенты: к.н.т., проф. Калинина В.Н.
к.э.н., доц. Бакуменко Л.П.

Содержание

1. Цели фирмы и необходимость распознавания ее кризисного состояния

Название курса “Распознавание кризисных состояний фирмы” отражает довольно узкую проблемы в сравнении с тем широким кругом задач, с которым приходится иметь дело менеджерам. Однако не все задачи в менеджменте имеют одинаковый приоритет. Его определяют цели, которые ставит перед собой руководство фирмы. Если цели фирмы адекватны сложившейся на рынке ситуации, а поставленные задачи, которые им соответствуют, удалось решить, то фирма достигает успеха. Организация считается добившейся успеха, если она добилась своих целей. Какую же цель преследует задача распознавания кризисных состояний или, проще говоря, зачем нужно вовремя определить наличие некого порога, за которым фирма перестает быть процветающей? Очевидно, решение этой задачи не является достаточным условием для реализации основной общей цели – миссии. Миссия – есть та причина, по которой фирма функционирует на рынке, и чтобы успешно на нем функционировать, необходимо разработать множество целей и стратегий, и адаптироваться к самым различным факторам окружающего мира. На самом деле, главная цель, выживание фирмы, но многие менеджеры почему-то игнорируют ее, считая само собой разумеющимся фактом. Между тем, в малом бизнесе, где высока степень конкуренции, многие фирмы просто не имеют возможности ставить перед собой более высокие задачи без риска разорения. В остальных случаях, конечно, миссия фирмы заключается чаще в росте прибыльности, завоевании рынка, вследствие чего задача не разориться становится слишком ограниченной. Но эта цель, если даже миссия фирмы усложняется, значения своего ничуть не теряет, поскольку является необходимым условием процветания фирмы. Вывод о наличии кризиса делается исследователем на основе созданной им модели для принятия решения, но окончательное решение принимает менеджер, использующий эту модель. Руководство должно знать, насколько сильна тенденция спада, т. е. возможность потери фирмой ее конкурентных преимуществ, а главное, когда совокупность неблагоприятных явлений ослабит фирму настолько, что ее состояние станет кризисным. Источники информации, которыми, как правило, мы можем оперировать: бизнес-план фирмы, стратегический план фирмы, финансовая отчетность, балансовый отчет. При этом предполагается, что существует доступ к финансовым показателям и показателям стратегического учета фирм-конкурентов.

Теперь о том, что мы будем распознавать. Смысл нашей задачи – зафиксировать порог, за которым складывается определенная комбинация показателей деятельности фирмы, определяющая общий неблагоприятный результат, который и будет кризисом (кризисным состоянием фирмы) Составляющих кризиса –множество. Поэтому

диагностика кризисного состояния является многомерной и сложной задачей.

Во избежание путаницы в терминологии сразу нужно разделить понятия “кризис” и “банкротство”. В отечественной литературе иногда отождествляются эти два неодинаковых понятия. Банкротство есть крайняя форма кризисного состояния, когда фирма, не имеет каких-либо возможностей оплатить кредиторскую задолженность и восстановить свою платежеспособность за счет собственных ресурсов. Если же проанализировать процесс спада фирмы, то становится очевидным, что между порогом кризиса и началом процедуры банкротства, как правило, существует значительный отрезок времени. За время от момента фиксации кризиса до начала банкротства фирма способна восстановить платежеспособность за счет собственных ресурсов (если, конечно, эти два момента не совпали). После начала процедуры банкротства это уже невозможно (за исключением случая, когда арбитражный суд признает фирму состоятельной): фирма либо ликвидируется, либо финансируется из других источников (бюджет, кредиторы). Поэтому при всех отрицательных аспектах кризиса не стоит переоценивать остроту ситуации. Целесообразно провести детальный анализ обстановки для выработки мер по ее улучшению.

Особенность задачи, решаемой в рамках данного курса, заключается в том, что при множестве различных показателей, отражающих результаты деятельности фирмы, существует всего две альтернатив при принятии решения: “ кризис” - “не кризис”. Достоинством математических моделей, все шире применяемых в менеджменте (как правило, в крупном бизнесе в США и Англии), является их способность вскрывать многие причинно-следственные механизмы, трудно распознаваемые методами неколичественного анализа. Очень хорошо себя зарекомендовало применительно к менеджменту теория принятия решений. Согласно этой теории задачи управления можно формально разделить на 3 категории; руководствуясь критерием неопределенности. Выделяют детерминированный случай, когда руководитель точно знает результаты каждого из альтернативных вариантов, которыми он располагает (ситуация определенности). Здесь эффективно применять методы линейного программирования, эти методы могут помочь однозначно определить результат, так как все входные данные имеются и могут быть использованы как исходные данные в математической постановке задачи. Решения, принимаемые в условиях риска, – это такие решения, результаты которых не могут выражаться точно, но известна вероятность каждого из них. Результат определяется конечным числом альтернатив, сумма вероятностей которых равна единице. Но при этом требуется, точный расчет вероятности на основе статистических данных. К задачам с риском относятся задачи, которые могут быть решены методами анализа временных рядов, распознавания образов, с помощью теории игр. В

условиях неопределенности, невозможно оценить степень вероятности результатов (исследователь не знает либо вообще, какие возможны результаты принятия решения, либо знает лишь некоторые из них). В этом случае модель может только весьма приближенно описать среду принятия решения, основываясь на значительных допущениях. В подобной ситуации менеджеры склонны полагаться на собственный опыт, хотя некоторые сложные математические модели (модель оптимального управления) целесообразно использовать применительно к нуждам менеджмента [12,14].

Задача диагностики кризисных состояний фирмы и является типичной задачей двухальтернативного принятия решений с риском и решается в рамках теории распознавания образов. Риск – это вероятность принятия ошибочного решения. В данном случае (в отличие, например, от игровых методов) эта вероятность является объективной, т. е. вычисляется методами интегрирования распределения оценки отношения правдоподобия. Следовательно, можно гарантировать любую желаемую достоверность правильного принятия решения. Из теории проверки гипотез (раздел математической статистики) известно, что « ... байесовский критерий отношения правдоподобия является оптимальным в том смысле, что он минимизирует риск вероятности ошибки» [16].

Методы распознавания образов занимают центральное место в курсе. Это объясняется тем, что принятие даже двухальтернативного решения “кризис” или “не кризис” требует охвата большого числа показателей деятельности фирмы, и влечет за собой необходимость использования многомерных статистических методов, т. к. 1) данные показатели являются случайными величинами; 2) данных показателей большое число; 3) показатели могут быть связаны между собой любым образом в любых сочетаниях. Распознавание образа – это отнесение объекта к тому или иному классу S1 или S2. Задача распознавания образов включает три этапа:

формирование признакового пространства;

обучение распознающей системы – создание обобщенных портретов (классов) убыточных S2 и процветающих S1 фирм для снятия неопределенности с помощью обучающих наблюдений;

принятие решений – отнесение фирмы к классу убыточных S2 или к классу процветающих S1..

Структура пособия полностью подчиняется вышеуказанным этапам решения задачи, следовательно начинать нужно с формулировки признаков, которые бы наиболее полно отразили разницу между процветающими и убыточными фирмами. В рамках экономической теории точного определения класса процветающих и класса убыточных фирм нет. Но разве не является критерием различия прибыль, получаемая фирмой? Ответ: “Нет.” С момента выбора миссии фирма ориентируется на удовлетворение нужд своих клиентов и на принципы

сосуществования на рынке с фирмами – конкурентами. «Прибыль, – пишут М. Мескон, М. Альберт и Ф. Хедоури, – представляет собой полностью внутреннюю проблему организации ... Она может выжить, только если будет удовлетворять какую-то потребность, находящуюся вне ее самой. Чтобы заработать прибыль, необходимую ей для выживания, фирма должна следить за средой, в которой функционирует. Поэтому именно в окружающей среде руководство подыскивает общую цель организации» [12]. Другие американские специалисты по менеджменту А. Томпсона и А. Стрикленда из университета штата Алабама высказывает ту же мысль:... прибыль – это скорее результат того, что делает компания. То, что мы собираемся иметь прибыль, не говорит ничего о том, в какой среде будет эта прибыль получена. Миссии компаний, ориентированные только на получение прибыли, не дают возможности отличить одно предприятие от другого ... Компания, которая говорит, что ее цель – получить прибыль, должна ответить на вопрос: “Что мы предпринимаем, чтобы получить прибыль” [14]. Следовательно, чтобы выяснить, удачен бизнес фирмы или неудачен, недостаточно уметь определять ситуацию внутри фирмы, менеджер обязан соотносить внутренние экономические показатели с показателями рыночной среды в отрасли. Выше было сказано, что важно вовремя распознавать кризисные ситуации. Но совершенно ясно, что никакая современная информация сама по себе из кризиса фирму не вытащит. Другое дело, что в случае несвоевременного принятия решения о наступлении кризиса, поправить положение будет очень трудно либо невозможно. Но даже если предположить, что менеджер своевременно распознал кризисную ситуацию, он обязан на будущее разработать такие стратегии, которые бы спасли фирму от дальнейшего спада и, в конечном итоге, от разорения. Поэтому принятие решения о том, что фирма находится в кризисе, повлечет за собой ряд ответственейших решений, непосредственно относящихся к функциям управления. Следует подчеркнуть, что в кризисных ситуациях, а также в начале деятельности, т. е. в тех случаях, когда бизнес фирмы наиболее ослаблен, стратегии управления меняются часто и быстро до тех пор, пока фирма на конкурентном рынке не приобретет устойчивое положение. Вступление в фазу кризиса – это вступление в новую ситуацию, характерную прежде всего острой нехваткой финансовых ресурсов. Чтобы выжить, необходима полная мобилизация всех имеющихся ресурсов фирмы и принятие нестандартных менеджерских решений. Приняв решение о кризисе, следует попытаться выделить главные причины, приведшие к нему, а также оценить реальную серьезность кризисного явления. С этого начинается пересмотр стратегии фирмы. Если же принято решение о том, что фирма по сравнению с конкурентами пока далека от кризиса, это означает, что существующая стратегия приносит свои плоды, что она эффективна, а менеджер, ее реализующий, вполне контролирует ситуацию справляется

со своими обязанностями. Пересмотр стратегии и ряд срочных мер, направленных на форсированное достижение конкурентного преимущества, успеха не принесет, так как вызовет быструю растрату финансовых ресурсов, усилит недоверие работников к переменам, не выгодным им, понизив их мотивацию, может отрицательно повлиять на организационную структуру, когда ответственные лица не сразу будут готовы нормально работать в изменившихся условиях. Корректировка стратегии была бы наиболее адекватна ситуации, в которой требуется лишь незначительная адаптация [13] .

Итак, мы выяснили, почему менеджер так остро нуждается в достоверной информации о наличии кризиса на фирме и что эта информация ему дает. Однако мы не выяснили, для чего следует пользоваться теорией распознавания образов. Почему менее надежно решение менеджера, пользующегося лишь своим опытом и интуицией? Почему ряд других моделей менее предпочтителен для решения этой задачи? На стадии анализа рыночной и внутрифирменной среды приходится сталкиваться, как уже говорилось, со множеством факторов, взаимодействующих между собой в любых сочетаниях с разной степенью связанности, зависимости друг от друга.

Ни один менеджер не сможет абсолютно верно указать такое правило, согласно которому пространству факторов среды, допустим размерности p, где каждый из p факторами может быть связан с остальными p - 1 факторами мерой связи, измеряемой от 0 до 1 (это может быть коэффициент корреляции, например), с заданной заранее гарантированной вероятностью ошибки однозначно бы соответствовала одна из альтернатив решения «кризис» – «не кризис». Объем информации, слишком велик и человеческий мозг не в состоянии ее обработать. Однако для распознающей системы эта задача вполне по силам. Ансамбль p-признаков по результатам обучения (сравнения с такими же p признаками у m фирм) обрабатывается с учетом всех возможных сочетаний между ними и полностью соответствует одному из двух вариантов решения, которое выдает распознающая система. При этом вероятность ошибки может быть задана любая! При наличии самого квалифицированного менеджера и распознающей системы принятие решения лучше доверить последней потому, что может быть обеспечена минимальная вероятность ошибки, недостижимая для человека. Возможность достижения самой высокой гарантированной достоверности принятого решения по сравнению с любой другой моделью ставит теорию распознавания образов и ее методы в наиболее выгодное положение в сравнении с остальными методами в рамках теории принятия решения. Кроме того, числа признаков (рост входной информации) повышает качество решения (выходной информации), вырабатываемое распознающей системой, т. е. система может работать с большими массивами данных отчего качество решения не ухудшается – это еще одно преимущество. Уникальное свойство распознающей

системы – способность обучаться – реализует третье преимущество: количество входной информации пропорционально качеству выходной. В этой ситуации уместно не согласиться с мнением известного американского теоретика в области научных методов управления Рассела Акофа о том, что избыточная информация может снизить качество управленческого решения [13,15]. Если бы это было так, то это бы противоречило выводам теории информации. Проблема скорее состоит в сложности процесса обработки и анализа информации. Менеджеры действительно чаще обладают ограниченными возможностями объективного анализа поступающей к ним информации, с осторожностью относятся к использованию математических моделей и, действительно, ошибаются. Однако причина не в количестве располагаемой информации, а в недостаточном умении ее использовать и интерпретировать. В настоящее время ситуации, требующие управленческих решений, усложняются, что связано с ростом изменчивости внешней среды, расширением номенклатуры товаров и услуг, увеличением числа новых фирм и даже новых отраслей экономики, приростом населения (а, значит, и потребителей) на Земле. Поэтому менеджер, опирающийся на свое суждение и суждения экспертов, принимая решения, будет все более и более подвержен риску ошибки. «Наиболее заметный и, возможно, наиболее значительный вклад школы научного управления заключается в разработке моделей, позволяющих принимать объективные решения в ситуациях, слишком сложных для простой причинно-следственной оценки альтернатив» [12].

2. Общая постановка задачи распознавания кризисных состояний фирмы

Проведенный в разделе 1 анализ кризисных явлений в экономике фирмы (предприятия) и экономического механизма возникновения кризисного состояния показывает, что главную роль в антикризисном управлении фирмой играет своевременное распознавание ее кризисного состояния с требуемым уровнем достоверности: D =1 – α = 1 – β (α, β – ошибки распознавания 1-го и 2-го рода) для своевременного принятия мер по предупреждению и предотвращению кризиса.

В общем виде можно полагать, что исследуемая фирма может принимать одно из двух взаимоисключающих состояний: S1 – нормальное (бескризисное) и S2 – кризисное. Распознавание представляет собой отнесение наблюдаемого неизвестного состояния, заданного совокупностью Xn наблюдений над его признаками X1, X2, ..., X_p

$$\bar{X}_n = \begin{pmatrix} X_{11}\ X_{12,}\ldots, X_{1n} \\ X_{21}\ X_{12},\ldots, X_{2n} \\ \ldots\ldots\ldots\ldots\ldots \\ X_{p_1}\ X_{p2},\ldots, X_{pn} \end{pmatrix} \qquad (2.1)$$

к одному из двух взаимоисключающих состояний S1 или S2.

$$\bar{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ . \\ . \\ . \\ x_{pi} \end{pmatrix} = \left(x_{1i}, x_{2i}, \ldots, x_{pi}\right)^T, i = 1,2,\ldots,n$$

Каждый столбец матрицы $\overline{X}_n$ представляет собой p - мерный вектор наблюдаемых значений p признаков X1, X2, ..., Xp, отражающих наиболее важные для распознавания свойства.

Набор признаков p, как правило, является одинаковым для всех распознаваемых классов S1, S2. Если каждый класс S1 и S2 описывается своим набором признаков, то задача распознавания становится тривиальной, поскольку однозначное отнесение имеющейся совокупности наблюдений к определенному классу легко осуществляется по набору составляющих ее признаков. Общая схема системы распознавания кризисных состояний фирмы приведена на рис. 1.

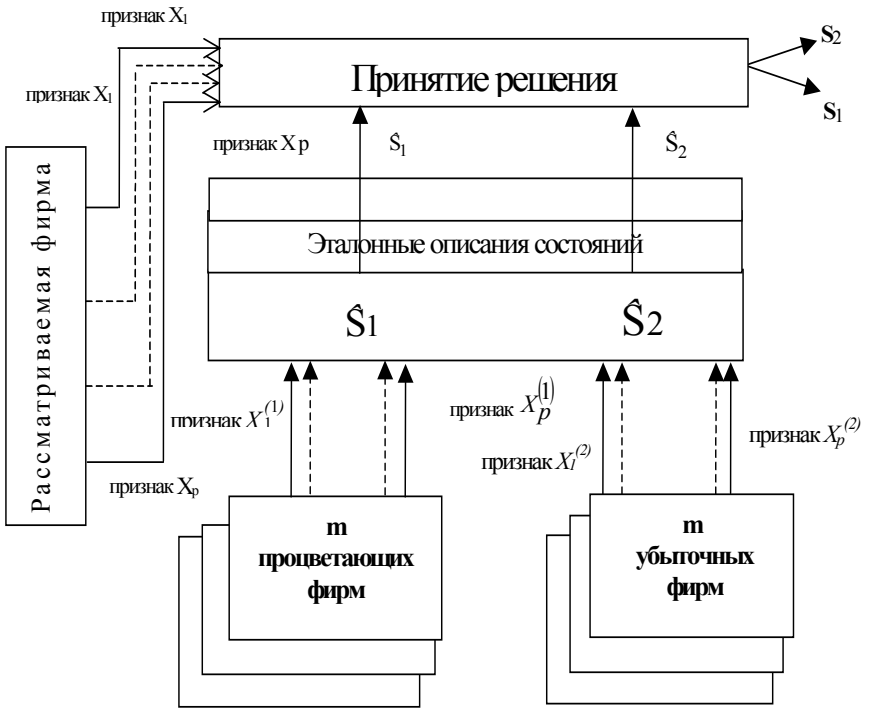

Рис. 1

Таким образом, рассматривается задача принадлежности наблюдаемого состояния к одному из конечного фиксированного числа классов s1, s2, описываемых одинаковым для всех классов набором признаков X1 , X2 , ..., Xp. При этом различие между классами будет проявляться только в том, что у разных объектов одни и те же признаки будут иметь различные характеристики (количественные, качественные и др.), и для любого набора признаков X1, ..., Xp можно задать правила, согласно которым двум классам S1 и S2 ставится в соответствие вектор

$$d12=\begin{vmatrix} d_1^{12} \\ . \\ . \\ . \\ d_p^{12} \end{vmatrix} \qquad (2.2)$$

состоящий из p скаляров, называемых межклассовыми расстояниями и выражающих степень отличия у этих классов характеристик данных признаков.

Определение набора признаков X1, X2, ..., Xp , т. е. формирование признакового пространства, является неотъемлемой составной частью распознающего процесса. С одной стороны, выбранная совокупность признаков должна в наибольшей степени отражать все те свойства состояний, которые важны для их распознавания, т. е. набор X1, X2, ..., Xp должен быть наиболее полным. С другой стороны, с увеличением

размерности p признакового пространства очень быстро возрастают вычислительная сложность процедур обучения и принятия решения, материальные и трудовые затраты на измерение необходимых характеристик объектов, т. е. на получение наблюдений на этапе обучения и принятия решений.

