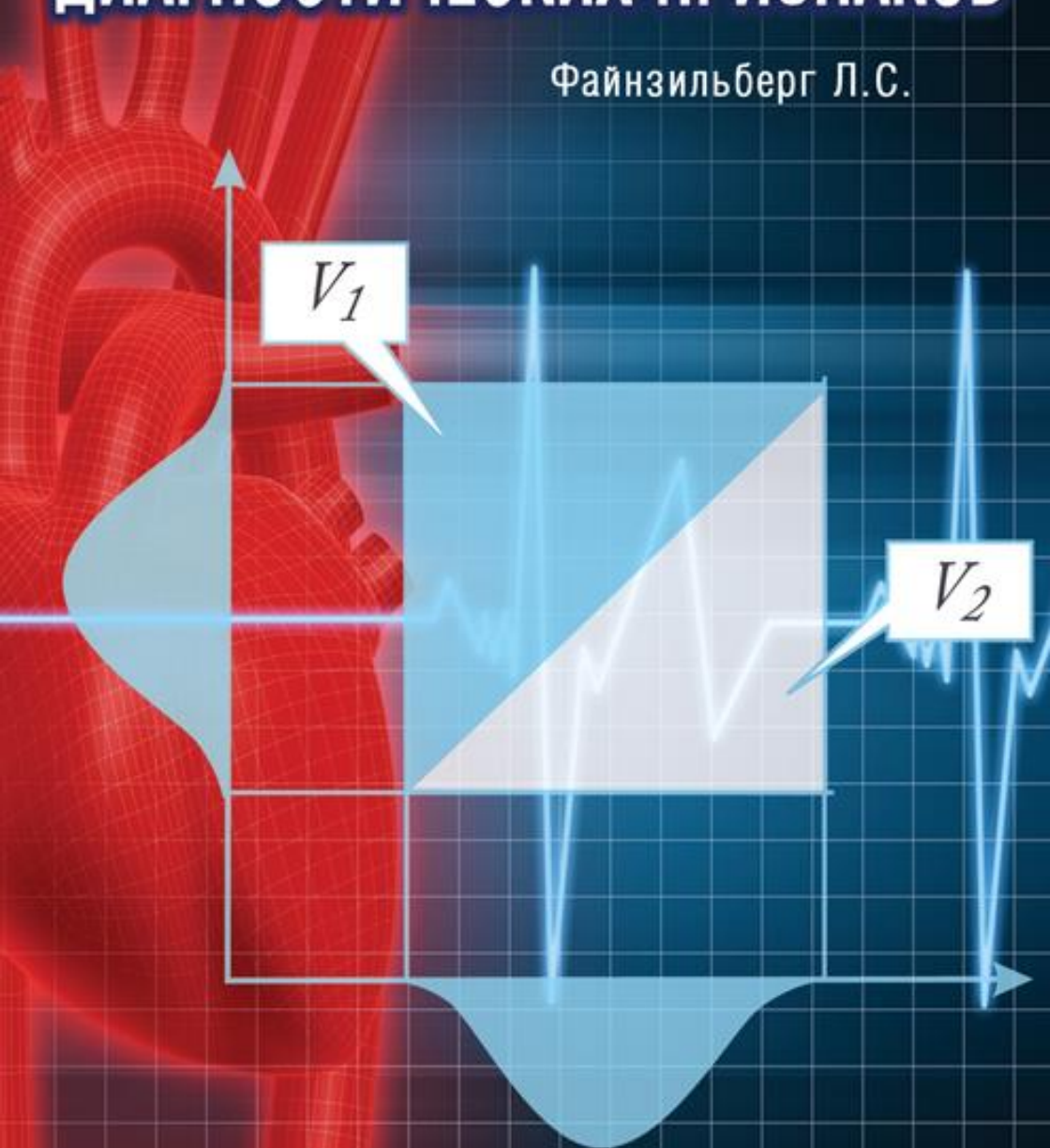


МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОЦЕНКИ ПОЛЕЗНОСТИ ДИАГНОСТИЧЕСКИХ ПРИЗНАКОВ

Файнзильберг Л.С.



Международный научно-учебный центр
информационных технологий и систем
НАН и МОН Украины

Л.С.Файнзильберг

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ
ОЦЕНКИ ПОЛЕЗНОСТИ
ДИАГНОСТИЧЕСКИХ ПРИЗНАКОВ**

Киев
«Освіта України»
2010

ББК
Ф 45

*Печатается по решению ученого совета
Международного научно-учебного центра
информационных технологий и систем НАН и МОН Украины
(протокол № 1 от 15 января 2010 г.)*

Рецензенты:

доктор технических наук, профессор В.И.Иваненко,
доктор физико-математических наук, профессор В.В.Новицкий

Ф 45 **Файнзильберг Л.С.**
**Математические методы оценки полезности
диагностических признаков.** Монография. – К.:
«Освита Украины», 2010. – 152 с.

ISBN 987-966-188-134-0

Монография посвящена одной из центральных проблем теории распознавания образов – построению эффективной системы признаков. Доказано, что не всякий признак, удовлетворяющий традиционному критерию информативности, является полезным с точки зрения уменьшения риска ошибочной диагностики.

Сформулированы условия, гарантирующие полезность диагностических признаков в байесовской схеме принятия решений. Показано, что при определенных условиях совокупность неинформативных в отдельности признаков может быть не только полезна, но и обеспечить безошибочную классификацию.

Приведены примеры практического использования полученных результатов в технических и медицинских приложениях.

Для читателей, интересующихся теорией статистического распознавания образов и практикой построения компьютерных систем диагностически.

ISBN 987-966-188-134-0

© Файнзильберг Л.С., 2010
© «Освита Украины», 2010

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
ГЛАВА 1	
ХАРАКТЕРИСТИКИ ДИАГНОСТИЧЕСКИХ ПРИЗНАКОВ.....	7
1.1. Связь задач диагностики с теорией распознавания образов ...	7
1.2. Критерии информативности признаков.....	13
1.3. Безусловная и условная независимость признаков.....	22
1.4. Безошибочное распознавание по совокупности пересекающихся признаков.....	31
ГЛАВА 2	
ДОСТАТОЧНЫЕ УСЛОВИЯ ПОЛЕЗНОСТИ ИНФОРМАТИВНЫХ ПРИЗНАКОВ.....	37
2.1. Условия полезности отдельного признака.....	37
2.2. Достаточные условия полезности признака в описании	45
2.3. Информационный критерий полезности признаков	61
ГЛАВА 3	
ОЦЕНКА ПОЛЕЗНОСТИ ДИАГНОСТИЧЕСКИХ ТЕСТОВ С ПОЗИЦИИ ТЕОРИИ СТАТИСТИЧЕСКИХ РЕШЕНИЙ.....	72
3.1. Достаточные условия полезности диагностического теста.....	72
3.2. Оценка полезности тестов в задачах скрининга.....	85
3.3. Традиционный и усиленный <i>ROC</i> -анализ	90
3.4. Интервальное обобщение результатов.....	100
ГЛАВА 4	
ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ РЕЗУЛЬТАТОВ.....	111
4.1. Анализ диагностических признаков термограммы	111
4.2. Анализ диагностических признаков электрокардиограммы....	124
СПИСОК ЛИТЕРАТУРЫ	142

ВВЕДЕНИЕ

На современном этапе развития общества все большую актуальность приобретает компьютерная диагностика. Диагностические системы используются в тех случаях, когда непосредственное определение состояния технического или биологического объекта затруднительно.

Например, агрессивность и высокая температура среды препятствуют созданию датчиков для непосредственного определения текущего состояния расплавленного металла в сталеплавильной ванне. Невозможно при массовых профилактических обследованиях определять состояние сердечно-сосудистой системы человека на основе метода коронарографии. Этот инвазивный метод хотя и является наиболее достоверным, но достаточно дорогой и, самое главное, небезопасный для обследуемого пациента. Список подобных примеров можно было бы продолжить.

Важнейшим этапом построения новых диагностических систем является выбор эффективной системы признаков, по которым можно надежно отличить одно состояние объекта от другого. Исходная совокупность признаков обычно задается субъективно, поскольку формальных методов выбора исходного пространства признаков не существует.

В одних случаях исходная система признаков заранее известна экспертам соответствующей предметной области, в других – признаки выбираются интуитивно на основании лишь скудных и порой противоречивых представлениях о процессах, происходящих в объекте.

При этом в исходное описание стремятся включить любой доступный для измерения параметр, который хоть в чем-то можно считать «потенциально» полезным для диагностики состояния объекта.

Но после того как исходное описание так или иначе построено, возникает необходимость в оценке уже формальным методом полезности любого признака.

подавляющее число научных публикаций, посвященных методам оценки полезности признаков, условно можно разделить на две большие группы.

В первую входят работы, в которых делаются достаточно серьезные и порой необоснованные предположения о вероятностных характеристиках признаков, например, гипотеза о нормальных многомерных распределениях признаков в классах. Часто такие предположения не выполняются для конкретной прикладной задачи.

Во вторую, намного более обширную группу публикаций, входят работы, в которых этап селекции признаков осуществляется уже на этапе обучения. Однако, как будет показано в главе 1, необоснованное увеличение числа диагностических признаков при малом объеме обучающей выборки может только ухудшить надежность распознавания состояния объекта.

Предлагаемая читателю книга преследует цель найти в некотором смысле компромисс между названными двумя подходами. В ней изложены результаты оригинальных исследований, которые направлены на формальную оценку полезности диагностических признаков *до этапа обучения* при весьма *ограниченных* представлениях о вероятностных характеристиках признаков, известных из *физических* представлений об объекте.

В монографии сформулированы теоремы, в которых определены достаточные условия полезности признаков с точки зрения уменьшения вероятности ошибочных решений или, в общем случае, среднего риска. Установлена взаимосвязь информативности и полезности признаков при статистическом распознавании двух классов состояний объекта. Рассмотрены условия полезности диагностических тестов с позиции теории статистических решений.

Проведенные исследования позволили усовершенствовать традиционный *ROC*-анализ за счет выделения в *ROC*-пространстве области гарантированно полезных тестов. Установлены граничные значения вероятностей ошибок пропуска цели и ложной тревоги для теста, гарантированно полезного с точки зрения уменьшения априорного риска.

Важно отметить, что многие из полученных результатов, например, достаточные условия полезности признаков (глава 2) и диагностических тестов (глава 3), оказались конструктивными и уже нашли практическое применение при решении прикладных задач (глава 4).

Другие же результаты, например, необходимые условия безошибочного распознавания двух классов по совокупности пересекающихся признаков (глава 1), условия инвариантности вероятности ошибки относительно исключения из описания информативного признака (глава 2), условия существования полезной совокупности «бесполезных» в отдельности признаков (глава 3) являются не менее важными, поскольку не только расширяют научные знания в теории статистического распознавания образов, но и стимулируют конструктора к поиску *нетривиальной* системы диагностических признаков при построении новых диагностических систем.

Автор считает своим приятным долгом выразить признательность рецензентам доктору технических наук, профессору В.И.Иваненко и доктору физико-математических наук, профессору В.В.Новицкому за ценные замечания.

ГЛАВА 1

ХАРАКТЕРИСТИКИ ДИАГНОСТИЧЕСКИХ ПРИЗНАКОВ

1.1. Связь задач диагностики с теорией распознавания образов

Диагностику состояния объектов часто рассматривают с позиции теории распознавания образов [5, 32]. Предполагается, что объект находится в одном из M возможных состояний V_1, \dots, V_M множества $V = \{V_1, \dots, V_M\}$. Имеется совокупность характеристик (свойств) объекта $x^{(N)} = (x_1, \dots, x_N)$, которые могут быть использованы в качестве диагностических признаков, позволяющих отличать одно состояние объекта от другого.

Требуется построить решающее правило $D(x^{(N)}) : X^{(N)} \rightarrow M$, которое отображает множество $X^{(N)} \triangleq \{x^{(N)}\}$ возможных значений признаков x_1, \dots, x_N на множество $M \triangleq \{1, \dots, M\}$ номеров возможных состояний (диагнозов) V_1, \dots, V_M . Другими словами, требуется построить алгоритм определения значений индикаторной переменной $s = D(x^{(N)})$ в виде

$$s = \begin{cases} 1, & \text{если } x^{(N)} \in \Omega_1 \\ \dots & \\ M, & \text{если } x^{(N)} \in \Omega_M \end{cases}, \quad (1.1)$$

где $\Omega_m, m = 1, \dots, M$ – непересекающиеся области пространства $X^{(N)}$, в которых принимаются решения в пользу класса V_m .

Математические методы распознавания можно разделить на две группы: детерминированные, которые в явном виде не используют статистические свойства наблюдаемых образов, и статистические, которые основываются на методах теории вероятностей и математической статистики.

Статистическая постановка задач распознавания возникает в тех практически важных случаях, когда при различных состояниях объекта $V_i \neq V_j$ могут совпадать значения каждого из имеющихся признаков x_n ($1 \leq n \leq N$), иными словами, допускается, что множества значений каждого из признаков, соответствующие различным классам, пересекаются.

При статистической постановке задачи распознавания наибольшую известность получил байесовский подход [58]. В соответствии с этим подходом состояния V_1, \dots, V_M объекта рассматриваются как случайные события с априорными вероятностями $P(V_m)$, $\sum_{m=1}^M P(V_m) = 1$, признаки – как случайные величины, для которых объективно существуют условные распределения $p(x^{(N)} | V_m)$, а множества

$$X_m^{(N)} = \{x^{(N)} : p(x^{(N)} | V_m) \neq 0\}, \quad X_1^{(N)} \cup \dots \cup X_M^{(N)} = X^{(N)}.$$

именуются собственными областями классов в пространстве признаков.

Замечание 1.1. Строго говоря, условные распределения признаков в i -м и j -м классах нужно обозначать разными функциями, например, так $p_i(x^{(N)} | V_i)$ и $p_j(x^{(N)} | V_j)$. Однако для упрощения записей мы будем опускать индексы и писать $p(x^{(N)} | V_i)$, $p(x^{(N)} | V_j)$ вместо $p_i(x^{(N)} | V_i)$, $p_j(x^{(N)} | V_j)$. Для обозначения совпадающих распределений будем использовать запись

$$p(x^{(N)} | V_i) \equiv p(x^{(N)} | V_j), \quad (1.2)$$

а для обозначения различных распределений в классах – запись вида

$$p(x^{(N)} | V_i) \neq p(x^{(N)} | V_j). \quad (1.3)$$

Замечание 1.2. Поскольку мы не налагаем никаких ограничений на множество $X^{(N)}$, то подобно тому, как это делается в работе [25] здесь и далее, где это особо не оговаривается, запись $p(x)$ будет обозначать обобщенную вероятностную плотность, представляющую собой плотность вероятности, если x – непрерывная величина, или вероятность дискретного значения x , если x – дискретная величина. При этом знак \sum будет обозначать операцию суммирования в первом случае или интегрирования во втором.

Интегрирование многомерного распределения $p(x^{(N)} | V_m)$ по множествам возможных значений всех признаков x_1, \dots, x_N , кроме признака x_n ($1 \leq n \leq N$), позволяет получить представление об одномерных (маргинальных) распределениях этого признака $p(x_n | V_1), \dots, p(x_n | V_M)$.

Множество $X_{nm} \triangleq \{x_n : p(x_n | V_m)\}$ представляет собой носитель маргинального распределения $p(x_n | V_m)$ признака x_n , $1 \leq n \leq N$ в классе V_m , $m = 1, \dots, M$.

Часто в задачах диагностики рассматриваются только два возможных диагноза – «исправен - неисправен», «здоров - болен» [6, 9, 24, 32]. Поэтому для упрощения дальнейшего изложения ограничимся случаем дихотомии, когда $M = 2$, тем более что к дихотомии чаще всего можно свести и общий случай $M \geq 2$, последовательно решая задачу распознавания одного класса и всех остальных.

На рис. 1.1. показана четыре возможных варианта расположения носителей X_1 , X_2 маргинальных распределений $p(x | V_1)$, $p(x | V_2)$ некоторого признака x в классах V_1 и V_2 .

Первый случай иллюстрирует «идеализированную» ситуацию, которая довольно редко встречается на практике: носители классов X_1 и X_2 полностью разделяются пороговым значением x_0 . В таком случае статистическая постановка задачи теряет смысл, поскольку при любом $x < x_0$ принимаются однозначные решения в пользу V_1 , а при любом $x > x_0$ – безошибочные решения в пользу V_2 .

В ситуации 2 множества X_1 и X_2 уже частично пересекаются. В этом случае безошибочные решения могут быть приняты только для

значений $x \in (X_1 \cup X_2) / (X_1 \cap X_2) \neq \emptyset$, в то время как для значений $x \in X_1 \cap X_2 \neq \emptyset$ не могут быть приняты однозначно верные решения x .

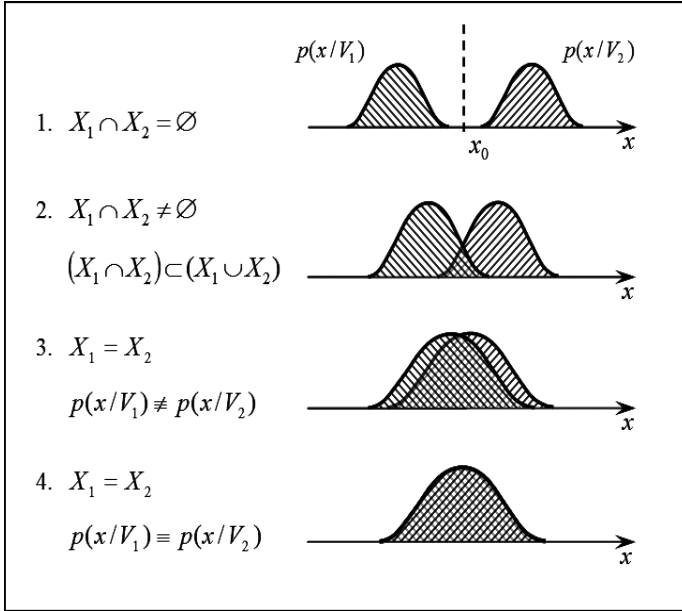


Рис.1.1. Расположения носителей условных распределений признака

В ситуации 3, когда носители распределений X_1 и X_2 совпадают, но не совпадают условные распределения $p(x|V_1)$ и $p(x|V_2)$ уже при любом значении x допускается ошибка. Условия, гарантирующие диагностическую ценность такого признака, будут исследованы в разделе 2.1.

И, наконец, в ситуации 4, когда совпадают не только множества X_1 и X_2 , но и условные распределения $p(x|V_1)$ и $p(x|V_2)$, признак сам по себе становится бесполезным. Однако в совокупности такие признаки могут быть не только полезны, но даже обеспечить безошибочное распознавание (см. рис. 1.2). Более детально математические

условия, обеспечивающие такую возможность, будут рассмотрены в разделе 1.4.

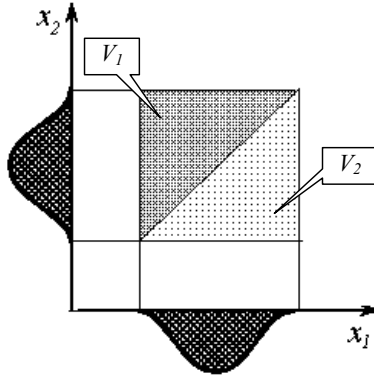


Рис. 1.2. Безошибочное распознавание классов V_1 и V_2 по совокупности бесполезных в отдельности признаков

Рассмотрим так называемый байесовский метод построения диагностического правила $s = D(x^{(N)})$, основанный на знаменитой формуле английского математика Томаса Байеса [5].

Требуется определить текущее состояние объекта $V_j \in \{V_1, V_2\}$ в предположении, что известны априорные вероятности $P(V_1), P(V_2)$ и условные распределения $p(x^{(N)} | V_1), p(x^{(N)} | V_2)$.

Пусть в результате обследования определены значения признаков $x_1 = \hat{x}_1, \dots, x_N = \hat{x}_N$. По формуле Байеса при известном $\hat{x}^{(N)} = (\hat{x}_1, \dots, \hat{x}_N)$ можно определить апостериорные вероятности

$$P(V_1 | \hat{x}^{(N)}) = \frac{P(V_1)p(\hat{x}^{(N)} | V_1)}{\sum_{m=1,2} P(V_m)p(\hat{x}^{(N)} | V_m)} \quad (1.2)$$

и

$$P(V_2 | \hat{x}^{(N)}) = \frac{P(V_2)p(\hat{x}^{(N)} | V_2)}{\sum_{m=1,2} P(V_m)p(\hat{x}^{(N)} | V_m)}. \quad (1.3)$$

В соответствии с правилом максимума апостериорный вероятностей решение в пользу V_1 принимается в том случае, когда

$$P(V_1 | \hat{x}^{(N)}) > P(V_1 | \hat{x}^{(N)}),$$

и в пользу V_2 , когда

$$P(V_1 | \hat{x}^{(N)}) < P(V_1 | \hat{x}^{(N)}).$$

Поскольку знаменатели в правых частях (1.2), (1.3) одинаковы, то такое правило эквивалентно правилу вида

$$s = \begin{cases} 1, & \text{если } \lambda(\hat{x}^{(N)}) > \lambda_0; \\ 2, & \text{если } \lambda(\hat{x}^{(N)}) < \lambda_0, \end{cases} \quad (1.4)$$

где

$$\lambda(\hat{x}^{(N)}) = \frac{p(\hat{x}^{(N)} | V_1)}{p(\hat{x}^{(N)} | V_2)} \quad (1.5)$$

– отношение правдоподобия при $x^{(N)} = \hat{x}^{(N)}$, а λ_0 – порог, определяемый отношением априорных вероятностей

$$\lambda_0 = \frac{P(V_2)}{P(V_1)}.$$

Понятно, что при фиксированном значении вектора признаков $x^{(N)} = \hat{x}^{(N)}$ условная вероятность ошибочного решения, принимаемого по правилу (1.4), определяется выражением

$$P(e | \hat{x}^{(N)}) \triangleq 1 - \max\{P(V_1 | \hat{x}^{(N)}), P(V_2 | \hat{x}^{(N)})\}. \quad (1.6)$$

Известно [17], что правило (1.4) обеспечивает минимум средней вероятности ошибки

$$P(e) \triangleq \sum_{x^{(N)} \in X^{(N)}} p(x^{(N)}) [1 - \max\{P(V_1 | x^{(N)}), P(V_2 | x^{(N)})\}] \quad (1.7)$$

на множестве $X^{(N)}$ возможных значений признаков [25].

Если потерей $L(i, j)$ ($i, j = 1, 2$) оценивать диагноз в пользу i -го класса, в то время как объект находится в j -м состоянии, то при из-

вестном отношении правдоподобия $\lambda(\hat{x}^{(N)})$ оптимальные решения, обеспечивающие минимум средних потерь (риска)

$$R = \sum_{x^{(N)} \in X^{(N)}} \sum_{j=1}^2 L(D(x^{(N)}, j)P(V_j)p(x^{(N)} | V_j)), \quad (1.8)$$

должны приниматься в соответствии с правилом [17]

$$s = \begin{cases} 1, & \text{если } \lambda(\hat{x}^{(N)}) > \theta_0; \\ 2, & \text{если } \lambda(\hat{x}^{(N)}) < \theta_0, \end{cases} \quad (1.9)$$

где

$$\theta_0 = \frac{P(V_2)(L_{21} - L_{22})}{P(V_1)(L_{12} - L_{11})}. \quad (1.10)$$

Легко видеть, что правило (1.4) является частным случаем правила (1.9), когда

$$L_{11} = L_{22} = 0 \text{ и } L_{12} = L_{21} = 1,$$

т.е. потери от правильных решений принимаются равными нулю, а от ошибочных – единице. В этом случае средний риск (1.8) вырождается в среднюю вероятность ошибочных решений (1.7).

При всей общности байесовского подхода существуют задачи, которые не могут быть выражены в его рамках, и поэтому получили известность небайесовские постановки статистической теории решений [66, с. 41]. Однако в монографии мы ограничимся лишь рассмотрением байесовских стратегий и сформулируем новые результаты, характерные именно для этой постановки задачи диагностики.

1.2. Критерии информативности признаков

Выбор адекватного множества диагностических признаков представляет собой одну из наиболее трудных проблем, с которой приходится сталкиваться конструктору систем, ориентированных на решение задач медицинской и технической диагностики [6, 12, 16, 24, 27, 94].

Главная цель селекции признаков состоит в переходе от исходного множества признаков $X^{(N_0)}$ к подмножеству $X^{(N)} \subset X^{(N_0)}$ из N

признаков ($N < N_0$) без уменьшения достоверности диагностики. В отличие от задачи конструирования исходного множества $X^{(N_0)}$ такая цель может быть достигнута с использованием формальных методов, исследованию которых посвящены многочисленные научные публикации.

Подавляющее число доступных публикаций, например, работы [10, 15, 28, 30, 38, 85, 88, 90, 92, 95, 96] посвящены исследованию процедур селекции признаков уже на этапе обучения по конечному числу наблюдений с известными состояниями объекта. Так в [15, с. 62] рассмотрен метод селекции признаков, основанный на анализе коэффициентов уравнения линейной дискриминантной функции

$$D(\cdot) = \sum_{n=1}^N \alpha_n x_n + \alpha_0,$$

построенной по имеющейся обучающей выборке наблюдений. Беспольными считаются признаки, которым соответствуют малые значения $|\alpha_n|$.

Разумеется, введение в описание признака, даже не несущего никакой информации, при неограниченной обучающей выборке не может ухудшить качество обучения. Однако при малом объеме выборки включение в описание бесполезных признаков может нанести вред.

Для иллюстрации сказанного рассмотрим наглядный пример (рис. 1.3), приведенный в работе [22, с.68].

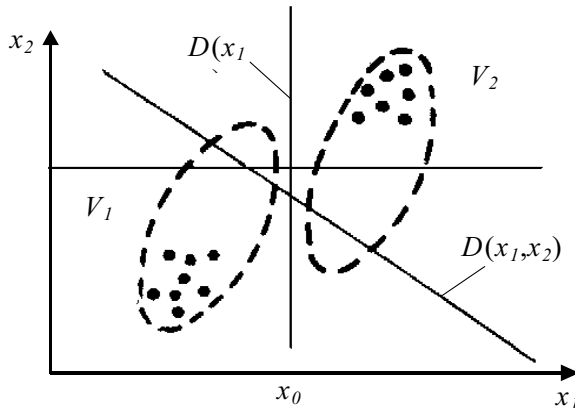


Рис.1.3. Одномерная и двумерная решающие функции $D(\cdot)$

Легко видно, что использование в линейной решающей функции $D(\cdot)$ только одного признака x_1 , гораздо лучше, чем совокупности $x = (x_1, x_2)$, которая разделяет выборку, но не разделяет генеральную совокупность объектов линейно разделимых классов V_1 и V_2 .

Поэтому важно иметь средства, позволяющие обоснованно подходить к выбору эффективных признаков еще *до этапа обучения*.

Для оценки эффективности признаков при статистической постановке задачи распознавания получил известность информационный подход [3-7, 18, 19, 86], в соответствии с которым полезность признака связывают с уменьшением неопределенности (шенноновской энтропии) [23].

При этом следует различать информативность *отдельного* признака и информативность признака *в совокупности* с другими.