Основным показателем качества распознающей системы является достоверность принимаемых ею решений [1, 2]. Если распознающая процедура допускает большой процент ошибочных решений, то подобно ненадежным компьютерам она делает практически непригодной любую, пусть даже очень совершенную в других отношениях систему, частью которой она является. Таким образом, практический интерес представляют только те системы, которые обеспечивают требуемый уровень достоверности распознавания.

Сокращение количества признаков уменьшает затраты на проведение измерений и вычислений, но может привести к снижению достоверности распознавания. Если время на обучение и принятие решения жестко ограничено, то повышение размерности признакового пространства может оказаться единственным средством увеличения достоверности. Таким образом, одновременное достижение минимума общей размерности признакового пространства и максимума достоверности распознавания оказывается, как правило, невозможным, и, следовательно, одной из основных задач синтеза распознающих систем является выбор из заданного множества признаков X1, X2, ..., Xp оптимального набора Xi1, Xi2, ... , Xp0 из p0 признаков, обеспечивающего требуемый по условиям решаемой задачи уровень достоверности распознавания и минимизирующего затраты на проведение измерений и вычислений.

Другой важной составной частью распознающего процесса является обучение, цель которого – восполнение недостатка априорных знаний о распознаваемых классах S1 и S2 путем использования информации о них, содержащейся в обучающих наблюдениях:

$$X_m^{(1)} = \begin{pmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \dots & x_{1m}^{(1)} \\ x_{21}^{(1)} & x_{22}^{(1)} & \dots & x_{2m}^{(1)} \\ . & . & . & . \\ x_{p1}^{(1)} & x_{p2}^{(1)} & \dots & x_{pm}^{(1)} \end{pmatrix}, \quad X_m^{(2)} = \begin{pmatrix} x_{11}^{(2)} & x_{12}^{(2)} & \dots & x_{1m}^{(2)} \\ x_{21}^{(2)} & x_{22}^{(2)} & \dots & x_{2m}^{(2)} \\ . & . & . & . \\ x_{p1}^{(2)} & x_{p2}^{(2)} & \dots & x_{pm}^{(2)} \end{pmatrix} \quad (2.3)$$

где m – количество обучающих наблюдений.

Хотя методы и подходы, используемые при обучении, могут быть разнообразными, конечный результат их использования, как правило, неизменен – это эталонные описания состояний $\hat{s}_1, \hat{s}_2$. Увеличение продолжительности обучения повышает достоверность распознавания за счет увеличения количества информации о распознаваемых классах, содержащейся в обучающих выборках и позволяющей уточнять их

эталонные описания $\hat{s}_1, \hat{s}_2$. В то же время увеличение времени обучения влечет за собой рост затрат на измерения и вычисления и, что самое главное, увеличение общего времени, требуемого для решения задачи распознавания. Сокращение же времени обучения может повлиять на качество эталонных описаний и в конечном итоге привести к снижению достоверности распознавания. Следовательно, определение минимального времени обучения, обеспечивающего заданный уровень достоверности распознавания, является одной из важных задач, возникающих при синтезе распознающих систем.

Реализация информации о распознаваемых классах, содержащейся в их эталонных описаниях $\hat{s}_1, \hat{s}_2$ и в совокупности наблюдений (2.1), осуществляется в процедуре принятия решений, занимающей центральное место в распознающем процессе. Процедура сводится к сопоставлению неклассифицированных наблюдений с эталонными описаниями и указанием номера класса l из множества 1, 2 номеров классов, к которому принадлежит рассматриваемая совокупность наблюдений. Таким образом, решающая процедура осуществляет отображение наблюдений на конечное множество натуральных чисел 1, 2 с использованием информации о классах, содержащейся в обучающих наблюдениях и отражаемой в эталонных описаниях классов $\hat{s}_1, \hat{s}_2$.

Увеличение продолжительности процедуры принятия решения в принципе повышает достоверность распознавания за счет вовлечения в процесс принятия решения большего количества информации о состоянии фирмы, содержащейся в описывающей совокупности наблюдений (2.1), которую в дальнейшем будем именовать контрольной выборкой. Однако для подавляющего большинства распознающих систем естественными являются требования минимальной продолжительности процедуры принятия решения как с точки зрения быстроты решения задач, так и с позиций минимизации затрат на измерения и вычисления. Таким образом, определение минимального времени принятия решения, обеспечивающего заданный уровень достоверности распознавания, также является одной из важных задач синтеза распознающих систем.

Итак, основными параметрами распознающей системы являются: количество признаков p, объемы выборок (обучающих m и контрольной n) и достоверность распознавания D. На практике при синтезе распознающей системы, заключающемся в выборе величин p, m, n и D, обеспечивающем решение задачи распознавания наилучшим образом, на значения всех или некоторых из перечисленных параметров накладываются ограничения, обусловливаемые либо необходимостью достижения высокого уровня достоверности принимаемых решений, либо жесткими требованиями на время обучения и распознавания, либо ограниченными возможностями по затратам на получение наблюдений, либо и тем, и другим, и третьим. В то же время отмеченный выше

сложный характер взаимосвязей между параметрами распознающей системы приводит к тому, что нередко удовлетворить всем налагаемым на них ограничениям можно при различных соотношениях между этими параметрами. В этих условиях появляется возможность выбора таких значений параметров p, m, n и D, которые удовлетворяют всем ограничениям и являются наилучшими (оптимальными) с точки зрения некоторого критерия, т. е. появляется возможность оптимизации распознающей системы [1].

Для обеспечения гарантированной достоверности распознавания важную роль играет получение в удобной для практического использования форме зависимости достоверности распознавания D от параметров распознающей системы p, m, n и межклассовых расстояний.

3. Анализ методов распознавания с точки зрения обеспечения гарантированной его достоверности

Детерминистские (перцептронные) методы основаны на использовании перцептрона и обучения на основе принципа подкрепления - наказания. Основная модель перцептрона, обеспечивающая отнесение образа к одному из двух классов, состоит из сетчатки S сенсорных элементов, соединенных с ассоциативными элементами сетчатки A, каждый элемент которой воспроизводит выходной сигнал только, если достаточное число сенсорных элементов, соединенных с его входом, находятся в возбужденном состоянии (рис. 2).

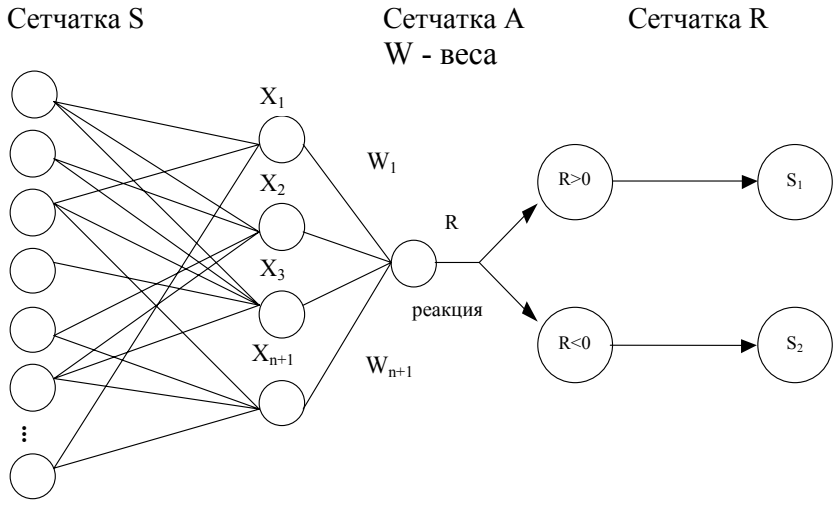

Рис.2

Реакция всей системы пропорциональна сумме взятых с определенными весами wi реакций xi элементов i ассоциативной сетчатки.

$$R=\sum_{i=1}^{n+1} w_i x_i = \overline{w}'\overline{x}\begin{cases} >0 \to S_1 \\ <0 \to S_2 . \end{cases} \qquad (3.1)$$

Разделение на несколько классов можно реализовать, добавив К элементов в R – сетчатку (К – число классов). Классификация проводится обычным способом: рассматриваются значения реакций R1, R2, ..., Rк , и образ причисляется к классу Si, если Ri > Rj для всех j = i (метод «чемпиона»).

Обучение перцептрона по принципу подкрепления-наказания. Обучающий алгоритм сводится к простой схеме итеративного определения вектора весов W. Заданы два обучающих множества, представляющие классы S1 и S2 соответственно.

Пусть W (1) – начальный вектор весов, выбираемый произвольно. Тогда на k-ом шаге обучения, если $\overline{x}(k) \in S_1$ и $\overline{w}(k)\,\overline{x}(k) \le 0$, то вектор весов $\overline{w}(k)$ заменяется вектором $\overline{w}(k+1) = \overline{w}(k) + c\overline{x}(k)$, где с - корректирующее приращение.

Если $\overline{x}(k) \in S_2$ и $\overline{w}(k)\,\overline{x}(k) \ge 0$, то вектор $\overline{w}(k)$ заменяется вектором $\overline{w}(k+1) = \overline{w}(k) - c\,\overline{x}(k)$. В противном случае w (k) не изменяется, т. е. $\overline{w}(k+1) = \overline{w}(k)$.

Таким образом, изменения в вектор весов W вносятся алгоритмом только в том случае, если образ, предъявляемый на k-ом шаге обучения, был при выполнении этого шага неправильно классифицирован, с помощью соответствующего вектора весов. Корректирующее приращение с должно быть положительным, и в данном случае предполагается, что оно постоянно.

Следовательно, алгоритм является процедурой типа “подкрепление-наказание”, причем подкреплением является отсутствие наказания, т. е. то, что в вектор весов $\overline{W}$ не вносится никаких изменений, если образ классифицирован правильно.

Если образ классифицирован неправильно, и произведение $\overline{w}'(k)\,\overline{x}(k)$ оказывается меньше нуля, когда оно должно быть больше нуля, система “наказывается” увеличением вектора весов $\overline{W}(k)$ на величину, пропорциональную $\overline{x}(k)$. Точно так же, если произведение $\overline{w}'(k)\,\overline{x}(k)$ оказывается больше нуля, когда оно должно быть меньше нуля, система наказывается противоположным образом.

Сходимость алгоритма наступает при правильной классификации всех образов с помощью некоторого вектора весов.

Основная задача, заключающаяся в выборе подходящего множества решающих функций, решается, в основном, методом проб и ошибок,

поскольку, как указано в [4], единственным способом оценки качества выбранной системы является прямая проверка. Совершенно ясно, что отсутствие аналитических методов оценки достоверности распознавания, увязанной с параметрами распознающей процедуры, не позволяет обеспечить гарантированную достоверность распознавания в системах детерминистского распознавания, основанных на использовании перцепторных алгоритмов.

В лингвистическом (синтаксическом, структурном) подходе признаками служат подобразы (непроизводные элементы) и отношения, характеризующие структуру образа. Для описания образов через непроизводные элементы и их отношения используется «язык» образов. Правила такого языка, позволяющие составлять образы из непроизводных элементов, называются грамматикой. Грамматика определяет порядок построения образа из непроизводных элементов. При этом образ представляется некоторым предложением в соответствии с действующей грамматикой.

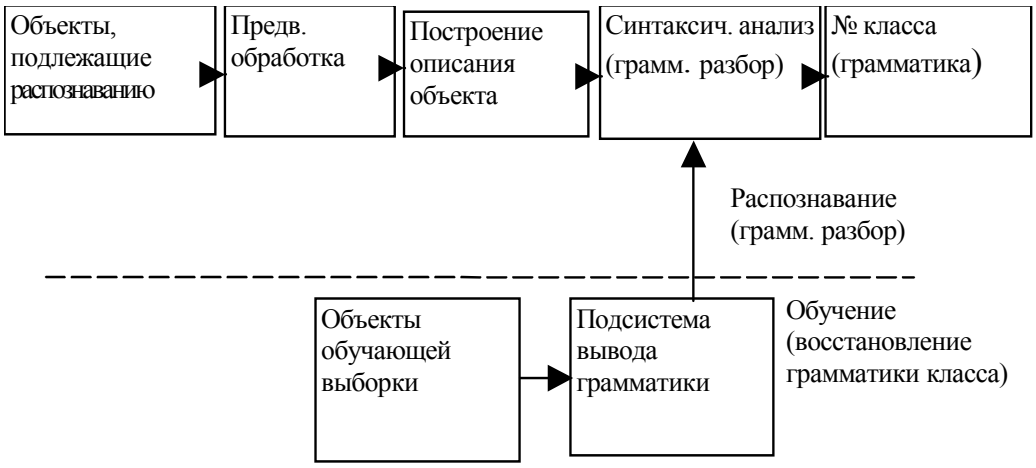

Рис. 3

Распознавание состоит из двух этапов (рис.3):

определение непроизводных элементов и их отношений для конкретных типов объектов и обучение;

проведение синтаксического анализа предложения, представляющего объект, чтобы установить какая из имеющихся фиксированных грамматик могла породить имеющееся описание объекта (грамматический разбор) .

Грамматики часто удается определять на основе априорных сведений об объекте, в противном случае грамматики всех классов восстанавливаются в ходе обучения, которое использует априорные сведения об объектах и обучающую выборку. Объект после предварительной обработки (например, черно-белое изображение можно закодировать с помощью сетки, или матрицы нулей и единиц) представляется некоторой структурой языкового типа (например,

цепочкой или графом). Затем он разбивается (сегментируется), определяются непроизводные элементы и отношения между ними. Так, при использовании операции соединения объект получает представление в виде цепочки соединенных непроизводных элементов.

Решение о синтаксической правильности представления объекта, т.е. о его принадлежности к определенному классу, задаваемому определенной грамматикой, вырабатывается синтаксическим анализатором (блоком грамматического разбора). Цепочка непроизводных элементов, представляющая поданный на вход системы объект, сопоставляется с цепочками непроизводных элементов, описывающими классы. Распознаваемый объект с помощью выбранного критерия согласия (подобия) относится к тому классу, с которым обнаруживается наилучшая близость.

Обучение: по заданному набору обучающих объектов, представленных описаниями структурного типа, делается вывод грамматики, характеризующей структурную информацию об изучаемом классе объектов. Структурное описание соответствующего класса формируется в процессе обучения на примере реальных объектов, относящихся к этому классу. Это эталонное описание в форме грамматики используется затем для синтаксического анализа. В более общем случае обучение может предусматривать определение наилучшего набора непроизводных элементов и получение соответствующего структурного описания классов.

Для распознавания двух классов объектов S1 и S2 необходимо описать их объекты с помощью признаков V (непроизводных элементов, подобразов). Каждый объект может рассматриваться как цепочка или предложение из V . Пусть существует грамматика Г1 такая, что порождаемый ею язык L(Г1) состоит из предложений (объектов), принадлежащих исключительно одному из классов (например, S1). Предъявляемый неизвестный объект можно отнести к S1, если он является предложением языка L (Г1), и к S2, если он является предложением языка L (Г2).

Пример. Распознавание прямоугольников на фоне других фигур (рис. 4)

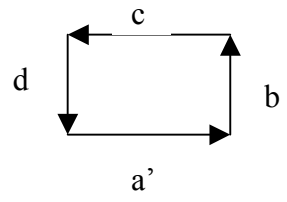

Рис. 4

Выбор непроизводных элементов:

a' -0° отрезок горизонтальной линии

b' -90° отрезок вертикальной линии

c' -180° отрезок горизонтальной линии

d' -270° отрезок вертикальной линии

Множество всех возможных прямоугольников задается с помощью одного предложения – цепочки $a'\,b'\,c'\,d'$. Составляем грамматику Г для прямоугольников, все другие грамматики – не прямоугольников, наблюдаемый объект сопоставляется с грамматиками и принимается решение о его принадлежности.

Если же требуется различение прямоугольников разных размеров, то приведенное описание не адекватно. В качестве непроизводных элементов необходимо использовать отрезки единичной длины. Тогда множество прямоугольников различных размеров можно описывать с помощью языка: $L=\{a^n, b^m, c^n, d^n, n, m = 1, 2, ...\}$.

Как указано в [5], структурный подход к распознаванию не располагает еще строгой математической теорией и рассматривается как комплекс практически работающих эвристических приемов. Это не позволяет увязать главный показатель качества распознавания – достоверность – другими параметрами распознавания и, следовательно, не позволяет осуществить синтез распознающей системы, обеспечивающей гарантированную достоверность распознавания.

В логических системах распознавания [5] классы и признаки объектов рассматриваются как логические переменные. Все априорные сведения о классах S1, S2 и признаках X1, X2, ..., Xp, присущих объектам классов S1, S2, полученные в результате проведения ряда экспериментов (обучения), также выражаются в виде булевых функций. Основным методом решения задач логического распознавания является метод построения сокращенного базиса с помощью алгоритмов получения произведения для булевых функций и отрицания булевой функции и приведения последней к тупиковой дизъюнктивной нормальной форме. Как указывается в [5], логические алгоритмы распознавания в ряде случаев не позволяют получить однозначное решение о принадлежности распознаваемого объекта к классу, а в тех случаях, когда такое решение удается найти, получить в аналитическом виде оценку достоверности распознавания через параметры распознающей системы оказывается невозможно, что делает необходимым использование метода Монте-Карло [5]. К тому же в системах логического распознавания основной упор делается на использование априорных знаний в ущерб процедуре обучения, количественная связь которой с достоверностью распознавания никак не установлена.

Дальнейшим развитием логических методов распознавания являются разработанные Ю. И. Журавлевым алгоритмы логического распознавания, основанные на вычислении оценок (АВО) [6] , которые, в отличие от указанных методов, обеспечивают возможность получения

однозначного решения о принадлежности распознаваемых объектов к определенному классу. АВО основаны на вычислении оценок сходства, количественно характеризующих близость распознаваемого объекта к эталонным описаниям классов, построенным на основе использования обучающей и априорной информации, задаваемой в виде таблицы обучения.

Пусть множество объектов $\{w\}$ поделено на классы S1 , S2 , и объекты описаны одним и тем же набором признаков x1, x2, ..., xp, каждый из которых может принимать значения из множества ($\{0,1,\ldots,d\}$), для простоты из $\{0,1\}$.

Априорная информация представляется в виде таблицы обучения, содержащей описания на языке признаков $\{x_1,\ldots,x_p\}$ всех имеющихся объектов, принадлежащих различным классам (табл. 1).

Пример: Задана таблица обучения (табл. 1)

Табл.1

Классы	Объекты	Значения признаков					
		X1	X2	X3	X4	X5	X6
S1	w1	0	0	0	0	0	0
	w2	0	1	0	0	1	1
	w3	1	1	0	1	1	1
S2	w4	0	1	0	1	0	1
	w5	1	1	1	1	1	1
	w6	1	1	0	0	0	1
		Z1		Z2		Z3	

и подлежащий распознаванию объект w'

w’ 1 1 0 0 0 0

Алгоритм распознавания сравнивает описание распознаваемого объекта w' с описанием всех объектов w1, ..., w6 и по степени похожести (оценки) принимается решение, к какому классу (S1 или S2) относится объект. Классификация основана на вычислении степени похожести (оценки) распознаваемого объекта w' на объекты, принадлежность которых к классам известна. Эта процедура включает в себя три этапа: сначала подсчитывается оценка для каждого объекта w1, ..., w6 из таблицы, а затем полученные оценки используются для получения суммарных оценок по каждому из классов S1 и S2. Чтобы учесть взаимосвязь признаков, степень похожести объектов вычисляется не последовательным сопоставлением признаков, а сопоставлением всех возможных (или определенных) сочетаний признаков, входящих в описание объектов.

Из полного набора признаков $\overline{X} = \{x_1, \ldots, x_p\}$ выделяется система подмножеств множества признаков Z1, ..., Zl (система опорных множеств признаков, либо все подмножества множества признаков фиксированной длины k, k = 2, ..., p - 1, либо вообще все подмножества множества признаков).

Для вычисления оценок по подмножеству Z1 выделяются столбцы, соответствующие признакам, входящим в Z1, остальные столбцы вычеркиваются. Проверяется близость строки $Z_1 \widetilde{w}'$ со строками $Z_1 w_1', \ldots, Z_1 w_r'$, принадлежащими объектам класса S1. Число строк этого класса, близких по выбранному критерию классифицируемой строке Z1 w' обозначается $\Gamma_{z_l}(w', S_l)$ – оценка строки w' для класса S1 по опорному множеству Z1.

Аналогичным образом вычисляются оценки для класса S2: $\Gamma_{z_1}(w', S_2)$, ... Применение подобной процедуры ко всем остальным опорным множествам алгоритма позволяет использовать систему оценок $\Gamma_{z_2}(w', S_1), \Gamma_{z_2}(w', S_2), \ldots, \Gamma_{z_1}(w', S_1), \Gamma_{z_1}(w', S_2)$.

Величины

$$\Gamma(w', S_1) = \Gamma_{z_1}(w', S_1) + \Gamma_{z_2}(w', S_1) + \ldots + \Gamma_{z_l}(w', S_1) = \sum_{z_A} \Gamma(w', S_1)$$
$$\Gamma(w', S_2) = \Gamma_{z_1}(w', S_2) + \Gamma_{z_2}(w', S_2) + \ldots + \Gamma_{z_l}(w', S_2) = \sum_{z_A} \Gamma(w', S_2) \qquad (3.2)$$

представляют собой оценки строки w' для соответ-ствующих классов по системе опорных множеств алгоритма ZA. На основе анализа этих величин принимается решение об отнесении объекта w' к классам S1 или S2 (например, к классу, которому соответствует максимальная оценка, либо эта оценка будет превышать оценку другого класса на определенную пороговую величину η и т. д.).

Так, в примере введем подмножества

Z1 = < x1,x2 >, Z2 = < x3,x4 >, Z3 = < x5 , x6 >

Строки будем считать близкими, если они полностью совпадают.

Тогда

$$Z_1 : \Gamma_{z_1}(w', S_1) = 1 \qquad \Gamma_{z_1}(w', S_2) = 2$$
$$Z_2 : \Gamma_{z_2}(w', S_1) = 2 \qquad \Gamma_{z_2}(w', S_2) = 1 \qquad (3.3)$$
$$Z_3 : \Gamma_{z_3}(w', S_1) = 1 \qquad \Gamma_{z_3}(w', S_2) = 0$$
$$\Gamma_{z_A}(w', S_1) = \Gamma_{z_1}(w', S_1) + \Gamma_{z_2}(w', S_1) + \Gamma_{z_3}(w', S_1) = 1 + 2 + 1 = 4$$
$$\Gamma_{z_A}(w', S_2) = \Gamma_{z_1}(w', S_2) + \Gamma_{z_2}(w', S_2) + \Gamma_{z_3}(w', S_2) = 2 + 1 + 0 = 3$$

Согласно решающему правилу, реализующему принцип простого большинства голосов, объект w' относится к классу S1, так как

$\Gamma(w', S_1) > \Gamma(w', S_2)$.

Дальнейшим обобщением АВО является алгебраический подход к решению задач распознавания и классификации [6], позволяющий преодолеть ограниченные возможности существующих алгоритмов

распознавания путем расширения их семейства с помощью алгебраических операций, введения алгебры на множестве решаемых и близких к ним задач распознавания и построения алгебраического замыкания семейства алгоритмов решения указанных задач. К сожалению, методы количественной оценки достоверности распознавания при использовании АВО и алгебраического подхода к решению задач распознавания, позволяющие в аналитическом виде увязать вероятности ошибок распознавания с параметрами распознающей системы (временем обучения и принятия решения, межклассовым расстоянием и размерностью признакового пространства), в настоящее время отсутствуют [6], что не позволяет обеспечить гарантированную достоверность в указанных системах распознавания.

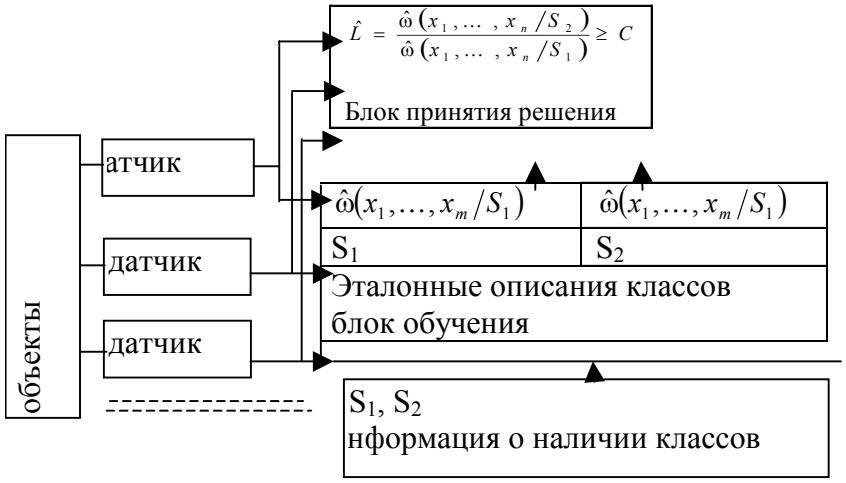

Рис. 5

Статистический метод распознавания. В статистическом методе распознавания (рис. 5) в ходе обучения формируются эталонные описания-оценки многомерных условных плотностей вероятности, которые содержат всю информацию, присутствующую в измерениях x11, ..., x1m, ..., xp1,..., xpm и о всех взаимосвязях между признаками X1, ..., Xp [1,2]. Оценка $\hat{w}(\bar{x}_1,\ldots,\bar{x}_m/S_i)$ является случайной величиной. Для принятия решения используется статистика отношения правдоподобия

$$L(\bar{x})=L(\bar{x}_1,\ldots,\bar{x}_n)=\frac{\hat{w}(\bar{x}_1,\ldots,\bar{x}_n/S_2)}{\hat{w}(\bar{x}_1,\ldots,\bar{x}_n/S_1)}, \quad (3.4)$$

представляющая неотрицательную случайную величину, получаемую функциональным преобразованием $Z=\hat{L}(\bar{x}_1,\ldots,\bar{x}_n)$, которое отображает точки n-мерного пространства выборок на действительную полуось. Таким образом, для вынесения решения, достаточно использовать значение одной случайной величины – статистики отношения правдоподобия $\hat{L}(x_1,\ldots,x_n)$, а не значения каждого элемента выборки (x1, x2, ..., xn) по отдельности. То есть отношение

правдоподобия несет всю статистическую информацию о классах, содержащуюся в данной выборке. Подобная статистика называется достаточной и приводит к редукции наблюдаемых данных: отображению выборочного n-мерного пространства X на действительную положительную полуось (рис. 6).

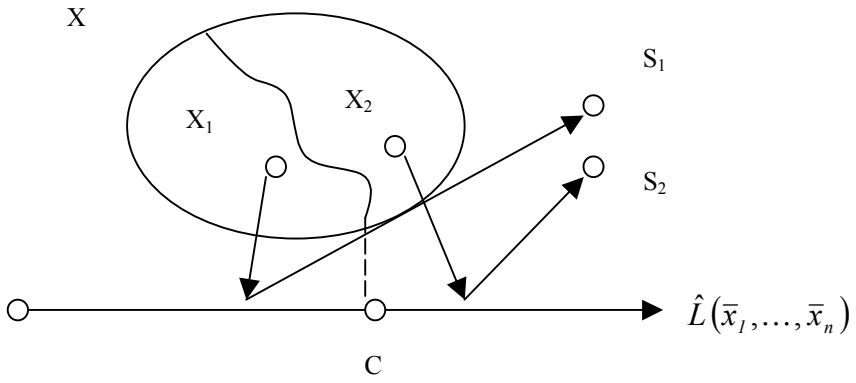

Рис.6 63.5

Поверхность в n-мерном выборочном пространстве, разделяющая пространство X на подпространства X1 и X2, Поверхность в n-мерном выборочном пространстве, разделяющая пространство X на подпространства X1 и X2 , отображается в точку C на оси $L \geq 0$. Принятие решения теперь состоит в отображении интервала 0 < L < C в точку S1 и интервала $L \geq C$ в точку S2. Сводим векторную задачу к скалярной.

Любое монотонное преобразование $\Psi\left[\hat{L}(x_1,\ldots,x_n)\right]$ достаточной статистики отношения правдоподобия также представляет достаточную статистику. Например, $\Psi\left(\hat{L}\right) = \ln\hat{L}(x_1,\ldots,x_n)$. Если элементы выборки независимы, то имеем сумму

$$ln\,\hat{L}(\bar{x}) = ln\,\hat{L}(\bar{x}_1,\ldots,\bar{x}_n) = \sum_{\ell=1}^{n} ln\,\hat{L}(x_\ell) = \sum_{\ell=1}^{n} ln\frac{\hat{w}(\bar{x}_\ell / S_2)}{\hat{w}(\bar{x}_\ell / S_1)}. \qquad (3.5)$$

Для нахождения вероятности ошибок распознавания достаточно располагать распределением отношения правдоподобия (или его логарифма), которое в свою очередь определяется по правилам нахождения функций от случайных величин.

Редукция данных позволяет преодолеть трудности, связанные с вычислением n-кратных интегралов, возникающие при прямом (без введения отношения правдоподобия) вычислении вероятностей ложных тревог α и пропуска цели β.

$$\alpha = \int_{X_2} \hat{w}(\bar{x} / S_1)d\bar{x} = \iint_{X_2} \hat{w}(\bar{x}_1,\ldots,\bar{x}_n / S_1)d\bar{x}_1,\ldots,d\bar{x}_n, \beta = \int_{X_1} \hat{w}(\bar{x} / S_2)d\bar{x} \qquad (3.6)$$

Так как событие $\bar{x} \in X_2$ эквивалентно событию $\hat{L}(\bar{x}) \geq C$, а событие $\bar{x} \in X_1$ – событию $\hat{L}(\bar{x}) < C$, то вероятности ошибок распознавания α и β представляются однократными интегралами.

$$\alpha = P\left\{\hat{L}(\bar{x}) \geq c / S_1\right\} = \int_c^{\infty} w_{L_{m,n}}(z / S_1)\, dz = 1 - F_{\hat{L}}(c / S_1), \quad (3.7)$$

$$\beta = P\left\{\hat{L}(\bar{x}) < c / S_2\right\} = \int_0^{c} w_{\hat{L}_{m,n}}(z / S_2)\, dz = F_{\hat{L}}(c / S_2), \quad (3.8)$$

или, переходя к логарифмам отношения правдоподобия $\ln \hat{L}(\bar{x})$ [(см. (3.5)]:

$$\alpha = P\left\{ln\, L(\bar{x}) \geq ln\, c / S_1\right\} = \int_{ln\, c}^{\infty} w_{ln\, L}(z / S_1)\, dz = 1 - F_{ln\, L}(ln\, c / S_1), \quad (3.9)$$

$$\beta = P\left\{ln\, L(\bar{x}) < ln\, c / S_2\right\} = \int_0^{ln\, c} w_{ln\, L}(z / S_2)\, dz = F_{ln\, L}(ln\, c / S_2), \quad (3.10)$$

где $w_{\hat{L}}(z/S_1), F_{\hat{L}}(z/S_1), w_{\hat{L}}(z/S_2), F_{\hat{L}}(z/S_2) \left(w_{ln\hat{L}}(z/S_1), F_{ln\hat{L}}(z/S_1) w_{ln\hat{L}}(z/S_2), F_{ln\hat{L}}(z/S_2)\right)$ – соответственно плотности вероятности и функции распределения статистики отношения правдоподобия $\hat{L}$ (или его логарифма $\ln \hat{L}$) при наличии классов S1 и S2.

Таким образом вероятности ошибок распознавания аналитически выражаются через объемы обучающей и контрольной выборок, размерность признакового пространства (размерность вектора $\bar{x}$) и межклассовые расстояния (в отношении правдоподобия), что позволяет выбрать параметры, гарантируя достоверность распознавания и используя всю информацию о классах, содержащуюся в измерениях.

Важнейшей особенностью реальных систем распознавания, которая практически не учитывается в других рассмотренных (кроме статистического) детерминистских (перцептронных), синтаксических (лингвистических), логических, в том числе с использованием АВО, алгебраических системах распознавания, является то, что наблюдения $\{\bar{x}_1, \ldots, \bar{x}_n\}$ неизбежно подвержены многочисленным случайным возмущениям, непредсказуемый, вероятностный характер которых проявляется на всех этапах, начиная с процесса получения самих наблюдений и кончая процессом принятия решения, который всегда является случайным. Дестабилизирующие факторы выступают в распознавании как погрешности измерительных приборов, неточности регистрации, шумы в каналах связи при передаче данных измерений, аппаратурные шумы, наконец, как ошибки округления при вычислениях, являющиеся следствием ограниченности разрядной сетки ЭВМ. Взаимодействуя между собой, указанные возмущения приводят к тому, что наблюдения неизбежно оказываются реализациями случайных величин. Отсюда видно, что разработка адекватных исследуемым

процессам методов распознавания неизбежно связана с исследованием случайных отображений, что оказывается возможным только на основе статистических методов. Следовательно, только статистические методы распознавания [1, 2] позволяют в полной мере отразить тонкую структуру и все особенности проявления распознаваемых объектов через описывающие их признаки как при обучении, так и при принятии решений с учетом всех дестабилизирующих факторов, и количественно описать указанные процессы, используя хорошо развитые методы математической статистики. Это создает основу для количественного выражения основных параметров распознающего процесса – размерности признакового пространства, времени обучения и принятия решения через главный показатель качества системы – достоверность распознавания, что в свою очередь позволяет реализовать в системах статистического распознавания гарантированную достоверность распознавания.

Таким образом, проведенное сопоставление наиболее распространенных методов распознавания с точки зрения гарантированной достоверности распознавания показывает, что единственным методом, обеспечивающим полное адекватное описание исследуемых объектов с учетом всех дестабилизирующих факторов и на этой основе позволяющим количественно выразить главный показатель качества – достоверность распознавания – через все основные параметры распознающей системы: время обучения и распознавания, размерность признакового пространства и межклассовые расстояния, является статистический метод распознавания, на основе которого и будет строиться все дальнейшее изложение.

4. Формирование признакового пространства

Выбор показателей состояния фирмы. Анализ состояния фирмы – это не только исследование процессов, происходящих в структуре самой фирмы, это, прежде всего, анализ той среды, в которой функционирует фирма. Обычно разделяют внутреннюю среду фирмы, факторы которой определяются целиком управленческими решениями (цели фирмы, ее структура, люди, технологии), и маркетинговую (внешнюю). Маркетинговую среду представляют факторы микро – и макросреды. “Микросреда представлена силами, имеющими непосредственное отношение к самой фирме и ее возможностям по обслуживанию клиентуры, т.е. поставщиками, маркетинговыми посредниками, клиентами, конкурентами и контактными аудиториями. Макросреда представлена силами более широкого социального плана, которые оказывают влияние на микросреду, такими, как факторы демографического, экономического, природного, технического, политического и культурного характера” [17]. Факторы микросреды часто называют факторами прямого воздействия на фирму, а факторы

макросреды – косвенного воздействия. Таким образом фирма определяет внутрифирменную среду, отрасль экономики определяет микросреду, рынок в широком смысле слова – макросреду.

При формировании признакового пространства мы сделаем одно допущение: в качестве обучающих фирм количества 2m (см. разделы 1, 2) мы имеем право взять любые фирмы, но возьмем те, которые работают в той же отрасли, что и фирма, состояние которой мы диагностируем. Это избавит нас от необходимости учитывать и факторы макросреды [17]. Факторы макросреды прямо влияют на микросреду, обуславливая, таким образом, межотраслевое различие на уровне макросреды. В пределах же одной отрасли и одного региона влияние надотраслевых факторов инвариантно относительно фирм, осуществляющих свой бизнес в этой отрасли и на этой территории. Поэтому, взяв все: обучающие и интересующую нас фирму в пределах одной отрасли, мы можем избавиться от сложной задачи анализа самой отрасли (возможностей и преимуществ отрасли, ее недостатков и слабостей и т.д.), нас интересуют только фирмы. Если бы мы взяли в качестве обучающих фирмы из разных отраслей, то учет факторов макросреды был бы обязателен, так как потребовалось бы сравнивать признаки, характеризующие разницу не только между фирмами, но и между отраслями, в которых они работают. Точно так же, поступают и менеджеры, сравнивая результаты деятельности своей фирмы с соответствующими показателями фирм-конкурентов. Таким образом, влияние макросреды при таком подходе будет учтено, но не прямо, а опосредованно, через внутриотраслевые факторы (например, мы не станем рассматривать факторы миграции, но при изменении этого фактора в сторону увеличения населения изменится предложение рабочей силы и число потенциальных потребителей для отрасли также в сторону увеличения, что отразится на наших признаках – значит и влияние макросреды будет учтено).

Особенность методов математического моделирования вообще заключается в том, что возможна обработка информации только тогда, когда она состоит из показателей, выраженных количественно. Если явления физики и техники на сегодняшний день практически полностью описываются количественно, то в экономике существует целый ряд процессов, который удачно перевести на язык формул пока не удается. Однако, к счастью, эти факторы (иррационального свойства) в экономической жизни не являются определяющими, а зачастую и прямо зависят от таких факторов, как производство, финансы, которые количественно могут быть описаны, особенно статистическими методами.

Исходя из этих ограничений здесь разработан условный перечень показателей, который разделен на три области, определяющие деятельность фирмы: финансовая деятельность, маркетинг, производство. Все они – неотъемлемая составляющая жизни фирмы и от

успеха в этих областях больше, чем от каких-либо других факторов, зависит успех фирмы в целом. Каждая из этих областей деятельности тесно связана с другими, и серьезный провал хотя бы в одной из них влечет за собой кризис и в остальных, а значит, и кризис всей фирмы.