Определение 1.1. Признак x_n , $1 \leq n \leq N$ информативен сам по себе, если

$$I(V, X_n) \stackrel{\Delta}{=} H(V) - H(V | X_n) > 0, \quad (1.11)$$

где $I(V, X_n)$ – количество информации, содержащееся во множестве $X_n = \{x_n : p(x_n) > 0\}$ относительно множества $V = \{V_1, V_2, \dots\}$,

$$H(V) \stackrel{\Delta}{=} - \sum_{m=1,2} P(V_m) \log P(V_m) \quad (1.12)$$

и

$$H(V | X_n) \stackrel{\Delta}{=} - \sum_{x_n \in X_n} p(x_n) \sum_{m=1,2} P(V_m | x_n) \log P(V_m | x_n) \quad (1.13)$$

соответственно начальная и средняя условная энтропии по Шеннону.

К проверке условия (1.11) сводится оценка полезности признака, предложенная впервые, по-видимому, в работе [86]. Однако такая оценка неправомерна, когда признаки статистически зависимы в классах.

Поэтому, если уж связывать полезность признака с его информативностью, то при $N \geq 2$ разумнее использовать другое определение.

Определение 1.2. Признак x_n , $1 \leq n \leq N$ информативен в совокупности с другими признаками, если при его исключении из набора

$x^{(N)} = (x_1, \dots, x_N)$ происходит изменение средней условной энтропии, т.е.

$$H(V | X^{(N-1)}) - H(V | X^{(N)}) > 0, \quad (1.14)$$

где

$$H(V | X^{(N)}) \triangleq - \sum_{x^{(N)} \in X^{(N)}} p(x^{(N)}) \sum_{m=1,2} P(V_m | x^{(N)}) \log P(V_m | x^{(N)}), \quad (1.15)$$

$$H(V | X^{(N-1)}) \triangleq - \sum_{x^{(N-1)} \in X^{(N-1)}} p(x^{(N-1)}) \sum_{m=1,2} P(V_m | x^{(N-1)}) \log P(V_m | x^{(N-1)}). \quad (1.16)$$

Здесь $x^{(N-1)}$ – сокращенный вектор признаков, который не включает x_n , а $X^{(N-1)} \triangleq \{x^{(N-1)} : p(x^{(N-1)}) > 0\}$.

Оценка (1.14) по существу была предложена в свое время в работах [3, 5]. Между тем, с точки зрения достоверности результатов диагностики более удачной оценкой полезности признака может служить не изменение средней условной энтропии $H(V | \cdot)$, а изменение средней вероятности ошибки $P(e)$ или, в общем случае, среднего риска.

Очевидно, при числе классов $M = 2$ для каждого фиксированного $x^{(N)} = \hat{x}^{(N)}$ условная вероятность ошибочных решений

$$P(e | \hat{x}^{(N)}) \triangleq 1 - \max \{P(V_1 | \hat{x}^{(N)}), P(V_2 | \hat{x}^{(N)})\}$$

однозначно определяет и частную условную энтропию

$$H(V | \hat{x}^{(N)}) \triangleq - P(e | \hat{x}^{(N)}) \log P(e | \hat{x}^{(N)}) - [1 - P(e | \hat{x}^{(N)})] \log [1 - P(e | \hat{x}^{(N)})]. \quad (1.17)$$

Но даже в этом случае между средней вероятностью ошибочных решений

$$P(e) \triangleq \sum_{x \in X} p(x) [1 - \max \{P(V_1 | x), P(V_2 | x)\}]$$

и средней условной энтропией

$$H(V | X) \triangleq - \sum_{x \in X} p(x) \sum_{m=1,2} P(V_m | x) \log P(V_m | x),$$

не существует однозначной связи [25].

В работе [25] В.А.Ковалевский определил точные верхнюю и нижнюю границы условной энтропия $H(V | X)$ при заданной вероятности ошибочных решений $P(e)$, которые понадобятся нам в дальнейших исследованиях. При $M = 2$ эти границы можно представить так

$$\sup H(V | X) = -P(e) \log P(e) - [1 - P(e)] \log [1 - P(e)], \quad (1.18)$$

$$\inf H(V | X) = 2P(e). \quad (1.19)$$

Как следует из [25], точная нижняя граница (1.19) может быть достигнута при любом фиксированном $P(e)$, $0 \leq P(e) \leq \min\{P(V_1), P(V_2)\}$, если X – множество непрерывных величин x .

Однако при решении прикладных задач часто используют дискретные признаки. Поэтому интересно знать: может ли точная нижняя граница (1.19) достигаться и в том случае, когда признак x имеет конечное число градаций.

Оказывается такая возможность существует. Этот факт в точных формулировках определяет следующая теорема.

Теорема 1.1. Точная нижняя граница (1.19) средней условной энтропии $H(V | X)$ достигается при любом значении средней вероятности ошибочных решений $P(e)$, $0 \leq P(e) \leq \min\{P(V_1), P(V_2)\}$, если X – множество дискретных значений x с числом градаций $\mu > 2$, либо только при $P(e) = 0$ и при $P(e) = \min\{P(V_1), P(V_2)\}$, если $\mu = 2$.

Доказательство. По теореме В.А.Ковалевского [25, с. 128] при числе классов $M = 2$ средняя условная энтропия $H(V | X)$ достигает своей точной нижней границы (1.19), если существует, по крайней мере, одно из двух непересекающихся подмножеств X' и X'' таких, что

$$P(e | x) = 0 \quad \forall x \in X', \quad (1.20)$$

$$P(e | x) = 0,5 \quad \forall x \in X'', \quad (1.21)$$

причем

$$X' \cup X'' = X. \quad (1.22)$$

где

$$P(e | x) \triangleq 1 - \max\{P(V_1 | x), P(V_2 | x)\}$$

по-прежнему условная вероятность ошибочных решений.

При $H(V | X) = \inf(H(V | X))$ в силу (1.19) и того, что по определению $X' \cap X'' = \emptyset$, для средней вероятности ошибочных решений $P(e)$ можно записать выражение

$$P(e) = \sum_{x \in X'} p(x)P(e | x) + \sum_{x \in X''} p(x)P(e | x),$$

которое с учетом (1.20) и (1.21) приобретает вид

$$P(e) = 0,5 \sum_{x \in X''} p(x). \quad (1.23)$$

Выражение (1.23) можно подвергнуть дальнейшему преобразованию, если принять во внимание, что условие (1.21) означает ни что иное как равенство апостериорных вероятностей классов в области X'' :

$$P(V_1 | x) = P(V_2 | x) \quad \forall x \in X''$$

или, в эквивалентной форме записи,

$$p(x | V_1)P(V_1) = p(x | V_2)P(V_2) \quad \forall x \in X''. \quad (1.24)$$

Тогда на основании формулы полной вероятности

$$p(x) = p(x | V_1)P(V_1) + p(x | V_2)P(V_2)$$

в силу (1.24) выражение (1.23) для $P(e)$ при $H(V | X) = \inf(H(V | X))$ окончательно преобразуется к виду

$$P(e) = P(V_1) \sum_{x \in X''} p(x | V_1) + P(V_2) \sum_{x \in X''} p(x | V_2). \quad (1.25)$$

Пусть X – множество дискретных x с числом градаций $\mu \geq 2$. Вполне очевидно, что независимо от числа градаций μ при выполнении условий (1.20) – (1.22) средняя условная энтропия $H(V | X)$ достигает своей точной нижней границы (1.19), совпадающей с ее точной верхней границей (1.18), для двух крайних значений $P(e)$, а именно $P(e) = 0$, когда $X'' = \emptyset$ (безошибочное распознавание) и $P(e) = 0,5$, когда $X' = \emptyset$ (случай совпадающих распределений $p(x | V_1) \equiv p(x | V_2)$ при равновероятных классах $P(V_1) = P(V_2) = 0,5$). Поэтому доказательство теоремы 2.1 сводится к определению усло-

вий, когда $H(V|X) = \inf(H(V|X))$ при $0 < P(e) < 0,5$, или, что то же самое, при ограничениях

$$X' \neq \emptyset \text{ и } X'' \neq \emptyset. \quad (1.26)$$

В соответствии с (1.20) условные распределения $P(x|V_1)$ и $p(x|V_2)$ в области X' должны быть такими, чтобы $\forall x \in X'$ одна из апостериорных вероятностей $P(V_1|x)$ или $P(V_2|x)$ обращалась в нуль. А это с учетом ограничения $X' \neq \emptyset$ возможно лишь в том случае, когда

$$p(x|V_1) > 0 \text{ и } p(x|V_2) = 0 \quad \forall x \in X' \quad (1.27)$$

или же

$$p(x|V_1) = 0 \text{ и } p(x|V_2) > 0 \quad \forall x \in X'. \quad (1.28)$$

Заметим, что при ограничениях (1.26) для числа μ_1 градаций x , образующих подмножество $X' \neq \emptyset$, справедлива оценка

$$1 \leq \mu_1 \leq \mu - 1. \quad (1.29)$$

Принимая во внимание, что $\mu \geq 2$ рассмотрим отдельно два возможных случая, когда $\mu = 2$ и $\mu > 2$.

Пусть вначале $\mu = 2$. Тогда в соответствии с (1.29) $\mu_1 = 1$. Очевидно, что для единственного значения $x \in X'$ правомерно потребовать либо выполнение условия (1.27) либо условия (1.28).

Не умаляя общности, положим для определенности, что $\forall x \in X'$ выполняется условие (1.27). В этом случае

$$\sum_{x \in X'} p(x|V_1) > 0 \text{ и } \sum_{x \in X'} p(x|V_2) = 0. \quad (1.30)$$

Поскольку $X' \cap X'' = \emptyset$ и $X' \cup X'' = X$, то, в силу очевидного тождества

$$\sum_{x \in X'} p(x|V_m) + \sum_{x \in X''} p(x|V_m) = 1, \quad m = 1, 2, \quad (1.31)$$

из (1.30) следует, что

$$\sum_{x \in X''} p(x|V_1) < 1, \quad \sum_{x \in X''} p(x|V_2) = 1. \quad (1.32)$$

На основании (1.25) с учетом (1.32) заключаем, что если $\mu = 2$, то при выполнении условий (1.27) $H(V | X) = \inf(H(V | X))$ лишь в том случае, когда

$$P(e) = P(V_2) < P(V_1). \quad (1.33)$$

Аналогичным образом можно показать, что если при $\mu = 2$ вместо (1.27) выполняется условие (1.28), то отсюда немедленно следует, что $H(V | X) = \inf(H(V | X))$ лишь в том случае, когда

$$P(e) = P(V_1) < P(V_2). \quad (1.34)$$

На основании (1.33) и (1.34) заключаем, что для случая $\mu = 2$ точная нижняя граница (1.19) достигается только при

$$P(e) = \min\{P(V_1), P(V_2)\}. \quad (1.35)$$

Рассмотрим теперь второй возможный случай, когда $\mu > 2$. В этом случае из (1.29) следует, что $\mu_1 \geq 1$, т.е. подмножество X' уже может быть образовано несколькими значениями x .

При $\mu_1 > 1$ правомерно потребовать, чтобы для одних значений $x \in X'$ выполнялось условие (1.27), а для остальных значений $x \in X'$ – условие (1.28). В этом случае

$$\sum_{x \in X'} p(x | V_1) > 0 \text{ и } \sum_{x \in X'} p(x | V_2) > 0$$

и, в силу (1.31),

$$\sum_{x \in X'} p(x | V_1) < 1, \quad \sum_{x \in X'} p(x | V_2) < 1. \quad (1.36)$$

Принимая во внимание неравенства (1.36) на основании (1.25) заключаем, что если в случае $\mu > 2$ для одних значений $x \in X'$ выполняется условие (1.27), а для остальных значений $x \in X'$ – условие (1.28), то средняя вероятность ошибочных решений $P(e)$, при которой $H(V | X) = \inf(H(V | X))$, одновременно удовлетворяет условиям

$$P(e) < P(V_1) \text{ и } P(e) < P(V_2),$$

а значит

$$P(e) < \min\{P(V_1), P(V_2)\}. \quad (1.37)$$

Если к тому же учесть, что условные распределения $p(x|V_1)$ и $p(x|V_2)$ в области X'' связаны лишь единственным ограничением (1.24), всегда можно задать эти распределения таким образом, чтобы величина $P(e)$, которая в рассматриваемом случае вычисляется в соответствии с (1.25), удовлетворяла условию $P(e) = \beta$, где β – любое наперед заданное положительное число. Теорема 1.1. доказана.

Примечательно, что для случая $\mu = 2$ Ю.Л. Барабаш [2, с. 112] получил другое выражение для точной нижней границы

$$\inf H(V | X) = -P_0(e) \log \frac{P_0(e)}{P_0(e) + P(e)} - P(e) \log \frac{P(e)}{P_0(e) + P(e)}, \quad \mu = 2, N = 1, \quad (1.38)$$

которая, в отличие от (1.19), достигается при любом значении $P(e)$.

Здесь, как и прежде, $P_0(e) \triangleq \min\{P(V_1), P(V_2)\}$.

Как и следовало ожидать (рис. 2.1.), оценки (1.19) и (1.38) совпадают только при $P(e) = 0$ и при $P(e) = P_0(e) \triangleq \min\{P(V_1), P(V_2)\}$, а при всех остальных значениях $P(e)$, $0 < P(e) < P_0(e)$ оценка (1.38) лежит выше оценки (1.19).

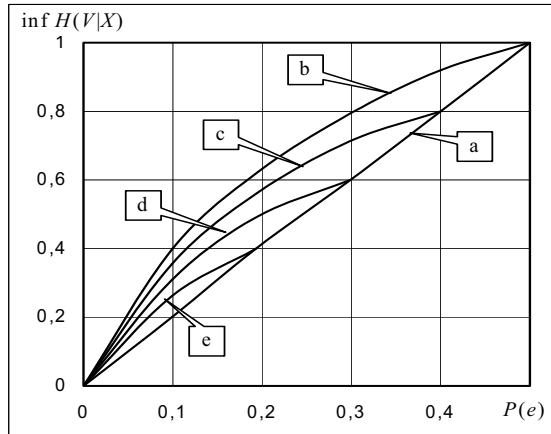


Рис.1.3. Нижние границы средней условной энтропии $H(V | X)$

а – согласно выражению (1.19) и выражению (1.38)

при б – $P_0(e)=0,5$; в – $P_0(e)=0,4$; д – $P_0(e)=0,3$; е – $P_0(e)=0,2$;

1.3. Безусловная и условная независимость признаков

Понятно, что при отсутствии полной априорной информации о многомерных распределениях $p(x^{(N)} | V_m)$ невозможно непосредственно воспользоваться приведенными выше соотношениями для расчета $P(e)$ и $I(V, X^{(N)})$. Максимум, на что можно рассчитывать при решении прикладных задач – получение оценок априорных вероятностей $\hat{P}(V_m)$, например, знание распространенности (преваленса) исследуемого заболевания [12], оценок одномерных эмпирических распределений (гистограмм), построенных для отдельных признаков [13] и, в лучшем случае – коэффициентов корреляций между отдельными признаками.

Как будет показано далее при решении ряда прикладных задач из физических закономерностей процессов, происходящих в объекте, могут быть известны дополнительные априорные сведения, которые и хотелось бы использовать для оценки полезности признаков.

В качестве таких ограниченных априорных сведений будем рассматривать следующие индивидуальные и совместные характеристики признаков:

а) имеет ли какой либо признак $x_n, 1 \leq n \leq N$ одинаковые или же различные распределения в классах, т.е.

$$p(x_n | V_1) \equiv p(x_n | V_2), \quad (1.38)$$

или же

$$p(x_n | V_1) \neq p(x_n | V_2), \quad (1.39)$$

б) существует ли статистическая связь между x_n и остальными $N - 1$ признаками в k -м классе, т.е. выполняется условие

$$p(x_n | x^{(N-1)}, V_k) \equiv p(x_n | V_k), \quad (1.40)$$

или же условие

$$p(x_n | x^{(N-1)}, V_k) \neq p(x_n | V_k). \quad (1.41)$$

Рассмотрим некоторые довольно любопытные свойства признаков, устанавливающие взаимосвязь (1.38)-(1.41).

Пусть x_n и x_m ($1 \leq n \leq N, 1 \leq m \leq N, n \neq m$) – любые два признака в наборе $x^{(N)} = (x_1, \dots, x_N)$. Следуя [2, 6, 65] будем различать *безусловную* статистическую независимость x_n и x_m , когда

$$p(x_n, x_m) \equiv p(x_n)p(x_m), \quad (1.42)$$

и *условную* статистическую независимость x_n и x_m в k -м классе, когда

$$p(x_n, x_m | V_k) \equiv p(x_n | V_k)p(x_m | V_k). \quad (1.43)$$

Известно [2] известно, что соотношения (1.42) и (1.43) не могут выполняться совместно, если признаки x_n и x_m имеют различные распределения в классах, т.е.

$$p(x_n | V_1) \neq p(x_n | V_2) \text{ и } p(x_m | V_1) \neq p(x_m | V_2). \quad (1.44)$$

Тем не менее, из этого результата той форме, в которой он был сформулирован в [2, с. 90] еще не ясно, обязательно ли из условий (1.42) и (1.44) вытекает статистическая зависимость x_n и x_m сразу в обоих классах.

Поэтому нам понадобилось провести дополнительные исследования [54], позволяющие несколько усилить формулировку этого результата. Попутно был обнаружен целый ряд других, не менее любопытных свойств, формулировке и доказательству которых предположим следующую лемму.

Лемма 1.1. Пусть в N -мерном пространстве R^N случайных векторов $x^{(N)} = (x_1, \dots, x_N)$ с заданным на нем многомерным распределением $p(x^{(N)})$ выделено множество

$$X^{(N)} = \{x^{(N)} : p(x^{(N)}) > 0\}, \quad (1.45)$$

включающее те и только те значения $x^{(N)} \in R^{(N)}$, для которых $p(x^{(N)})$ отлично от нуля. Тогда, для того чтобы компоненты x_1, \dots, x_N векторов $x^{(N)}$ были статистически независимы необходимо, чтобы множество $X^{(N)}$ представляло собой прямое произведение всех своих проекций X_1, \dots, X_N соответственно на множества значений x_1, \dots, x_N , т.е. чтобы

$$X^{(N)} = X_1 \times \dots \times X_N. \quad (1.46)$$

Доказательство. Если x_1, \dots, x_N статистически независимы, то непременно

$$p(x^{(N)}) \triangleq p(x_1, \dots, x_N) = p(x_1) \cdots p(x_N), \quad (1.47)$$

где $p(x_n)$ – одномерное распределение n -й ($n = 1, \dots, N$) компоненты x_n вектора $x^{(N)}$.

Поскольку любой вектор $x^{(N)}$, n -я компонента которого принадлежит проекции X_n множества $X^{(N)}$, заведомо принадлежит самому множеству $X^{(N)}$, а распределение $p(x_n)$ является маргинальным по отношению к $p(x^{(N)})$, то в силу определения (1.45) имеем

$$p(x_n) > 0 \quad \forall x_n \in X_n, \quad n = 1, \dots, N. \quad (1.48)$$

На основании (1.47) и (1.48) заключаем, что

$$p(x^{(N)}) > 0 \quad \forall x^{(N)} \in X_1 \times \cdots \times X_N. \quad (1.49)$$

Поскольку же всегда

$$X^{(N)} \subseteq X_1 \times \cdots \times X_N, \quad (1.50)$$

а по определению (1.45) множество $X^{(N)}$ включает те и только те значения $x^{(N)}$, для которых $p(x^{(N)}) > 0$, то из (1.49) с учетом (1.50) как раз и следует соотношение (1.46). Лемма 1.1 доказана.

Свойство 1.1. Если признаки безусловно независимы и, кроме того, каждый из этих признаков имеет различные распределения в классах, то, по крайней мере, в одном из классов эти признаки будут статистически зависимыми, т.е.

$$p(x_n, x_m | V_1) \neq p(x_n | V_1)p(x_m | V_1) \quad (1.51)$$

и (или)

$$p(x_n, x_m | V_2) \neq p(x_n | V_2)p(x_m | V_2). \quad (1.52)$$

Доказательство. Воспользовавшись формулой полной вероятности, выразим распределения, фигурирующие в (1.42), через соответствующие распределения в классах

$$p(x_n, x_m) \equiv P(V_1)p(x_n, x_m | V_1) + P(V_2)p(x_n, x_m | V_2), \quad (1.53)$$

$$p(x_n) \equiv P(V_1)p(x_n | V_1) + P(V_2)p(x_n | V_2), \quad (1.54)$$

$$p(x_m) \equiv P(V_1)p(x_m | V_1) + P(V_2)p(x_m | V_2). \quad (1.55)$$

Подстановка выражений $p(x_n, x_m)$, $p(x_n)$ и $p(x_m)$ из (1.53) – (1.55) в (1.42) дает

$$\begin{aligned} & P(V_1)p(x_n, x_m | V_1) + P(V_2)p(x_n, x_m | V_2) \equiv \\ & \equiv P(V_1)[P(V_1)p(x_n | V_1)p(x_m | V_1) + P(V_2)p(x_n | V_1)p(x_m | V_2)] + \\ & + P(V_2)[P(V_1)p(x_n | V_2)p(x_m | V_1) + P(V_2)p(x_n | V_2)p(x_m | V_2)] \end{aligned} \quad (1.56)$$

На основании условия (1.44) можно заключить, что

$$X_n^0 \neq \emptyset, \quad X_m^0 \neq \emptyset, \quad (1.57)$$

где

$$X_n^0 \triangleq \{x_n : p(x_n | V_1) - p(x_n | V_2) \neq 0\}, \quad (1.58)$$

$$X_m^0 \triangleq \{x_m : p(x_m | V_1) - p(x_m | V_2) \neq 0\}. \quad (1.59)$$

Поскольку в силу (1.42) признаки x_n и x_m – независимы, то, согласно лемме 1.2, множество $X^{(2)}$, на котором $p(x_n, x_m) > 0$, удовлетворяет условию

$$X^{(2)} = X_n \times X_m. \quad (1.60)$$

Из условия (1.60) с учетом (1.57) и того, что $X_n^0 \subset X_n$ и $X_m^0 \subset X_m$, следует $X_0^{(2)} \neq \emptyset$, где

$$X_0^{(2)} \triangleq \{x^{(2)} \equiv (x_n, x_m) : x_n \in X_n^0, x_m \in X_m^0\},$$

а это означает, что непременно имеет место соотношение

$$[p(x_n | V_1) - p(x_n | V_2)][p(x_m | V_1) - p(x_m | V_2)] \neq 0. \quad (1.61)$$

Соотношение (1.61) в силу того, что $P(V_1) + P(V_2) = 1$, после несложных преобразований может быть представлено в виде

$$\begin{aligned}
& P(V_1)[P(V_1)p(x_n | V_1)p(x_m | V_1) + P(V_2)p(x_n | V_1)p(x_m | V_1)] + \\
& + P(V_2)[P(V_1)p(x_n | V_2)p(x_m | V_2) + P(V_2)p(x_n | V_2)p(x_m | V_2)] \neq \\
& \neq P(V_1)[P(V_1)p(x_n | V_1)p(x_m | V_1) + P(V_2)p(x_n | V_1)p(x_m | V_2)] + \\
& + P(V_2)[P(V_1)p(x_n | V_2)p(x_m | V_1) + P(V_2)p(x_n | V_2)p(x_m | V_2)]. \quad (1.62)
\end{aligned}$$

Как видно, в правых частях (1.56) и (1.62) фигурирует одно и то же выражение. Отсюда можно заключить, что левые части (1.56) и (1.62) связаны соотношением

$$\begin{aligned}
& P(V_1)p(x_n, x_m | V_1) + P(V_2)p(x_n, x_m | V_2) \neq \\
& \neq P(V_1)[P(V_1)p(x_n | V_1)p(x_m | V_1) + P(V_2)p(x_n | V_1)p(x_m | V_1)] + \\
& + P(V_2)[P(V_1)p(x_n | V_2)p(x_m | V_2) + P(V_2)p(x_n | V_2)p(x_m | V_2)],
\end{aligned}$$

которое после очевидных преобразований можно записать так:

$$\begin{aligned}
& P(V_1)p(x_n, x_m | V_1) + P(V_2)p(x_n, x_m | V_2) \neq \\
& \neq P(V_1)p(x_n | V_1)p(x_m | V_1) + P(V_2)p(x_n | V_2)p(x_m | V_2). \quad (1.63)
\end{aligned}$$

Предположим, что в одном из классов, например, в классе V_1 , признаки x_n и x_m статистически независимы, т.е.

$$p(x_n, x_m | V_1) \equiv p(x_n | V_1)p(x_m | V_1). \quad (1.64)$$

При таком предположении из (1.63) с учетом (1.64) сразу же следует, что

$$p(x_n, x_m | V_2) \neq p(x_n | V_2)p(x_m | V_2). \quad (1.65)$$

А это как раз и означает, что в другом классе признаки x_n и x_m оказываются уже статистически зависимыми.

Разумеется, соотношение (1.63) может выполняться и в том случае, когда совместно с соотношением (1.65) выполняется и соотношение

$$p(x_n, x_m | V_1) \neq p(x_n | V_1)p(x_m | V_1),$$

т.е. когда и в одном, и в другом классе признаки x_n и x_m окажутся условно зависимыми. Именно такая трактовка свойства 2.1 дана была в работе [2, с. 90]. Между тем оказывается, что наличие статистической зависимости между x_n и x_m сразу в *обоих* классах вовсе не явля-

ется обязательным: при выполнении условий (1.42) и (1.44) условие (1.43) может иметь место при $k=1$ или же при $k=2$. Свойство 1.2 доказано.

Пример 1.2. Пусть $P(V_1) = 0,8$; $P(V_2) = 0,2$. Пусть признаки x_n и x_m имеют две градации x_n^1, x_n^2 и x_m^1, x_m^2 . В таблице 1.1 представлено совместное распределение $p(x_n, x_m)$, которое удовлетворяет условию безусловной независимости (1.42).

Таблица 1.1.

Безусловное распределение $p(x_n, x_m)$

x_m	x_n		$p(x_m)$
	$x_n = x_n^1$	$x_n = x_n^2$	
$x_m = x_m^1$	0,45	0,15	0,60
$x_m = x_m^2$	0,30	0,10	0,40
$p(x_n)$	0,75	0,25	-

Условные распределения $p(x_n, x_m | V_1)$, $p(x_n, x_m | V_2)$ представлены в таблицах 1.2, 1.3. Легко убедиться в том, что признаки остаются условно независимыми в первом классе, но условно зависимы во втором.

Таблица 1.2.

Условное распределение $p(x_n, x_m | V_1)$

x_m	x_n		$p(x_m V_1)$
	$x_n = x_n^1$	$x_n = x_n^2$	
$x_m = x_m^1$	0,44	0,11	0,55
$x_m = x_m^2$	0,36	0,09	0,45
$p(x_n V_1)$	0,80	0,20	-

Таблица 1.3.

Условное распределение $p(x_n, x_m | V_2)$

x_m	x_n		$p(x_m V_2)$
	$x_n = x_n^1$	$x_n = x_n^2$	
$x_m = x_m^1$	0,49	0,31	0,80
$x_m = x_m^2$	0,06	0,14	0,20
$p(x_n V_2)$	0,55	0,45	-

Свойство 1.2. Пусть признаки x_n и x_m безусловно независимы и, по крайней мере, один из этих признаков имеет одинаковые распределения в классах. Тогда, если x_n и x_m условно независимы в одном классе, то они непременно будут независимыми и в другом классе, и наоборот, если x_n и x_m условно зависимы в одном классе, то они будут зависимыми и в другом классе.