Примерный перечень показателей экономической деятельности фирмы:

1. Финансовая деятельность:	Чистая прибыль
	Балансовая прибыль
	Балансовая прибыль от продаж
	Балансовая прибыль на инвестированный капитал
	Балансовая прибыль от акций
	Балансовая прибыль на основные средства
	Прирост прибыли балансовой чистой
	Прирост дохода в расчете на акцию
	Коэффициент абсолютной ликвидности
	Коэффициент покрытия
	Оборотный капитал
	Коэффициент маневренности
	Коэффициент финансовой устойчивости
	Коэффициент финансирования
	Коэффициент инвестирования
	Коэффициент общей оборачиваемости
	Рентабельность собственного капитала по балансовой прибыли
	Рентабельность собственного капитала по частной прибыли
	Рентабельность инвестиций
	Рентабельность всех операций по балансовой прибыли
	Рентабельность всех операций по частной прибыли
	Чистый доход
2.Маркетинг:	Общий объем продаж
	Доля рынка в отрасли
	Прирост объема продаж
	Прирост доли рынка
	Расходы на рекламу и пропаганду
	Цена товара A
	Цена товара B
	... ...
	... ...
	Цена товара N
	Прирост цены товара A
	Прирост цены товара B
	... ...
	... ...
	Прирост цены товара N
	Расходы на пред- и послепродажное обслуживание клиентов

3.Производство:	Валовые издержки производства Объем производства в ценах соответствующего периода Издержки по выплате зарплат и премий постоянному штату Издержки хранения готовой продукции Издержки хранения полуфабрикатов и материала Издержки по транспортировке материалов и полуфабрикатов Издержки по транспортировке готовой продукции Издержки ведения портфеля отложенных заказов Издержки по найму, оформлению и увольнению рабочих Издержки по использованию труда подрядчиков Расходы на амортизацию основных средств Расходы на коммунальные услуги по содержанию основных средств Фонды отдачи Фондовооруженность труда Производительность труда Индекс снижения себестоимости продукции

Много ли мы теряем от невозможности использовать качественные показатели (хотя бы даже применительно к нашей задаче)? В принципе немного – количественные показатели всегда имеют приоритет над качественными. Подавляющее большинство условных оценок, таких как «заметный», «высокий», «сильный» основывается на результатах сравнительного анализа количественных показателей. Чаще это бывает сравнительный анализ показателей деятельности некоторого количества фирм, когда определяется некий усредненный вариант по совокупности показателей фирмы, вокруг которого ранжируются оценки показателей разных фирм. Сравнительный анализ может проводиться и при сопоставлении показателей фирмы с общеотраслевыми показателями. И совершенно очевидно, что правило приоритета количественно выраженных величин над субъективными факторами выполняется, когда, например: мотивация работников определяется исходя из среднеотраслевой зарплаты, авторитет фирмы определяется из сравнительного анализа среднеотраслевого объема продаж и показателя объема продаж данной фирмы, предпочтения клиентов определяются путем сравнения средней цены по отрасли и цены товара, установленной рассматриваемой фирмой. Количественные показатели помогают нам устранить отрицательное влияние фактора неопределенности, что практически не могут сделать субъективные категории. Количественные показатели можно сопоставлять, строить на их основе прогнозы, агрегировать новые показатели, а качественные обладают лишь хорошей наглядностью, не более того. Специфика работы менеджеров, правда, пока не позволяет им полностью пренебречь качественными факторами, так как многие процессы (взаимодействие и иерархия полномочий,

лидерство и групповые отношения, социокультурные ценности и этика, массовая психология) приходится брать в расчет, несмотря на то, что эти явления пока слабо описываются математическими категориями. Но предпочтение иррациональных факторов рациональным было бы большой ошибкой, поскольку привело бы к неоправданному риску, который конечно же выше, чем риск неправильно истолковать цифры.

Формирование признакового пространства. Первоначальный ансамбль признаков Y = (Y1, Y2, ... , Yq) формируется из числа доступных измерению характеристик распознаваемых объектов таким образом, чтобы наиболее полно и всесторонне отразить все существенные для распознавания свойства. Однако увеличение размерности признакового пространства повышает вычислительную сложность распознающей процедуры и общие затраты на измерение характеристик объектов, т. е. на получение необходимого числа наблюдений. Поскольку время обучения и в особенности принятия решения, как правило, ограничено, повышение размерности признакового пространства может оказаться единственным способом увеличения достоверности. Следовательно, требования к размерности признакового пространства с точки зрения повышения достоверности распознавания и минимизации затрат на получение наблюдений (измерений) являются, как уже подчеркивалось в разделе 2, противоречивыми. Отсюда вытекает большая важность проблемы оптимизации размерности признакового пространства, которая может быть сформулирована как замена первоначального набора q признаков Y = (Y1, Y2, ... , Yq] таким набором X = (X1, X2,..., Xp] (где p – новое число признаков), который минимизирует некоторый критерий J(X).

Наиболее распространенными способами формирования ансамбля признаков X являются селекция (выбор) признаков из исходного набора Y = (Y1, Y2, ... , Yq].

X=(X1,X2,...,Xp)= $=\left(Y_{i_1}, Y_{i_2}, ..., Y_{i_p}\right), p \le q; 1 \le i_i \le q; j = 1, ..., p, i_j \ne i_k$ и выделение признаков, т. е. проведение ортогонального линейного преобразования исходного пространства признаков Y = (Y1, Y2, ... , Yq] в новое пространство X = (X1, X2, ..., Xp).

$$X = AY \qquad (4.1)$$

Преобразование (4.1), как правило, является декоррелирующим, поэтому в качестве столбцов матрицы преобразования A выбирают собственные векторы общей ковариационной матрицы M распознаваемых совокупностей. Сама ковариационная матрица M* в этом случае становится диагональной с собственными числами λ_i на диагонали

$$M^* = ATMA = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ . & . & . & . \\ 0 & 0 & \dots & \lambda_p \end{pmatrix}. \qquad (4.2)$$

После указанного преобразования отбирают p (p < q) новых признаков, соответствующих тем собственным числам λ_i матрицы M*, которые оказывают набольшее влияние на значение выбранного критерия J (X).

При выборе признаков методом минимизации внутриклассового разброса наблюдений критерий

$$J = \operatorname{tr} M^* = \sum_{j=1}^{p} \lambda_j . \qquad (4.3)$$

Здесь вместо q характеристик или старых признаков Y1, Y2, ..., Yq выбирают p новых признаков X1, X2,..., Xp, p < q , которые соответствуют минимальным собственным числам λ_j.

Другим критерием, который может быть использован при формировании новых признаков, является критерий наилучшей аппроксимации межклассового расстояния. Необходимо выбрать p новых признаков, соответствующих, в отличие от предыдущего случая, максимальным собственным числам λ_j, так, чтобы значение J сократилось не более чем на заданное или на минимально возможное значение.

Последний критерий наилучшей аппроксимации можно, вообще говоря, применить непосредственно к выражению (4.3) для внутриклассового разброса, но в этом случае новые признаки должны уже соответствовать не минимальным, а максимальным собственным числам в сумме (4.3). Таким образом, различные критерии могут приводить к противоположным по смыслу рекомендациям по выбору признаков.

Существует усовершенствованный критерий, объединяющий оба предыдущих. Сущность его состоит в совместной минимизации внутриклассового разброса наблюдений и максимизации межклассового расстояния. Этот критерий принят в дискриминантном анализе, в котором наблюдения разных классов проектируются на заданное пространство в пространстве признаков Y1, Y2, ..., Yq (например, на прямую линию) таким образом, чтобы расстояние между центрами классов стало максимальным, а разброс наблюдений внутри каждого класса – минимальным. Количественным выражением критерия является функция от матриц T2 и T1 , характеризующих внутриклассовый и межклассовый разбросы наблюдений

$$J = \mathrm{tr}\,(T2\text{-}1\ T1)\ . \qquad (4.4)$$

Формирование признакового пространства по данному критерию производится аналогично описанному ранее с использованием декоррелирующего преобразования путем выбора признаков, соответствующих максимальным собственным значениям λ_j матрицы T2-1 T1 .

С целью масштабирования вклада новых признаков в критерий J можно произвести кластеризацию, т. е. применить к новым признакам X1, X2, ..., Xp преобразование UX , матрица которого диагональна:

$$U = \begin{pmatrix} U_{11} & 0 & \dots & 0 \\ 0 & U_{22} & \dots & 0 \\ . & . & . & . \\ 0 & 0 & \dots & U_{pp} \end{pmatrix}. \qquad (4.5)$$

Отсутствие связи перечисленных методов формирования признакового пространства с основными показателями качества распознавания, в первую очередь с главным из них – достоверностью, приводит, в ряде случаев, как уже указывалось выше, к противоречиям и не позволяет однозначно осуществить оптимальный выбор признаков, гарантирующий достижение требуемой достоверности распознавания.

В общем виде задачу формирования признакового пространства необходимо ставить, исходя из требований к распознающей системе в целом. В реальных условиях обычно требуется, чтобы принимаемые системой решения имели гарантированную достоверность, которая достигалась бы при минимуме используемого оборудования, энергетических затрат, времени обучения системы, времени принятия решений и т. д. Поэтому характеристики достоверности неизбежно должны быть увязаны с количеством обучающих наблюдений, используемых для задания классов, объемом контрольных выборок, необходимых для принятия решений, и размерностью признакового пространства [1, 2] .

Каждое обучающее и контрольное наблюдение, очевидно, требует проведения p актов измерения значений признаков. Поэтому выбор признаков является составной частью минимизации общей размерности задачи распознавания. Если предположить, что трудоемкость задачи распознавания складывается из трудоемкостей задач обучения и принятия решений, то минимизации подлежит уже не просто размерность признакового пространства p, а общее количество наблюдений (измерений) [1]

$$\rho = p(2m + n), \qquad (4.6)$$

где m – объем обучающей выборки

n – объем контрольной выборки.

5. Обучение

Непараметрическое обучение (оценивание неизвестных плотностей вероятностей наблюдений). Источником информации о распознаваемых образах является совокупность результатов независимых наблюдений (выборочных значений), составляющих обучающие (xi(1))1m = (x1(1), x2(1), ... , xm(1)), $\left(\mathbf{x}_i^{(2)}\right)_1^m = \left(\mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)}, ..., \mathbf{x}_m^{(2)}\right)$ и контрольную (экзаменационную) (xi)1n = (x1, x2, ..., xn) выборки. В зависимости от характера задачи распознавания (одномерной или многомерной) xi может быть либо одномерной, либо p-мерной величиной. Основной целью обучения являются преодоление априорной неопределенности о распознаваемых классах S1 и S2 путем использования информации о них, содержащейся в обучающих выборках и построение эталонных описаний классов – оценок условных плотностей вероятностей $\hat{w}_n\left(x_1, x_2, ..., x_m / S_1\right)$ и $\hat{w}_n\left(x_1, x_2, ..., x_m / S_2\right)$.

Решающее значение для выбора метода распознавания имеет вид априорной неопределенности, для преодоления которой используется обучение.

В наиболее общем случае отсутствия априорных сведений не только о параметрах, но и о самом виде закона распределения наблюдаемой совокупности выборочных значений, априорная неопределенность носит название непараметрической [2] , а сами методы распознавания, применяемые в этих условия, именуются непараметрическими. Таким образом случай непараметрического обучения является самым общим и его содержанием является статистическое оценивание неизвестных условных плотностей вероятностей $\hat{w}_n\left(\bar{x}_1, ..., \bar{x}_m / S_i\right), i = \overline{1,2}$ признаков X1, ... , Xp. Наиболее распространенными методами статистического оценивания неизвестных плотностей вероятностей скалярных и векторных наблюдений являются гистограммный, полигональный методы, представление плотности вероятности линейной комбинацией базисных функцийи др.

Гистограммная оценка неизвестной плотности вероятности [7] строится в виде ступенчатой кривой: над каждым отрезком оси абсцисс, изображающим интервал значений наблюдаемой величины (значения признака Xi), строится прямоугольник, площадь которого пропорциональна частоте попаданий наблюдений в этот интервал. При равной ширине интервалов (что обычно и бывает) высоты прямоугольников пропорциональны частотам. Гистограммный метод обобщается и на многомерный случай. Так, для представления двумерного распределения строятся трехмерные фигуры. Горизонтальная плоскость делится на клетки как шахматная доска. В центре клетки восстанавливается перпендикуляр, пропорциональный по своей длине частоте, отвечающей интервалу. На нем строится

прямоугольный параллелепипед, по объему пропорциональный частоте, соответствующей этой клетке. Полученная объемная фигура является двумерной гистограммой.

Полигональные оценки получают путем сглаживания гистограммы, соединяя прямыми линиями крайние левые, средние или крайние правые точки верхних столбиков, в результате получается кусочно-линейные функции (ломаные). Иногда кусочно-линейные функции состоят из отрезков прямых, проведенных с учетом разности высот соседних столбиков. Подобные аппроксимации не обязательно имеют вид обычных ломаных, концы отдельных прямых могут не совпадать, а соединяться вертикальными прямыми. Такой оценкой, в частности, является полигон Смирнова [7]. Полигональные оценки, как и гистограммные, обобщаются и на многомерный случай.

Оценивание плотности вероятности может быть также осуществлено путем представления ее линейной комбинацией базисных функций. Одномерная плотность вероятности w(x) может быть представлена разложением в ряд по базисным ортогональным функциям {Qn(x)} с весовой функцией $\varphi(x)$.

$$w(x) = \varphi(x)\sum_{k=0}^{\infty} C_k Q_k(x) \qquad (5.1)$$

Коэффициенты Ck можно определить, умножив обе части (5.1) на функцию Qn(x) и проинтегрировав с использованием условия ортогональности:

$$\int_{-\infty}^{\infty} \varphi(x) Q_k(x)\, Q_n(x) dx = \delta_{kn} \qquad (5.2),$$

где $\delta_{kn} = \begin{cases} 1 & , k = n \\ 0 & , k \neq n \end{cases}$ – символ Кронекера (5.3).

При этом в сумме все члены, за исключением одного при k = n, равны нулю и, следовательно

$$C_n = \int_{-\infty}^{\infty} w(x) Q_n(x) dx \quad . \qquad (5.4)$$

Если {Qn(x)} совокупность ортогональных номиналов, то

$$Q_n(x) = \sum_{u=0}^{n} a_u x^u \quad , \qquad (5.5)$$

тогда, так как по определению момента mr случайной величины r-ого порядка

$$m_r = \int_{-\infty}^{\infty} w(x) x^r dx \quad , \qquad (5.6)$$

то

$$C_n = \sum_{r=0}^{n} a_r m_r \qquad (5.7)$$

и, следовательно, подставляя значение Cn из (5.7) в (5.1)

$$w(x) = \varphi(x) \sum_{k=0}^{\infty} \sum_{r=0}^{k} Q_k(x) a_r m_r \qquad (5.8)$$

при условии, что моменты mr существуют.

В качестве примера рассмотрим наблюдения ξ с нулевым средним и единичной функцией (т.е. нормированные и центрированные [– переход к произвольным наблюдениям с a и σ^2 дает $w\left(\frac{x-a}{\sigma}\right)/\sigma$, и произведем разложение неизвестной плотности вероятности w(x) в ряд по полиномам Эрмита Hn(x)

$$\mathrm{H}_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}} \quad , n = 0, 1, 2, \ldots \qquad (5.9)$$

Учитывая условия нормировки (5.2), получаем из (5.1) с учетом того, что $\varphi(x) = \left(1/\sqrt{2\pi}\right)\exp\left\{-x^2/2\right\}$ – нормальная плотность вероятности:

$$w(x) = \frac{1}{\sqrt{2\pi}} \; e^{-\frac{x^2}{2}} \sum_{k=0}^{\infty} \frac{C_k}{\sqrt{k!}} H_k(x) \quad , \qquad (5.10)$$

где

$$C_k = \frac{1}{\sqrt{k!}} \int_{-\infty}^{\infty} w_\xi(x) \, , H_k(x) \, dx = \frac{1}{\sqrt{k!}} \, m_1\left\{H_k(\xi)\right\} \quad , \qquad (5.11)$$

причем C0 = 1, а вследствие принятой нормировки случайной величины ξ имеем C1 = C2 = 0.

Используя определение полиномов Эрмита (5.9), можно (5.10) переписать в виде:

$$w_\xi(x) = \varphi(x) + \sum_{k=3}^{\infty} (-1)^k \frac{C_k}{\sqrt{k!}} \varphi^{(k)}(x) \quad , \qquad (5.12)$$

где $\varphi^{(k)}(x)$ – k-я производная нормальной плотности распределения.

Вычислим несколько коэффициентов Ck в ряду (5.12) по формуле (5.11).

$$C_3 = \frac{1}{\sqrt{3!}} \int_{-\infty}^{\infty} \left(x^3 - 3x\right) w_\xi(x)\, dx = \frac{k}{\sqrt{3!}}, \qquad (5.13)$$

$$C_4 = \frac{1}{\sqrt{4!}} \int_{-\infty}^{\infty} \left(x^4 - 6x^2 + 3\right) w_\xi(x) dx = \frac{\gamma}{\sqrt{4!}}, \qquad (5.14)$$

где k и γ – соответственно коэффициенты асимметрии и эксцесса

$$k = \frac{\mu_3}{\sqrt{\mu_2{}^3}}, \qquad (5.15)$$

$$\gamma = \frac{\mu_4}{\mu_2{}^2} - 3, \qquad (5.16)$$

а μ_2, μ_3, μ_4 – соответственно второй, третий, четвертый центральный моменты распределения, в качестве которых можно использовать выборочные центральные моменты μ_2, μ_3, μ_4, полученные по выборке наблюдений. Подставляя (5.13) и (5.14) в (5.12), получаем в результате приближенную оценку неизвестной плотности вероятности $w_\xi(x)$:

$$w_\xi(x) = \varphi(x) - \frac{k}{3!}\varphi^{(3)}(x) + \frac{\gamma}{4!}\varphi^{(4)}(x) - \ldots \qquad (5.17)$$

Оценивание неизвестной плотности вероятности линейной комбинацией базисных функций обобщается и на многомерные плотности вероятности. Однако, в этом случае, отыскание универсальной системы базисных функций и вычисление коэффициентов разложения становится трудной задачей. Одним из методов нахождения коэффициентов разложения является метод последовательных приближений, известный под названием метода потенциальных функций.

Существуют и другие методы оценивания плотности вероятности, в том числе метод Парзена и метод k – ближайших соседей, основанные на суммировании наблюдений с некоторыми весовыми функциями, называемыми обычно ядром и выбираемыми таким образом, чтобы возможно больше “размазать” столбики и в итоге получить более гладкую аппроксимацию неизвестной плотности вероятности.

Представляется более обоснованным исходить при оценивании плотности вероятности $w(\bar{x})$ из ее определения как производной от функции распределения F(x) с использованием известных методов численного дифференцирования [2, 8]. В настоящее время достаточно хорошо развиты методы оценивания функций распределения эмпирическими ступенчатыми функциями, определена точность оценивания для конечных объемов выборок при различных способах задания расстояния между F(x) и ее оценкой $\hat{F}(x)$ [2].

Нахождение производной представляет собой линейную операцию с последующим переходом к пределу. Ввиду этого можно попытаться построить линейную комбинацию значений эмпирической функции,

которая при асимптотическом росте объемов обучающих выборок $m \to \infty$ сходилась бы по вероятности к F(x), а при конечных фиксированных m позволяла оценить погрешность аппроксимации плотности по известным характеристикам эмпирической функции распределения. При этом мы можем пользоваться значениями эмпирической функции распределения $\hat{F}(x)$ во всех точках x области ее определения. Если раньше при оценивании плотности мы могли использовать в формулах только конечное множество обучающих наблюдений, то теперь получаем в свое распоряжение бесконечную выборку значений $\hat{F}(x)$. Таким путем ликвидируется основной источник трудностей непараметрического оценивания плотности вероятности w(x): в отличие от $\{\hat{w}(x)\}$ множество исходных данных $\{\hat{F}(x)\}$ становится равномощным множеству значений оцениваемой функции распределения F(x).

Пусть для обучения используются одномерные (p = 1) классифицированные обучающие выборки $(x_1^{(1)},...,x_m^{(1)})$ и $(x_1^{(2)},...,x_m^{(2)})$.

Для получения оценок условных функций распределения $\hat{F}(x/S_1)$ и $\hat{F}(x/S_2)$ рассмотрим одномерные условные случайные процессы $\xi^{(1)}(t)$ и $\xi^{(2)}(t)$, относительно которых мы будем полагать (в некоторых случаях, быть может, с определенным приближением), что они удовлетворяют условию эргодичности [2,3]. Эти процессы представляют собой случайные изменения во времени значений признака X при условии, что наблюдаемая совокупность принадлежит одному из классов соответственно S1 или S2. Тогда отношение суммарного времени $\sum_k t_k$ пребывания реализации случайного процесса ξ(t) под некоторым уровнем x к длительности реализации T ($0 < T < \infty$).

$$\hat{F}(x)=\frac{1}{T}\sum_k t_k, \qquad (5.18)$$

(здесь tk – длительность k-го выброса ξ(t) под уровнем x) может рассматриваться как оценка $\hat{F}(x)$ функции распределения F(x) случайного процесса ξ(t), которая является несмещенной и состоятельной [2].

Пусть теперь для обучения используются p-мерные классифицированные обучающие выборки,

$$(\bar{x}_1^{(1)},...,\bar{x}_m^{(1)})=\begin{pmatrix} x_{11}^{(1)}, & ..., & x_{1m}^{(1)} \\ ... & ... & ... \\ x_{p1}^{(1)}, & ..., & x_{pm}^{(1)} \end{pmatrix} \text{ и } (\bar{x}_1^{(2)},...,\bar{x}_m^{(2)})=\begin{pmatrix} x_{11}^{(2)}, & ..., & x_{1m}^{(2)} \\ ... & ... & ... \\ x_{p1}^{(2)}, & ..., & x_{pm}^{(2)} \end{pmatrix}.$$

Для получения оценок условных многомерных функций распределения $\hat{F}_p(x_1,x_2,...,x_p/S_1)$ и $\hat{F}_p(x_1,x_2,...,x_p/S_2)$ рассмотрим векторные p - мерные условные случайные процессы $\bar{\xi}^{(1)}{}_p(t)$ и $\xi^{(2)}{}_p(t)$, которые, как и

в рассмотренном выше одномерном случае, мы будем полагать, хотя бы приближенно, эргодическими [2, 3].

Тогда отношение суммарного времени $\sum_k t_k$ пребывания реализации p-мерного векторного случайного процесса $\overline{\xi}_p(t)$ внутри области, ограниченной некоторой гиперплоскостью Q(x1, x2, ..., xp) (т.е. пребывания процесса ξ1(t) ниже уровня x1, процесса ξ2(t) ниже x2, ... , процесса ξp(t) ниже xp) к длительности реализации T(0<T<∞),

$$\hat{F}_p(x_1,x_2,...,x_p)=\frac{1}{T}\sum_k t_k \qquad (5.19)$$

может рассматриваться как оценка $\hat{F}_p(x_1,x_2,...,x_p)$ p-мерной функции распределения случайного процесса $\overline{\xi}_p(t)$, которая является несмещенной и состоятельной [2].

Функции распределения $\hat{F}(\overline{x}/S_k)=\hat{F}_k(\overline{x}), k=1,2,$ содержат всю информацию о классах образов. Поэтому наиболее естественным было бы построение правила принятия решения на основе эмпирических функций распределения $\hat{F}_k(\overline{x})$. Указанные функции концентрируют всю информацию, содержащуюся в обучающих выборках $\{\overline{X}_m^{(k)}\}$, и позволяют оценивать точность аппроксимации функций Fk(x) при любых объемах m. Однако на сегодняшний день все правила принятия решений в статистическом распознавании строятся с использованием плотностей вероятностей.

Для приведения в соответствие структуры решающего правила, использующего плотности вероятностей признаков, и вида исходных данных, представленных выражениями эмпирических функций распределения, необходимо по эмпирическим функциям распределения $\hat{F}_k(x)$ сформировать оценки плотностей $\hat{w}_k(x)$ и подставить их в решающее правило вместо априорно неизвестных плотностей wk(x). При этом требуется, чтобы оценки $\hat{w}_k(x)$ сходились к априорно неизвестным плотностям wk(x) при асимптотическом увеличении объемов обучающих выборок. В результате мы приходим к двухэтапной процедуре обучения. На первом этапе по обучающим выборкам строятся эмпирические функции распределения для всех классов образов. На втором этапе по эмпирическим функциям распределения формируют оценки плотностей вероятностей с использованием известных методов численного дифференцирования [8] и известных соотношений:

$$w(x)=dF(x)/dx, F(x)=\int_{-\infty}^{x} w(y)dy \qquad (5.20)$$

$$w_p(x_1,x_2,...,x_p)=\frac{\partial^p F_p(x_1,x_2,...,x_p)}{\partial x_1 \partial x_2 ... \partial x_p}, \qquad (5.21)$$

$$F_p(x_1,x_2,...,x_p)=\int_{-\infty}^{x_1}...\int_{-\infty}^{x_p} w_p(y_1,y_2,...,y_p)dy_1dy_2...dy_p,$$

В качестве примера можно привести полученную с использованием методов численного дифференцирования оценку Розенблатта [2] $\hat{\omega}(x)$ одномерной плотности вероятности ω(x)

$$\hat{\omega}(x)=\hat{F}(x+h)-\hat{F}(x-h)/2h,$$

где h > 0 – некоторый параметр.

ПРИБЛИЖЕННЫЙ МЕТОД СВЕДЕНИЯ НЕПАРАМЕТРИЧЕСКОЙ АПРИОРНОЙ НЕОПРЕДЕЛЕННОСТИ К ПАРАМЕТРИЧЕСКОЙ. Если в результате предварительного анализа наблюдаемой совокупности выборочных значений можно хотя бы с некоторым приближением установить вид закона их распределения, то априорная неопределенность относится лишь к параметрам этого распределения, так что целью обучения в этом случае становится получение оценок этих параметров. Подобная априорная неопределенность носит название параметрической, а методы распознавания, применяемые в этих условиях, именуются параметрическими. Хотя с формальной точки зрения закон распределения выборочных значений может быть произвольным, на практике в параметрическом распознавании почти всегда используется нормальный закон. Дело в том, что если при распознавании одномерных совокупностей их распределение всегда может быть описано одним (например, нормальным, биноминальным, экспоненциальным, пуассоновским и др.) законом, то при распознавании многомерных совокупностей каждая компонента вектора выборочных значений (т. е. наблюдаемые значения каждого признака) может иметь свой отличный от других компонент закон распределения (что не может рассматриваться как аномалия, поскольку сам ансамбль признаков формируется, таким образом, чтобы возможно полнее охарактеризовать различные свойства распознаваемых явлений). Но тогда многомерное совместное распределение совокупности выборочных значений должно описываться некоторым многомерным законом, включающим в себя компоненты с различными законами распределения.

В литературе аналитические выражения подобных «разнокомпонентных» законов отсутствуют. К этому следует добавить, что, как указано в [8], «современный уровень знаний таков, что пока точному многомерному анализу, за редкими исключениями, поддаются лишь задачи, где рассматривается нормальный случай» и, следовательно, как указывается в [9], «почти все выводы многомерной статистики опираются на предположения о нормальности рассматриваемых распределений». Отсюда следует, что на сегодняшний день параметрические методы распознавания, по-существу, являются методами распознавания нормально распределенных совокупностей, так что задачей параметрического обучения в этих условиях является

оценивание параметров (средних, дисперсий, ковариационных матриц) нормальных плотностей вероятностей, используемых в решающем правиле.

Большие вычислительные сложности и трудности математического порядка, связанные с вычисление непараметрической оценки плотностей вероятностей делает целесообразными попытки сведения непараметрической априорной неопределенности к параметрической.

Если для обеспечения достоверности, равной $1 - \alpha = 0{,}9$ требуемая сумма объемов обучающей и контрольной выборки составляет при расстоянии между совокупностями $\alpha\varepsilon = 0{,}1$ $m + n = 2200$, то при применении непараметрического подхода для достижения такой же достоверности необходимо располагать объемом $m + n = 9000 \div 11600$, т. е. $\approx$ в 5 раз больше [1]. Следовательно, если бы мы смогли ограничивать затраты на переход от непараметрической к параметрической неопределенности 5-кратным увеличением выборок, это было бы вполне оправданно.

Пусть $\xi 1, ..., \xi q$ – последовательность независимых одинаково распределенных случайных величин, имеющих конечные средние $m1\{\xi k\} = a$ и дисперсии $\mu 2\{\xi k\} = \sigma 2$. Тогда последовательность нормированных и центрированных сумм

$$\eta_q = \frac{1}{\sigma\sqrt{q}} \sum_{k=1}^{q} (\xi_k - a) \qquad (5.22)$$

сходится по распределению к стандартной гауссовской нормальной величине, что равносильно утверждению

$$\lim_{q\to\infty} P\left\{ \frac{1}{\sigma\sqrt{q}} \sum_{k=1}^{q} (\xi_k - a) \le x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right) du = F(x), \qquad (5.23)$$

т. е. последовательность функций распределения сумм ηq независимых одинаково распределенных случайных величин ξk при $q \to \infty$ сходится к гауссовской (нормальной) функции распределения с параметрами (0; 1). Эта формула является аналитическим выражением центральной предельной теоремы теории вероятностей, которая легко обобщается на многомерный случай. Пусть $\overline{\xi}_1, ..., \overline{\xi}_q$ – последовательность p-мерных независимых векторных случайных величин с одинаковыми p-мерными функциями распределения, компоненты которых могут быть распределенными по разным законам (разнораспределенными) с вектором средних $\overline{a}$ и ковариационными матрицами $\overline{M}$. Тогда последовательность p-мерных функций распределения сумм

$$\overline{\eta}_q = \frac{1}{\sqrt{q}} \sum_{k=1}^{q} (\overline{\xi}_k - \overline{a}) \qquad (5.24)$$

при $q \to \infty$ сходится к p-мерной гауссовской (нормальной) функции распределения с нулевым вектором средних и ковариационной матрицей M.

Приближенное выражение плотности распределения суммы ηq с точностью до малых порядка O (1/q3/2) получается с использованием оценки (5.17) [1]:

$$W_{\eta_q}(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\left[1+\frac{k}{6\sqrt{q}}H_3(x)+\frac{\gamma}{24q}H_4(x)+\frac{k^2}{72q}H_6(x)+\cdots\right], \quad (5.25)$$

где κ и γ – коэффициенты асимметрии и эксцесса, вычисляемые по формулам (5.15) и (5.16), а Нк(х) – полиномы Эрмита (см. формулу (5.9)).

Как видно из анализа формулы (5.25) приближенное выражение плотности вероятности суммы ηq представляет собой очень быстро сходящийся ряд, что свидетельствует о том, что приближенная нормализация суммы ηq наступает уже при достаточно малых значениях q, в особенности, если исходное распределение Wξ(x) является симметричным (в этом случае коэффициент асимметрии k, как известно, равен нулю и, следовательно вносящие ощутимый вклад в значение ω_{η_q} второй и четвертый члены суммы исчезают).

Детальный анализ влияния числа q членов суммы ηq на скорость ее нормализации осуществлен в работе [1], в результате чего установлено, что для наиболее распространенных в практических приложениях законов распределения приближенная нормализация суммы ηq наступает при q = 3 ÷ 5. Рассмотрим два наиболее сильно отличающихся как от нормального закона, так и друг от друга закона распределения:

равномерный

$$\omega_p(\xi)=\begin{cases} b^{-1} \text{ при } 0\le\xi\le b; \\ 0 \quad \text{при } \xi<0, \xi>b, \end{cases} \quad (5.26)$$

и экспоненциальный

$$\omega_э(\xi)=\begin{cases} \rho^{-1}\exp(-\xi/\rho) \quad \text{при } \xi\ge 0; \\ 0 \qquad\qquad\quad \text{при } \xi<0; \\ \qquad\quad \rho>0. \end{cases} \quad (5.27)$$

На рис. 7 представлена полученная в[1] зависимость уровня нормальности суммарных распределений ω_{η_q} (х) суммы ηq независимых равномерно (точечная кривая) и экспоненциально (сплошная кривая) распределенных случайных величин от количества суммарных членов q. Как видно из рисунка, допустимый уровень нормальности p = 0,85 достигается в случае равномерного распределения уже при q = 2 (это объясняется тем, что оно является симметричным), а в случае чрезвычайно асимметричного (и, следовательно, наиболее трудного для осуществления нормализации) экспоненциального распределения при вполне допустимом в практических приложениях значении q = 5.

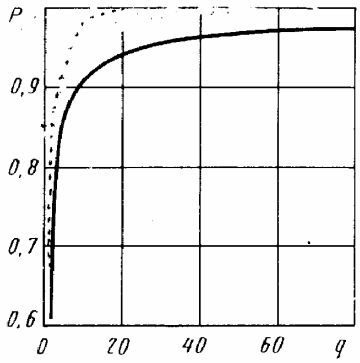

Рис. 7

Таким образом, поставленная нами ранее задача ограничить затраты на переход от непараметрической априорной неопределенности к параметрической пятикратным увеличением объемов выборок оказывается выполнимой, что очень упрощает процедуру обучения. Действительно, образуя в ходе предварительной обработки обучающих и контрольной выборок новые выборки, каждая из которых представляет собой нормированную сумму из пяти исходных скалярных (при p = 1) или векторных (при p >1) наблюдений, мы получаем совокупность новых выборок (скалярных или векторных), которые независимо от вида законов распределения исходных наблюдений всегда приближенно распределены по гауссовскому (нормальному) закону. Это особенно важно в многомерном (векторном) случае, поскольку позволяет, не исследуя конкретные законы распределения каждого признака, которые в большинстве случаев могут отличаться друг от друга, описать совместное распределение полученных новых суммарных наблюдений многомерным гауссовским (нормальным) законом распределения, преодолев к тому же упомянутое выше отсутствие в математической литературе аналитических выражений многомерных законов распределения, включающих в себя компоненты с различными законами распределения («разнокомпонентных»).

При нормальном распределении признака для построения эталонных описаний классов достаточно вычислить выборочные средние и дисперсии по классифицированным обучающим выборкам $\left(x_i^{(k)}\right)_1^m$

$$\hat{a}_k = \frac{1}{m}\sum_{i=1}^{m} x_i^{(k)}, \hat{\sigma}_k^2 = \frac{1}{m}\sum_{i=1}^{m}\left(x_i^{(k)} - \hat{a}_k\right)^2, k = 1,2 \quad , \qquad (5.28)$$

которые представляют собой [2] оценки максимального правдоподобия указанных параметров.

Как известно [1], выборочные среднее и дисперсия (5.28) являются состоятельными оценками среднего и дисперсии, а $\hat{a}_k$ является к тому же и несмещенной оценкой среднего. Для устранения небольшого смещения оценки (5.28) дисперсии достаточно умножить ее на m/(m – 1) и получить следующее выражение выборочной дисперсии:

$$\hat{\sigma}_k^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(x_i^{(k)} - \hat{a}_k\right)^2, \qquad (5.29)$$

которое дает не только состоятельную, но и несмещенную оценку дисперсии нормального распределения [2].

В многомерном случае процесс построения эталонных описаний классов при нормальном распределении совокупности признаков X1, X2, ..., Xp также упрощается, так как вместо весьма громоздких и трудоемких процедур формирования оценок условных p-мерных плотностей вероятности достаточно лишь вычислить по выборке $\left(\overline{x}_i^{(k)}\right)_1^m$ из sk выборочный вектор средних

$$\hat{\overline{a}}_k = \frac{1}{m}\sum_{i=1}^{m}\overline{x}_i^{(k)},\ x_i^T = \left(x_{1i}, x_{2i}, \ldots, x_{pi}\right),\ k = 1, 2 \qquad (5.30)$$

и выборочную ковариационную матрицу

$$\hat{\overline{M}}_k = \frac{1}{m}\sum_{i=1}^{m}\left(\overline{x}_i^{(k)} - \hat{\overline{a}}_k\right)\left(\overline{x}_i^{(k)} - \hat{\overline{a}}_k\right)^T, \qquad (5.31)$$

которые являются оценками максимального правдоподобия вектора средних $\overline{a}_k$ и ковариационной матрицы $\overline{M}_k$ рассматриваемой нормальной совокупности [2]. Здесь и далее знак Т означает операцию транспонирования.