Доказательство. Положим для определенности, что признаком, который в соответствии с условием свойства имеет одинаковые распределения в классах, является признак x_n , т.е.

$$p(x_n | V_1) \equiv p(x_n | V_2). \quad (1.66)$$

Тогда, повторяя выкладки, используемые при доказательстве свойства 1, из условия (1.42) с учетом (1.53) – (1.55) приходим к соотношению (1.56), которое в силу (1.66) теперь уже может быть преобразовано к виду

$$\begin{aligned} & P(V_1)p(x_n, x_m | V_1) + P(V_2)p(x_n, x_m | V_2) \equiv \\ & \equiv P(V_1)p(x_n | V_1)p(x_m | V_1) + P(V_2)p(x_n | V_2)p(x_m | V_2) \end{aligned} \quad (1.67)$$

или, что то же самое,

$$\begin{aligned} & P(V_1)[p(x_n, x_m | V_1) - p(x_n | V_1)p(x_m | V_1)] \equiv \\ & \equiv P(V_2)[p(x_n, x_m | V_2) - p(x_n | V_2)p(x_m | V_2)]. \end{aligned} \quad (1.68)$$

Предположим, что в одном из классов, например, в классе V_1 , признаки x_n и x_m статистически независимы, т.е.

$$p(x_n, x_m | V_1) \equiv p(x_n | V_1)p(x_m | V_1).$$

Тогда из (1.68) сразу же следует, что

$$p(x_n, x_m | V_2) \equiv p(x_n | V_2)p(x_m | V_2),$$

т.е. x_n и x_m оказываются независимыми и в классе V_2 .

Если же предположить, что в классе V_1 , признаки x_n и x_m статистически зависимы, т.е.

$$p(x_n, x_m | V_1) \not\equiv p(x_n | V_1)p(x_m | V_1),$$

то в этом случае из (1.68) немедленно следует зависимость x_n и x_m в классе V_2 , поскольку

$$p(x_n, x_m | V_2) \neq p(x_n | V_2)p(x_m | V_2).$$

Тем самым завершается доказательство свойства 1.2.

Заметим, что в соответствии со свойством 1.2 безусловно независимые признаки x_n и x_m могут оказаться условно зависимыми (причем непременно в обоих классах), даже если оба эти признака имеют одинаковые распределения в классах, т.е.

$$p(x_n | V_1) \equiv p(x_n | V_2), \quad p(x_m | V_1) \equiv p(x_m | V_2). \quad (1.69)$$

В этом особом случае справедливо такое дополнительное свойство.

Свойство 1.3. Пусть признаки x_n и x_m безусловно независимы и оба имеют одинаковые распределения в классах, т.е. пусть выполняется условие (1.69). Тогда если эти признаки условно зависимы в классах, то

$$p(x_m | x_n, V_1) \neq p(x_m | x_n, V_2), \quad (1.70)$$

иначе говоря, зависимость между x_n и x_m различна в классах.

Доказательство. Соотношение (1.67), полученное при доказательстве свойства 1.2., с учетом того, что $p(x_n | V_1) \equiv p(x_n | V_2)$ можно представить так

$$\begin{aligned} P(V_1)p(x_n, x_m | V_1) + P(V_2)p(x_n, x_m | V_2) &\equiv \\ &\equiv P(V_1)p(x_n | V_k)p(x_m | V_k), \quad k = 1, 2. \end{aligned}$$

Отсюда в силу известного тождества

$$p(x_n, x_m | V_k) \equiv p(x_n | V_k)p(x_m | x_n, V_k)$$

с учетом (1.69) имеем

$$P(V_1)p(x_m | x_n, V_k) + P(V_2)p(x_m | x_n, V_k) \equiv p(x_m | V_k), \quad k = 1, 2. \quad (1.71)$$

Предположим, что сформулированное свойство неверно, т.е.

$$p(x_m | x_n, V_1) \equiv p(x_m | x_n, V_2). \quad (1.72)$$

При таком предположении из (1.71) с учетом (1.72) получим

$$p(x_m | x_n) \equiv p(x_m | x_n, V_k), \quad k = 1, 2.$$

Но это означало бы, что признаки x_n и x_m статистически независимы в классах, что противоречит условию свойства.

Полученное противоречие подтверждает справедливость свойства 1.3.

Применяя аналогичную технику доказательства нетрудно убедиться в справедливости и таких свойств.

Свойство 1.4. Если признаки x_n и x_m безусловно независимы и, кроме того, условно независимы в обоих классах, то по крайней мере один из этих признаков имеет одинаковые распределения в классах, т.е.

$$p(x_n | V_1) \equiv p(x_n | V_2) \quad \text{и / или} \quad p(x_m | V_1) \equiv p(x_m | V_2). \quad (1.73)$$

Свойство 1.5. Если признаки x_n и x_m условно независимы в обоих классах, т.е.

$$p(x_n, x_m | V_k) \equiv p(x_n | V_k)p(x_m | V_k), \quad k = 1, 2$$

и оба эти признака имеют различные распределения в классах, т.е.

$$p(x_n | V_1) \not\equiv p(x_n | V_2) \quad \text{и} \quad p(x_m | V_1) \not\equiv p(x_m | V_2),$$

то x_n и x_m безусловно зависимы

$$p(x_n, x_m) \not\equiv p(x_n)p(x_m). \quad (1.74)$$

Свойство 1.6. Если любой из признаков x_n и x_m имеет одинаковые распределения в классах и, кроме того, x_n и x_m условно независимы только в одном из классов, то эти признаки безусловно зависимы в смысле (1.74).

Свойство 1.7. Если любой из признаков x_n и x_m имеет одинаковые распределения в классах и, кроме того, x_n и x_m условно независимы в обоих классах, то эти признаки безусловно независимы.

Свойство 1.8. Если признаки x_n и x_m условно независимы в обоих классах, но безусловно зависимы в смысле (1.74), то каждый из этих признаков имеет различные распределения в классах

$$p(x_n | V_1) \neq p(x_n | V_2) \quad \text{и} \quad p(x_m | V_1) \neq p(x_m | V_2).$$

Все приведенные выше свойства естественным образом могут быть обобщены на любые две совокупности признаков исходного набора $x^{(N)} = (x_1, \dots, x_N)$. В частности, эти свойства можно распространить на случай, когда рассматривается какой либо один признак x_n , $1 \leq n \leq N$ и сокращенный набор $x^{(N-1)}$, который не содержит этот признак.

1.4. Безошибочное распознавание по совокупности пересекающихся признаков

Пусть каждый признак x_n , $1 \leq n \leq N$ в наборе $x^{(N)} = (x_1, \dots, x_N)$ имеет пересекающиеся условные распределения $p(x_n | V_1)$ и $p(x_n | V_2)$ в классах V_1 и V_2 , иначе говоря, пусть множества

$$X_{1n} \triangleq \{x_n : p(x_n | V_1) > 0\} \tag{1.75}$$

и множества

$$X_{2n} \triangleq \{x_n : p(x_n | V_2) > 0\} \tag{1.76}$$

заведомо пересекаются

$$X_{1n} \cap X_{2n} \neq \emptyset \quad \forall n = 1, \dots, N \tag{1.77}$$

Ясно, что безошибочное распознавание по каждому в отдельности такому признаку невозможно. В то же время при $N \geq 2$, вообще говоря, возможны случаи, когда совокупность признаков $x^{(N)} = (x_1, \dots, x_N)$ позволяет провести безошибочное распознавание, несмотря на ограничение (1.77) [59]. Иллюстрацией такой возможности служит приведенный выше рисунок 1.2.

Возникает вопрос: каким формальным условиям должны удовлетворять пересекающиеся признаки, чтобы можно было рассчитывать на возможность безошибочного распознавания по их совокупности.

Покажем, что даже если каждый из N признаков набора $x^{(N)} = (x_1, \dots, x_N)$ имеет одинаковые распределения в классах, когда

$$p(x_n | V_1) \equiv p(x_n | V_2) \quad \forall n = 1, \dots, N \quad (1.78)$$

и даже если сами по себе эти признаки статистически независимы, когда

$$p(x^{(N)}) \stackrel{\Delta}{=} p(x_1, \dots, x_N) \equiv \prod_{n=1}^N p(x_n), \quad (1.79)$$

то и в этом, казалось бы, совсем «безнадежном» случае существует принципиальная возможность безошибочного распознавания классов V_1 и V_2 по совокупности $x^{(N)} = (x_1, \dots, x_N)$ [54].

Теорема 1.2. Пусть выполняется условие (1.77), т.е. каждый из N признаков набора $x^{(N)} = (x_1, \dots, x_N)$ имеет *пересекающиеся* распределения в классах.. Тогда для безошибочного распознавания классов V_1 и V_2 классов по совокупности $x^{(N)} = (x_1, \dots, x_N)$ необходимо, чтобы, по крайней мере, в одном из классов признаки были статистически зависимыми, т.е.

$$P(x^{(N)} | V_k) \neq \prod_{n=1}^N p(x_n | V_k), \quad k = 1 \text{ и/или } k = 2. \quad (1.80)$$

Доказательство. Рассмотрим множества

$$S_1 = X_{11} \times \dots \times X_{1N}, \quad (1.81)$$

$$S_2 = X_{21} \times \dots \times X_{2N}, \quad (1.82)$$

образованные прямым произведением всех N множеств (1.75) и (1.76) соответственно.

Поскольку условия (1.77) выполняются для всех $n = 1, \dots, N$, то множества S_1 и S_2 , образованные в соответствии с (1.82), (1.83), непрерывно пересекаются:

$$S_1 \cap S_2 \neq \emptyset. \quad (1.83)$$

Допустим, что в одном из классов, например в классе V_1 , признаки x_1, \dots, x_N статистически независимы, т.е.

$$p(x^{(N)} | V_1) \equiv \prod_{n=1}^N p(x_n | V_1). \quad (1.84)$$

Тогда в силу Леммы 1.1, доказанной в предыдущем разделе, собственная область этого класса $X_1^{(N)} = \{x^{(N)} : p(x^{(N)} | V_1) > 0\}$ должна быть образована прямым произведением всех своих проекций на множества значений x_1, \dots, x_N , иными словами, прямым произведением $X_{11} \times \dots \times X_{1N}$. Принимая во внимание (1.81), заключаем, что в этом случае $S_1 \setminus X_1^{(N)} = \emptyset$, откуда на основании (1.83) имеем

$$X_1^{(N)} \cap S_2 \neq \emptyset. \quad (1.85)$$

Вполне очевидно, что для безошибочного распознавания классов V_1 и V_2 классов по совокупности $x^{(N)} = (x_1, \dots, x_N)$ необходимо, чтобы множество $X_1^{(N)}$ не пересекалось с множеством

$$X_2^{(N)} = \{x^{(N)} : p(x^{(N)} | V_2) > 0\},$$

представляющим собой собственную область второго класса в многомерном пространстве $X^{(N)}$, т.е. чтобы

$$X_1^{(N)} \cap X_2^{(N)} = \emptyset. \quad (1.86)$$

Легко видеть, что условия (1.85) и (1.86) могут выполняться совместно, в том и только лишь в том случае, когда

$$S_2 \setminus X_2^{(N)} \neq \emptyset.$$

А это в силу леммы 1.2 с учетом определения (1.82) означает, что в классе V_2 признаки x_1, \dots, x_N должны уже быть статистически зависимыми. Теорема 1.2 доказана.

Теорема 1.3. Пусть выполняется условие (1.78), т.е. каждый из N признаков набора $x^{(N)} = (x_1, \dots, x_N)$ имеет *одинаковые* распределения в классах. Тогда для безошибочного распознавания классов V_1 и V_2 классов по совокупности $x^{(N)} = (x_1, \dots, x_N)$ необходимо, чтобы сразу в обоих классах признаки были статистически зависимыми, т.е.

$$P(x^{(N)} | V_k) \neq \prod_{n=1}^N p(x_n | V_k), \quad k = 1 \text{ и } k = 2. \quad (1.87)$$

Доказательство. Прежде всего заметим, что условия (1.78) могут выполняться в том случае, когда

$$X_{1n} \setminus X_{2n} = \emptyset \quad \forall n = 1, \dots, N. \quad (1.88)$$

Действительно, если хотя бы для одного из признаков множества X_{1n} и X_{2n} не совпадали, то отсюда следовало бы, что существуют такие значения этого признака, при которых условное распределение в одном классе равно нулю, а в другом – отлично от нуля. Но это противоречило бы (1.78).

Принимая во внимание (1.88) на основании (1.81) и (1.82) заключаем

$$S_1 \setminus S_2 \neq \emptyset. \quad (1.89)$$

Допустим, что в одном из классов, например в классе V_1 , признаки x_1, \dots, x_N статистически независимы, т.е. выполняется условие (1.84). Тогда в силу леммы 1.2. с учетом (1.81) имеем

$$X_1^{(N)} \setminus S_2 \neq \emptyset. \quad (1.90)$$

Поскольку в соответствии с определением (1.82) множества S_2 непременно $X_2^{(N)} \subseteq S_2$ и, разумеется, $X_2^{(N)} \neq \emptyset$, то из условия (1.90) следует

$$X_1^{(N)} \cap X_2^{(N)} \neq \emptyset. \quad (1.91)$$

Но это означает, что безошибочное распознавание V_1 и V_2 классов по совокупности $x^{(N)} = (x_1, \dots, x_N)$ невозможно. Следовательно, в условиях данной теоремы безошибочное распознавание становится невозможным, если хотя бы в одном из классов признаки x_1, \dots, x_N статистически независимы. Теорема 1.3 доказана.

Следует обратить внимание, что для выполнения условий безошибочного распознавания, о которых речь шла в теоремах 1.2. и 1.3, вовсе не обязательно наличие безусловной статистической зависимости между пересекающимися признаками x_1, \dots, x_N . В самом деле, если каждый из этих признаков имеет различные распределения в классах, а сами признаки безусловно независимы, то, согласно Свойству 1.1, по крайней мере, в одном из классов эти признаки будут условно за-

висимыми. А это как раз в силу теоремы 1.2 и является необходимым условием безошибочного распознавания.

Точно так же, даже если выполняются условия (1.78), то, согласно Свойству 1.2, безусловно независимые признаки могут оказаться условно зависимыми сразу в обоих классах, причем в силу Свойства 1.3 эта зависимость непременно различна в классах. Следовательно, и в этом случае может выполняться необходимое условие безошибочного распознавания, сформулированное теоремой 1.3.

Пример 1.2. Пусть $P(V_1) = 0,8$; $P(V_2) = 0,2$, $N = 2$, т.е. в нашем распоряжении имеется два признака x_1 и x_2 , Пусть каждый из них имеет четыре градации: $x_1^1, x_1^2, x_1^3, x_1^4$ и $x_2^1, x_2^2, x_2^3, x_2^4$.

В таблицах 1.4, 1.5 представлены Условные распределения $p(x_1, x_2 | V_1)$, $p(x_1, x_2 | V_2)$.

Таблица 1.4.

Условное распределение $p(x_1, x_2 | V_1)$

x_2	x_1				$p(x_2 V_1)$
	$x_1 = x_1^1$	$x_1 = x_1^2$	$x_1 = x_1^3$	$x_1 = x_1^4$	
$x_2 = x_2^1$	0	0	0,1	0,025	0,1250
$x_2 = x_2^2$	0	0	0,3	0,075	0,3750
$x_2 = x_2^3$	0,15	0,0375	0	0	0,1875
$x_2 = x_2^4$	0,25	0,0625	0	0	0,3125
$p(x_1 V_1)$	0,4	0,1	0,4	0,1	-

Таблица 1.5.

Условное распределение $p(x_1, x_2 | V_2)$

x_2	x_1				$p(x_2 V_2)$
	$x_1 = x_1^1$	$x_1 = x_1^2$	$x_1 = x_1^3$	$x_1 = x_1^4$	
$x_2 = x_2^1$	0,1	0,025	0	0	0,1250
$x_2 = x_2^2$	0,3	0,075	0	0	0,3750
$x_2 = x_2^3$	0	0	0,15	0,0375	0,1875
$x_2 = x_2^4$	0	0	0,25	0,0625	0,3125
$p(x_1 V_2)$	0,4	0,1	0,4	0,1	-

Легко видно, что оба признака имеют одинаковые маргинальные распределения в классах. Тем не менее по совокупности таких признаков можно провести *безошибочное* распознавание классов, используя следующее правило

$$s \stackrel{\Delta}{=} D(x_1, x_2) = \begin{cases} 1, & \text{если } (x_1, x_2) \in \Omega_1, \\ 2, & \text{если } (x_1, x_2) \in \Omega_2, \end{cases}$$

где

$$\Omega_1 = \{(x_1^1, x_2^3), (x_1^1, x_2^4), (x_1^2, x_2^3), (x_1^2, x_2^4), (x_1^3, x_2^1), (x_1^3, x_2^2), (x_1^4, x_2^1), (x_1^4, x_2^2)\},$$

$$\Omega_2 = \{(x_1^1, x_2^1), (x_1^1, x_2^2), (x_1^2, x_2^1), (x_1^2, x_2^2), (x_1^3, x_2^3), (x_1^3, x_2^4), (x_1^4, x_2^3), (x_1^4, x_2^4)\}$$

В таблице 1.6 представлено совместное распределение $p(x_1, x_2)$, полученное по формуле полной вероятности.

Таблица 1.6.

Безусловное распределение $p(x_1, x_2)$

x_2	x_1				$p(x_2)$
	$x_1=x_1^1$	$x_1=x_1^2$	$x_1=x_1^3$	$x_1=x_1^4$	
$x_2=x_2^1$	0,050	0,01250	0,050	0,01250	0,1250
$x_2=x_2^2$	0,150	0,03750	0,150	0,03750	0,3750
$x_2=x_2^3$	0,075	0,01875	0,075	0,01875	0,1875
$x_2=x_2^4$	0,125	0,03125	0,125	0,03125	0,3125
$p(x_1)$	0,4	0,1	0,4	0,1	-

Как видно из таблицы 1.6, признаки сами по себе статистически независимы

$$p(x_1, x_2) \equiv p(x_1)p(x_2). \quad (1.92)$$

Тем не менее из-за статистической зависимости между признаками в классах (см. табл. 1.4 и 1.5) они позволили безошибочно распознавать классы. Этот лишний раз показывает насколько важно в каждом конкретном случае исследовать вопрос об условной статистической зависимости между признаками прежде, чем принимать решение об их бесполезности.

ГЛАВА 2

ДОСТАТОЧНЫЕ УСЛОВИЯ ПОЛЕЗНОСТИ ИНФОРМАТИВНЫХ ПРИЗНАКОВ

2.1. Условия полезности отдельного признака

В работе [68] Бен-Бассат дал такое определение.

Определение 2.1. Признак x называется *нерелевантным*, если апостериорная вероятность k -го класса сохраняется неизменной при всех возможных значениях x , т. е.

$$P(V_k | x) = P(V_k) \quad \forall x \in X, \quad (2.1)$$

где $X \triangleq \{x : p(x) > 0\}$ – носитель распределения $p(x)$ признака x .

Введем несколько иное определение [49].

Определение 2.2. Признак x назовем *бесполезным*, если средняя вероятность ошибочных решений

$$P(e) = \sum_{x \in X} p(x) [1 - \max_k P(V_k | x)], \quad (2.2.)$$

принимаемых по правилу максимума апостериорных вероятностей $P(V_k | x)$, совпадает с вероятностью ошибочного решения

$$P_0(e) = 1 - \max_k P(V_k), \quad (2.3)$$

принимаемого по априорным вероятностям $P(V_1), \dots, P(V_M)$, т.е. если

$$P(e) = P_0(e). \quad (2.4)$$

Легко показать, что при числе классов $M = 2$ условие (2.1) эквивалентно тому, что признак x имеет тождественно совпадающие условные распределения в классах, т.е.

$$p(x | V_1) \equiv p(x | V_2). \quad (2.5)$$

При этом заведомо

$$I(V, X_n) \stackrel{\Delta}{=} H(V) - H(V | X) = 0, \quad (2.6)$$

где $H(V)$ и $H(V | X)$ – соответственно начальная и средняя условная энтропии по Шеннону.

Следовательно, всякий нерелевантный по Бен-Бассату признак одновременно является и неинформативным в смысле определения 1.2. Совершенно очевидно также, что при выполнении условия (2.5) признак оказывается и бесполезным в смысле Определения 2.2.

Покажем [49], что в общем случае обратное утверждение неверно: релевантный и информативный признак, имеющий различные распределения в классах

$$p(x | V_1) \neq p(x | V_2), \quad (2.7)$$

может оказаться бесполезным, несмотря на то, что при различных значениях $x \in X$ апостериорные вероятности классов изменяются.

Для того чтобы формально определить условие бесполезности релевантного признака, докажем следующую лемму [74].

Лемма 2.1. Пусть $P(V_1) \neq P(V_2)$. Тогда признак x – бесполезен, т.е. $P(e) = P_0(e)$ в том и только в том случае, когда

$$[P(V_1) - P(V_2)][P(V_1 | x) - P(V_2 | x)] > 0 \quad \forall x \in X, \quad (2.8)$$

другими словами, для каждого возможного значения $x \in X$ знак разности апостериорных вероятностей

$$\Delta(x) = P(V_1 | x) - P(V_2 | x)$$

совпадает со знаком разности априорных вероятностей

$$\Delta_0 = P(V_1) - P(V_2).$$

Доказательство. Предположим, что выполняется неравенство (2.8), например, что

$$P(V_1) > P(V_2) \text{ и } P(V_1 | x) > P(V_2 | x) \quad \forall x \in X.$$

Это значит, что

$$\min\{P(V_1), P(V_2)\} = P(V_2)$$

и

$$\min\{P(V_1 | x), P(V_2 | x)\} = P(V_2 | x) \quad \forall x \in X.$$

Следовательно, если выполняется условие (2.8), то в соответствии с (2.2) с учетом того, что $P(V_1) + P(V_2) = 1$, имеем

$$P(e) = \sum_{x \in X} p(x)P(V_2 | x) = P(V_2) = \min\{P(V_1), P(V_2)\} = P_0.$$

Предположим теперь, что условие (2.8) не выполняется, например, $P(V_1) > P(V_2)$, но существует подмножество $X^* \subset X$ значений x , таких, что

$$P(V_1 | x) < P(V_2 | x) \quad \forall x \in X^*.$$

Тогда в соответствии с (2.2) имеем

$$P(e) = \sum_{x \in X \setminus X^*} p(x)P(V_2 | x) + \sum_{x \in X^*} p(x)P(V_1 | x). \quad (2.9)$$

Заменяя в правой части (2.9) $P(V_1 | x)$ на $P(V_2 | x)$ с учетом того, что $P(V_1 | x) < P(V_2 | x)$, $\forall x \in X^*$ заключаем, что если $X^* \neq \emptyset$, то

$$P(e) < \sum_{x \in X} p(x)P(V_2 | x),$$

т. е. $P(e) < P_0(e)$. Лемма 2.1 доказана.

Отметим, что условия леммы 2.1 вовсе не предполагают выполнения условия (2.5). Отсюда как раз и следует, что релевантный признак может быть бесполезным.

Продемонстрируем такую возможность на примерах дискретных и непрерывных распределений диагностических признаков.

Пусть x - бинарный признак (симптом), принимающий только два возможных значения $x = x^+$ и $x = x^-$ с условными вероятностями $p(x^+ | V_k)$ и $p(x^- | V_k) = 1 - p(x^+ | V_k)$, $k = 1, 2$.

Теорема 2.1. Если априорные вероятности классов не равны, т.е. $P(V_1) \neq P(V_2)$, то для любой заданной условной вероятности $p(x^+ | V_1)$ может быть указана такая условная вероятность $p(x^+ | V_2)$, что

$$p(x^+ | V_1) \neq p(x^+ | V_2), \quad (2.10)$$

но при этом выполняется равенство $P(e) = P_0(e)$, т.е. x – релевантный, но бесполезный признак.

Доказательство. Предположим вначале, что $P(V_2) < P(V_1)$. Пусть Ω – множество возможных значений $p(x^+ | V_1)$ и $p(x^+ | V_2)$, которые удовлетворяют неравенствам

$$p(x^+ | V_1) > \lambda_0 p(x^+ | V_2), \quad (2.11)$$

$$p(x^+ | V_1) < \lambda_0 p(x^+ | V_2) + (1 - \lambda_0), \quad (2.12)$$

где

$$\lambda_0 = \frac{P(V_2)}{P(V_1)}$$

– отношение априорных вероятностей классов.

Поскольку $\lambda_0 < 1$ при $P(V_2) < P(V_1)$, то для заданного значения вероятности $p(x^+ | V_1)$ всегда можно указать соответствующее значение вероятности $p(x^+ | V_2) \neq p(x^+ | V_1)$, при которых неравенства (2.11) и (2.12) выполняются совместно, т.е. $\Omega \neq \emptyset$.

По формуле Байеса с учетом того, что $p(x^- | V_k) = 1 - p(x^+ | V_k)$ для $k = 1, 2$, из системы неравенств (2.11) и (2.12) следует

$$p(V_1 | x^+) > p(V_2 | x^+) \text{ и } p(V_1 | x^-) > p(V_2 | x^-).$$

Но это свидетельствует о том, что выполняется условие (2.8), т.е. в соответствии с Леммой 2.1 выполняется равенство $P(e) = P_0(e)$.

Доказательство Теоремы 2.1 для случая $P(V_2) > P(V_1)$ выполняется аналогичным образом, но в этом случае вместо (2.11), (2.12) рассматриваются неравенства

$$p(x^+ | V_1) < \lambda_0 p(x^+ | V_2), \quad (2.13)$$

$$p(x^+ | V_1) > \lambda_0 p(x^+ | V_2) + (1 - \lambda_0), \quad (2.14)$$

которые определяют множество $\Omega \neq \emptyset$ при $\lambda_0 > 1$.

Заметим, что $\Omega = \emptyset$ при $\lambda_0 = 1$, поскольку в этом случае неравенства (1.11), (2.12) так же как и неравенства (2.13), (2.14) не могут выполняться совместно. Этот факт позволяет сформулировать такое следствие.

Следствие 2.1. Если априорные вероятности классов равны, т.е. $P(V_1) = P(V_2)$, то при выполнении условия (2.10) релевантный бинарный признак заведомо полезен в смысле условия $P(e) < P_0(e)$.

На рис. 2.1 показаны границы областей Ω бесполезности релевантного признака при различных значениях λ_0 . Как видно область Ω уменьшается по мере приближения λ_0 к единице, и $\Omega = \emptyset$, когда $\lambda_0 = 1$.

Пример 2.1. Пусть $P(V_1) = 0,8$; $P(V_2) = 0,2$. Тогда $\lambda_0 = 0,25$. Предположим, что $p(x^+ | V_1) = 0,7$. Выберем соответствующее значение $p(x^+ | V_2) = 0,2$, при котором точка с координатами $p(x^+ | V_2) = 0,2$ и $p(x^+ | V_1) = 0,7$ принадлежит области Ω (рис.2.1, а). Заметим, что при этом бинарный признак x имеет различные распределения в классах: $p(x^+ | V_1) \neq p(x^+ | V_2)$.

Для того чтобы убедиться в бесполезности признака, вычислим по формуле Байеса апостериорные вероятности классов при различных значениях признака:

$$P(V_1 | x^+) = 0,933; P(V_2 | x^+) = 0,067; P(V_1 | x^-) = 0,6; P(V_2 | x^-) = 0,4.$$

Легко видеть, что

$$P(V_1 | x^+) > P(V_2 | x^+) \text{ и } P(V_1 | x^-) > P(V_2 | x^-),$$

т. е. при любом значении признака оптимальным является решение в пользу класса V_1 , совпадающее с априорным решением. Отсюда следует, что $P(e) = P_0(e) = 0,2$.

Полученные результаты позволяют сформулировать необходимые и достаточные условия полезности признака-симптома.

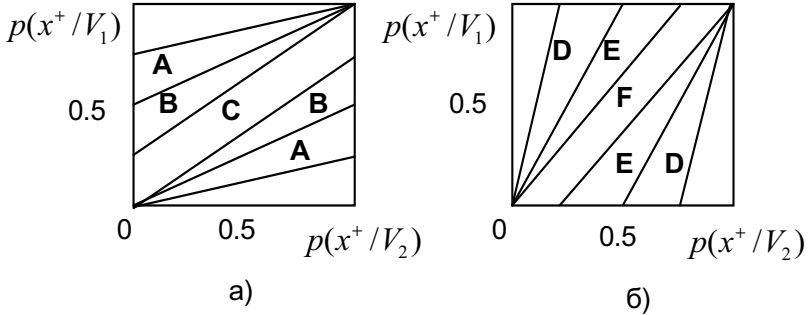


Рис.2.1. Области бесполезности бинарного признака

А — $\lambda_0 = 0,25$; В — $\lambda_0 = 0,5$; С — $\lambda_0 = 0,75$;

Д — $\lambda_0 = 4,0$; Е — $\lambda_0 = 2,0$; F — $\lambda_0 = 1,33$.

Утверждение 2.1. Бинарный признак x , принимающий два возможных значения x^+ и x^- , обеспечивает диагностику состояний V_1 и V_2 со средней вероятностью ошибок меньше априорной ошибки в том и только в том случае, когда выполняется одно из следующих условий:

а) $p(x^+ | V_1) \neq p(x^+ | V_2)$, если $\lambda_0 = 1$;

б) $p(x^+ | V_1) < \lambda_0 p(x^+ | V_2)$ или $p(x^+ | V_1) > \lambda_0 p(x^+ | V_2) + (1 - \lambda_0)$, если $\lambda_0 < 1$;

в) $p(x^+ | V_1) > \lambda_0 p(x^+ | V_2)$ или $p(x^+ | V_1) < \lambda_0 p(x^+ | V_2) + (1 - \lambda_0)$, если $\lambda_0 > 1$, где $\lambda_0 = P(V_2) / P(V_1)$ — отношение априорных вероятностей классов.

Покажем теперь, что условия Леммы 2.1 могут выполняться и для непрерывных признаков.

Теорема 2.2. Пусть признак x имеет нормальные распределения в классах

$$p(x | V_1) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left[\frac{-(x - M_1)^2}{2\sigma_1^2} \right], \quad (2.15)$$

$$p(x | V_2) = \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left[\frac{-(x - M_2)^2}{2\sigma_2^2} \right], \quad (2.16)$$

с математическими ожиданиями M_1, M_2 и среднеквадратическими отклонениями σ_1, σ_2 . Пусть априорные вероятности классов удовлетворяют условию $P(V_1) > P(V_2)$.

Тогда, если выполняются условия

$$\sigma_1 > \sigma_2 \quad (2.17)$$

и

$$\frac{\sigma_1}{\sigma_2} \exp \left[-\frac{(M_1 - M_2)^2}{2(\sigma_1^2 - \sigma_2^2)} \right] > \lambda_0, \quad (2.18)$$

где $\lambda_0 = P(V_2) / P(V_1)$ – отношение априорных вероятностей классов, то

$$P(e) = P_0(e).$$

Доказательство. Отметим, что в силу (2.15), (2.16) с учетом (2.17) признак x имеет различные распределения в классах, т.е. является релевантным и информативным.

В соответствии с Леммой 2.1 при $P(V_1) > P(V_2)$ признак x будет бесполезным в том и только в том случае, когда

$$p(x | V_1) > \lambda_0 p(x | V_2) \quad \forall x \in X. \quad (2.19)$$

Подстановка распределений из (2.15), (2.16) в (2.19) дает

$$\exp \left[\frac{(x - M_2)^2}{2\sigma_2^2} - \frac{(x - M_1)^2}{2\sigma_1^2} \right] > \lambda_0 \frac{\sigma_1}{\sigma_2} \quad \forall x \in X. \quad (2.20)$$

Логарифмируя левую и правую часть (2.20) получим соотношение

$$\frac{(x - M_2)^2}{2\sigma_2^2} - \frac{(x - M_1)^2}{2\sigma_1^2} > \ln \left[\lambda_0 \frac{\sigma_1}{\sigma_2} \right] \quad \forall x \in X,$$

которое после очевидных преобразования можно представить в виде

$$[\sigma_1^2 - \sigma_2^2]x^2 + 2(M_1\sigma_2^2 - M_2\sigma_1^2)x + \left[M_2^2\sigma_1^2 - M_1^2\sigma_2^2 - 2\sigma_1^2\sigma_2^2 \ln\left(\lambda_0 \frac{\sigma_1}{\sigma_2}\right) \right] > 0 \quad \forall x \in X. \quad (2.21)$$

Таким образом, в рассматриваемом случае условие бесполезности признака x сводится к условию положительности при любом $x \in X$. квадратного трехчлена, фигурирующего в левой части (2.21). Как видно из (2.21) это возможно в том и только в том случае, когда параметры распределений $p(x|V_1)$ и $p(x|V_2)$ удовлетворяют условиям

$$\sigma_1^2 - \sigma_2^2 > 0, \quad (2.22)$$

$$4(M_1\sigma_2^2 - M_2\sigma_1^2)^2 - 4(\sigma_1^2 - \sigma_2^2) \left[M_2^2\sigma_1^2 - M_1^2\sigma_2^2 - 2\sigma_1^2\sigma_2^2 \ln\left(\lambda_0 \frac{\sigma_1}{\sigma_2}\right) \right] < 0. \quad (2.23)$$

Из (2.22) немедленно следует (2.17), а условие (2.23) после очевидных упрощений сводится к неравенству

$$\ln\left(\lambda_0 \frac{\sigma_2}{\sigma_1}\right) < -\frac{(M_1 - M_2)^2}{2(\sigma_1^2 - \sigma_2^2)}, \quad (2.24)$$

которое, с учетом того, что $\exp(\cdot) > 0$, эквивалентно условию (2.18).

Теорема 2.2 доказана.

Следствие 2.2. При выполнении условия $P(V_1) > P(V_2)$ релевантный признак, имеющий нормальные распределения в классах с одинаковыми математическими ожиданиями, бесполезен, если отношение среднеквадратических отклонений превышает отношение соответствующих априорных вероятностей классов, т.е.

$$\frac{\sigma_2}{\sigma_1} > \frac{P(V_2)}{P(V_1)}.$$

Следствие прямо вытекает из условий Теоремы 2.2, если в (2.18), положить $M_1 = M_2$.

Заметим также, что при $\sigma_1 = \sigma_2$ квадратный трехчлен (2.21) переходит в линейную функцию x , которая не может быть положи-

тельна для любого $x \in X$, если $M_1 \neq M_2$. Отсюда немедленно вытекает важное для практического использования следствие.

Следствие 2.3. Всякий диагностический признак, имеющий нормальные распределения в классах с различными математическими ожиданиями и равными дисперсиями, полезен в смысле строгого неравенства $P(e) < P_0(e)$.

В то же время, при $\sigma_1 \neq \sigma_2$ нельзя однозначно сделать вывод о полезности признака, нормально распределенного в классах, даже если $M_1 \neq M_2$.

2.2. Достаточные условия полезности признака в описании

Рассмотрим теперь достаточные условия полезности признака x_n , $1 \leq n \leq N$ в совокупности с другими признаками N -мерного описания $x^{(N)} = (x_1, \dots, x_N)$, введя такое определение [72].

Определение 2.3. Признак x_n , $1 \leq n \leq N$ полезен в совокупности с другими $N-1$ признаками, если при его исключении из набора $x^{(N)} = (x_1, \dots, x_N)$ происходит изменение средней вероятности ошибки, т.е. выполняется *строгое* неравенство

$$P_N(e) < P_{N-1}(e), \quad (2.25)$$

где

$$P_N(e) \triangleq \sum_{x^{(N)} \in X^{(N)}} p(x^{(N)}) [1 - \max\{P(V_1 | x^{(N)}), P(V_2 | x^{(N)})\}], \quad (2.26)$$

$$P_{N-1}(e) \triangleq \sum_{x^{(N-1)} \in X^{(N-1)}} p(x^{(N-1)}) [1 - \max\{P(V_1 | x^{(N-1)}), P(V_2 | x^{(N-1)})\}].$$

Здесь $x^{(N-1)}$ – сокращенный вектор признаков, который не включает x_n , а $X^{(N-1)} \triangleq \{x^{(N-1)} : p(x^{(N-1)}) > 0\}$.

Доказательству теорем, формулирующих достаточные условия полезности x_n в совокупности с другими, предположим леммы.

Лемма 2.2. Пусть средняя вероятность ошибочных решений $P(e)$, принимаемых по описанию x , удовлетворяет условию

$$P(e) < \min \{P(V_1), P(V_2)\}, \quad (2.27)$$

а собственные области классов

$$X_k = \{x : p(x | V_k) > 0\}, \quad k = 1, 2 \quad (2.28)$$

пересекаются, т.е.

$$X_1 \cap X_2 \neq \emptyset. \quad (2.29)$$

Пусть далее выполняются условия

а) множества (2.28) односвязные,

б) условные распределения $p(x | V_k), k = 1, 2$ представляют собой непрерывные функции от x .

Тогда найдется такое значение $x_0 \in X_2 \cap X_1$, при котором

$$P(V_1 | x_0) = P(V_2 | x_0).$$

Доказательство. Введем в рассмотрение функцию

$$f(x) = P(V_1 | x) - P(V_2 | x),$$

которую на основании формулы Байеса можно представить в эквивалентной форме записи

$$f(x) = \frac{P(V_1)p(x | V_1) - P(V_2)p(x | V_2)}{p(x)}, \quad (2.30)$$

где

$$p(x) = P(V_1)p(x | V_1) + P(V_2)p(x | V_2) \quad \forall x \in X_1 \cup X_2.$$

В силу условия (2.27) с учетом (2.30) заключаем, что

$$\begin{aligned} \exists x^+ : x^+ \in X_1 \cup X_2, f(x^+) > 0, \\ \exists x^- : x^- \in X_1 \cup X_2, f(x^-) < 0. \end{aligned} \quad (2.31)$$

Поскольку по условию множества X_1 и X_2 односвязные, то в силу (2.29) односвязным будет и множество $X_1 \cup X_2$. Из условия б) с учетом (2.30) заключаем, что функция $f(x)$ непрерывна на всем од-

носвязном множестве $X_1 \cup X_2$. Принимая во внимание (2.31) по теореме Больцано-Коши [57] заключаем

$$\exists x_0 : f(x_0) = 0, \quad x_0 \in X_1 \cup X_2,$$

или, с учетом определения (2.30),

$$\exists x_0 : P(V_1 | x_0) = P(V_2 | x_0), \quad x_0 \in X_1 \cup X_2. \quad (2.32).$$

В том случае, когда $X_1 = X_2$, а значит $(X_1 \cap X_2) = (X_1 \cup X_2)$, из условия (2.32) немедленно следует справедливость теоремы. Поэтому для завершения доказательства остается показать, что $x_0 \in X_1 \cap X_2$ и в том случае, когда множества X_1 и X_2 частично пересекаются, т.е. $(X_1 \cap X_2) \subset (X_1 \cup X_2)$.

В этом нетрудно убедиться, если принять во внимание, что в силу (2.30), с учетом определений множеств X_1 и X_2 , имеем

$$f(x) = +1 \quad \forall x \in X_1 \setminus X_2,$$

$$f(x) = -1 \quad \forall x \in X_2 \setminus X_1.$$

Следовательно, условие $f(x_0) = 0$ может выполняться только при $x \in X_1 \cap X_2$. Лемма 2.2 доказана.

Замечание 2.1. Лемма 2.2 определяет условия существования границы x_0 между областями оптимальных байесовских решений в пространстве $X = \{x\}$, которая, вообще говоря, может и не существовать, если не выполняются все перечисленные условия.

Лемма 2.3. Для того чтобы при исключении из описания $x^{(N)} = (x_1, \dots, x_N)$ признака x_n , $1 \leq n \leq N$ средняя вероятность ошибки сохранялась неизменной, т.е. чтобы $P_N(e) = P_{N-1}(e)$, необходимо и достаточно, чтобы

$$\Omega_k^{(N)} = \Omega_k^{(N-1)} \times X_n, \quad k = 1, 2, \quad (2.33)$$

где $\Omega_m^{(N)}, \Omega_m^{(N-1)}$ – области оптимальных решений соответственно в N -мерном и $(N-1)$ -мерном пространствах признаков, а $X_n = \{x_n\}$ – множество возможных значений x_n .

Доказательство. Следуя [25, с.106] будем рассматривать переход от полного описания $x^{(N)} = (x_1, \dots, x_N)$ к сокращенному описанию $x^{(N-1)}$ как преобразование описаний, состоящее в объединении нескольких различных N -мерных описаний в одно $(N-1)$ -мерное описание.

Легко видно, что при выполнении условия (2.33) решения, принимаемые по полному $x^{(N)} = (x_n, x^{(N-1)})$ и сокращенному $x^{(N-1)}$ описаниям, будут совпадать для любого значения $x_n \in X_n$. Если же условие (2.33) не выполняется, то при исключении признака x_n и переходе к сокращенному описанию $x^{(N-1)}$ неизбежно будут объединяться такие описания $x^{(N)}$, по которым следует принимать разные решения. В результате некоторые решения, принимаемые по сокращенному описанию $x^{(N-1)}$, будут отличаться от оптимальных решений, принимаемых по полному описанию $x^{(N)}$, что и приводит к увеличению вероятности ошибки, т.е. к выполнению строгого неравенства (2.25). Лемма 2.3 доказана.

Замечание 2.2. Признак x_n может быть полезен в совокупности с другими признаками набора $x^{(N)} = (x_1, \dots, x_N)$ в смысле *строгого* неравенства (2.25) лишь в том случае, когда заведомо

$$P_N(e) < \min\{P(V_2)P(V_2)\} \quad (2.34)$$

и

$$P_{N-1}(e) > 0, \quad (2.35)$$

т.е. если N -мерное описание $x^{(N)} = (x_1, \dots, x_N)$ в целом является полезным, а сокращенное описание $x^{(N-1)}$ еще позволяет провести безошибочное распознавание.

Теорема 2.3. Для того чтобы при ограничениях (2.34), (2.35) признак x_n , $1 \leq n \leq N$ был полезен в совокупности с другими $N-1$ признаками описания $x^{(N)} = (x_1, \dots, x_N)$ достаточно, чтобы

а) признак x_n имел различные распределения в классах, т.е.

$$p(x_n | V_1) \neq p(x_n | V_2); \quad (2.36)$$

б) в каждом классе отсутствовала статистическая связь между x_n и остальными $N-1$ признаками, т.е.

$$p(x_n | x^{(N-1)}, V_k) \equiv p(x_n | V_k), \quad k = 1, 2; \quad (2.37)$$

в) условные распределения $p(x^{(N-1)} | V_k)$, $k = 1, 2$ были непрерывными функциями от $x^{(N-1)}$, а собственные области классов $X_k^{(N-1)} = \{x^{(N-1)} : p(x^{(N-1)} | V_k) > 0\}$, $k = 1, 2$ – односвязными множествами.

Доказательство. Поскольку для вероятности ошибки $P_{N-1}(e)$ всегда справедлива оценка

$$P_{N-1}(e) \leq \min\{P(V_1), P(V_2)\}, \quad (2.38)$$

а $P_N(e) < \min\{P(V_2)P(V_2)\}$ в силу ограничения (2.34), то строгое неравенство (2.25) заведомо выполняется, если $P_{N-1}(e) = P_0(e)$.

Поэтому нам остается показать справедливость сформулированной теоремы ли в том случае, когда (2.38) переходит в строгое неравенство

$$P_{N-1}(e) < P_0(e). \quad (2.39)$$

Рассмотрим функцию

$$\Phi(x^{(N-1)}) \triangleq \frac{p(x^{(N-1)} | V_1)}{p(x^{(N-1)} | V_2)} - \frac{P(V_2)}{P(V_1)}, \quad (2.40)$$

которая в силу условия в) непрерывна $\forall x^{(N-1)} \in X^{(N-1)}$.

На основании формулы Байеса заключаем, что при выполнении условия (2.39) существует непустая область $\Omega_1^{(N-1)} \subset X^{(N-1)}$ оптимальных решений в пользу класса V_1 , для которой с учетом определения (2.40) справедливо соотношение

$$\Phi(x^{(N-1)}) > 0 \quad \forall x^{(N-1)} \in \Omega_1^{(N-1)}. \quad (2.41)$$

Заметим, что согласно лемме 2.2. при выполнении условия (2.39) и условия в) существует граница между областями оптимальных решений¹, на которой $\Phi(x^{(N-1)}) = 0$.

¹ Отметим, что по определению точка, принадлежащая границе множества $\Omega_1^{(N-1)}$, не принадлежит самому множеству, иначе говоря, $\Omega_1^{(N-1)}$ – открытое множество.

Покажем теперь, что на множестве $X_n \triangleq \{x_n\}$ возможных значений x_n найдется такое значение x_n^1 , при котором

$$\frac{p(x_n^1 | V_2)}{p(x_n^1 | V_1)} < 1. \quad (2.42)$$

Действительно, если допустить противное, т.е. что

$$p(x_n | V_2) \geq p(x_n | V_1) \quad \forall x_n \in X_n, \quad (2.43)$$

и учесть, что на основании (2.36) неравенства (2.43) не могут перейти в равенства $\forall x_n \in X_n$, то после суммирования неравенств (2.43) получим

$$\sum_{x_n \in X_n} p(x_n | V_2) > \sum_{x_n \in X_n} p(x_n | V_1).$$

Но это невозможно, так как всегда

$$\sum_{x_n \in X_n} p(x_n | V_2) = \sum_{x_n \in X_n} p(x_n | V_1) = 1.$$

Полученное противоречие подтверждает справедливость (2.42).

Аналогичным образом можно убедиться и в том, что найдется такое значение $x_n^2 \in X_n$, при котором

$$\frac{p(x_n^2 | V_2)}{p(x_n^2 | V_1)} > 1. \quad (2.44)$$

Введем обозначения

$$\varepsilon_1 \triangleq \frac{p(x_n^1 | V_2)P(V_2)}{p(x_n^1 | V_1)P(V_1)} - \frac{P(V_2)}{P(V_1)}, \quad (2.45)$$

$$\varepsilon_2 \triangleq \frac{p(x_n^2 | V_2)P(V_2)}{p(x_n^2 | V_1)P(V_1)} - \frac{P(V_2)}{P(V_1)}. \quad (2.46)$$

На основании (2.42) и (2.44) из (2.45) и (2.46) имеем

$$\varepsilon_1 < 0, \varepsilon_2 > 0. \quad (2.47)$$

Поскольку на границе области $\Omega_1^{(N-1)} \subset X^{(N-1)}$ функция $\Phi(x^{(N-1)})$ обращается в нуль, а при остальных значениях $x^{(N-1)} \in \Omega_1^{(N-1)}$ эта функция положительна, то в силу непрерывности $\Phi(x^{(N-1)})$ непременно найдется такое непустое подмножество $\Omega^* \subset \Omega_1^{(N-1)}$, что для любых $\varepsilon_1 < 0$ и $\varepsilon_2 > 0$ выполняется соотношение

$$\varepsilon_1 < 0 < \Phi(x^{(N-1)}) < \varepsilon_2 \quad \forall x^{(N-1)} \in \Omega^*. \quad (2.48)$$

Подстановка выражений для $\Phi(x^{(N-1)})$, ε_1 , ε_2 из (2.40), (2.45), (2.46) в (2.48) после несложных преобразований дает

$$\frac{p(x^{(N-1)} | V_1)}{p(x^{(N-1)} | V_2)} > \frac{p(x_n^1 | V_2)P(V_2)}{p(x_n^1 | V_1)P(V_1)} \quad \forall x^{(N-1)} \in \Omega^*, \quad (2.49)$$

$$\frac{p(x^{(N-1)} | V_1)}{p(x^{(N-1)} | V_2)} < \frac{p(x_n^2 | V_2)P(V_2)}{p(x_n^2 | V_1)P(V_1)} \quad \forall x^{(N-1)} \in \Omega^*. \quad (2.50)$$

Поскольку в соответствии с условием (2.37)

$$p(x^{(N-1)}, x_n^1 | V_k) \equiv p(x^{(N-1)} | V_k)p(x_n^1 | V_k), \quad k = 1, 2$$

и

$$p(x^{(N-1)}, x_n^2 | V_k) \equiv p(x^{(N-1)} | V_k)p(x_n^2 | V_k), \quad k = 1, 2,$$

то из (2.49) и (2.50) имеем

$$\frac{p(x^{(N-1)}, x_n^1 | V_1)}{p(x^{(N-1)}, x_n^1 | V_2)} > \frac{P(V_2)}{P(V_1)} \quad \forall x^{(N-1)} \in \Omega^*, \quad (2.51)$$

$$\frac{p(x^{(N-1)}, x_n^2 | V_1)}{p(x^{(N-1)}, x_n^2 | V_2)} < \frac{P(V_2)}{P(V_1)} \quad \forall x^{(N-1)} \in \Omega^*, \quad (2.52)$$

Из (2.51), (2.52) непосредственно следует, что $\forall x^{(N-1)} \in \Omega^* \subset \Omega_1^{(N-1)}$ при одном значении $x_n = x_n^1$ оптимальные решения по N признакам должны приниматься в пользу класса V_1 , а при другом значении $x_n = x_n^2$ – в пользу класса V_2 . Следовательно, область $\Omega_1^{(N)}$ оптимальных решений в N -мерном пространстве призна-

ков в пользу класса V_1 не удовлетворяет условиям леммы 2.3, а значит выполняется строгое неравенство (2.25). Теорема 2.3 доказана.

Замечание 2.3. Согласно теореме 2.3 одно лишь условие (2.36), которое означает, что x_n информативен сам по себе, вовсе не гарантирует полезность этого признака в описании $x^{(N)} = (x_1, \dots, x_N)$. Более того, как было показано выше, при $P(V_1) \neq P(V_2)$ такой признак может быть бесполезен сам по себе. Но независимо от того, будет ли этот признак полезен сам по себе, является он непрерывным² или дискретным признаком, его полезность гарантирована в описании, если выполняются другие условия теоремы 2.3.

Теорема 2.4. Для того чтобы при ограничениях (2.34), (2.35) признак x_n , $1 \leq n \leq N$ был полезен в совокупности с другими $N-1$ признаками описания $x^{(N)} = (x_1, \dots, x_N)$ достаточно, чтобы

а) x_n был статистически связан с остальными $N-1$ признаками хотя бы в одном из классов, т.е.

$$p(x_n | x^{(N-1)}, V_k) \neq p(x_n | V_k) \quad k=1 \text{ и/или } k=2; \quad (2.53)$$

б) статистическая связь между x_n и $x^{(N-1)}$ была различной в классах при каждом фиксированном значении $x^{(N-1)}$, т.е.

$$p(x_n | x^{(N-1)}, V_1) \neq p(x_n | x^{(N-1)}, V_2) \quad \forall x^{(N-1)} \in X^{(N-1)}; \quad (2.54)$$

в) условные распределения $p(x^{(N-1)} | V_k)$, $k=1,2$ были непрерывными функциями от $x^{(N-1)}$, а собственные области классов $X_k^{(N-1)} = \{x^{(N-1)} : p(x^{(N-1)} | V_k) > 0\}$, $k=1,2$ – односвязными множествами.

Доказательство. Покажем вначале, что в силу условия (2.54) найдется такое значение $x_n^1 \in X_n$, при котором

$$\frac{p(x_n^1 | x^{(N-1)}, V_2)}{p(x_n^1 | x^{(N-1)}, V_1)} < 1 \quad \forall x^{(N-1)} \in X^{(N-1)}. \quad (2.55)$$

² Доказательство справедливости теоремы 2.3 существенно упрощается, если наложить дополнительное ограничение о том, что X_n – непрерывный признак. Именно этот случай был нами рассмотрен в работе [53]. Однако такое дополнительное ограничение не обязательно.

С этой целью предположим противное, т.е. что при каждом фиксированном $x^{(N-1)} \in X^{(N-1)}$ выполняются неравенства

$$p(x_n | x^{N-1}, V_2) \geq p(x_n | x^{N-1}, V_1) \quad \forall x_n \in X_n. \quad (2.56)$$

Если учесть, что в силу условия (2.54) знак равенства в (2.56) не может иметь места $\forall x_n \in X_n$, то после суммирования неравенств (2.56) получим

$$\sum_{x_n \in X_n} p(x_n | x^{N-1}, V_2) > \sum_{x_n \in X_n} p(x_n | x^{N-1}, V_1).$$

Но это противоречит тому, что

$$\sum_{x_n \in X_n} p(x_n | x^{N-1}, V_k) = \sum_{x_n \in X_n} \frac{p(x_n, x^{N-1} | V_k)}{p(x^{N-1} | V_k)} = \frac{p(x^{(N-1)} | V_k)}{p(x^{(N-1)} | V_k)} \equiv 1.$$

Полученное противоречие подтверждает справедливость неравенства (2.55).

Аналогичным образом можно убедиться в том, что $\exists x_2 \in X_n$, когда

$$\frac{p(x_n^2 | x^{(N-1)}, V_2)}{p(x_n^2 | x^{(N-1)}, V_1)} > 1 \quad \forall x^{(N-1)} \in X^{(N-1)}. \quad (2.57)$$

Введем в рассмотрение функции

$$\alpha_1(x^{(N-1)}) \triangleq \frac{p(x_n^1 | x^{(N-1)}, V_2)P(V_2)}{p(x_n^1 | x^{(N-1)}, V_1)P(V_1)} - \frac{P(V_2)}{P(V_1)} \quad (2.58)$$

и

$$\alpha_2(x^{(N-1)}) \triangleq \frac{p(x_n^2 | x^{(N-1)}, V_2)P(V_2)}{p(x_n^2 | x^{(N-1)}, V_1)P(V_1)} - \frac{P(V_2)}{P(V_1)}, \quad (2.59)$$

в которых $x^{(N-1)}$ выступает в качестве независимой переменной, а x_n^1 и x_n^2 – в качестве фиксированных параметров.