Выборочные вектор средних (5.30) и ковариационная матрица (5.31) являются состоятельными оценками, причем (5.30), кроме того, несмещенная, тогда как оценка (5.31) ковариационной матрицы смещенная, поскольку ее среднее значение

$$m_1\left\{\hat{\overline{M}}_k\right\} = \left[(m-1)/m\right]\hat{\overline{M}}_k. \qquad (5.32)$$

Несмещенная оценка ковариационной матрицы получается умножением (5.31) на m /(m — 1):

$$\hat{\overline{M}}_k = \frac{1}{m-1}\sum_{i=1}^{m}\left(\overline{x}_i^{(k)} - \hat{\overline{a}}_k\right)\left(\overline{x}_i^{(k)} - \hat{\overline{a}}_k\right)^T,\ k = 1, 2 \qquad (5.33)$$

6. Принятие решений

Выбор оптимального решающего правила, позволяющего наилучшим образом относить контрольную выборку наблюдений к одному из взаимоисключающих классов s1 и s2, производится в соответствии с теорией статистических решений [2, 8] с использованием характеристик, полученных в процессе обучения. В рамках этой теории все виды решающих правил основаны на формировании отношения правдоподобия L (или его логарифма ln L) и его сравнении с определенным порогом c, (или ln c) (значение которого определяется выбранным критерием качества [2])

$$L = \frac{\omega_n(x_1, x_2, \ldots, x_n \mid s_2)}{\omega_n(x_1, x_2, \ldots, x_n \mid s_1)} \begin{matrix} \gamma_2 \\ > \\ < \\ \gamma_1 \end{matrix} c, \quad \ln L = \ln \frac{\omega_n(x_1, x_2, \ldots, x_n \mid s_2)}{\omega_n(x_1, x_2, \ldots, x_n \mid s_1)} \begin{matrix} \gamma_2 \\ > \\ < \\ \gamma_1 \end{matrix} \ln c, \qquad (6.1)$$

где ωn (x1, x2, ..., xn | sj) – условная совместная плотность вероятности векторов выборочных значений x1, x2, ..., xn (функция правдоподобия) при условии их принадлежности к классу Sj, j = 1, 2. Однако если в теории статистических решений указанные плотности ωn (x1, x2, ..., xn | sj) являются априорно известными, то в статистическом распознавании они в принципе не известны, вследствие чего в решающее правило подставляются не сами плотности вероятности ωn (x1, x2, ..., xn | sj), а их оценки $\hat{\omega}_n(x_1, x_2, \ldots, x_n \mid s_j)$, получаемые в процессе обучения, поэтому в решающем правиле с порогом с сравнивается уже не само отношение правдоподобия L, а его оценка $\hat{L}$:

$$\hat{L} = \frac{\hat{\omega}_n(x_1, x_2, \ldots, x_n \mid s_2)}{\hat{\omega}_n(x_1, x_2, \ldots, x_n \mid s_1)} \begin{matrix} \gamma_2 \\ > \\ < \\ \gamma_1 \end{matrix} c. \qquad (6.2)$$

При $\hat{L} \geq c$ принимается решение γ2: контрольная выборка принадлежит классу s2, в противном случае (при $\hat{L} < c$) она считается принадлежащей классу s1 и, следовательно, принимается решение γ1.

На практике помимо обучающих выборок иногда имеется и другая дополнительная информация о классах образов, могут выдвигаться различные требования к продолжительности, стоимости обучения и распознавания, достоверности решений и т. д. Дополнительные сведения влияют на выбор порога и способ сравнения оценок отношений правдоподобия с порогами. Так, в ряде случаев известно, что некоторый класс sk чаще предъявляется для распознавания, чем другие. Целесообразно тогда при формировании отношений правдоподобия L придать больший вес функции правдоподобия $\hat{\omega}_n(x_1, x_2, \ldots, x_n \mid s_k)$ по сравнению с другими.

Указанная дополнительная информация учитывается путем выбора наиболее подходящего решающего правила из имеющегося в теории статистических решений широкого ассортимента критериев: байесовского, Неймана-Пирсона, минимаксного, Вальда, максимума апостериорной вероятности, максимального правдоподобия и др.

В теории статистических решений полный комплект априорных данных включает в себя априорные вероятности P1 = P(S1) и P2 = P(S2) классов S1 и S2 и матрицу потерь (платежную матрицу) П:

$$П = \begin{pmatrix} П_{11} & П_{12} \\ П_{21} & П_{22} \end{pmatrix}, \qquad (6.3)$$

где Пkl – потери от принятия решения о том, что имеет место класс k, тогда как на самом деле имеет место класс l (k, l = 1, 2).

В качестве ориентировочных прикидочных значений априорных вероятностей P1 и P2 классов S1 и S2 можно в первом приближении принять к примеру соответственно отношение числа процветающих p и убыточных r фирм к общему числу p + r рассматриваемых фирм в отрасли. При рассмотрении функций потерь Пkl можно учесть то очевидное обстоятельство, что потери П12 от принятия решения γ1 о том, что фирма находится в процветающем состоянии, тогда как на самом деле имеет место класс S2 – фирма убыточна (кризис), должны быть приняты намного большими, чем потери П21 от принятия решения о том, что фирма убыточна (кризис), тогда как на самом деле имеет место класс S1 – фирма находится в процветающем состоянии.

Байесовский алгоритм. При наличии полного комплекта данных: априорных вероятностей классов p1 и p2 и матрицы П (6.3) по определению среднего значения дискретной случайной величины

$$m_1 = \sum_r x_r p_r \qquad (6.4)$$

можно записать общее выражение среднего риска [2]

$$R = \sum_{j=1}^{2} \sum_{k=1}^{2} П_{jk} P\{\gamma_k \cap S_j\}, \qquad (6.5)$$

где $P\{\gamma_k \cap S_j\} = P\{\gamma_k = П_{kj}\}$ – совместная вероятность принятия решения γk, тогда как на самом деле имел место класс Sj.

С учетом правила умножения вероятностей

$$P\{\gamma_k \cap S_j\} = P\{S_j\} P\{\gamma_k \mid S_j\} = P_j P\{\gamma_k \mid S_j\} \qquad (6.6)$$

средний риск R будет иметь следующий вид:

$$R = \sum_{j=1}^{2} \sum_{k=1}^{2} П_{jk} P_j P\{\gamma_k \mid S_j\} =$$

$$P_1 \left[П_{11} P\{\gamma_1 \mid S_1\} + П_{12} P\{\gamma_2 \mid S_1\}\right] + P_2 \left[П_{21} P\{\gamma_1 \mid S_2\} + П_{22} P\{\gamma_2 \mid S_2\}\right] \qquad (6.7)$$

Вероятности ошибок распознавания 1-го рода α (ложных тревог) и 2- го рода β (пропуска цели), т. е. соответственно вероятности α того, что будет принято решение γ2 о том, что фирма убыточна, тогда как на самом деле она находится в процветающем состоянии S1 и вероятности

β (пропуск цели) того, что будет принято решение γ1 о том, что фирма процветает, тогда как на самом деле она убыточна, т. е. находится в состоянии S2, по определению равны (см. также (3.6)):

$$\alpha = P\{\gamma_2 | S_1\} = \int_{X_2} \omega(\bar{x}/S_1) d\bar{x} \quad (6.8)$$

$$P\{\gamma_1 | S_1\} = \int_{X_1} \omega(\bar{x}/S_1) d\bar{x} = 1 - \alpha \quad (6.9)$$

$$\beta = P\{\gamma_1 | S_2\} = \int_{X_1} \omega(\bar{x}/S_2) d\bar{x} \quad (6.10)$$

$$P\{\gamma_2 | S_2\} = 1 - \beta = \int_{X_2} \omega(\bar{x}/S_2) d\bar{x} \quad (6.11)$$

Подставляя (6.8) – (6.11) в (6.7), получаем

$$R = P_1 П_{11} + P_2 П_{21} + P_1(П_{12} - П_{11})\alpha - P_2(П_{21} - П_{22})(1-\beta) = \\ = P_1 П_{11} + P_2 П_{21} - [P_2(П_{21} - П_{22})(1-\beta) - P_1(П_{12} - П_{11})\alpha] \quad (6.12)$$

или, введя обозначения

$$r_1^f = П_{11}(1-\alpha) + П_{12}\alpha \quad (6.13)$$

$$r_2^f = П_{21}\beta + П_{22}(1-\beta), \quad (6.14)$$

выражаем R через r_1^f и r_2^f (для использования при рассмотрении минимаксного алгоритма в последующем материале):

$$R = P_1 r_1^f + P_2 r_2^f \quad (6.15)$$

В качестве критерия оптимальности алгоритма принятия решения принимается минимальное значение среднего риска R (баейсовский критерий). Тот или иной алгоритм определяется выбором области X2 (или ее дополнения в выборочном пространстве X1), что проявляется через величины α и 1 – β. Подставляя значения этих величин из выражений (6.8) и (6.11) в (6.12), получаем

$$R = P_1 П_{11} + P_2 П_{21} - \int_{X_2} [P_2(П_{21} - П_{22})\omega(x|S_2) - P_1(П_{12} - П_{11})\omega(x|S_1)] d\bar{x}, \quad (6.16)$$

где

$$R_0 = P_1 П_{11} + P_2 П_{21} \quad (6.17)$$

неотрицательная известная константа и R ≥ 0.

Обозначим

$$f(x) = P_2(П_{21} - П_{22})\omega(x|S_2) - P_1(П_{12} - П_{11})\omega(x|S_1), \quad (6.18)$$

тогда

$$R = R_0 - \int_{X_2} f(\bar{x}) d\bar{x} \quad . \quad (6.19)$$

Так как для любого подмножества A множества $\bar{X}_2$ при $f(\bar{x}) \geq 0$ $\int_{X_2} f(\bar{x}) d\bar{x} \geq \int_{A} f(\bar{x}) d\bar{x}$ имеет место неравенство , то интеграл в правой части (6.19) достигает максимума тогда и только тогда, когда в область интегрирования включаются все члены выборочного пространства, для

которых подинтегральная функция f(x) неотрицательна. Отсюда следует, что минимальное значение среднего риска достигается при условии, что в область X2 принятия решения γ2 включаются все выборки, для которых функция f(x) из (6.18) неотрицательна, а в область X1 принятия решения γ1 – все выборки, для которых функция f(x) – отрицательна, т. е.

$$P_2(П_{21}-П_{22})\hat{\omega}(\bar{x}\,|\,S_2)-P_1(П_{12}-П_{11})\hat{\omega}(\bar{x}\,|\,S_1)\begin{matrix}\gamma_2\\ >\\ <\\ \gamma_1\end{matrix}0, \qquad (6.20)$$

откуда

$$\hat{L}(\bar{x})=\frac{\hat{\omega}(\bar{x}\,|\,S_2)}{\hat{\omega}(\bar{x}\,|\,S_1)}\begin{matrix}\gamma_2\\ >\\ <\\ \gamma_1\end{matrix}\frac{П_{12}-П_{11}}{П_{21}-П_{22}}\frac{P_1}{P_2}, \qquad (6.21)$$

Если граница между областями X2 и X1 выбраны согласно (6.21), то минимальный средний (байесовский) риск определяется по формуле (6.12), в которой условные вероятности ошибок αБ и βБ вычисляются согласно (6.8) и (6.10), где в интегралах фигурируют области интегрирования X2 и X1, определенные согласно (6.21). Т. е. байесовский алгоритм (6.21) запишется в виде:

$$\hat{L}(\bar{x})\begin{matrix}\gamma_2\\ >\\ <\\ \gamma_1\end{matrix}C_Б=\frac{П_{12}-П_{11}}{П_{21}-П_{22}}\frac{P_1}{P_2}, \qquad (6.22)$$

где

$$\alpha_Б=\int_{X_2}\hat{\omega}(\bar{x}\,|\,S_1)d\bar{x}, \quad \beta_Б=\int_{X1}\hat{\omega}(\bar{x}\,|\,S_2)d\bar{x}. \qquad (6.23)$$

И из (6.12) имеем байесовский риск:

$$R_Б=P_1П_{11}-P_2П_{21}+P_1(П_{12}-П_{11})\alpha_Б-P_2(П_{21}-П_{22})(1-\beta_Б). \qquad (6.24)$$

Минимальная величина R называется байесовским риском, поэтому и правило (6.22) также носит название байесовского.

Поскольку событие $\bar{x}\in X_2$ эквивалентно $\hat{L}(\bar{x})\ge C_Б$, а $\bar{x}\in X_1$ – событию $\hat{L}(\bar{x})<C_Б$, то с учетом (3.7) и (3.8) вероятности ошибок распознавания αБ и βБ можно выразить однократными интегралами от плотности вероятности оценки отношения правдоподобия $\omega_{\hat{L}}(z\,|\,S_i),\ i=\overline{1,2}$:

$$\alpha_Б=P\{\hat{L}(\bar{x})\ge C_Б/S_1\}=\int_{C_Б}^{\infty}\omega_{\hat{L}}(z\,|\,S_1)dz=1-F_{\hat{L}}(C_Б/S_1), \qquad (6.25)$$

$$\beta_{Б} = P\left\{\hat{L}(\bar{x}) < C_{Б}/S_2\right\} = \int_0^{C_{Б}} \omega_{\hat{L}}(z|S_2)dz = F_{\hat{L}}(C_{Б}/S_2), \tag{6.26}$$

Алгоритм максимальной апостериорной вероятности.

Предположим, матрица потерь П неизвестна, т.е. П12=П21= 1, П11=П22 = 0 и, следовательно П = 1.

По формуле Байеса апостериорная вероятность гипотезы Bj при условии отсутствия события A (Bk – полная группа)

$$P\{B_j/A\} = \frac{P\{B_j\}P\{A/B_j\}}{\sum_{k=1}^{n} P\{B_k\}P\{A/B_k\}}. \tag{6.27}$$

Здесь у нас $P_1 = P_1(S_1|\bar{x})$ и $P_2 = P(S_2|\bar{x})$ составляют полную группу: P1+P2=1. Следовательно, по формуле Байеса находим апостериорные вероятности гипотез S1и S2, если в результате наблюдений получена выборка $\bar{x}$:

$$P\{S_1/\bar{x}\} = \frac{P_1\,\omega(\bar{x}|S_1)}{P_2\,\omega(\bar{x}|S_1) + P_2\,\omega(\bar{x}|S_2)}, \tag{6.28}$$

$$P\{S_2/\bar{x}\} = \frac{P_2\,\omega(\bar{x}|S_2)}{P_2\,\omega(\bar{x}|S_1) + P_2\,\omega(\bar{x}|S_2)}, \tag{6.29}$$

откуда:

$$\frac{P(S_2|\bar{x})}{P(S_1|\bar{x})} = \hat{L}(\bar{x})\frac{P_2}{P_1}. \tag{6.30}$$

Теперь устанавливаем правило решения: принимается S2, если $P\{S_2/\bar{x}\} \geq P\{S_1/\bar{x}\}$ и S1, если $P\{S_1/\bar{x}\} > P\{S_2/\bar{x}\}$, тогда

$$\hat{L}(\bar{x}) \begin{matrix} \gamma_2 \\ > \\ < \\ \gamma_1 \end{matrix} \frac{P_2}{P_1}, \tag{6.31}$$

т.е. условие $P\{S_1/\bar{x}\} + P\{S_2/\bar{x}\} = 1$ равносильно принятию той гипотезы, для которой апостериорная вероятность больше 1/2.

С другой стороны, алгоритм максимальной апостериорной вероятности (6.31) можно получить и непосредственно подстановкой в (6.21) значений П12=П21=1 и П11=П22=0.

При этом из (6.24) имеем средний риск RMAB

$$R_{MAB} = P_2 + P_1\alpha_{Б} - P_2(1-\beta_{Б}) = P_1\alpha_{Б} + P_2\beta_{Б}, \tag{6.32}$$

который, как видно из (6.32), равен априорной вероятности ошибочного решения. Следовательно, алгоритм максимальной апостериорной вероятности минимизирует априорную вероятность ошибок, т.е. в длинной последовательности решений обеспечивает максимальную частоту правильных решений.

Минимаксный алгоритм

Если потери $П = \begin{pmatrix} П_{11} & П_{12} \\ П_{21} & П_{22} \end{pmatrix}$ известны, но неизвестны априорные вероятности классов P1 и P2, то принимающий решение опасается, что ему попадется именно тот случай, при котором P1 и P2 таковы, что дают максимум величине минимального байесовского риска. Тогда он предполагает что P1 и P2 распределены наименее благоприятно, и им соответствует максимум байесовского риска – минимаксный риск. Так как события S1 и S2 составляют полную группу, то достаточно определить наименее благоприятное значение P1=P{S1}=P1M, которому соответствует максимум байесовского риска – минимаксный риск.

$$R_M(P_{1M}) = \max_{0<P_1<1} R_Б(P_1) \qquad (6.33)$$

Зависимость байесовского риска RБ(P1) от вероятности P1 изображена на рис. 8.

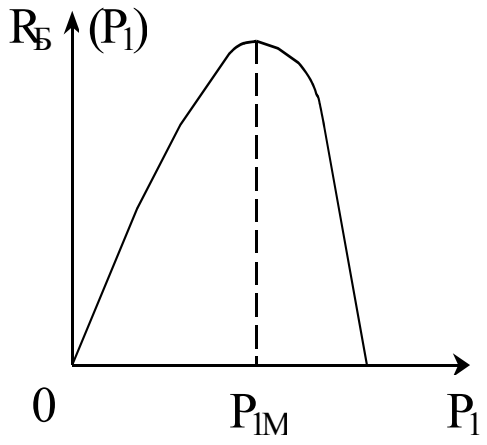

Рис. 8

Уравнение прямой касательной к RБ(P1) в точке P1 имеет вид:

$$У(P_1) = r_2^f + P_1(r_1^f - r_2^f), \qquad (6.34)$$

где r_1^f и r_1^f в соответствии с (6.13) и (6.14) равны

$$r_1^f = П_{11}(1-\alpha_Б) + П_{12}\alpha_Б \qquad (6.35)$$

$$r_2^f = П_{21}\beta_Б + П_{22}(1-\beta_Б). \qquad (6.36)$$

В точке P1=P1M максимума функции RБ(P1) касательная к кривой байесовского риска параллельна оси абсцисс и, следовательно, Y(P1) = const, т. е. не зависит от переменной P1. Согласно (6.34) это условие максимума функции RБ(P1) выполняется, если коэффициент при P1, равен нулю, т.е. значение P1M удовлетворяет уравнению:

$$r_1^f(P_{1M}) = r_2^f(P_{1M}). \qquad (6.37)$$

Следовательно минимаксный риск

$$R_M(P_{1M}) = r_2^f(P_{1M}) = r_1^f(P_{1M}) \quad (6.38)$$

Уравнение для нахождения порога СМ при минимаксном алгоритме находим, подставляя в (6.38) значения r_1^f и r_2^f из (6.35) и (6.36)

$$П_{11}(1-\alpha_M) + П_{12}\alpha_M = П_{21}\beta_M + П_{22}(1-\beta_M) \quad (6.39)$$

или

$$П_{11} + (П_{12} - П_{11})\alpha_M = П_{22} + (П_{21} - П_{22})\beta_M, \quad (6.40)$$

но, как известно (см. (6.25) и (6.26)):

$$\alpha_M = 1 - F_{\hat{L}}(C_M \mid S_1), \quad (6.41)$$

$$\beta_M = F_{\hat{L}}(C_M \mid S_2), \quad (6.42)$$

следовательно, искомое трансцендентное уравнение для определения СМ:

$$П_{11} + (П_{12} - П_{11})[1 - F_{\hat{L}}(C_M \mid S_1)] = П_{22} + (П_{21} - П_{22})[F_{\hat{L}}(C_M \mid S_2)]. \quad (6.43)$$

Алгоритм, оптимальный по критерию Неймана-Пирсона.

При отсутствии данных о потерях П и априорных вероятностях классов Р1 и Р2 может применяться алгоритм Неймана-Пирсона, который обеспечивает минимальную вероятность ошибок β при условии, что вероятность ошибок α не больше заданного значения α0.