Нетрудно заметить, что в силу (2.55) и (2.57)

$$\alpha_1(x^{(N-1)}) < 0 \quad \forall x^{(N-1)} \in X^{(N-1)} \quad (2.60)$$

и

$$\alpha_2(x^{(N-1)}) > 0 \quad \forall x^{(N-1)} \in X^{(N-1)}. \quad (2.61)$$

Вернемся теперь снова к функции $\Phi(x^{(N-1)})$, которая определяется соотношением (2.40). Как было показано при доказательстве теоремы 2.3

$$\Phi(x^{(N-1)}) > 0 \quad \forall x^{(N-1)} \in \Omega_1^{(N-1)}, \quad (2.62)$$

где $\Omega_1^{(N-1)}$ – область оптимальных решений по сокращенному описанию $x^{(N-1)}$ в пользу класса V_1 , причем $\Phi(x^{(N-1)}) = 0$ на границе этой области.

С другой стороны из (2.61) с учетом того, что $\Omega_1^{(N-1)} \subset X^{(N-1)}$ следует

$$\alpha_2(x^{(N-1)}) > 0 \quad \forall x^{(N-1)} \in \Omega_1^{(N-1)}. \quad (2.63)$$

Поскольку функции $\Phi(x^{(N-1)})$ и $\alpha_2(x^{(N-1)})$ непрерывны, причем $\Phi(x^{(N-1)}) = 0$ в точке $x^{(N-1)} = x_0^{(N-1)}$, нетрудно показать, что существует непустое множество $\Omega^* \subset \Omega_1^{(N-1)}$ значений $x^{(N-1)}$, отличных от $x_0^{(N-1)}$, таких, что

$$0 < \Phi(x^{(N-1)}) < \alpha_2(x^{(N-1)}) \quad \forall x^{(N-1)} \in \Omega^*. \quad (2.64)$$

С другой стороны, в силу (2.60) имеем

$$\Phi(x^{(N-1)}) > \alpha_1(x^{(N-1)}) \quad \forall x^{(N-1)} \in \Omega^*. \quad (2.65)$$

Неравенства (2.64) и (2.65) с учетом определений (2.40), (2.60), (2.61) функции $\Phi(x^{(N-1)})$, $\alpha_1(x^{(N-1)})$, $\alpha_2(x^{(N-1)})$ могут быть преобразованы к виду

$$\frac{p(x^{(N-1)} | V_1)}{p(x^{(N-1)} | V_2)} > \frac{p(x_n^1 | x^{(N-1)}, V_2)P(V_2)}{p(x_n^1 | x^{(N-1)}, V_1)P(V_1)} \quad \forall x^{(N-1)} \in \Omega^*, \quad (2.66)$$

$$\frac{p(x^{(N-1)} | V_1)}{p(x^{(N-1)} | V_2)} < \frac{p(x_n^2 | x^{(N-1)}, V_2)P(V_2)}{p(x_n^2 | x^{(N-1)}, V_1)P(V_1)} \quad \forall x^{(N-1)} \in \Omega^*. \quad (2.67)$$

Принимая во внимание (2.66), (2.67) с учетом того, что

$$p(x^{(N)} | V_k) \equiv p(x^{(N-1)} | V_k)p(x_n | x^{(N-1)}, V_k), \quad k = 1, 2$$

мы снова приходим к соотношениям (2.51),(2.52), из которых, как было показано при доказательстве теоремы 2.3, немедленно следует выполнение строгого неравенства (2.25). Теорема 2.4 доказана.

Замечание 2.4. Следует обратить внимание, что неинформативный сам по себе признак x_n , удовлетворяющий условию

$$p(x_n | V_1) \equiv p(x_n | V_2), \quad (2.68)$$

оказывается гарантировано полезным в совокупности с другими признаками, если выполняются условия теоремы 2.4. Этот факт следует из того, что условия теоремы 2.4 могут, вообще говоря, выполняться совместно с условием (2.68).

Доказательству следующей теоремы предпошлем еще одну лемму [50].

Лемма 2.4. Пусть $\xi = (\xi_1, \dots, \xi_N)$ – N – мерный случайный вектор с непрерывной плотностью распределения $p_\xi(x, \theta)$, зависящей от параметра $\theta \in \Theta$. Тогда, для того, чтобы при изменении параметра θ плотность $p_\xi(x, \theta)$, не меняя своего вида, сдвигалась в пространстве R^N , т.е. чтобы при любых фиксированных $\theta_1 \in \Theta, \theta_2 \in \Theta$

$$p_\xi(x, \theta_2) \equiv p_\xi(x - \delta, \theta_1), \quad (2.69)$$

где $\delta \equiv (\delta_1, \dots, \delta_N) = \delta(\Delta\theta)$ – некоторая вектор-функция, зависящая от разности $\Delta\theta = \theta_2 - \theta_1$, необходимо и достаточно выполнение следующих условий:

а) математическое ожидание $M(\xi)$ вектора ξ зависит от параметра θ , причем

$$M_{\theta_2} \xi = M_{\theta_1} \xi + \delta, \quad (2.70)$$

б) центральные смешанные моменты величин ξ_1, \dots, ξ_N всех порядков не зависят от параметра θ , т.е. для любых натуральных чисел k_1, \dots, k_N выполняется равенство

$$G_1(\xi) = G_2(\xi), \quad (2.71)$$

где $G_j(\xi) = M_{\theta_j} (\xi_1 - M_{\theta_j} \xi_1)^{k_1} \dots (\xi_N - M_{\theta_j} \xi_N)^{k_N}$.

Доказательство. Необходимость доказывается непосредственной проверкой. Для доказательства достаточности применим формулу бинома Ньютона и, проведя ряд преобразований, представим выражение для G_j в виде

$$G_j(\xi) = \sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} a_{i_1 \dots i_N}^{(j)}(\xi) \int_{R^N} \dots \int_{R^N} x_1^{i_1} \dots x_N^{i_N} p_\xi(x, \theta_j) dx,$$

где

$$a_{i_1 \dots i_N}^{(j)}(\xi) = (-1)^L \binom{i_1}{k_1} \dots \binom{i_N}{k_N} (M_{\theta_j} \xi_1)^{k_1 - i_1} \dots (M_{\theta_j} \xi_N)^{k_N - i_N},$$

$$L = (k_1 + \dots + k_N) - (i_1 + \dots + i_N),$$

$$\binom{i_l}{k_l} = \frac{k_l!}{i_l!(k_l - i_l)!}, \quad l = \overline{1, N} \quad - \text{число сочетаний из } k_l \text{ по } i_l.$$

Из условия (2.70) следует, что $a_{i_1 \dots i_N}^{(2)}(\xi) = a_{i_1 \dots i_N}^{(1)}(\xi + \delta)$. Принимая во внимание этот факт, преобразуем $G_1(\xi)$ с учетом очевидного выражения $p_\xi(x - \delta, \theta) = p_{\xi + \delta}(x, \theta)$. В результате условие (2.71) примет вид:

$$\begin{aligned} & \sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} a_{i_1 \dots i_N}^{(2)}(\xi) \int_{R^N} \dots \int_{R^N} x_1^{i_1} \dots x_N^{i_N} p_\xi(x, \theta_2) dx = \\ & = \sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} a_{i_1 \dots i_N}^{(2)}(\xi) \int_{R^N} \dots \int_{R^N} x_1^{i_1} \dots x_N^{i_N} p_\xi(x - \delta, \theta_1) dx. \end{aligned}$$

Отсюда, пользуясь методом математической индукции, легко показать, что для любых натуральных чисел k_1, \dots, k_N выполняется равенство

$$\int_{R^N} \dots \int_{R^N} x_1^{k_1} \dots x_N^{k_N} p_\xi(x, \theta_2) dx = \int_{R^N} \dots \int_{R^N} x_1^{k_1} \dots x_N^{k_N} p_\xi(x - \delta, \theta_1) dx$$

или в эквивалентной форме записи

$$\int_{R^N} \dots \int_{R^N} x_1^{k_1} \dots x_N^{k_N} g(x) dx = 0 \quad \forall k_1, \dots, k_N = 0, 1, \dots, \quad ,$$

где $g(x) = p_\xi(x, \theta_2) - p_\xi(x - \delta, \theta_1)$.

Последнее выражение означает не что иное, как ортогональность непрерывной функции $g(x)$ к любому полиному вида $x_1^{k_1} \dots x_N^{k_N}$. От-

сюда, в силу полноты системы полиномов вида $x_1^{k_1} \dots x_N^{k_N}$ в пространстве непрерывных функций [31] следует, что $g(x) \equiv 0$. Но последнее соотношение как раз и эквивалентно выполнению равенства (2.69). Лемма 2.4 доказана.

Теорема 2.5. Для того чтобы при ограничениях (2.34), (2.35) признак x_n , $1 \leq n \leq N$ был полезен в описании $x_n = (x_1, \dots, x_N)$ достаточно выполнение следующих условий

а) x_n был одинаково распределен в классах, т.е.

$$p(x_n | V_1) \equiv p(x_n | V_2); \quad (2.72)$$

б) x_n и остальные $N-1$ признаков были статистически независимы в одном из классов и корреляционно связаны в другом классе, т.е.

$$p(x^{(N)} | V_1) \equiv p(x^{(N-1)} | V_1)p(x_n | V_1), \quad (2.73)$$

$$M\{x^{(N)} | V_2\} \neq M\{x^{(N-1)} | V_2\}M\{x_n | V_2\}, \quad (2.74)$$

причем эта связь такова, что корреляционные моменты

$$G_2 \triangleq M \left\{ \prod_{\substack{l=1 \\ l \neq n}}^N (x_l - M\{x_l | x_n, V_2\})^{k_l} \right\}$$

при всех $k_l = 0, 1, \dots$ не зависели от x_n ;

в) условные распределения $p(x^{(N)} | V_m)$, $m = 1, 2$ были непрерывными функциями, производные от которых по любому направлению отличны от нуля за исключением, быть может, изолированных точек, а собственные области классов $X_m^{(N)} = \{x^{(N)} : p(x^{(N)} | V_m) > 0\}$ – односвязными множествами.

Доказательство. Как уже показывалось при доказательстве теоремы 2.3, учитывая (2.34) нам нужно показать справедливость сформулированной теоремы лишь для случая $P_{N-1}(e) < P_0(e)$, имея в виду, что в силу ограничений (2.35) и условия в) существует граница между областями решений в пространстве $X^{(N-1)}$, т.е. $\exists x_0^{(N-1)} \in X^{(N-1)}$ такое, что

$$\frac{p(x_0^{(N-1)} | V_1)}{p(x_0^{(N-1)} | V_2)} = \frac{P(V_2)}{P(V_1)} \quad \forall x_0^{(N-1)} \in X_0^{(N-1)}.$$

Принимая во внимание условия (2.72), (2.73) представим выражение для отношения правдоподобия $\lambda(x^{(N)}) \triangleq \frac{p(x^{(N)} | V_1)}{p(x^{(N)} | V_2)}$ в точке $x_0^{(N-1)} \in X_0^{(N-1)}$ так:

$$\lambda(x_0^{(N-1)}, x_n) = \lambda(x_0^{(N-1)}) \frac{p(x_0^{(N-1)} | V_2)}{p(x_0^{(N-1)} | x_n, V_2)}. \quad (2.75)$$

На основании условия б) в силу леммы 2.4 имеем

$$p(x^{(N-1)} | x_n, V_2) = p(x^{(N-1)} + \delta^{(N-1)} | x_n^{(1)}, V_2),$$

где $\delta^{(N-1)}$ – некоторая вектор-функция, зависящая от $x_n \in X_n$, $x_n^{(1)} \in X_n$. В то же время в соответствии с формулой конечных приращений

$$p(x^{(N-1)} + \delta^{(N-1)} | x_n^{(1)}, V_2) = p(x^{(N-1)} | x_n^{(1)}, V_2) + \sum_{\substack{i=1 \\ i \neq n}}^N \left. \frac{\partial p(x^{(N-1)} / x_n^{(1)}, V_2)}{\partial x_i} \right|_{x_*^{(N-1)}} \delta_i^{(N-1)},$$

где $x_*^{(N-1)} = x^{(N-1)} + \gamma \delta^{(N-1)}$, $\gamma \in (0, 1)$.

Предположим от противного, что признак x_n не является полезным в описании $x^{(N-1)}$. Тогда, на основании леммы 2.3 с учетом (2.75), имеем:

$$p(x_0^{(N-1)} | V_2) = p(x_0^{(N-1)} | x_n, V_2) \quad \forall x_n \in X_n.$$

А это значит, что

$$\text{grad} [p(x_0^{(N-1)} + \gamma \delta^{(N-1)} | x_n^{(1)}, V_2) \delta^{(N-1)}] = 0.$$

Поскольку же $p(x^{(N)}) = p(x^{(N-1)} | x_n) \cdot p(x_n)$, то отсюда следует равенство нулю производной плотности $p(x^{(N)} | V_2)$ по направлению $\delta^{(N-1)}$ для всех $x^{(N)} = (x_0^{(N-1)} + \gamma \delta^{(N-1)}, x_n)$, что противоречит условию в).

Полученное противоречие подтверждает справедливость сформулированной теоремы.

В тех же случаях, когда правомерна гипотеза о нормальном законе распределения признаков в классах, разумно использовать следствие, вытекающее из теоремы 2.5.

Следствие 2.4. Пусть описание $x^{(N)} = (x_1, \dots, x_N)$ имеет невырожденное многомерное нормальное распределение в классах V_1 и V_2 .

Тогда признак x_n ($1 \leq n \leq N$) полезен в описании, если выполняются условия:

а) x_n некоррелирован ни с одним из других $N - 1$ признаков в одном классе и коррелирован хотя бы с одним из признаков в другом классе;

б) математическое ожидание и дисперсия этого признака одинаковы в классах, т.е.

$$M\{x_n | V_1\} = M\{x_n | V_2\}, \quad (2.76)$$

$$D\{x_n | V_1\} = D\{x_n | V_2\}, \quad (2.77)$$

то этот признак x_n заведомо полезен в описании $x^{(N)}$.

Доказательство. Если

$$p(x^{(N)} | V_m) = (2\pi)^{\frac{N}{2}} |D_m^{(N)}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x^{(N)} - \mu_m^{(N)})^T (D_m^{(N)})^{-1} (x^{(N)} - \mu_m^{(N)})\right]$$

где

$$\mu_m^{(N)} = (M\{x_1 | V_m\}, \dots, M\{x_N | V_m\})$$

– вектор математических ожиданий, а

$$D_m^{(N)} = (M\{x_i - M\{x_i | V_m\}\})(M\{x_i - M\{x_i | V_m\}\}), \quad i = 1, \dots, N, \quad j = 1, \dots, N$$

– ковариационная матрица, то маргинальная плотность x_n имеет вид

$$p(x_n | V_m) = (2\pi D\{x_n | V_m\})^{-\frac{1}{2}} \exp\left\{-\frac{(x_n - M\{x_n | V_m\})^2}{2D\{x_n | V_m\}}\right\}.$$

Отсюда следует, что при выполнении условий (2.76), (2.77) будет иметь место условие (2.72).

Не умаляя общности, положим для определенности, что x_n некоррелирован ни с одним из других $N-1$ признаков в классе V_1 . Тогда $p(x^{(N)}|V_1)$ представима так:

$$p(x^{(N)}|V_1) = (2\pi)^{\frac{N-1}{2}} (2\pi)^{\frac{1}{2}} |D_1^{(N-1)}|^{-\frac{1}{2}} (\mathbf{D}\{x_n|V_1\})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x^{(N-1)} - \mu_1^{(N-1)})^T (D_1^{(N-1)})^{-1} (x^{(N-1)} - \mu_1^{(N-1)}) - \frac{(x_n - \mathbf{M}\{x_n|V_1\})^2}{2\mathbf{D}\{x_n|V_1\}}\right\}, \quad (2.78)$$

где $D_1^{(N-1)}$ – матрица, получаемая из матрицы $D_1^{(N)}$ вычеркиванием n -й строки и n -го столбца, $\mu_1^{(N-1)}$ – вектор $\mu_1^{(N)}$ с исключенным n -м элементом.

Легко видно, что правая часть (2.78) представляет собой произведение $p(x^{(N-1)}|V_1)p(x_n|V_1)$, а значит имеет место условие (2.73). Кроме того, условное распределение $p(x^{(N-1)}|x_n, V_m)$ является нормальным распределением с вектором математических ожиданий

$$z_m^{(N-1)} = \mu_m^{(N-1)} + D_m^n \frac{x_n - \mathbf{M}\{x_n|V_m\}}{\mathbf{D}\{x_n|V_m\}}$$

и ковариационной матрицей

$$S_m^{(N-1)} = D_m^{(N-1)} - \frac{D_m^n (D_m^n)^T}{\mathbf{D}\{x_n|V_m\}},$$

где D_m^n – вектор, представляющий собой n -й столбец матрицы $D_m^{(N-1)}$ без диагонального элемента.

Поскольку по условию x_n коррелирован хотя бы с одним из $N-1$ признаков в классе V_2 , то вектор D_2^n не является нулевым, а значит $z_2^{(N-1)}$ зависит от x_n , т.е. имеет условие (2.74).

В то же время из выражения для $S_m^{(N-1)}$ видно, что ковариационная матрица условного распределения $p(x^{(N-1)}|x_n, V_m)$ не зависит от x_n .

Поскольку согласно [14] для нормально распределенных случайных величин центральные моменты нечетных порядков равны нулю,

а четных – определяются лишь элементами ковариационной матрицы, то из независимости $S_m^{(N-1)}$ от x_n следует и независимость от x_n всех центральных моментов, фигурирующих в условии б) теоремы 2.5.

Если к тому же учесть, что для нормальных распределений заведомо выполняется условие в) теоремы 2.5, то в силу этой теоремы признак x_n полезен в описании $x^{(N)}$. Следствие 2.4 доказано.

2.3. Информационный критерий полезности признаков

Сформулируем условие инвариантности вероятности ошибочных решений относительно исключения из описания информативного признака.

Будем как и прежде рассматривать переход от $x^{(N)} = (x_1, \dots, x_N)$ к $x^{(N-1)}$ как некое преобразование описаний [25], состоящее в объединении множества описаний $x^{(N)} = (x^{(N-1)}, x_n)$ с различным значениями x_n в одно описание $x^{(N-1)}$.

В соответствии с правилом максимума апостериорных вероятностей при $M = 2$ решения о принадлежности к классам V_1, V_2 принимаются по знакам разностей апостериорных вероятностей

$$\Delta = P(V_1 | x^{(N-1)}, x_n) - P(V_1 | x^{(N-1)}, x_n). \quad (2.79)$$

Поэтому средняя вероятность ошибочных решений сохранится неизменной при переходе $x^{(N)} = (x^{(N-1)}, x_n)$ к $x^{(N-1)}$, если для каждого фиксированного $x^{(N-1)} \in X^{(N-1)}$ разность (2.79) будет иметь один и тот же знак $\forall x_n \in X_n$. При этом, естественно, допускается, что сами знаки разностей Δ могут быть различными при разных $x^{(N-1)} \in X^{(N-1)}$.

Ясно, что при единственном ограничении

$$P(V_1 | x^{(N-1)}, x_n) + P(V_1 | x^{(N-1)}, x_n) = 1,$$

налагаемом на апостериорные вероятности классов, условие сохранения знака разности (2.79) для $\forall x^{(N-1)} \in X^{(N-1)}$ может быть выполнено

не только в том тривиальном случае, когда апостериорные вероятности $P(V_k | x^{(N-1)}, x_n)$ не зависят от x_n , но и в более общем случае, когда имеет место соотношение

$$P(V_1 | x^{(N-1)}, x_n) \neq P(V_1 | x^{(N-1)}, x_n) = 1.$$

В этой связи условие, состоящее в том, что $\forall x^{(N-1)} \in X^{(N-1)}$ разность (2.79) является знакопостоянной функцией, зависящей от x_n , разумно интерпретировать как условие инвариантности вероятности ошибочных решений относительно исключения из описания информативного признака.

Пример 2.2. Пусть $P(V_1) = 0,4$, $P(V_1) = 0,6$, $N = 2$, а каждый и признаков имеет две градации, т.е. $x_1 = x_1^1, x_1^2$ и $x_2 = x_2^1, x_2^2$. В таблицах 2.1 и 2.2 представлены условные распределения этих признаков.

Таблица 2.1.

Условное распределение $p(x_1, x_2 | V_1)$

x_2	x_1		$p(x_2 V_1)$
	$x_1 = x_1^1$	$x_1 = x_1^2$	
$x_2 = x_2^1$	0,56	0,24	0,80
$x_2 = x_2^2$	0,14	0,06	0,20
$p(x_1 V_1)$	0,70	0,30	-

Таблица 2.2.

Условное распределение $p(x_1, x_2 | V_2)$

x_2	x_1		$p(x_2 V_1)$
	$x_1 = x_1^1$	$x_1 = x_1^2$	
$x_2 = x_2^1$	0,02	0,18	0,20
$x_2 = x_2^2$	0,08	0,72	0,80
$p(x_1 V_1)$	0,10	0,90	-

Как видно признаки статистически независимы в обоих классах, и каждый из них имеет различные распределения в классах.

В соответствии с исходными данными вычислим средние условные энтропии по Шеннону: $H(V | X_1, X_2) = 0,512$, $H(V | X_1) = 0,678$, откуда

$$H(V | X_1) - H(V | X_1, X_2) = 0,166 > 0 .$$

Следовательно x_2 - информативен в совокупности с признаков $x^{(2)} = (x_1, x_2)$ с точки зрения строгого неравенства (1.14).

В то же время нетрудно убедиться в том, что при переходе от описания $x^{(2)} = (x_1, x_2)$ к сокращенному описанию $x^{(1)} = x_1$ вероятность ошибочных решений остается неизменной:

$$P_2(e) = P_1(e) = 0,18.$$

Замечание 2.5. Разность (2.79) может оказаться не только знакопостоянной функцией от x_n при каждом фиксированном значении $x^{(N-1)} \in X^{(N-1)}$, но и иметь один и тот же знак $\forall x^{(N-1)} \in X^{(N-1)}$. В этом особом случае апостериорные вероятности $P(V_k | x^{(N-1)}, x_n)$ по-прежнему могут зависеть от x_n , однако вероятность ошибочных решений при исключении x_n из описания $x^{(N)} = (x^{(N-1)}, x_n)$ остается неизменной и равной априорной вероятности ошибки $P_0(e) = \min\{P(V_1), (P(V_2)\}$. Для случая $N = 1$ такая ситуация рассматривалась в разделе 2.1 (пример 2.1).

Принимая во внимание ограничения (2.34) и (2.35) сформулируем и докажем еще одну теорему [20].

Теорема 2.6. Если при исключении из описания $x^{(N)} = (x_1, \dots, x_N)$ признака x_n ($1 \leq n \leq N$) средняя условная энтропия изменяется настолько, что

$$H(V | X^{(N-1)}) - H(V | X^{(N)}) > I^*, \tag{2.80}$$

где

$$I^* = -H(V | X^{(N)})[1 + 0,5 \log 0,5H(V | X^{(N)})] - [1 - 0,5H(V | X^{(N)})] \log [1 - 0,5H(V | X^{(N)})], \tag{2.81}$$

то x_n - гарантировано полезный признак в описании $x^{(N)} = (x^{(N-1)}, x_n)$ с точки зрения строгого неравенства (2.25).

Доказательство. Для доказательства воспользуемся приведенными в главе 1 выражениями, которые при числе классов $M = 2$ определяют точные верхнюю (1.18) и нижнюю (1.19) границы средней условной энтропии $H(V | X)$ при данной средней вероятности ошибочных решений $P(e)$.

На основании (1.19) заключаем, что при данной средней условной энтропии $H(V | X^{(N)})$ для вероятности ошибочных решений $P_N(e)$ справедлива оценка

$$P_N(e) \leq 0,5H(V | X^{(N)}). \quad (2.82)$$

В самом деле, если допустить противное, то приходим к тому, что

$$H(V | X^{(N)}) < 2P_N(e),$$

но это противоречит оценке (1.19).

Будем теперь исходить из того, что выполняется условие (2.80). Тогда с учетом (2.81) получим

$$\begin{aligned} H(V | X^{(N-1)}) &> -0,5H(V | X^{(N)}) \log 0,5H(V | X^{(N)}) - \\ &- [1 - 0,5H(V | X^{(N)}) \log [1 - 0,5H(V | X^{(N)})]]. \end{aligned} \quad (2.83)$$

Покажем, что при выполнении этого условия

$$P_{N-1}(e) > 0,5H(V | X^{(N)}). \quad (2.84)$$

Действительно, предположим противное, т.е. что

$$P_{N-1}(e) \leq 0,5H(V | X^{(N)}).$$

При таком предположении в силу того, что $H(V | X^{(N)}) \leq 1$, а $\sup H(V | X)$ согласно (1.18) является монотонно возрастающей функцией от $P(e)$ при $P(e) \leq 0,5$ имеем

$$\begin{aligned} \sup H(V | X^{(N-1)}) &\leq -0,5H(V | X^{(N)}) \log 0,5H(V | X^{(N)}) - \\ &- [1 - 0,5H(V | X^{(N)}) \log [1 - 0,5H(V | X^{(N)})]]. \end{aligned} \quad (2.85)$$

Поскольку же всегда $H(V | X^{(N-1)}) \leq \sup H(V | X^{(N-1)})$, то усиление (2.85) приводит к неравенству

$$H(V | X^{(N-1)}) \leq -0,5H(V | X^{(N)}) \log 0,5H(V | X^{(N)}) - \\ - [1 - 0,5H(V | X^{(N)}) \log [1 - 0,5H(V | X^{(N)})],$$

которое противоречит неравенству (2.83). Тем самым подтверждается справедливость соотношения (2.84).

Сравнение (2.82) и (2.84) окончательно позволяет заключить, что из условия (2.80) следует выполнение строгого неравенства (2.25).

Теорема 2.6 доказана.

Величину I^* , фигурирующую в правой части условия (2.80), уместно назвать информационным «барьером», преодоление которого гарантирует полезность признака x_n в описании $x^{(N)} = (x^{(N-1)}, x_n)$.

График зависимости I^* от $H(V | X^{(N)})$ представлен на рис. 2.2.

Наибольшее значение

$$I_{\max}^* = \log_2 1,25$$

достигается при $H(V | X^{(N)}) = 0,4$.

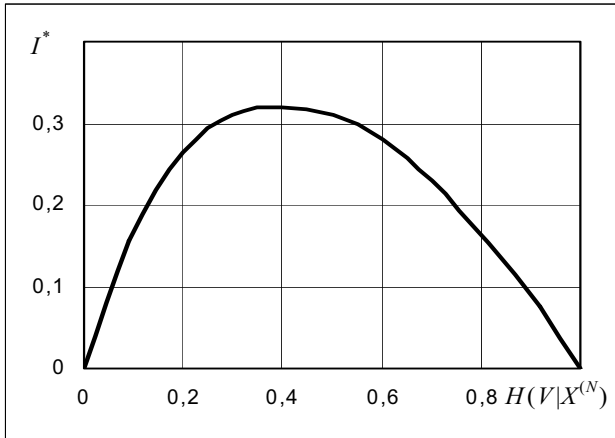


Рис.2.2. График зависимости информационного барьера I^*

Следствие 2.5. Если признак x_n статистически не зависит от совокупности остальных $N - 1$ признаков в описании $x^{(N)} = (x_1, \dots, x_N)$, т.е.

$$p(x^{(N)}) \equiv p(x_n)p(x^{(N-1)}), \quad (2.86)$$

а количество информации $I(V, X_n)$, содержащееся в множестве $X_n = \{x_n\}$ относительно множества классов $V = \{V_1, V_2\}$, превышает информационный барьер I^* , т.е.

$$I(V, X_n) > I^*, \quad (2.87)$$

то этот признак x_n заведомо полезен в описании $x^{(N)}$. Следствие прямо вытекает из теоремы 2.6 в силу того, что при выполнении условия (2.86) для количества информации $I(V, X_n)$ справедлива оценка

$$I(V, X_n) \leq I(V, X^{(N-1)}) - I(V, X^{(N)}),$$

являющаяся простым обобщением одного из результатов работы [4].

Теорема 2.7. Для распознавания классов V_1 и V_2 по описанию x с вероятностью ошибочных решений $P(e) < P_0(e)$ достаточно, чтобы количество информации

$$I(V, X) \triangleq H(V) - H(V | X), \quad (2.88)$$

содержащееся в множестве описаний $X = \{x\}$ относительно множества классов $V = \{V_1, V_2\}$, превышало некоторую величину

$$I_0^* = \log(1 + \lambda_0) - \frac{\lambda_0}{1 + \lambda_0} - 2 \min \left\{ \frac{1}{1 + \lambda_0}, \frac{\lambda_0}{1 + \lambda_0} \right\}, \quad (2.89)$$

зависящую исключительно от отношения априорных вероятностей

$$\lambda_0 = \frac{P(V_2)}{P(V_1)}. \quad (2.90)$$

Доказательство. Пусть выполняется условие

$$I(V, X) > I_0^*,$$

которое с учетом (2.89), (2.90) и того, что

$$H(V) \stackrel{\Delta}{=} -P(V_1) \log P(V_1) - P(V_2) \log P(V_2),$$

а

$$P_0(e) \stackrel{\Delta}{=} \min(P(V_1), P(V_2)),$$

может быть преобразовано к виду

$$I(V, X) > H(V) - 2P_0(e). \quad (2.91)$$

На основании (2.88) условие (2.91) можно представить так

$$H(V) < 2P_0(e). \quad (2.92)$$

Покажем что при выполнении этого условия непременно $P(e) < P_0(e)$. В самом деле, предположим противное, т.е. что $P(e) = P_0(e)$ (случай $P(e) > P_0(e)$ вообще невозможен при решениях, принимаемых по правилу максимума апостериорных вероятностей).

При таком предположении в соответствии с (1.19) имеем

$$\inf H(V | X) = 2P_0(e).$$

Однако последнее равенство противоречит условию (2.92). Теорема 2.7 доказана.

На рис. 2.3. показан график зависимости I_0^* от λ_0 при $0 \leq \lambda_0 \leq 1$. Если учесть, что согласно (2.89)

$$I_0^*(\lambda_0) = I_0^*(\lambda_0)^{-1},$$

то этой зависимостью можно пользоваться и при $\lambda_0 > 1$, поскольку в этом случае $0 \leq \lambda_0^{-1} \leq 1$.

Следствие 2.6. Если априорные вероятности классов равны, т.е. $P(V_1) = P(V_2)$, то всякий информативный признак является полезным. Следствие прямо вытекает из (2.88) с учетом (2.89) в силу того, что $I_0^* = 0$ при $\lambda_0 = 1$.

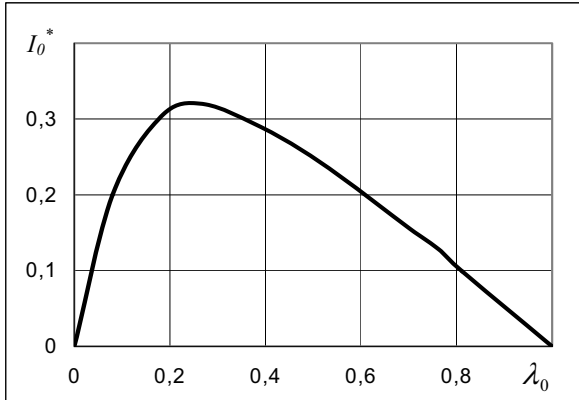


Рис. 2.3. График зависимости I_0^* от λ_0

Примечательно, что точно такой же результат эквивалентности критериев информативности и полезности признака при $P(V_1) = P(V_2)$ был выше получен несколько иным путем и сформулирован в следствии 2.1 теоремы 2.1.

Рассмотрим применение полученных условий для одного специального случая.

Пусть выполняются все условия теоремы 2.3. Тогда в соответствии с этой теоремой x_n — заведомо полезный признак в описании $x^{(N)}$, а значит для вероятности ошибочных решений $P_{N-1}(e)$, принимаемых по сокращенному описанию $x^{(N-1)}$, не содержащему x_n , справедлива оценка

$$P_N(e) < P_{N-1}(e) \leq_0 P_0(e), \quad (2.93)$$

где как и прежде $P_N(e)$ — вероятности ошибочных решений принимаемых по исходному описанию $x^{(N)}$, а $P_0(e) \stackrel{\Delta}{=} \min\{P(V_1), P(V_2)\}$.

Предположим теперь, что этот признак x_n сам по себе бесполезен (такая возможность обсуждалась нами в замечании 2.3).

Покажем, что при таком предположении можно получить другую, более сильную оценку сверху $P_{N-1}(e)$, чем оценка (2.93), а именно

$$P_{N-1}(e) \leq P_{N-1}^*(e) \leq P_0(e), \quad (2.94)$$

где

$$P_{N-1}(e) = -0,5P_N(e) \log P_N(e) - 0,5[1 - P_N(e)] \log[1 - P_N(e)] + \\ + 0,5 \log(1 + \lambda_0) - 0,5 \frac{\lambda_0}{1 + \lambda_0} \log \lambda_0 - \min \left\{ \frac{1}{1 + \lambda_0}, \frac{\lambda_0}{1 + \lambda_0} \right\}. \quad (2.95)$$

С этой целью заметим, что если x_n сам по себе бесполезен, то, согласно теореме 2.7,

$$I(V, X_n) \leq I_0^*, \quad (2.96)$$

где $I(V, X)$ – количество информации, содержащееся в множестве $X_n = \{x_n\}$ относительно множества $V = \{V_1, V_2\}$, а I_0^* – величина, определяемая соотношением (2.89). В противном случае, если бы $I(V, X) > I_0^*$, то признак был бы гарантировано полезным, что противоречит условию рассматриваемого случая.

С другой стороны, по условия (2.36) теоремы 2.3 признак x_n сам по себе информативен, т.е.

$$I(V, X_n) > 0. \quad (2.97)$$

Объединение (2.96) и (2.97) дает оценку

$$0 < I(V, X_n) \leq I_0^*. \quad (2.98)$$

На основании выражение (1.18) и (1.19) для точных верхней и нижней границ средней условной энтропии с учетом того, что

$$I(V, X^{(N)}) \triangleq H(V) - H(V | X^{(N)}),$$

для количества информации $I(V, X^{(N)})$ содержащемуся в множестве $X^{(N)} = \{x^{(N)}\}$ относительно множества классов $V = \{V_1, V_2\}$, имеем оценку

$$H(V) + P_N(e) \log P_N(e) + [1 - P_N(e)] \log[1 - P_N(e)] \leq \\ \leq I(V, X^{(N)}) \leq H(V) - 2P_N(e). \quad (2.99)$$

Поскольку по условию теоремы 2.3 x_n статистически не зависит от остальных признаков в обоих классах, т.е.

$$p(x_n | x^{(N-1)}, V_k) \equiv p(x_n | V_k), \quad k = 1, 2,$$

то в силу одного из результатов работы [4] имеем

$$I(V, X^{(N)}) \leq I(V, X^{(N-1)}) + I(V, X_n),$$

откуда

$$I(V, X^{(N-1)}) \geq I(V, X^{(N)}) - I(V, X_n). \quad (2.100)$$

Усиление неравенства (2.100) с учетом (2.98) (2.99) дает

$$I(V, X^{(N-1)}) \geq H(V) + P_N(e) \log P_N(e) + [1 - P_N(e)] \log [1 - P_N(e)] - I_0^*,$$

что равносильно неравенству

$$H(V, X^{(N-1)}) \leq -P_N(e) \log P_N(e) - [1 - P_N(e)] \log [1 - P_N(e)] + I_0^*.$$

Принимая во внимание выражение (2.89) для величины I_0^* и используя введенное обозначение (2.95) для величины $P_{N-1}^*(e)$, последнее неравенство можно записать в компактной форме

$$H(V | X^{(N-1)}) \geq 2P_{N-1}^*(e). \quad (2.101)$$

Покажем, что из неравенства (2.101) следует оценка (2.94). С этой целью предположим противное, т.е. что

$$P_{N-1}(e) > P_{N-1}^*(e).$$

При таком предположении согласно (1.19), с учетом того, что $\inf H(V | X)$ является монотонно возрастающей функцией от $P(e)$, получим

$$\inf H(V | X) > 2P_{N-1}^*.$$

Усиление последнего неравенства в силу того, что всегда $H(V | X^{(N-1)}) \geq \inf H(V | X^{(N-1)})$, дает

$$H(V | X^{(N-1)}) > 2P_{N-1}^*(e).$$

Но это противоречит неравенству (2.101). Полученное противоречие как раз и подтверждает справедливость оценки (2.94).

Для иллюстрации на рис. 2.4. представлены графики зависимости $P_{N-1}^*(e)$ от $P_N(e)$ при различных значениях λ_0 . Применимость оценки $P_{N-1}^*(e)$ ограничена областью, определяемой неравенством

$$P_{N-1}^*(e) \leq P_0(e).$$

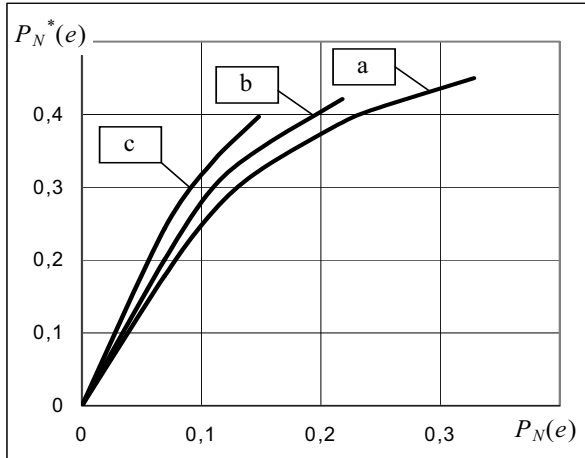


Рис. 2.4. Графики зависимости $P_{N-1}^*(e)$ от $P_N(e)$
а — $\lambda_0 = 0,9$; б — $\lambda_0 = 0,8$; в — $\lambda_0 = 0,7$;

ГЛАВА 3

ОЦЕНКА ПОЛЕЗНОСТИ ДИАГНОСТИЧЕСКИХ ТЕСТОВ С ПОЗИЦИИ ТЕОРИИ СТАТИСТИЧЕСКИХ РЕШЕНИЙ

3.1. Достаточные условия полезности диагностического теста

В предыдущей главе были сформулированы достаточные условия, гарантирующие полезность отдельного признака x_n или же признака x_n в совокупности с другими признаками описания $x^{(N)} = (x_n, x^{(N-1)})$ для тех задач, когда алгоритм диагностики построен по оптимальной схеме – правилу максимума апостериорных вероятностей $P(V_1 | x_n)$, $P(V_2 | x_n)$ или же $P(V_1 | x^{(N)})$, $P(V_2 | x^{(N)})$.

Однако не всегда алгоритм диагностики (диагностический тест) строится формальным путем. Довольно часто предлагаются новые диагностические тесты, основанные лишь на разумных гипотезах и интуиции конструктора. В этих случаях также возникает необходимость оценить полезность диагностического теста прежде, чем рекомендовать его для практического применения [12].

Рассмотрим следующую постановку задачи. Пусть для диагностики состояния объекта Z с двумя возможными состояниями V_1, V_2 , $P(V_1) + P(V_2) = 1$, предполагается использовать диагностический тест, который на основании измерения доступной информации принимает решения в виде индикаторной переменной

$$s(x) = \begin{cases} 1, & \text{если принято решение в пользу } V_1; \\ 2, & \text{если принято решение в пользу } V_2. \end{cases} \quad (3.1)$$

В данном случае нас не будет интересовать то, как построен сам алгоритм принятия решений (3.1), т.е. будем рассматривать диагностический тест как «черный ящик».

Предполагается, что тест может ошибаться, допуская ложноотрицательные и ложноположительные ошибки. Такие ошибки, которые в теории распознавания образов принято называть ошибками пропуска цели и ложной тревоги, будем считать случайными событиями

$$E_1 = (Z \in V_1) \wedge (\delta = 2),$$

$$E_2 = (Z \in V_2) \wedge (\delta = 1).$$

Предположим, что по некоторой представительной выборке наблюдений, для которых известны истинные состояния Z , получены оценки вероятностей $P(E_1)$ и $P(E_2)$. Будем, как это принято в теории статистических решений, характеризовать возможные потери платежной матрицей вида

$$L = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}, \quad (3.2)$$

в которой L_{11}, L_{22} – потери, связанные с правильными решениями, а L_{12}, L_{21} – потери от ошибок пропуска цели и ложной тревоги соответственно.

Тогда математическое ожидание потерь (средний риск) определяет взвешенная сумма указанных потерь с учетом вероятностей их появления:

$$R = \sum_{k=1}^2 \sum_{j=1}^2 L_{kj} P(V_k, \delta = j). \quad (3.3)$$

где величина $P(V_k, \delta = j)$ обозначает вероятность совместного выполнения двух случайных событий: Z находится в состоянии V_k , $k=1,2$, а алгоритм принял решение $\delta = j$, отнеся его к классу V_j , $j=1,2$.

Введем следующее определение [44].

Определение 3.1. Диагностический тест полезен, если средний риск R принимаемых на его основе решений, меньше риска R_0 априорных решений, принимаемых по априорным вероятностям $P(V_1)$ и $P(V_2)$, т.е. выполняется строгое неравенство

$$R < R_0. \quad (3.4)$$

Ставится задача получить условия, гарантирующие выполнение неравенства (3.4).

Обозначим

$$\omega = \frac{L_{12} - L_{11}}{L_{21} - L_{22}}, \quad (3.5)$$

полагая, что $L_{12} > L_{11}$ и $L_{21} > L_{22}$, а значит $\omega > 0$. Такое предположение вполне логично, поскольку потери от ошибочных результатов всегда превышают потери от правильных решений.

Замечание 3.1. В частном случае, когда $L_{11} = L_{22} = 0$, величина ω просто определяет соотношение потерь от ошибок пропуска цели и ложной тревоги. Однако в дальнейших исследованиях такое дополнительное предположение не обязательно.

Докажем справедливость такой теоремы.

Теорема 3.1. Диагностический тест полезен в смысле строгого неравенства (3.4), в том и только в том случае, когда

$$[(1 - P(V_1))P(E_2) < \omega P(V_1)[1 - P(E_1)] \quad (3.6)$$

при $P(V_1)(1 + \omega) < 1$ либо

$$[1 - P(V_1)][1 - P(E_2)] > \omega P(V_1)P(E_1) \quad (3.7)$$

при $P(V_1)(1 + \omega) > 1$.

Доказательство. Запишем совместные вероятности $P(V_k, \delta = j)$, фигурирующие в (3.3), в эквивалентной форме

$$P(V_k, \delta = j) = P(V_k)P(\delta = j | V_k), \quad \delta, j, k \in \{1, 2\}.$$

Поскольку по определению

$$P(\delta = 2 | V_1) \triangleq P(E_1) \quad \text{и} \quad P(\delta = 1 | V_2) \triangleq P(E_2)$$

причем

$$P(\delta = 1 | V_1) = 1 - P(E_1) \text{ и } P(\delta = 2 | V_2) = 1 - P(E_2)$$

то, согласно (3.3), средний риск решений (3.1) можно представить в виде

$$R = L_{11}P(V_1)[1 - P(E_1)] + L_{22}[1 - P(V_1)][1 - P(E_2)] + L_{12}P(V_1)P(E_1) + L_{21}[1 - P(V_1)]P(E_2). \quad (3.8)$$

Если же не использовать диагностический тест, то априорная «стратегия» решений о неизвестных состояниях Z сводится к выбору одного из двух вариантов. Либо всегда считать, что $Z \in V_1$ и тогда риск будет равен

$$L_{11}P(V_1) + L_{21}[1 - P(V_1)],$$

либо всегда считать, что $Z \in V_2$ и тогда риск будет равен

$$L_{22}[1 - P(V_1)] + L_{12}P(V_1).$$

Отсюда с учетом обозначение (3.5) заключаем, что если выполняется условие

$$P(V_1)(1 + \omega) < 1, \quad (3.9)$$

то для обеспечения минимума априорного риска следует принимать решение $Z \in V_2$, а если выполняется условие

$$P(V_1)(1 + \omega) > 1, \quad (3.10)$$

то лучшими оказываются решения $Z \in V_1$.

Следовательно, априорный риск R_0 , фигурирующий в правой части (3.4), можно представить в виде

$$R_0 = \begin{cases} (L_{22}[1 - P(V_1)] + L_{12}P(V_1)), & \text{если } P(V_1)(1 + \omega) < 1 \\ (L_{11}P(V_1) + L_{21}[1 - P(V_1)]), & \text{если } P(V_1)(1 + \omega) > 1 \end{cases}. \quad (3.11)$$

Подстановка (3.8) и (3.11) в (3.4) после простых преобразований приводит к условиям (3.6), (3.7). Теорема 3.1 доказана.

На рис. 3.1 показаны границы областей полезного теста, определяемых соотношениями (3.6) и (3.7) соответственно.

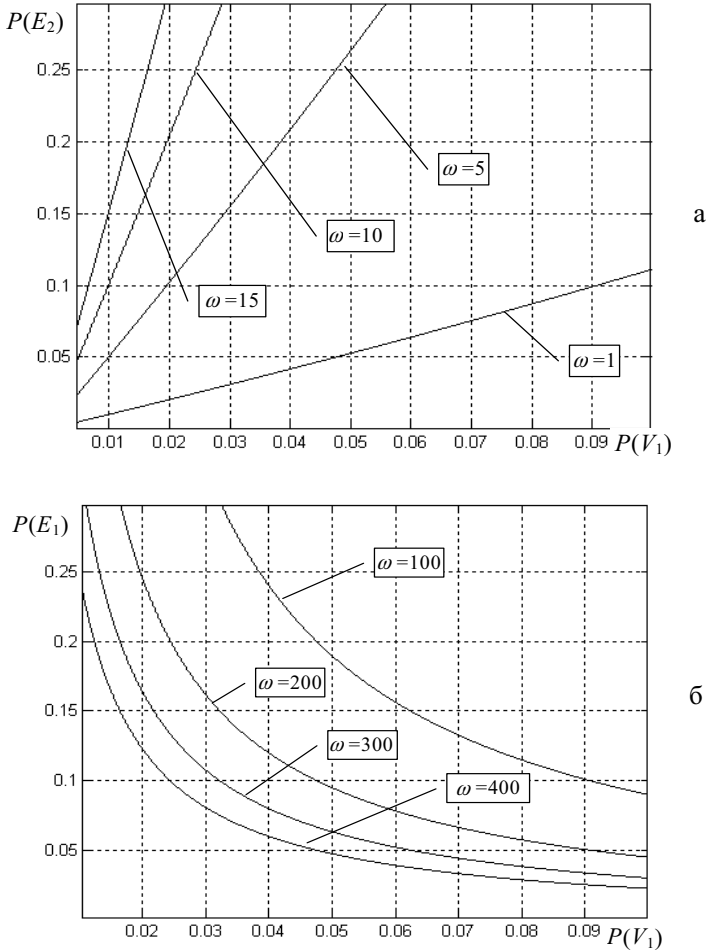


Рис. 3.1. Области полезности теста
 а – при $P(V_1)(1 + \omega) > 1$; б – при $P(V_1)(1 + \omega) < 1$

Из теоремы 3.1 следует, что если не выполняются условия (3.6) или (3.7), то с точки зрения уменьшения априорного риска диагностический тест абсолютно бесполезен. Этот факт позволяет сформулировать еще две не менее важных теоремы.

Теорема 3.2. Пусть выполняется условие $P(V_1)(1 + \omega) < 1$. Тогда при *сколь угодно малой* вероятности ошибки пропуска цели $P(E_1)$ диагностический тест заведомо бесполезен, если вероятность ошибки ложной тревоги удовлетворяет условию

$$P(E_2) \geq \frac{\omega P(V_1)}{1 - P(V_1)}. \quad (3.12)$$

Доказательство. Принимая во внимание, что $P(V_1) \neq 1$, запишем условие (3.6), гарантирующее полезность теста в области $P(V_1)(1 + \omega) < 1$, следующим образом

$$P(E_2) < \frac{\omega P(V_1)[1 - P(E_1)]}{1 - P(V_1)}. \quad (3.13)$$

Поскольку всегда $P(E_1) \geq 0$, то усиление (3.13) путем подстановки $P(E_1) = 0$ приводит к оценке сверху допустимого значения вероятности ошибки ложной тревоги полезного теста:

$$\sup_{0 \leq P(E_1) \leq 1} P(E_2) = \frac{\omega P(V_1)}{1 - P(V_1)} \quad (3.14)$$

для области $P(V_1)(1 + \omega) < 1$. Теорема 3.2 доказана.

Теорема 3.3. Пусть выполняется условие $P(V_1)(1 + \omega) > 1$. Тогда при *любой сколь угодно малой* вероятности ошибки ложной тревоги $P(E_2)$ диагностический тест заведомо бесполезен с точки зрения уменьшения априорного риска, если вероятность ошибки пропуска цели удовлетворяет условию

$$P(E_1) \geq \frac{1 - P(V_1)}{\omega P(V_1)}. \quad (3.15)$$

Доказательство. Поскольку $P(V_1) \neq 0$, то условие (3.7), гарантирующее полезность теста для области $P(V_1)(1 + \omega) > 1$, можно представить в эквивалентной форме:

$$P(E_1) < \frac{[1 - P(V_1)][1 - P(E_2)]}{\omega P(V_1)}. \quad (3.16)$$

Если учесть, что всегда $P(E_2) \geq 0$, то усиление (3.16) путем подстановки $P(E_2) = 0$ приводит к оценке сверху вероятности ошибки пропуска цели полезного теста

$$\sup_{0 \leq P(E_1) \leq 1} P(E_1) = \frac{1 - P(V_1)}{\omega P(V_1)} \quad (3.17)$$

для области $P(V_1)(1 + \omega) > 1$. Теорема 3.3 доказана.

Опираясь на полученные результаты, введем еще одно определение полезности признаков в совокупности.

Определение 3.2. Диагностический признак x_n ($1 \leq n \leq N$) полезен в совокупности с остальными $N - 1$ признаками, если его исключение из описания $x^{(N)} = (x_1, \dots, x_N)$ приводит к увеличению среднего риска, т.е.

$$R_{N-1} > R_N, \quad (3.18)$$

где R_N и R_{N-1} – средний риск диагностики по описанию $x^{(N)}$ и сокращенному описанию $x^{(N-1)}$, которое не содержит x_n .

Рассмотрим условия, гарантирующие выполнение строгого неравенства (3.18), для бинарных признаков (симптомов), характерных для задач медицинской диагностики.

Пусть для диагностики некоторого заболевания используется симптом x_1 , который может отсутствовать (случайное событие x_1^-) или присутствовать (случайное событие x_1^+) у пациента Z . Предположим, что симптом x_1 характерен для всех без исключения больных (класс V_1) и только для части здоровых (класс V_2). Формально это означает, что условные вероятности $P(x_1^+ | V_1)$, $P(x_1^+ | V_2)$ удовлетворяют соотношениям

$$P(x_1^+ | V_1) = 1, \quad 0 < P(x_1^+ | V_2) < 1. \quad (3.19)$$

Из (3.19) следует, что $P(x_1^- | V_1) = 0$. Значит при отсутствии симптома ($x_1 = x_1^-$) однозначно нужно принимать решение в пользу V_2 . Ясно, что в этом случае вероятность ошибочного отнесения больного к группе здоровых (вероятность ошибки «пропуска цели») равна нулю, т.е.

$$P(E_1) = 0.$$

Предположим теперь, что вероятность $P(x_1^+ | V_2)$ удовлетворяет условию

$$P(x_1^+ | V_2) \geq \frac{\omega P(V_1)}{1 - P(V_1)}. \quad (3.20)$$

Заметим, что для фиксированных значений $P(V_1)$ и ω , при которых выполняется условие $P(V_1)(1 + \omega) < 1$, неравенство (3.20) не противоречит второму из ограничений (3.19).

Если принимать решение в пользу класса V_1 , когда $x_1 = x_1^+$, т.е. считать пациента больным при наличии симптома, то величина $P(x_1^+ | V_2)$ определит вероятность ошибки ложной тревоги, т.е.

$$P(E_2) = P(x_1^+ | V_2).$$

Но тогда при выполнении условия (3.20) в соответствии с теоремой 3.2 признак x_1 заведомо бесполезен с точки зрения уменьшения априорного риска.

Для того чтобы еще раз убедиться в этом, сравним условные риски возможных решений, которые принимаются при $x_1 = x_1^+$.

Риск решения $Z \in V_1$ равен

$$R_1 = L_{11}P(V_1)P(x_1^+ | V_1) + L_{21}[1 - P(V_1)]P(x_1^+ | V_2),$$

а риск решения $Z \in V_2$ –

$$R_2 = L_{22}[1 - P(V_1)]P(x_1^+ | V_2) + L_{12}P(V_1)P(x_1^+ | V_1).$$

Если учесть, что $P(x_1^+ | V_1) = 1$, то с учетом обозначения (3.5) разность $\Delta R = R_2 - R_1$ определяется соотношением

$$\Delta R = (L_{21} - L_{22})[\omega P(V_1) - [1 - P(V_1)]P(x_1^+ | V_2)], \quad (3.21)$$

из которого в силу (3.20) и того, что $L_{21} > L_{22}$ следует $R_2 \leq R_1$.

Отсюда можно сделать вывод, что и при наличии симптома, т.е. при $x_1 = x_1^+$ необходимо принимать решение $Z \in V_2$.

Таким образом, при любом возможном значении признака $x_1 = x_1^-$ и $x_1 = x_1^+$ принимается одно и то же решение в пользу класса

V_2 , которое для фиксированных значений $P(V_1)$ и ω в области $P(V_1)(1 + \omega) < 1$ совпадает с априорным решением. Естественно, что такой признак-симптом сам по себе абсолютно бесполезен.

Предположим теперь, что в нашем распоряжении имеются два таких «бесполезных» в отдельности признака. Возникает естественный вопрос: в каких случаях их совокупность может оказаться полезной с точки зрения определения 3.3.

Теорема 3.4. Пусть x_1, x_2 – два признака, каждый из которых принимает два возможных значения x_1^-, x_1^+ и x_2^-, x_2^+ , причем

$$P(x_i^+ | V_1) = 1, \quad i = 1, 2, \quad (3.22)$$

Пусть, кроме того, признаки статистически независимы в обоих классах, т.е.

$$P(x_1, x_2 | V_k) \equiv P(x_1 | V_k)P(x_2 | V_k), \quad k = 1, 2. \quad (3.23)$$

Тогда для любых $P(V_1)$ и ω , удовлетворяющих условию

$$P(V_1)(1 + \omega) < 1,$$

существуют такие значения вероятностей $P(x_i^+ | V_2)$, $i = 1, 2$, при которых каждый из этих признаков бесполезен сам по себе, но их совокупность полезна.