Задача синтеза оптимального алгоритма принятия решения по указанному критерию состоит в определении минимума функционала:

$$Ф = \beta + C\alpha, \quad (6.44)$$

в котором вероятность β зависит от правила выбора решения, вероятность α фиксирована, и С – неопределенный множитель Лагранжа. Но сравнивая (6.44) с выражением для среднего риска (6.12)

$$R = P_1 П_{11} + P_2 П_{21} + P_1(П_{12} - П_{11})\alpha - P_2(П_{21} - П_{22})(1-\beta), \quad (6.45)$$

замечаем, что функционал Ф совпадает со средним риском R при Р2=Р1=1/2, П11=П22=0, П21=2, П12=2С (плата за ошибку первого рода α в С раз больше, чем за ошибку 2-го рода β). В этом случае легко убедиться, что последнее выражение для R становится равным Ф:

$$R = \beta + C\alpha = Ф. \quad (6.46)$$

Следовательно, минимум функционала Ф достигается при использовании байесовского алгоритма для Р1=Р2, П11=П22=0, П21=2, П12=2С, тогда он совпадает с минимальным байесовским риском R, определяемым выражением (6.46).

Тогда из (6.21) находим следующий оптимальный по критерию Неймана-Пирсона алгоритм:

$$\hat{L}(\bar{x}) = \frac{\hat{\omega}(\bar{x} \mid S_2)}{\hat{\omega}(\bar{x} \mid S_1)} \begin{matrix} > \\ < \\ \end{matrix} \begin{matrix} \gamma_2 \\ \\ \gamma_1 \end{matrix} \frac{П_{12} - П_{11}}{П_{21} - П_{22}} \frac{P_2}{P_1} = \frac{2c \cdot \frac{1}{2}}{2 \cdot \frac{1}{2}} = C, \quad (6.47)$$

где порог С находится из граничного условия (заданного значения вероятности ошибки 1-го рода α0 (6.25) и (6.26):

$$P\left\{\hat{L}(\bar{x}) \geq C/S_1\right\} = 1 - F_{\hat{L}}(C/S_1) = \alpha_0 \qquad (6.48)$$

Минимальная по критерию Неймана-Пирсона вероятность ошибки 2-го рода β получается из (6.26):

$$\beta = P\left\{\hat{L}(\bar{x}) < C/S_2\right\} = F_{\hat{L}}(C/S_2), \qquad (6.49)$$

где С определяется согласно (6.48).

Последовательный алгоритм Вальда. При последовательном анализе Вальда, применяемом как и в предыдущем алгоритме, при отсутствии данных об априорных вероятностях классов P1 и P2 и потерях П на каждом этапе пространство значений отношения правдоподобия $\hat{L}(\bar{x})$ разделяется на три области: допустимую G1, критическую G2 и промежуточную Gnp. Если значение отношения правдоподобия $\hat{L}(x)$ попадает в промежуточную область Gnp, то делается следующее наблюдение, и так до тех пор, пока при некотором значении n размера выборки это значение $\hat{L}(\bar{x})$ не попадает в одну из областей G1 или G2, после чего принимается решение о наличии класса S1 (при попадании в допустимую область G1) или S2 (при попадании в критическую область G2).

Критерием качества последовательного правила выбора решения обычно является минимум среднего значения размера выборки, необходимой для принятия решения (после чего процедура последовательного анализа завершается) при заданных значениях вероятностей ложной тревоги α и пропуска сигнала β. А. Вальдом показано [10], что среди всех правил выбора решения (в том числе и непоследовательных и, в частности, известных критериев байесовского, максимума апостеорной вероятности, максимума правдоподобия, Неймана-Пирсона, минимаксного), для которых условные вероятности ложной тревоги и пропуска сигнала не превосходят α и β, последовательное правило выбора решения, состоящее в сравнении отношения правдоподобия $L(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)$ с двумя порогами, нижним c1 и верхним c2, приводит к наименьшим средним значениям размера выборок m1{n | S1} (при наличии класса S1) и m1{n | S2} (при наличии класса S2).

Аналитически процедура последовательного анализа может быть выражена следующим образом: при n-м наблюдении принимается решение о наличии класса S1, если

$c_1 < \hat{L}(\bar{x}_1, \ldots, \bar{x}_k) < c_2$, $L(\bar{x}_1, \ldots, \bar{x}_n) \leq c_1$, k=1, …, n-1

и решение о наличии класса S2, если

$c_1 < \hat{L}(\bar{x}_1, \ldots, \bar{x}_k) < c_2$, $L(\bar{x}_1, \ldots, \bar{x}_n) \geq c_2$, k=1, …, n-1 . (6.50)

Нижний и верхний пороги c1 и c2 с некоторым приближением могут быть выражены через заданные значения вероятностей ложной тревоги α и пропуска сигнала β [10]

$$c_1 = \frac{\beta}{1-\alpha}, \quad c_2 = \frac{1-\beta}{\alpha}. \qquad (6.51)$$

Таким образом, последовательное правило выбора решения, в отличие от алгоритмов: байесовского, максимума апостериорной вероятности, максимума правдоподобия, Неймана-Пирсона, минимаксного предусматривает сравнение отношения правдоподобия с порогами c1 и c2, не зависящими от априорных вероятностей наличия или отсутствия сигнала и от потерь.

Алгоритм максимального правдоподобия. Если как и в рассмотренных последних двух алгоритмах принятия решений (Неймана-Пирсона и последовательном вальдовском) данные о потерях П и априорных вероятностях классов Р1 и Р2 отсутствуют, то может также применяться алгоритм максимального правдоподобия, который получается из байесовского алгоритма (6.21) при потерях П11=П22=0; П12=П21=1 и априорных вероятностях классов Р1=Р2=1/2

$$\hat{L}(x) = \frac{\hat{\omega}(\bar{x}|S_2)}{\hat{\omega}(\bar{x}|S_1)} \begin{matrix} > \gamma_2 \\ < \gamma_1 \end{matrix} 1 \qquad (6.52)$$

и заключается в принятии решения о наличии того класса S1 и S2, которому соответствует большее значение функции правдоподобия $\hat{\omega}(\bar{x}|S_1)$ или $\hat{\omega}(\bar{x}|S_2)$.

Подстановкой значений потерь П=1 и априорных вероятностей классов Р1 и Р2 в выражение для минимального байесовского риска (6.24) получаем выражение для минимального риска RМП, получающееся при использовании алгоритма максимального правдоподобия

$$R_{МП} = \alpha_{МП} + \beta_{МП}, \qquad (6.53)$$

где вероятности ошибок αМП и βМП получаются из (6.25) и (6.26):

$$\alpha_{МП} = \int_1^{\infty} \omega_{\hat{L}}(z/S_1)dz \qquad (6.54)$$

$$\beta_{МП} = \int_0^{1} \omega_{\hat{L}}(z/S_2)dz. \qquad (6.55)$$

Из (6.53) видно, что алгоритм максимального правдоподобия минимизирует суммарную вероятность ошибок распознавания RМП, в чем также проявляются его оптимальные свойства.

Выбор алгоритма, сохраняющего свои оптимальные свойства при использовании в нем оценок отношения правдоподобия. В начале раздела уже указывалось, что в рассмотренные решающие правила вместо отношения правдоподобия $L(\bar{x}) = \frac{\omega(\bar{x}|S_2)}{\omega(\bar{x}|S_1)}$ подставляется его оценка $\hat{L}(\bar{x}) = \frac{\hat{\omega}(\bar{x}|S_2)}{\hat{\omega}(\bar{x}|S_1)}$. Проведенная в работах [1, 2] проверка оптимальности рассмотренных алгоритмов: байесовского, максимума апостериорной вероятности, минимаксного, Неймана-Пирсона, последовательного вальдовского и максимума правдоподобия при подстановке в них оценок $\hat{L}(\bar{x})$ отношения правдоподобия вместо $L(\bar{x})$ показала, что при указанной подстановке оптимальные свойства сохраняет только алгоритм максимального правдоподобия, который, вследствие этого, практически во всех случаях будет использоваться в последующем изложении. При этом используется решающее правило (6.52), либо эквивалентное указанному правилу решающее правило для логарифма отношения правдоподобия [см. также (6.1)]:

$$\ln \hat{L}(\bar{x}) = \ln \frac{\hat{\omega}(x|S_2)}{\hat{\omega}(x|S_1)} \begin{matrix} > \gamma_2 \\ < \gamma_1 \end{matrix} 0 \qquad (6.56)$$

7. Одномерное распознавание

Распознавание одномерных образов с неизвестными средними a1 и a2 и общей дисперсией σ2. Неизвестные средние определяются в результате обучения из (5.28):

$$\hat{a}_1 = \frac{1}{m}\sum_{i=1}^{m} x_i^{(1)}, \quad \hat{a}_2 = \frac{1}{m}\sum_{i=1}^{m} x_i^{(2)} \tag{7.1}$$

и представляют собой несмещенные и состоятельные оценки максимального правдоподобия средних по обучающим выборкам $\left(x_i^{(1)}\right) = \left(x_i^{(1)}, \ldots, x_m^{(1)}\right)$ из S1 и $\left(x_i^{(2)}\right) = \left(x_i^{(2)}, \ldots, x_m^{(2)}\right)$ из S2. Оценка логарифма отношения правдоподобия будет иметь вид:

$$\ln \hat{L}(x_1, \ldots, x_n) = \ln \frac{\hat{\omega}_n(\bar{x} \mid S_2)}{\hat{\omega}_n(\bar{x} \mid S_1)} = \ln \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_i - \hat{a}_2)^2\right\}}{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - \hat{a}_1)^2\right\}}$$

$$= \ln \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}\left[(x_k - \hat{a}_2)^2 - (x_k - \hat{a}_1)^2\right]\right\} = \frac{1}{2\sigma^2}\sum_{k=1}^{n}\left[(x_k - \hat{a}_2)^2 - (x_k - \hat{a}_1)^2\right] =$$

$$= -\frac{1}{2\sigma^2}\sum_{k=1}^{n}\left[x_k^2 - 2x_k\hat{a}_2 + \hat{a}_2^2 - x_k^2 + 2x_k\hat{a}_1 - \hat{a}_1^2\right] =$$

$$= -\frac{1}{2\sigma^2}\sum_{k=1}^{n}\left\{2x_k\left[-(\hat{a}_2 - \hat{a}_1)\right] + \hat{a}_2^2 - \hat{a}_1^2\right\} =$$

$$= \frac{(\hat{a}_2 - \hat{a}_1)}{\sigma^2}\sum_{k=1}^{n} x_k - \frac{n(\hat{a}_2^2 - \hat{a}_1^2)}{2\sigma^2} =$$

$$= \frac{n}{2}\frac{\hat{a}_2 - \hat{a}_1}{\sigma}\frac{1}{\sigma}\left[\frac{2}{n}\sum_{k=1}^{n} x_k - \hat{a}_2 - \hat{a}_1\right] = \frac{n}{2}YZ \tag{7.2}$$

где обозначены случайные величины Y и Z:

$$Y = \frac{\hat{a}_2 - \hat{a}_1}{\sigma}, \quad Z = \frac{1}{\sigma}\left[\frac{2}{n}\sum_{k=1}^{n} x_k - (\hat{a}_2 + \hat{a}_1)\right]. \tag{7.3}$$

Решающее правило получается подстановкой значения $\ln \hat{L}(\bar{x})$ из (7.2) в (6.56)

$$\frac{n}{2}\frac{\hat{a}_2 - \hat{a}_1}{\sigma}\frac{1}{\sigma}\left[\frac{2}{n}\sum_{k=1}^{n} x_k - \hat{a}_2 - \hat{a}_1\right] \begin{matrix} > \gamma_2 \\ < \gamma_1 \end{matrix} \ln C, \ \hat{a}_2 > \hat{a}_1$$

или

$$\frac{1}{n}\sum_{k=1}^{n} x_k \begin{matrix} > \gamma_2 \\ < \gamma_1 \end{matrix} \frac{\hat{a}_1 + \hat{a}_2}{2} + \sigma^2 \frac{\ln C}{n(\hat{a}_2 - \hat{a}_1)}, \ \hat{a}_2 > \hat{a}_1 \tag{7.4}$$

для алгоритма максимума правдоподобия c = 1, ln c =0 и

$$\frac{1}{n}\sum_{k=1}^{n} x_k \begin{matrix} > \\ < \\ \end{matrix} \frac{\hat{a}_1 + \hat{a}_2}{2} \quad \begin{matrix} \gamma_2 \\ \\ \gamma_1 \end{matrix} . \tag{7.5}$$

Вероятность ошибок распознавания одномерных образов. Для нахождения вероятностей ошибок распознавания 1 и 2 рода α и β, найдем сначала распределение оценки логарифма отношения правдоподобия $\omega_{\ln \hat{L}}(l)$, которое выражается через Y и Z как распределение произведения этих случайных величин [1, 2]:

$$\omega_{\ln \hat{L}}(l) = \int_{-\infty}^{\infty} \omega_Y(u)\omega_Z(l/u)du/|u| \tag{7.6}$$

или

$$\omega_{\ln \hat{L}}(l) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left[\frac{(u+a)^2}{\sigma_1^2} + \frac{(l/u-a)^2}{\sigma_2^2}\right]\frac{du}{|u|}, \tag{7.7}$$

где обозначено

$$a = \frac{a_1 - a_2}{\sigma}, \quad \sigma_1^2 = \frac{2}{m}, \quad \sigma_2^2 = \frac{2}{m} + \frac{4}{n}. \tag{7.8}$$

Вероятности ошибок 1-го и 2-го рода α и β одномерного распознавания определяются подстановкой значения $\omega_{\ln \hat{L}}(l)$ в формулы (3.9) и (3.10):

$$\alpha = \beta = \int_0^{\infty} \omega_{\ln \hat{L}}(l)dl = \frac{1}{2\pi\sigma_1\sigma_2}\int_0^{\infty}\left\{\int_0^{\infty}\exp\left\{-\frac{1}{2}\left[\frac{(u+a)^2}{\sigma_1^2} + \frac{(l/u-a)^2}{\sigma_2^2}\right]\right\}\frac{du}{u} + \right.$$
$$\left. + \int_0^{\infty}\exp\left\{-\frac{1}{2}\left[\frac{(u-a)^2}{\sigma_1^2} + \frac{(l/u+a)^2}{\sigma_2^2}\right]\right\}\frac{du}{u}\right\}dl . \tag{7.9}$$

Меняя порядок интегрирования в (7.9), получаем:

$$\alpha = \beta = \frac{1}{\sigma_1\sqrt{2\pi}}\int_0^{\infty}\left[\exp\left\{-\frac{(u-a)^2}{2\sigma_1^2}\right\}\frac{1}{\sigma_2\sqrt{2\pi}}\int_0^{\infty}\exp\left\{-\frac{(l/u-a)^2}{\sigma_2^2}\right\}d(l/u) + \right.$$
$$\left. + \exp\left\{-\frac{(u+a)^2}{2\sigma_1^2}\right\}\frac{1}{\sigma_2\sqrt{2\pi}}\int_0^{\infty}\exp\left\{-\frac{(l/u+a)^2}{\sigma_2^2}\right\}d(l/u)\right]du . \tag{7.10}$$

Внутренние интегралы можно заменить их значениями $F\left(-\frac{a}{\sigma_2}\right)$ и $F\left(\frac{a}{\sigma_2}\right)$, выражающимися через табулированный интеграл Лапласа F(Z):

$$F\left(-\frac{a}{\sigma_2}\right) = \frac{1}{\sigma_2\sqrt{2\pi}}\int_0^{\infty}\exp\left\{-\frac{(l/u-a)^2}{\sigma_2^2}\right\}d(l/u), \tag{7.11}$$

$$F\left(\frac{a}{\sigma_2}\right) = \frac{1}{\sigma_2\sqrt{2\pi}}\int_0^{\infty}\exp\left\{-\frac{(l/u+a)^2}{\sigma_2^2}\right\}d(l/u), \tag{7.12}$$

где [11]

$$F(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{Z}e^{-\frac{x^2}{2}}dx \quad . \qquad (7.13)$$

Подставляя (7.11) и (7.12) в (7.10), получаем для $\alpha = \beta$:

$$\alpha=\beta=\frac{1}{\sigma_1\sqrt{2\pi}}\int_0^{\infty}\left[\exp\left\{-\frac{(u-a)^2}{2\sigma_1^2}\right\}F\left(-\frac{a}{\sigma_2}\right)+\exp\left\{-\frac{(u+a)^2}{2\sigma_1^2}\right\}F\left(\frac{a}{\sigma_2}\right)\right]du. \quad . \qquad (7.14)$$

Оставшиеся интегралы также можно заменить их значениями $F\left(\frac{a}{\sigma_1}\right)$ и $F\left(-\frac{a}{\sigma_1}\right)$, выражающимися через интеграл Лапласа F(z) (7.13):

$$F\left(\frac{a}{\sigma_1}\right)=\frac{1}{\sigma_1\sqrt{2\pi}}\int_0^{\infty}\exp\left\{-\frac{(u-a)^2}{2\sigma_1^2}\right\}du, \qquad (7.15)$$

$$F\left(-\frac{a}{\sigma_1}\right)=\frac{1}{\sigma_1\sqrt{2\pi}}\int_0^{\infty}\exp\left\{-\frac{(u+a)^2}{2\sigma_1^2}\right\}du \; . \qquad (7.16)$$

Сопоставлением (7.15) и (7.16) с (7.14) получаем для вероятностей ошибок распознавания $\alpha = \beta$ их выражение через табулированный интеграл Лапласа:

$$\alpha=\beta=F\left(\frac{a}{\sigma_1}\right)F\left(-\frac{a}{\sigma_2}\right)+F\left(-\frac{a}{\sigma_1}\right)F\left(\frac{a}{\sigma_2}\right), \qquad (7.17)$$

Во многих практически важных случаях целесообразно иметь выражение вероятностей ошибок распознавания $\alpha = \beta$ через другой табулированный интеграл – интеграл вероятностей Φ(x) [11].

$$\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_0^{x}e^{-Z^2}dZ \qquad (7.18)$$

который связан с интегралом Лапласа (7.13) формулами

$$F(x)=\frac{1}{2}\left[\left(\Phi\left(x/\sqrt{2}\right)\right)+1\right] \qquad (7.19)$$

$$\Phi(x)=2F\left(x\sqrt{2}\right)-1 \quad . \qquad (7.20)$$

Подставляя значение F(x), выраженное через интеграл вероятностей Φ(x) согласно (7.19) в (7.17), получаем выражение вероятностей ошибок распознавания 1-го и 2-го рода $\alpha = \beta$ через табулированный интеграл вероятностей (7.18)

$$\alpha=\beta=\frac{1}{2}-\frac{1}{2}\Phi\left(\frac{|a|}{2\sqrt{2/n+1/m}}\right)\Phi\left(|a|\sqrt{m}/2\right), \qquad (7.21)$$

Результаты вычисления зависимости вероятностей ошибок распознавания $\alpha = \beta$ по формуле (7.21) при a = 1 от объема m обучающих выборок при различных объемах n контрольных выборок представлены на рис. 9. Как видно из рисунка, влияние объема обучающих выборок особенно сильно проявляется в области малых m ($m \le 30$), где, в частности, увеличение m от 5 до 20 (при n = 30)

приводит к уменьшению вероятности ошибок распознавания от 0, 1 до 0,02. При дальнейшем увеличении объема обучающих выборок (m ≥ 50) их влияние на вероятность ошибок распознавания становится менее ощутимым, поскольку эталонные описания при таких значениях m уже достаточно хорошо сформированы и дальнейшее обучение мало что к ним может добавить. Аналогичным образом влияет на вероятности ошибок α = β объем контрольных выборок n это влияние сильно проявляется при малых n (n ≤ 20) и становится мало ощутимым при n > 30.