Доказательство. Предположим, что

$$P(x_i^+ | V_2) \geq \rho, \quad i = 1, 2, \quad (3.24)$$

где

$$\rho = \frac{\omega P(V_1)}{1 - P(V_1)}. \quad (3.25)$$

Как было показано ранее, в этом случае рассматриваемые признаки сами по себе бесполезны.

Рассмотрим условные риски возможных решений при комбинации значений признаков $(x_1 = x_1^+) \wedge (x_2 = x_2^+)$. Принимая во внимание (3.23) заключаем, что при такой комбинации значений симптомов риск решения $Z \in V_1$ равен

$$R_1(x_1^+, x_2^+) = L_{11}P(V_1)P(x_1^+ | V_1)P(x_2^+ | V_1) +$$

$$+ L_{21}[1 - P(V_1)]P(x_1^+ | V_2)P(x_2^+ | V_2), \quad (3.26)$$

а риск решения $Z \in V_2$ –

$$R_2(x_1^+, x_2^+) = L_{22}[1 - P(V_1)]P(x_1^+ | V_2)P(x_2^+ | V_2) + L_{12}P(V_1)P(x_1^+ | V_1)P(x_2^+ | V_1). \quad (3.27)$$

При выполнении условия $P(V_1)(1 + \omega) < 1$ для минимизации априорного риска всегда следует принимать решение $Z \in V_2$. Потребуем, чтобы при комбинации значений признаков $(x_1 = x_1^+) \wedge (x_2 = x_2^+)$ оптимальным было противоположное решение, т.е. $Z \in V_1$. Другими словами, потребуем, чтобы выполнялось строгое неравенство

$$R_1(x_1^+, x_2^+) < R_2(x_1^+, x_2^+).$$

На основании (3.26), (3.27) с учетом обозначения (3.25) последнее неравенство можно представить в виде

$$P(x_1^+ | V_2)P(x_2^+ | V_2) < \rho. \quad (3.28)$$

С другой стороны, в соответствии с (3.24),

$$P(x_1^+ | V_2)P(x_2^+ | V_2) \geq \rho^2. \quad (3.29)$$

Поскольку при выполнении условий $P(V_1) \neq 0$ и $P(V_1)(1 + \omega) < 1$ величина ρ лежит в пределах $0 < \rho < 1$, то всегда найдутся такие значения вероятностей $P(x_1^+ | V_2)$ и $P(x_2^+ | V_2)$, удовлетворяющие (3.24), при которых неравенства (3.28), (3.29) выполняются совместно. Теорема 3.4 доказана.

На рис. 3.2 в координатах ρ и $\chi = P(x_1^+ | V_2)P(x_2^+ | V_2)$ показана область Ω значений, удовлетворяющих (3.28), (3.29).

Пример 3.1. Пусть $P(V_1) = 0.04$, $L_{11} = L_{22} = 0$, $L_{12} = 4$, $L_{21} = 1$. Следовательно $\omega = 4$, а $\rho \approx 0,167$. Предположим, что имеются два бинарных признака $x_1 = x_1^+, x_1^-$ и $x_2 = x_2^+, x_2^-$, которые статистически независимы в обоих классах. Условные распределения признаков представлены в таблицах 3.1 и 3.2.

Легко видеть, что для каждого из признаков выполняется условие (3.24), а, значит, эти признаки сами по себе бесполезны. В этом

легко убедиться, если сравнить условные риски возможных решений при различных значениях этих признаков.

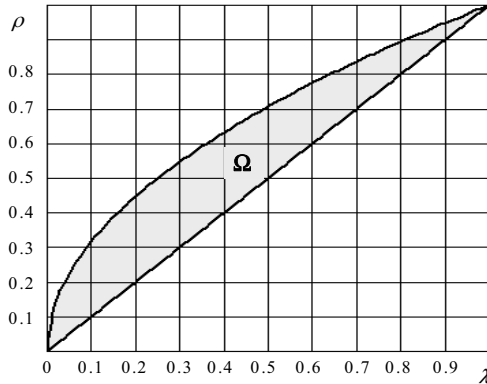


Рис. 3.2. Область полезной совокупности признаков

Таблица 3.1.

Условные распределения симптомов в классе V_1

x_1	x_2		$p(x_1 V_1)$
	$x_2 = x_2^-$	$x_2 = x_2^+$	
$x_1 = x_1^-$	0	0	0
$x_1 = x_1^+$	0	1	1
$p(x_2 V_1)$	0	1	-

Таблица 3.2.

Условные распределения симптомов в классе V_2

x_m	x_2		$p(x_1 V_2)$
	$x_2 = x_2^-$	$x_2 = x_2^+$	
$x_1 = x_1^-$	0,49	0,21	0,7
$x_1 = x_1^+$	0,21	0,09	0,3
$p(x_2 V_2)$	0,7	0,3	-

Например, принятие решения $Z \in V_1$ по признаку x_1 приводит к риску $R_1(x_1^+) = 0,288$ при наличии симптома ($x_1 = x_1^+$) и рис-

ку $R_1(x_1^-) = 0,672$ при отсутствии симптома ($x_1 = x_1^-$), в то время как решение $Z \in V_2$ приводит к риску $R_2(x_1^+) = 0,16$ при $x_1 = x_1^+$ и риску $R_2(x_1^-) = 0$ при $x_1 = x_1^-$.

Следовательно, и при наличии, и при отсутствии симптома x_1 с точки зрения минимума среднего риска должны приниматься решения в пользу класса V_2 , которые при заданных значениях $P(V_1) = 0,04$ и $\omega = 4$ совпадают с априорными решениями. Аналогичная ситуация и для признака x_2 .

Вместе с тем из условия данного примера следует, что

$$\chi = P(x_1^+ | V_2)P(x_2^+ | V_2) = 0,09.$$

Поскольку $\rho \approx 0,167$, то выполняется условие $\chi < \rho \leq \sqrt{\chi}$. А это, как следует из приведенного выше доказательства, свидетельствует о том, что совокупность признаков полезна.

В самом деле, при обнаружении обоих симптомов, когда $(x_1 = x_1^+) \wedge (x_2 = x_2^+)$ решение $Z \in V_1$ приводит к риску

$$R_1(x_1^+, x_2^+) = 0,0864 ,$$

тогда как решение $Z \in V_2$ приводит к большему риску

$$R_2(x_1^+, x_2^+) = 0,16 .$$

Следовательно, при такой комбинации признаков оптимальное решение не совпадает с априорным, что и обеспечивает уменьшение среднего риска.

Рассмотрим теперь один любопытный частный случай.

Следствие 3.1. Если, в дополнение к условиям теоремы 3.4, допустить, что $L_{11} = L_{22} = 0$, а $L_{12} = L_{21} = 1$, а вероятности наличия каждого из симптомов для объектов второго класса одинаковы, т.е.

$$P(x_1^+ | V_2) = P(x_2^+ | V_2) = P_2^+ ,$$

то, как следует из доказательства, при выполнении условия

$$\frac{P(V_1)}{1 - P(V_1)} \leq P_2^+ < \sqrt{\frac{P(V_1)}{1 - P(V_1)}} \quad (3.30)$$

такие признаки будут бесполезны сами по себе, но полезны в совокупности.

На рис. 3.3 в координатах $P_2^+ - P(V_1)$ показана область Ω значений, удовлетворяющих условию (3.30).

Пример 3.2. Пусть $P(V_1) = 0,1$ и симптом $x = x_1$ или симптом $x = x_2$ присутствует для всех объектов первого класса ($P_1^+ = 1$) и для 20% объектов второго класса ($P_2^+ = 0,2$). Как видно из рис. 3.3, точка с такими координатами попадает в область Ω .

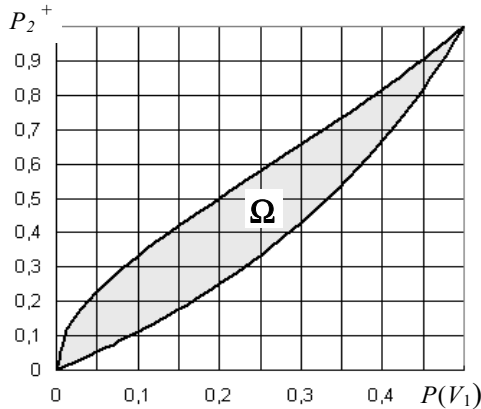


Рис. 3.3. Полезная совокупность «бесполезных» симптомов

Непосредственной проверкой по формуле Байеса можно убедиться в том, что $P(V_1 | x^+) \approx 0.357$, в то время как $P(V_2 | x^+) \approx 0.642$. С другой стороны, $P(V_1 | x^-) = 0$, а $P(V_2 | x^-) = 1$, т.е.

$$P(V_2 | x) > P(V_1 | x) \quad \forall x = x^+, x^-.$$

Следовательно, при любом значении отдельно взятого симптома оптимальным решением по правилу максимума апостериорных вероятностей $P(V_1 | x)$, $P(V_2 | x)$ оказывается решение $Z \in V_2$, совпадающее с решением, принимаемым по максимуму априорных вероятностей $P(V_1)$ и $P(V_2) = 1 - P(V_1)$ распознаваемых классов. Значит такой признак-симптом сам по себе бесполезен.

В то же время совместное использование таких условно независимых симптомов полезно: при обнаружении обоих симптомов оптимальным оказывается решение в пользу класса V_1 , поскольку $P(V_1 | x_1^+, x_2^+) \approx 0.74$, а $P(V_2 | x_1^+, x_2^+) \approx 0.26$.

3.2. Оценка полезности тестов в задачах скрининга

Переосмыслим теперь полученные результаты применительно к задаче скрининга, которая сводится к выявлению больных людей в большой группе обследуемых, где есть и больные (класс V_1) и здоровые (класс V_2). При этом подразумевается, что класс V_2 включает не абсолютно здоровых людей, а лиц, у которых нет рассматриваемого заболевания.

Пусть для выявления больных предполагается использовать диагностический тест. По-прежнему будем рассматривать тест как «черный ящик», который на основании доступной информации принимает одно из двух решений (диагнозов): положительное решение $\delta = 1$ (обследуемый «Болен») или отрицательное решение $\delta = 2$ (обследуемый «Здоров»).

Для оценки диагностической эффективности теста проводится его «экзамен» на некоторой репрезентативной группе пациентов с заранее известными диагнозами. Результаты такого экзамена можно представить в виде 2×2 матрицы (таблицы 3.3).

Таблица 3.3

Результаты тестирования репрезентативной группы

ИСТИННЫЙ ДИАГНОЗ	РЕЗУЛЬТАТ ТЕСТИРОВАНИЯ	
	Решение – «Болен» ($\delta=1$)	Решение – «Здоров» ($\delta=2$)
Класс V_1 – «Болен»	<i>TP</i>	<i>FN</i>
Класс V_2 – «Здоров»	<i>FP</i>	<i>TN</i>

В таблице приняты следующие обозначения:

– *TP* (*True Positive*) – число правильных диагнозов «Болен» (истинноположительный результат);

– TN (*True Negative*) – число правильных диагнозов «Здоров» (истинноотрицательный результат);

– FP (*False Positive*) – число здоровых, ошибочно признанных больными (ложноположительный результат или ошибка ложной тревоги);

– FN (*False Negative*) – число больных, ошибочно признанных здоровыми (ложноотрицательный результат или ошибка пропуска цели).

По данным табл. 3.3 легко оценить принятые в медицинской диагностике операционные характеристики теста – чувствительность и специфичность [13].

Чувствительность (*Sensitivity*) S_E определяется долей (процентом) предварительно верифицированных больных, которые тест признал больными

$$S_E = \frac{TP}{TP + FN}, \quad (3.31)$$

а специфичность (*Specificity*) S_P определяется долей (процентом) здоровых, которые тест отнес к группе здоровых

$$S_P = \frac{TN}{TN + FP}. \quad (3.32)$$

Очевидно, что если выборка репрезентативна, то S_E является оценкой величины $1 - P(E_1)$, а S_P оценкой величины $1 - P(E_2)$, где $P(E_1), P(E_2)$ – вероятности ошибок пропуска цели и ложной тревоги.

Из теоремы 3.1 немедленно следует, что для оценки полезности теста по описанной выше схеме достаточно проверить выполнение одного из условий

$$S_E > \theta(1 - S_P), \quad \text{если } \theta \geq 1, \quad (3.33)$$

$$S_E > 1 - \theta + \theta(1 - S_P), \quad \text{если } \theta < 1, \quad (3.34)$$

где параметр θ определяется по формуле

$$\theta = \frac{1 - P}{\omega P},$$

ω – соотношение потерь (3.5), а $P \neq 0$ – частота (преваленс) заболевания.

Поскольку численные примеры часто бывают более убедительными, чем формальные рассуждения, рассмотрим модельный пример.

Пример 3.3. Пусть планируется провести скрининг заболевания, встречаемость которого составляет 2 %, т.е. $P = 0,02$. Для диагностики предлагается использовать тест, имеющий чувствительность $S_E = 0,9$ и специфичность $S_p = 0,9$. Заметим, что такие операционные характеристики обычно считаются достаточно высокими для медицинских диагностических тестов.

Предположим, что потери L_{11}, L_{22} от правильных решений одинаковы и равны нулю, а потери от ложноотрицательных и ложноположительных результатов составляют соответственно $L_{12} = 5$ и $L_{21} = 1$, т.е. $\omega = 5$.

Поскольку при таких исходных данных $\theta > 1$, то проверке подлежит условие полезности (3.33). Легко убедиться в том, что это условие не выполняется, а значит тест абсолютно бесполезен для скрининга.

На рис. 3.5 представлены ожидаемые результаты тестирования группы в 10000 человек.

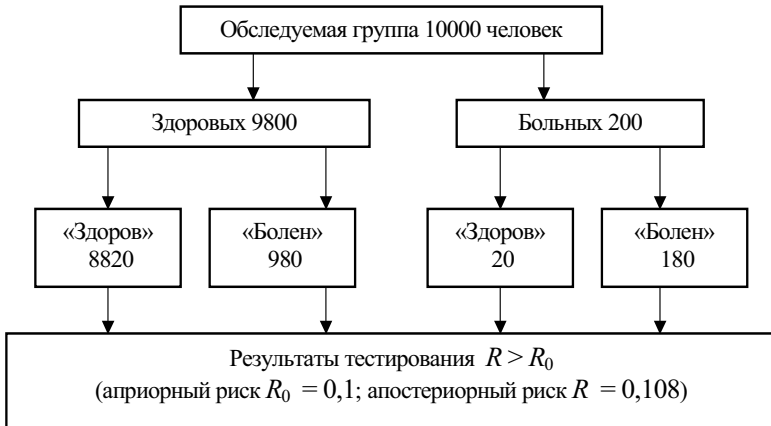


Рис.3.5. Ожидаемые результаты скрининга

Легко видеть, что априорные потери, связанные с тем, что 200 больных не будут выявлены, составляет $5 \cdot 200 = 1000$, т.е. $R_0 = 0,1$. В то же время суммарные потери, связанные с ложноположительными результатами диагностики для 980 здоровых пациентов и ложноотрицательными результатами для 20 человек, составят $1 \cdot 980 + 5 \cdot 20 = 1080$, т.е. $R = 0,108$.

Следовательно, несмотря на высокие показатели чувствительности и специфичности, скрининг на основе такого теста только увеличивает риск.

Если изменить условия примера, положив $L_{12} = L_{21} = 1$, то увеличение риска от применения теста оказывается еще более впечатляющим. В этом случае ожидаемый средний риск диагностики составит $R = 0,1$, в то время как априорный риск равен $R_0 = 0,02$.

Несмотря на кажущуюся «парадоксальность» того, что $R \gg R_0$ полученный результат имеет вполне обоснованное пояснение, приведенное нами в работе [47]: тест не является оптимальным с точки зрения байесовской стратегии. При указанных $P(V_1)$, L_{12} , L_{21} , S_E и S_P решения “Болен” являются не обоснованными с точки зрения минимизации апостериорного риска, а значит для *любого* обследуемого всегда должен приниматься диагноз “Здоров”, совпадающий с априорными решениями.

Важной особенностью скрининга и профилактических обследований является низкие значения преваленса больных в обследуемой группе. В частности, согласно [13] преваленс сахарного диабета составляет 2%, а рака молочной железы в возрасте более 50 лет – 1% [27].

Если к тому же допустить, что соотношение потерь ω от ложноотрицательных и ложноположительных ошибок не очень велико, то выполняется условие $\theta \geq 1$. В этих предположениях достаточно проверять лишь условие (3.33), чтобы убедиться в полезности теста.

На рис. 3.6 показаны границы областей полезности теста, построенные согласно (3.33) при $\omega = 5$ для преваленса $0 < P \leq 0,4$.

Из условия теоремы 3.2 прямо следует, что для значений P и ω , удовлетворяющих условию $P(1 + \omega) < 1$, или, что то же самое, $\theta > 1$, диагностический тест абсолютно бесполезен, если его специфичность удовлетворяет условию

$$S_p \leq S_p^0, \quad (3.35)$$

где

$$S_p^0 = \frac{1 - P(1 + \omega)}{1 - P}. \quad (3.36)$$

Как видно по графикам (рис.3.6) при малых значениях преваленса ($P < 0,02$) специфичность полезного теста должна быть выше 90 % уже при $\omega \leq 5$, а при условии равенства потерь от ложноположительных и ложноотрицательных результатов ($\omega = 1$) – выше 98 %.

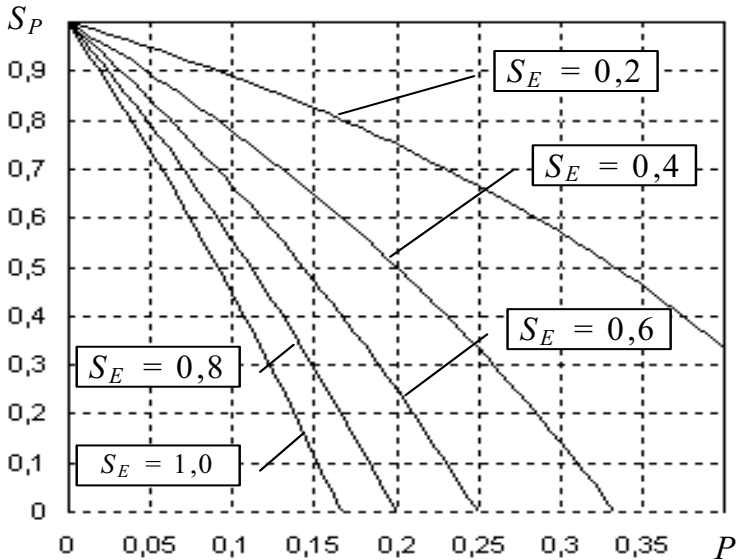


Рис. 3.6. Границы областей полезности теста

Поскольку для реальных тестов $S_E < 1$, причем, по мере снижения чувствительности, растет число ложноотрицательных результатов, то для обеспечения полезности теста величина специфичности должна быть выше приведенных граничных величин, что еще более сужает область полезности теста.

3.3. Традиционный и усиленный *ROC*-анализ

Удобным средством оценки эффективности диагностического теста является метод, основанный на анализе так называемой операционной характеристической кривой (*ROC* – *Receiver Operating Characteristic curve*), который получил широкую известность в научных публикациях, особенно за рубежом [70, 78, 79, 87, 93]. Традиционный *ROC*-анализ предусматривает сравнение операционных характеристик теста – чувствительности и специфичности [82, 92].

Диагностический тест с фиксированными операционными характеристиками (3.31) и (3.32) отображается точкой в *ROC*-пространстве, в котором по оси ординат откладываются значения S_E , а по оси абсцисс – значения $1 - S_p$ (рис. 3.7).

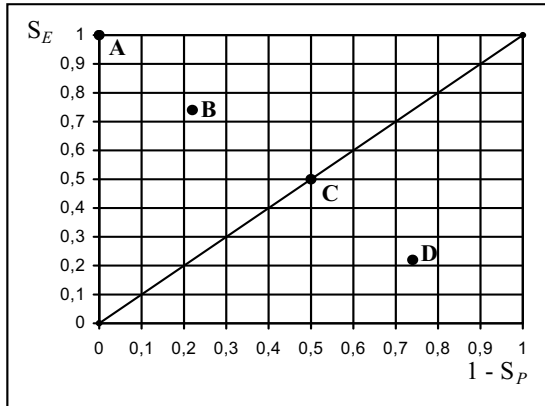


Рис. 3.7. Отображение характеристик теста в *ROC*-пространстве

ROC-пространство дает наглядное графическое представление о диагностической ценности теста и позволяет сравнивать эффективность различных тестов.

Идеальный тест (тест A на рис.3.7) расположен в точке с координатами (0, 1). Такой тест всегда принимает истинноположительный результат без ошибок ложной тревоги, а значит, относит всех больных к классу больных, не причисляя здоровых к больным. Отсюда следует, что диагностически ценные тесты находятся в левом верх-

нем углу ROC -пространства: чем ближе точка, определяемая операционными характеристиками S_E и S_P , к точке A , тем эффективнее тест.

Тесты, расположенные в левом нижнем углу ROC -пространства, являются «консервативными», поскольку при малом проценте ошибок ложной тревоги имеют низкую чувствительность. Понятно, что если консервативный тест, имеющий высокое значение специфичности, принял решение «Болен», то к такому решению следует относиться с высокой степенью доверия. В то же время решение «Здоров» может быть и не верным из-за низкой чувствительности.

Тесты, расположенные в правом верхнем углу, являются «либеральными», так как при больших значениях истинноположительных результатов дают также большой процент ошибок ложной тревоги. Поскольку такие тесты имеют высокое значение чувствительности, то его решение «Здоров», скорее всего, верно, в то время как противоположное решение «Болен» может быть ошибочным из-за низкой специфичности.

Тесты, расположенные на диагонали $S_E = 1 - S_P$ ROC -пространства (например, тест C на рис. 3.7), не дают дополнительную диагностическую информацию о классах, а потому являются *бесполезными*. Такие тесты эквивалентны классификаторам, использующим стратегию «случайного угадывания» классов.

Тесты, расположенные ниже диагонали (например, тест D на рис. 3.7), еще хуже тех, которые используют стратегию случайного угадывания. Однако, такой тест легко можно «перевести» в точку B , симметричную относительно диагонали, поменяв стратегию принятия решений на противоположную [81].

При решении задач медицинской диагностики часто используют пороговые решающие правила [27], например, правило вида

$$\begin{aligned} &\text{принимаем решение «Болен», если } x > x_0, \\ &\text{принимаем решение «Здоров», если } x \leq x_0, \end{aligned} \quad (3.37)$$

где x – некоторый измеряемый параметр (диагностический признак), x_0 – пороговое значение.

В этом случае появляется возможность оценить чувствительность и специфичность теста при различных значениях порога x_0 , что соот-

ветствует последовательности точек бинарного классификатора в ROC -пространстве (рис. 3.8).

Из правила (3.37) следует, что если при фиксированном пороговом значении x_0 пациент отнесен к классу «Здоров», то для любого более низкого значения признака x пациент также будет отнесен к этому же классу. Отсюда вытекает следующий простой алгоритм построения экспериментальной ROC -кривой [78]:

- отсортировать наблюдаемые значения признака x в порядке убывания;
- последовательно уменьшать (с некоторым шагом) порог x_0 , перемещаясь вниз по списку отсортированных значений x , и по экзаменационной выборке наблюдений вычислять соответствующие пары значений S_E и S_P ;
- отобразить полученные таким образом последовательности пар значений S_E и S_P в ROC -пространстве.

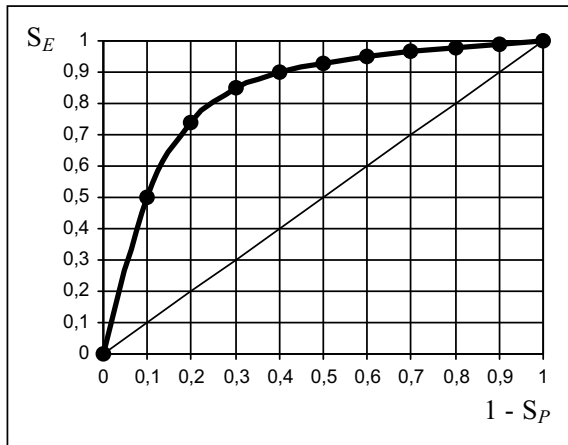


Рис. 3.8. ROC -кривая теста, основанного на правиле (3.37)

Заметим, что для операционных характеристик тестов, лежащих выше диагонали ROC -пространства, выполняется условие

$$S_E > 1 - S_P, \quad (3.38)$$

тогда как для бесполезных тестов справедливо равенство $S_E = 1 - S_P$.

Для сравнения эффективностей диагностических тестов удобно использовать некоторую интегральную характеристику, в качестве которой часто используют площадь AUC (*Area Under Curve*), вычисленную под ROC -кривой [80].

Очевидно, что величина AUC определяет среднюю чувствительность теста \bar{S}_E при возможных значениях специфичности $0 \leq S_p \leq 1$ или среднюю специфичность теста \bar{S}_p при возможных значениях чувствительности $0 \leq S_E \leq 1$, причем $AUC = 1$ для идеального теста и $AUC = 0,5$ для бесполезного теста, операционные характеристики которого не удовлетворяют условию (3.38). Поэтому считается, что чем больше AUC приближается к единице, тем эффективнее тест [97].

Площадь под ROC -кривой может быть вычислена с помощью правила трапеций [97], U -критерия Манна-Уитни [84] или коэффициента Джини [83].

ROC -анализ используют для нахождения оптимального значения порога x_0 , или, как часто говорят, точки отсечения (*Cut-off value*). Понятно, что оптимальное значение x_0 зависит от предъявляемых к тесту требований, т.е. от выбранного критерия оптимальности [99].

Например, в работе [98] оптимальной предлагается считать точку на ROC -кривой, которая имеет наибольшее значение чувствительности при наименьшем значении доли ложноположительных результатов. Эта точка может быть найдена с помощью индекса Йодена (*Youden index*) как максимальное отличие ординат ROC -кривой и диагонали (рис. 3.9) [91]:

$$J = \max(S_E + S_p - 1). \quad (3.39)$$

Мы рассмотрели лишь некоторые особенности традиционного ROC -анализа для случая двух классов. В общем же случае число возможных диагнозов $M \geq 2$. Очевидно, что для каждого истинного диагноза V_1, \dots, V_M диагностический тест может принять одно из M решений, но только одно из них будет верным. Отсюда следует, что в общем случае существует $N_M = M^2 - M$ возможных ошибок диагностического теста.

В рассмотренном случае ($M = 2$) число ошибок $N_2 = 2$, при числе классов $M = 3$ число возможных ошибок $N_3 = 6$ и т.д. Следовательно, уже при $M = 3$ ROC -анализ становится достаточно громозд-

ким и не столь наглядным: вместо двумерной *ROC*-кривой требуется проводить анализ 6-мерной *ROC*-поверхности.