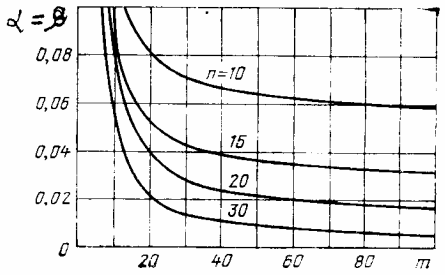

Рис. 9

Распознавание одномерных образов с неизвестными средними и неизвестными дисперсиями. Наиболее общим случаем одномерного распознавания является определение принадлежности выборки $(x_i)^n = (x_1,\ldots,x_n)$ независимых наблюдений к одному из двух классов S1 и S2 характеризующихся неизвестными средними a1 и a2, и неизвестными дисперсиями σ_1^2 и σ_2^2. В ходе обучения вычисляются оценки неизвестных средних $\hat{a}_1$ и $\hat{a}_2$

$$\hat{a}_1 = \frac{1}{m}\sum_{i=1}^{m} x_i^{(1)} \quad \hat{a}_2 = \frac{1}{m}\sum_{i=1}^{m} x_i^{(2)}, \tag{7.22}$$

и дисперсий $\hat{\sigma}_1^2$ и $\hat{\sigma}_2^2$

$$\hat{\sigma}_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(x_i^{(1)} - \hat{a}_1\right)^2 \quad \hat{\sigma}_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(x_i^{(2)} - \hat{a}_2\right)^2, \tag{7.23}$$

Оценка логарифма отношения правдоподобия будет очевидно иметь следующий вид

$$\ln \hat{L}(x_1,\ldots,x_n) = \ln\frac{\hat{\omega}(x_1,\ldots,x_n/S_2)}{\hat{\omega}(x_1,\ldots,x_n/S_1)} =$$

$$= \ln\frac{\dfrac{1}{\left(2\pi\hat{\sigma}_2^2\right)^{n/2}}\exp\left\{-\dfrac{1}{2\hat{\sigma}_2^2}\sum_{i=1}^{n}\left(x_i-\hat{a}_2\right)^2\right\}}{\dfrac{1}{\left(2\pi\hat{\sigma}_1^2\right)^{n/2}}\exp\left\{-\dfrac{1}{2\hat{\sigma}_1^2}\sum_{i=1}^{n}\left(x_i-\hat{a}_1\right)^2\right\}}. \tag{7.24}$$

Решающее правило получается подстановкой значения $\ln \hat{L}(x_1,\ldots,x_n)$ из (7.24) в (6.56):

$$\frac{1}{2}\sum_{i=1}^{n}\left[\frac{1}{\hat{\sigma}_1^2}(x_i-\hat{a}_1)^2-\frac{1}{\hat{\sigma}_2^2}(x_i-\hat{a}_2)^2\right]+\frac{n}{2}\ln\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\begin{array}{c}\gamma_2\\ >\\ <\\ \gamma_1\end{array}\ln C\,, \qquad (7.25)$$

для алгоритма максимального правдоподобия C = 1 , ln C = 0 и решающее правило:

$$\frac{1}{2}\sum_{i=1}^{n}\left[\frac{1}{\hat{\sigma}_1^2}(x_i-\hat{a}_1)^2-\frac{1}{\hat{\sigma}_2^2}(x_i-\hat{a}_2)^2\right]+\frac{n}{2}\ln\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\begin{array}{c}\gamma_2\\ >\\ <\\ \gamma_1\end{array}0\,, \qquad (7.26)$$

Введем параметры распознавания:

$$r = \sigma_2^2 / \sigma_1^2\,,$$
$$d^2 = (a_2 - a_1)^2 / \sigma_1^2\,. \qquad (7.27)$$

На рис 10 (а и б) приведены графики зависимости вероятности ошибок α=β от значений параметров d2 и r, вычисленные в работе [1] для n=m=10. Их анализ позволяет утверждать: с ростом расстояния d2 между классами, объемов выборок m и n вероятности ошибок α и β убывают; по мере увеличения r вероятности ошибок α и β сначала незначительно возрастают, а затем начинают быстро уменьшаться (при d2 = 0 сразу уменьшаются).

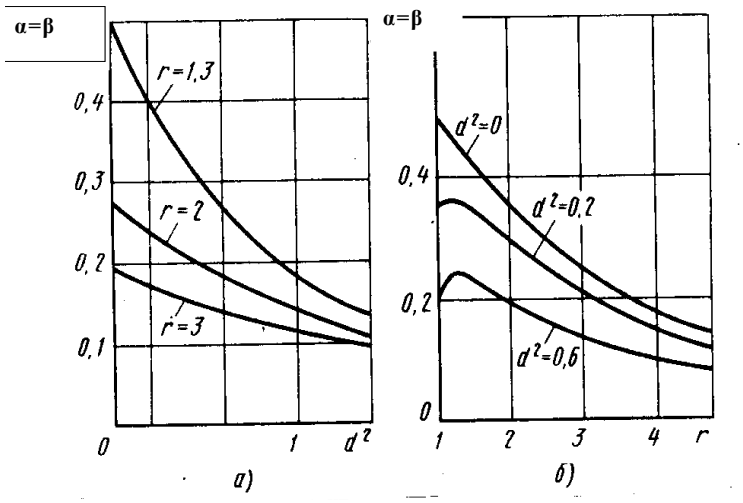

Рис.10

Это объясняется тем, что рост r фактически означает увеличение дисперсий случайных величин, составляющих обучающие и контрольные выборки из класса S2, что должно приводить при неизменных значениях других параметров к увеличению вероятностей ошибок α и β. С другой стороны, чем больше r, тем сильнее отличие распределений у классов S1 и S2 друг от друга и тем меньше, следовательно, должны быть вероятности ошибок α и β. Таким образом, характер изменения вероятностей α и β с ростом r определяется противоположным влиянием этих двух тенденций. Так, увеличение r с 1,01 до 1,3 при m = 10, n = 10, d2 =0,6 сопровождается увеличением вероятности ошибки α с 0,2 до 0,24. Однако при дальнейшем увеличении r до 2,0 вероятность ошибки α падает до 0,196. Это объясняется тем, что с ростом r усиливается влияние тенденции, ведущей к уменьшению вероятностей ошибок α и β, и начиная с некоторого значения r*, ее влияние становится доминирующим. При этом величина r* тем меньше, чем меньше d2. Так, при d2 = 0,6r* $\approx$ 1,3, при d2 = 0,2r* $\approx$ 1,25 при d2$\leq$0,01r* < 1,01.

8. Многомерное распознавание

Распознавание многомерных образов, различающихся векторами средних. Пусть на вход распознающей системы поступают многомерные (векторные) наблюдения $(\bar{x}_i)_1^n$, принадлежащие одному из двух классов s1 и s2, различающихся только своими неизвестными векторами средних $\hat{a}_1$ и $\hat{a}_2$ (и, следовательно, имеющие общую ковариационную матрицу M). Оценки неизвестных векторов средних $\hat{\bar{a}}_1$ и $\hat{\bar{a}}_2$ определяются в результате обучения из (5.30):

$$\hat{\bar{a}}_1 = \frac{1}{m}\sum_{i=1}^{m}\bar{x}_i^{(1)}, \quad \hat{\bar{a}}_2 = \frac{1}{m}\sum_{i=1}^{m}\bar{x}_i^{(2)}. \qquad (8.1)$$

Оценка логарифма отношения правдоподобия $\ln\hat{L}(\bar{x}_1,\ldots,x_n)$ будет иметь следующий вид:

$$\ln\hat{L}(\bar{x}_1,\ldots,\bar{x}_n) = \frac{1}{2}\sum_{i=1}^{n}\left\{(\bar{x}_i-\hat{a}_1)^T M^{-1}(\bar{x}_i-\hat{a}_2)-(\bar{x}_i-\hat{a}_2)^T M^{-1}(\bar{x}_i-\hat{a}_2)\right\}=$$

$$=\frac{n}{2}(\bar{a}_2-\bar{a}_1)^T M^{-1}\left[\frac{2}{n}\sum_{i=1}^{n}\bar{x}_i-\hat{a}_2-\hat{a}_1\right]=\frac{n}{2}\bar{y}^T M^{-1}\bar{z}, \qquad (8.2)$$

где обозначены случайные величины y и z

$$\bar{y}=\hat{a}_2-\hat{a}_1; \quad \bar{z}=(2/n)\sum_{i=1}^{n}x_i-\hat{a}_2-\hat{a}_1. \qquad (8.3)$$

Решающее правило получается подстановкой значения $\ln\hat{L}(\bar{x}_1,\ldots,x_n)$ из (8.2) в (6.56):

$$\frac{n}{2}\left(\hat{\bar{a}}_2-\hat{\bar{a}}_1\right)^T\bar{M}^{-1}\left(\frac{2}{n}\sum_{i=1}^{n}\bar{x}_i-\hat{\bar{a}}_2-\hat{\bar{a}}_1\right)\begin{matrix}\gamma_2\\ >\\ <\\ \gamma_1\end{matrix}\ln C, \qquad (8.4)$$

где порог ln C в связи с предпочтением алгоритму максимального правдоподобия, сохраняющему свои оптимальные свойства при подстановке в него оценок логарифма правдоподобия (см. разд. 6), выбираем, как правило, равным: ln C = 0 (т. к. в этом случае C = 1).

Вероятность ошибок распознавания многомерных образов с разными векторами средних. Нахождение вероятности ошибок распознавания 1-го и 2-го рода α и β осуществляется [1, 2] по той же методологии, что и нахождение вероятностей ошибок распознавания одномерных образов в разделе 7 (см. формулы (7.2) – (7.21)), предусматривающей вычисление плотности вероятности оценок логарифма отношения правдоподобия (8.2), которое выражается по формуле (7.6) через введенные в (8.3) случайные величины y и z как распределение произведения этих случайных величин. В результате выполненного в работах [1, 2] достаточно трудоемкого и сложного процесса интегрирования общее выражение для вероятности ошибок

многомерного распознавания первого и второго рода α и β удается свести к следующему двойному интегралу (при C = 1)

$$\alpha = \beta = \left[\Theta(p)\exp\left\{-d^2/2\sigma_1^2\right\}/\sqrt{2\pi}(p-3)!!\right]\int_0^{\infty}\int_{-\pi/2}^{\pi/2} t^{p-1}\times$$
$$\times\cos^{p-2}\varphi\exp\left\{(-1/2)\left[t^2 - 2d\sigma_1^{-1}t\sin\varphi\right]\right\}\times$$
$$\times F\left[-d\sin\varphi/\sigma_2\right]dtd\varphi, \qquad (8.5)$$

где

$$\Theta(p) = \begin{cases} 1 & ,если\ p = 2k+1 \\ \sqrt{\dfrac{2}{\pi}} & ,если\ p = 2k, \end{cases} \qquad (8.6)$$

F(x) – табулированный интеграл Лапласа (7.13), σ_1^2 и σ_2^2 выражаются через объемы контрольных n и обучающих m выборок по формулам (7.8):

$\sigma_1^2 = 2/m$; $\sigma_2^2 = 2/m + 4/n$,

а d – скалярная величина – расстояние Махаланобиса [17]:

$d^2 = (\bar{a}_2 - \bar{a}_1)^T \bar{M}^{-1}(\bar{a}_2 - \bar{a}_1)$. (8.7)

Зависимость вероятности ошибок распознавания $\alpha = \beta$ от размерности признакового пространства P и объемов обучающих m = M и контрольных n = N выборок, вычисленная по формуле (8.5), приведена на рис. 11

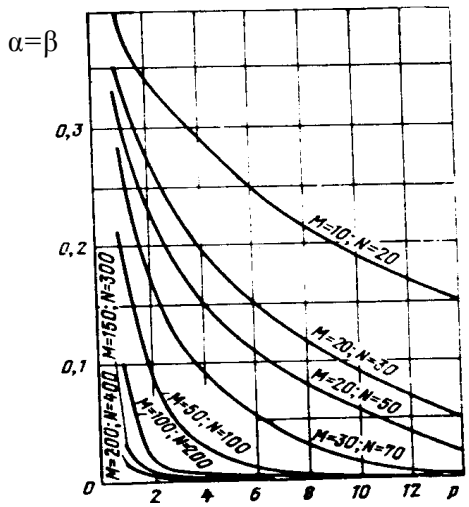

Рис. 11

Как видно из приведенных па рисунке 11 данных с уменьшением значения объемов обучающих m=M и контрольной n=N выборок требуемое значение размерности признакового пространства P, обеспечивающее заданный уровень достоверности распознавания $1 - \alpha = 1 - \beta$, увеличивается. Аналогичный характер носит взаимосвязь выбранной размерности признакового пространства P с требуемыми объемами обучающих m=M и контрольной n=N выборок: сокращение

размерности P должно компенсироваться увеличением объемов m=M и n=N.

Таким образом, в тех случаях, когда по условиям функционирования систем распознавания увеличение с целью обеспечения требуемой достоверности значения какого-либо из ее параметров (к примеру, объемов обучающих m = M и (или) контрольной n = N выборок) оказывается невозможным, заданный уровень может быть достигнут увеличением другого параметра (к примеру, размерности признакового пространства P).

Возвращаясь к общему выражению (8.5) вероятности ошибок многомерного распознавания α и β, следует заметить, что аналитически выразить $\alpha = \beta$ удается при $p = 2k + 1$, $k = 1, 2, \ldots$; однако лишь в трехмерном случае (k=1) получается сравнительно компактное выражение через табулированный интеграл Лапласа (7.13):

$$\alpha_3 = \beta_3 = F(d/\sigma_1)F(-d/\sigma_2) + F(-d/\sigma_1)F(d/\sigma_2) + \\ + \left[\sigma_1\sigma_2/\sqrt{2\pi}d\left(\sigma_1^2 - \sigma_2^2\right)\right]\{\sigma_2 \exp\{-d^2/2\sigma_1^2\}[F(d/\sigma_2) - \\ - F(-d/\sigma_2)] - \sigma_1 \exp\{-d^2/2\sigma_2^2\}[F(d/\sigma_1) - F(-d/\sigma_1)]\}. \quad (8.8)$$

Распознавание многомерных ансамблей с неизвестными векторами средних и неизвестными разными ковариационными матрицами. В наиболее общем случае априори неизвестными оказываются как векторы средних $\bar{a}_1$, $\bar{a}_2$, так и ковариационные матрицы $\bar{M}_1$, $\bar{M}_2$ распознаваемых ансамблей. В ходе обучения вычисляются оценки $\hat{a}_1$ и $\hat{a}_2$ неизвестных векторов средних по формулам (5.30)

$$\hat{\bar{a}}_1 = \frac{1}{m}\sum_{i=1}^{m}\bar{x}_i^{(1)}, \quad \hat{\bar{a}}_2 = \frac{1}{m}\sum_{i=1}^{m}\bar{x}_i^{(2)},$$

.

И оценки $\hat{M}_1$ и $\hat{M}_2$ неизвестных ковариационных матриц по формулам (5.33):

$$\hat{\bar{M}}_1 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\bar{x}_i^{(1)} - \hat{\bar{a}}_1\right)\left(\bar{x}_i^{(1)} - \hat{\bar{a}}_1\right)^T;$$

$$\hat{\bar{M}}_2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\bar{x}_i^{(2)} - \hat{\bar{a}}_2\right)\left(\bar{x}_i^{(2)} - \hat{\bar{a}}_2\right)^T. \quad (8.9)$$

Решающее правило будет иметь следующий вид:

$$\frac{1}{2}\sum_{i=1}^{n}\left[\left(\bar{x}_i - \hat{\bar{a}}_1\right)^T \bar{M}_1^{-1}\left(\bar{x}_i - \hat{\bar{a}}_1\right) - \left(\bar{x}_i - \hat{\bar{a}}_2\right)^T \hat{\bar{M}}_2^{-1}\left(\bar{x}_i - \hat{\bar{a}}_2\right)\right] + \frac{n}{2}\ln\frac{\det\hat{\bar{M}}_1}{\det\hat{\bar{M}}_2} \begin{matrix}\gamma_2 \\ > \\ < \\ \gamma_1\end{matrix} \ln C \quad (8.10)$$

где как и ранее в связи с предпочтением, отдаваемым алгоритму максимального правдоподобия (см. раздел 6), порог ln C = 0, т. к. в этом случае C = 1.

Автор выражает глубокую благодарность студенту 3-го курса МЭСИ Я. Я. Фомину за активное участие в написании этого пособия.

9. Список литературы

1. Фомин Я. А. Савич А. В. Оптимизация распознающих систем – М. Машиностроение, 1993.

2. Фомин Я. А., Тарловский Г. Р. Статистическая теория распознавания образов. — М.: Радио и связь, 1986. 264 с.

3. Фомин Я. А. Теория выбросов случайных процессов.– М.: Связь, 1980.

4. Ту Дж., Гонсалес Р. Принципы распознавания образов/Пер. с англ. — М.: Мир. 1978. 412 с.

5. Горелик А. Л., Скрипкин В. А. Методы распознавания.– М. Высшая школа, 1984. 208 с.

6. Журавлев Ю. И. Об алгебраическом подходе к решению задач распознавания и классификации//Проблемы кибернетики. – М.: Наука. 1978. Вып. 33, с. 5—68.

7. Классификация и кластер / Под ред. Дж. Вэн Райзина: Пер. с англ.– М., Мир, 1980 – 390 с.

8. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды/Пер. с англ. под ред. А. Н. Колмогорова, Ю. В. Прохорова. — М.: Наука, 1976. 736 с.

9. Прохоров Ю. В. Многомерные распределения: неравенства и предельные теоремы//Итоги науки и техники. 1973. Т. 10. С. 5—24. Сер. «Теория вероятностей, математическая статистика, теоретическая кибернетика»).

10 . Вальд А. Последовательный анализ - М.: Физматгиз, 1960. 328 с.

11. Большев Л. М., Смирнов Н. В. Таблицы математической статистики. — М.: Наука. 1983. 416 с.

12. Мескон М., Альберт М., Хедоури Ф. “Основы менеджмента”, Пер. с Англ. – М. Дело, 1998.

13. Акоф Р., Сасиени М. “Основы исследования операций” М., Пер. с англ. – Мир, 1971.

14. Томсон А., Стрикленд А. “Стратегический менеджмент” М. Пер. с англ. – ЮНИТИ, 1998.

15. Acoff R. E. “Management Information System” Management Science, 1967.

16. Фукунага К. “Введение в статистическую теорию распознавания образов. Пер. с англ. – М. Наука, 1979.

17. Котлер Ф. Основы маркетинга. – М. Пер. с англ. – Прогресс, 1992.