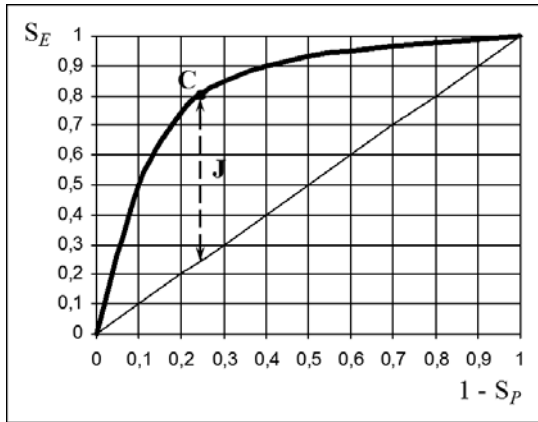


Рис. 3.9. Оптимальная точка отсечения *C* (по индексу Йордена)

Но можно пойти и другим путем, последовательно решая задачу распознавания одного класса $V_i \in \{V_1, \dots, V_M\}$ и всех остальных. При таком подходе, вместо N_M -мерной *ROC*-поверхности достаточно построить и проанализировать M обычных двумерных *ROC*-кривых, отображающих взаимосвязь истинноположительных и ложноположительных результатов, полученных при распознавании каждого класса $V_i \in \{V_1, \dots, V_M\}$ и объединенного класса

$$\bar{V}_i = \bigcup_{j \neq i} V_j, \quad i, j = 1, \dots, M.$$

Рассмотрим теперь некоторые усовершенствования традиционного *ROC*-анализа, позволяющие оценивать эффективность диагностического теста с точки зрения ожидаемых потерь (среднего риска).

Пусть P – вероятность появления больного в группе обследуемых (преваленс заболевания). Будем как и прежде характеризовать возможные ошибки диагностического теста величинами L_{12}, L_{21} , которые определяют плату (потери) за одну ошибку пропуска цели (не

выявление больного) и ложной тревоги (отнесение здорового пациента к группе больных).

В этом случае, как было показано выше, ожидаемые потери (средний риск) от неправильной диагностики определит соотношение

$$R = P(V_1) \cdot (1 - S_E) \cdot L_{12} + [1 - P(V_1)] \cdot (1 - S_P) \cdot L_{21}. \quad (3.40)$$

Из (3.40) непосредственно следует, что два теста, имеющие операционные характеристики $S_E^{(1)}, S_P^{(1)}$ и $S_E^{(2)}, S_P^{(2)}$, будут иметь одинаковые потери, если выполняется условие [89]

$$\frac{S_E^{(2)} - S_E^{(1)}}{(1 - S_P^{(2)}) - (1 - S_P^{(1)})} = m, \quad (3.41)$$

где

$$m = \frac{L_{21} [1 - P(V_1)]}{L_{12} P(V_1)}. \quad (3.42)$$

Семейство параллельных прямых

$$S_E = m(1 - S_P) + \text{const}$$

с положительным угловым коэффициентом m , зависящим от $P(V_1)$, L_{12} и L_{21} , образует в ROC -пространстве изолинии эффективности (*Iso-performance lines*) [89].

Все тесты, соответствующие точкам на изолинии, имеют одни и те же ожидаемые потери, причем, чем больше значение пересечения изолинии с осью ординат S_E , тем меньше потери. Поэтому, если изолинию эффективности провести как касательную к ROC -кривой, то в точке (или на отрезке) касания изолинии и ROC -кривой потери будут минимальными [99].

Для сравнения эффективностей нескольких тестов с точки зрения ожидаемых потерь в работе [89] предложен так называемый метод выпуклой ROC -оболочки (*ROC convex hull method*), суть которого состоит в следующем:

- определяем операционные характеристики каждого i -го бинарного теста и отображаем в ROC -пространстве точки с координатами $S_E[i]$ и $1 - S_P[i]$, а для каждого j -го теста, основанного на пороговом решающем правиле (3.37), последовательно изменяем порог

$x_0^{(j)}$ и отображаем множество пар точек $S_E^{(j)}[i]$, $1 - S_P^{(j)}[i]$ на j -й ROC-кривой;

- строим выпуклую оболочку множества точек, что для n точек может быть сделано с помощью алгоритма QuickHull за время $O[n \log(n)]$ [67];

- по формуле (3.42) вычисляем угловой коэффициент m наклона изолинии эффективности и отображаем в ROC-пространстве семейство изолиний эффективности, соответствующих заданным значениям $P(V_1)$, L_{12} и L_{21} ;

- точка (или совокупность точек) касания изолинии эффективности к построенной выпуклой оболочке определит операционные характеристики S_E^{opt} и S_P^{opt} наиболее эффективного теста, поскольку именно такой тест ближе всего к «идеальному» тесту с координатами $1 - S_P = 0$ и $S_E = 1$.

Для иллюстрации приведенного алгоритма на рисунке 3.10 представлены примеры экспериментальных ROC-кривых различных тестов и соответствующая выпуклая оболочка (отмечена пунктирной линией). Различные заданные значения $P(V_1)$, L_{12} и L_{21} определяют наклоны трех изолиний эффективности с угловыми коэффициентами m_1 , m_2 и m_3 .

В первом случае (рис. 3.10, а) ROC-кривые всех трех тестов «касаются» выпуклой оболочки. Поэтому при условии $m = m_1$ потенциально наилучшим будет тест A , при $m = m_2$ – тест B , а при $m = m_3$ – тест C .

Во втором случае (рис. 3.10, б) ROC-кривая теста C «не касается» выпуклой оболочки. Поэтому в такой ситуации для условия $m = m_3$ наилучшим с точки зрения ожидаемых потерь будет уже не тест C , как в первом случае, а тест A .

Еще один пример показан на рис. 3.10, в. При условии $m = m_1$ наилучшим будет тест A , при $m = m_2$ – тест B , а в случае $m = m_3$ оба теста A и B не обеспечивают заданные ожидаемые потери, поскольку не существует точек касания соответствующей изолинии эффективности с выпуклой оболочкой.

Понятно, что после того, как для некоторой совокупности тестов построена выпуклая оболочка, исследование эффективности «нового» теста предполагает перепостроение выпуклой оболочки. Однако

из добавляемых тестов потенциально оптимальными могут быть лишь те, которые «расширяют» выпуклую оболочку.

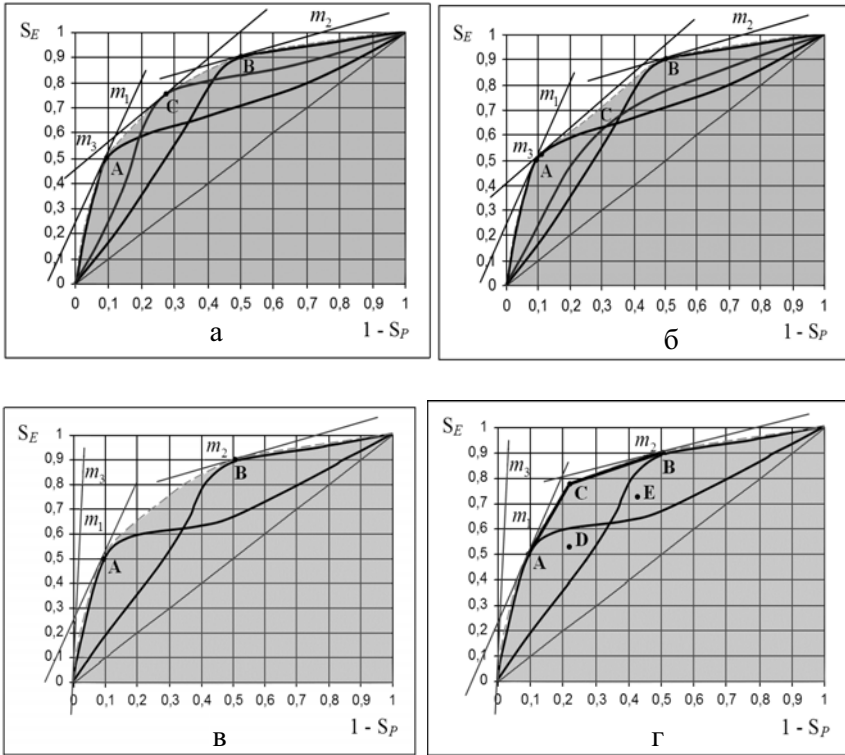


Рис. 3.10. Иллюстрация метода выпуклой ROC-оболочки (пояснения в тексте)

Например, добавим к выпуклой оболочке, построенной по двум ROC-кривым, три бинарных теста C , D и E (рис. 3.10, г). Легко видеть, что при условии $m \in [m_1, m_2]$ наилучшими могут быть только тесты, находящиеся на ломаной ACB . Следовательно, тест C может быть оптимальным, поскольку он расширяет выпуклую ROC-оболочку, в то время как тесты D и E не являются потенциально оптимальными.

В отличие от метода выпуклой ROC-оболочки, полученные нами достаточные условия (3.33) и (3.34) не требуются привлечения визу-

ального анализа результатов в ROC -пространстве и позволяют непосредственно оценить пригодности нового диагностического теста с известными операционными характеристиками S_E , S_P для скрининга заболеваний с преваленсом P при заданном соотношением потерь $\omega = L_{12}L_{21}^{-1}$ от ошибок пропуска цели и ложной тревоги.

В то же время на основе условий теоремы 3.1 можно усилить традиционный ROC -анализ, ограничив соответствующую область анализа ROC -кривой [52].

Рис. 3.11 иллюстрирует предлагаемый подход для случаев $m \geq 1$ (рис. 3.11, а) и $m < 1$ (рис. 3.11, б). Серым цветом показаны области гарантированной полезности теста.

В соответствии с (3.33) прямая, ограничивающая область полезности теста для случаев $m \geq 1$, проходит через точку с координатами $(0,0)$ и точку с координатами $(1/m, 1)$, а для случая $m < 1$ – через точку с координатами $(1,1)$ и точку с координатами $(0, 1 - m)$.

Отрезок OA (рис. 3.11, а) соответствует случаю, когда $P(V_1) = 0,15$ и $\omega = L_{12}L_{21}^{-1} = 1$. Видно, что в данном случае тест бесполезен, поскольку ROC -кривая расположена ниже отрезка OA .

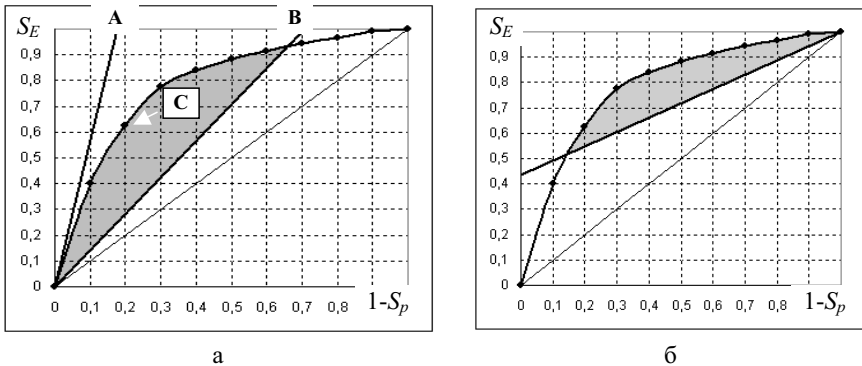


Рис. 3.11. Иллюстрация идеи усиленного ROC -анализа для случаев а – $m \geq 1$ и б – $m < 1$

Но если допустить, что потери от пропуска цели в четыре раза выше потерь от ложной тревоги, т.е. допустить, что $L_{12}L_{21}^{-1} = 4$, то соответствующая прямая OB уже пересекает ROC -кривую.

Следовательно, тест, имеющий чувствительность $C_E = 62,5\%$ и специфичность $C_P = 80\%$, окажется гарантированно полезным с точки зрения строгого неравенства (3.34), поскольку точка C , соответствующая таким операционным характеристикам, расположена на допустимой части ROC -кривой.

Заметим, что условия (3.33 и (3.34) позволяют решить и обратную задачу: определить допустимый интервал соотношения потерь $\omega = L_{12}L_{21}^{-1}$, при котором тест с чувствительностью S_E и специфичностью S_P можно считать полезным для скрининга заболевания с известным преваленсом P .

Этот интервал определяется формулой

$$\frac{1-P}{P} \frac{1-S_P}{S_E} \leq \omega \leq \frac{1-P}{P} \frac{S_P}{1-S_E}. \quad (3.43)$$

Например, тест, имеющий чувствительность $S_E = 62,5\%$ и специфичность $S_P = 80\%$, можно рекомендовать для выявления заболеваний с преваленсом $P(V_1) = 0,15$, если допустимо соотношение потерь от ложноотрицательных и ложноположительных ошибок в пределах $2,66 \leq \omega \leq 17,7$.

Примечательно, что полученные результаты можно сопоставить с изолиниями эффективности, о которых шла речь. Действительно, из выражения (3.41) следует, что уравнение изолинии с ожидаемыми потерями $R = R^*$ имеет вид:

$$S_E = m(1 - S_P) + 1 - \frac{R^*}{L_{12}P}. \quad (3.44)$$

Из сравнения (3.44) с (3.33) и (3.34) заключаем, что изолинии эффективности (3.44) совпадет с прямыми, ограничивающими области полезных тестов, если выполняется условие

$$R^* = \begin{cases} L_{12}P, & \text{при } m \geq 1, \\ L_{21}(1 - P), & \text{при } m < 1, \end{cases} \quad (3.45)$$

правая часть которого согласно (3.11) определяет априорный риск R_0 при $L_{11} = L_{22} = 0$.

Таким образом, область полезных тестов ограничивает одна изолиния эффективности, соответствующая уровню *априорного* риска R_0 , которая проходит через точку с координатами $(0,0)$ при $m \geq 1$ или точку с координатами $(1,1)$ при $m < 1$, причем, если такая изолиния

пересекает <i>ROC</i> -кривую,	то $R < R_0$,
касается <i>ROC</i> -кривой,	то $R = R_0$,
не пересекает <i>ROC</i> -кривую,	то $R > R_0$.

3.4. Интервальное обобщение результатов

При решении практических задач чаще всего неизвестны точные значения вероятностей ошибок

$$P(E_1) \triangleq P_{12} = P\{\delta = 1 | V_2\}, \quad (3.46)$$

$$P(E_2) \triangleq P_{21} = P\{\delta = 2 | V_1\}, \quad (3.47)$$

фигурирующих в условиях (3.6), (3.7) полезности теста. Ясно также и то, что лишь при достаточно большом числе наблюдений n_1 и n_2 правомерна замена этих вероятностей точечными оценками

$$P_{jk}^* = \frac{E_j}{n_k}, \quad k, j = 1, 2, \quad (3.48)$$

где n_k , E_j – соответственно число случаев, когда объект заведомо находился в состоянии V_k , а диагностический тест принял неверное решение $j \neq k$.

Поэтому представляет интерес обобщение условий (3.6), (3.7) на случай, когда вместо точечных значений вероятностей P_{12} и P_{21} используются их доверительные интервалы.

Согласно [11, с. 332], частота P_X^* случайного события X , вычисленная на выборке объемом n , с доверительной вероятностью β определяет доверительный интервал I_X для вероятности P_X , который можно записать в виде

$$I_X = \langle P_X^c, r_X \rangle, \quad (3.49)$$

где

$$P_X^c = \frac{P_X^* + t_\beta^2/2n}{1 + t_\beta^2/n}, \quad (3.50)$$

$$r_X = \frac{t_\beta \sqrt{\frac{P_X^*(1-P_X^*)}{n} + \frac{t_\beta^2}{4n^2}}}{1 + t_\beta^2/n} \quad (3.51)$$

– соответственно центр и радиус доверительного интервала I_X .

В этих соотношениях

$$t_\beta = \arg \Phi^* \left(\frac{1 + \beta}{2} \right),$$

где $\Phi^*(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{\tau^2}{2}} d\tau$ – нормальная функция распределения.

Принимая во внимание (3.47) – (3.49) и используя обозначение

$$\theta = \frac{\omega P}{1 - P}, \quad (3.52)$$

где $P = P(V_1)$ – априорная вероятность первого класса (в частности, преваленс заболевания в задачах скрининга), а ω – безразмерная величина определяется выражением (3.5), докажем следующую теорему [21].

Теорема 3.5. При фиксированном значении θ с вероятностью β диагностический тест полезен в смысле определения 3.1, если выполняется одно из условий

$$P_{12}^c + r_{12} < \theta(1 - P_{21}^c - r_{21}) \text{ при } \theta \leq 1, \quad (3.53)$$

$$P_{12}^c + r_{12} < 1 - \theta(P_{21}^c + r_{21}) \text{ при } \theta > 1, \quad (3.54)$$

где P_{12}^c, P_{21}^c и r_{12}, r_{21} – соответственно центры и радиусы доверительных интервалов $I_{12} = \langle P_{12}^c, r_{12} \rangle, I_{21} = \langle P_{21}^c, r_{21} \rangle$ условных вероятностей (3.46), (3.47), которые, согласно (3.50), (3.51), зависят исключительно

от частот (3.48), вычисленных по n_k наблюдениям с известным состоянием объекта, и доверительной вероятности β .

Доказательство. Поскольку по определению доверительного интервала $P_{12} \in I_{12}$ и $P_{21} \in I_{21}$, то при выполнении условия (3.53) с доверительной вероятностью β условные вероятности (3.46), (3.47) удовлетворяют неравенству

$$P_{12} < \theta(1 - P_{21}), \quad (3.55)$$

а при выполнении условия (3.54) – неравенству

$$P_{12} < 1 - \theta P_{21}. \quad (3.56)$$

Неравенства (3.55) и (3.56) с учетом обозначения (3.52) эквивалентны условиям (3.6), (3.7), которые, в соответствии с теоремой 3.1, гарантируют полезность диагностического теста с точки зрения определения 3.1. Теорема 3.5 доказана.

Повторяя рассуждения, использованные при доказательстве теоремы 3.5, легко показать, что при заданном значении θ с вероятностью β диагностический тест заведомо бесполезен, если выполняются условия

$$P_{12}^c - r_{12} > \theta(1 - P_{21}^c + r_{21}) \text{ при } \theta \leq 1, \quad (3.57)$$

$$P_{12}^c - r_{12} > 1 - \theta(P_{21}^c - r_{21}) \text{ при } \theta > 1. \quad (3.58)$$

Замечание 3.2. Если параметры доверительных интервалов $I_{12} = \langle P_{12}^c, r_{12} \rangle$, $I_{21} = \langle P_{21}^c, r_{21} \rangle$, удовлетворяют условиям (3.53), (3.54) или же условиям (3.57), (3.58), то с вероятностью β гарантировано выполнение строгого неравенства (3.4). Однако обратное утверждение неверно: вопрос о полезности диагностического теста остается открытым, если не выполняется ни одно из условий (3.53), (3.54) и (3.57), (3.58).

В качестве иллюстрации на рис. 3.12 представлены области значений частот ошибок диагностический теста P_{12}^* и P_{21}^* , которые с вероятностью $\beta = 0,99$ подтверждают (область A) и опровергают (область B) квалификацию диагностический теста. Указанные области

построены для следующих фиксированных величин: $P = 0,15$; $L_{12} = 5$; $L_{21} = 1$, $L_{11} = L_{22} = 0$, т.е. $\omega = 5$ и $\theta = 0,882$.

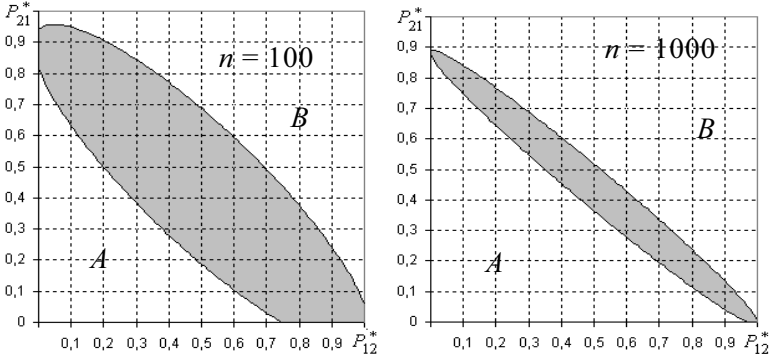


Рис. 3.12. Области неопределенности относительно полезности диагностического теста

Как и следовало ожидать, область неопределенной полезности диагностический теста (серая область) уменьшается по мере увеличения объема n наблюдений.

Для решения практических задач представляет интерес определение необходимого объема экспериментальной выборки, по которой должны быть оценены частоты ошибок диагностический теста (3.48) для того, чтобы с заданной вероятностью β однозначно подтвердить или опровергнуть полезность диагностический теста.

Прежде чем доказать теорему, определяющую такие условия, докажем вспомогательную лемму.

Лемма 3.1. Функции

$$f_1(n) = P_X^c(n) - r_X(n),$$

$$f_2(n) = 1 - P_X^c(n) - r_X(n),$$

строго монотонно возрастают с ростом n , а функции

$$f_3(n) = P_X^c(n) + r_X(n),$$

$$f_4(n) = 1 - P_X^c(n) + r_X(n),$$

строго монотонно убывают с ростом n .

Доказательство. Рассмотрим $f_1(n)$ как функцию от непрерывного аргумента n . Покажем, что первая производная этой функции положительна при любом $n > 0$, т.е.

$$f_1'(n) = \frac{t_\beta^2}{2n^2} \left(\frac{P_X^*(1-P_X^*) + \frac{t_\beta^2}{2n}}{t_\beta \sqrt{\frac{P_X^*(1-P_X^*)}{n} + \frac{t_\beta^2}{4n^2}}} - 1 \right) > 0.$$

Так как величина $\frac{t_\beta^2}{2n^2}$ заведомо положительна, для этого достаточно показать, что

$$\frac{P_X^*(1-P_X^*) + \frac{t_\beta^2}{2n}}{t_\beta \sqrt{\frac{P_X^*(1-P_X^*)}{n} + \frac{t_\beta^2}{4n^2}}} - 1 > 0,$$

или

$$P_X^*(1-P_X^*) + \frac{t_\beta^2}{2n} > t_\beta \sqrt{\frac{P_X^*(1-P_X^*)}{n} + \frac{t_\beta^2}{4n^2}}.$$

Поскольку обе части последнего неравенства положительны, оно эквивалентно неравенству

$$\left(P_X^*(1-P_X^*) + \frac{t_\beta^2}{2n} \right)^2 > \left(t_\beta \sqrt{\frac{P_X^*(1-P_X^*)}{n} + \frac{t_\beta^2}{4n^2}} \right)^2,$$

которое, после элементарных преобразований, приобретает вид

$$(P_X^*(1-P_X^*))^2 + \frac{t_\beta^2}{n} P_X^*(1-P_X^*) + \frac{t_\beta^2}{4n^2} > \frac{t_\beta^2}{n} P_X^*(1-P_X^*) + \frac{t_\beta^4}{4n^2}.$$

В результате приходим к неравенству $(P_X^*(1-P_X^*))^2 > 0$, справедливость которого доказывает, что функция $f_1(n)$ строго монотонно возрастает для любого $n > 0$.

Аналогично можно показать, что функция $f_2(n)$ строго монотонно возрастает, а функции $f_3(n), f_4(n)$ строго монотонно убывают при любых непрерывных $n > 0$.

Очевидно, что утверждения леммы остаются справедливыми и при целочисленных (дискретных) значениях $n > 1$. Лемма 3.1 доказана.

Теорема 3.6. Для любой доверительной вероятности β и фиксированном значении θ существует такое число $n_0 > 0$, что после оценки частот P_{jk}^* , $j, k = 1, 2, j \neq k$ ошибок диагностической теста по репрезентативной выборке объемом $n > n_0$ можно с вероятностью β подтвердить или опровергнуть полезность диагностической теста.

Доказательство. Для доказательства теоремы достаточно показать, что независимо от значений θ и β для любых P_{jk}^* , $j, k = 1, 2, j \neq k$ всегда возможно за счет увеличения количества испытаний обеспечить выполнение одного из условий (3.53), (3.57) при $\theta \leq 1$, либо одного из условий (3.54), (3.58) при $\theta > 1$.

Пусть при некотором значении $n > 0$ получены оценки частот P_{jk}^* такие, что для заданных значений $\theta \leq 1$ и β совместно выполняются условия

$$P_{12}^c + r_{12} > \theta(1 - P_{21}^c - r_{21}), \quad (3.59)$$

$$P_{12}^c - r_{12} < \theta(1 - P_{21}^c + r_{21}). \quad (3.60)$$

В этом случае невозможно с вероятностью β определить является диагностический тест полезным или нет.

Поскольку $\theta > 0$ и, в соответствии с леммой 3.1, при увеличении числа наблюдений n функции $P_{12}^c(n) - r_{12}(n)$ и $\theta[1 - P_{21}^c(n) - r_{21}(n)]$ строго монотонно возрастают, а функции $P_{12}^c(n) + r_{12}(n)$ и $\theta[1 - P_{21}^c(n) + r_{21}(n)]$ строго монотонно убывают, то всегда найдется такое значение $n_0 > n$, что одно из строгих неравенств (3.59), (3.60) нарушится, т.е.

$$P_{12}^c + r_{12} \leq \theta(1 - P_{21}^c - r_{21}), \quad (3.61)$$

или

$$P_{12}^c - r_{12} \geq \theta(1 - P_{21}^c + r_{21}). \quad (3.62)$$

Отсюда следует, что для всех $n > n_0$ будет выполняться условие (3.53), которое в соответствии с теоремой 3.5 с доверительной вероятностью β гарантирует полезность диагностический теста при $\theta \leq 1$, либо условие (3.57), которое с доверительной вероятностью β опровергает полезность диагностический теста при $\theta \leq 1$.

Аналогичные рассуждения позволяют доказать справедливость теоремы и для случая $\theta > 1$. Теорема 3.6 доказана.

Важным для практики следствием теоремы 3.6 является возможность численной оценки объема n экспериментальной выборки, необходимого для формальной оценки квалификации диагностический теста. Для заданного значения θ необходимо число экспериментов n_0 , при котором с доверительной вероятностью β квалификация диагностический теста может быть подтверждена или опровергнута, определяется решением уравнений

$$P_{12}^c(n) \pm r_{12}(n) = \theta[1 - P_{21}^c(n) \mp r_{21}(n)], \quad \text{если } \theta \leq 1,$$

$$P_{12}^c(n) \mp r_{12}(n) = 1 - \theta[P_{21}^c(n) \pm r_{21}(n)], \quad \text{если } \theta > 1.$$

Сформулируем теперь формальные условия, позволяющие сравнивать эффективность двух диагностический тестов с точки зрения такого определения.

Определение 3.3. Диагностический тест A_1 эффективнее диагностического теста A_2 , если средний риск R_1 , основанный на решениях A_1 , меньше среднего риска R_2 , основанного на решениях A_2 , т.е. выполняется строгое неравенство

$$R_1 < R_2. \quad (3.63)$$

Теорема 3.7. С доверительной вероятностью β диагностический тест A_1 эффективнее, чем диагностический тест A_2 , если

$$\theta(P_{21}^{c(1)} + r_{21}^{(1)} - P_{21}^{c(2)} + r_{21}^{(2)}) < P_{12}^{c(2)} - r_{12}^{(2)} - P_{12}^{c(1)} - r_{12}^{(1)}, \quad (3.64)$$

и, наоборот, с доверительной вероятностью β диагностический тест A_2 эффективнее, чем диагностический тест A_1 , если

$$\theta(P_{21}^{c(1)} - r_{21}^{(1)} - P_{21}^{c(2)} - r_{21}^{(2)}) > P_{12}^{c(2)} + r_{12}^{(2)} - P_{12}^{c(1)} + r_{12}^{(1)}, \quad (3.65)$$

где $P_{12}^{c(i)}$, $P_{21}^{c(i)}$ и $r_{12}^{(i)}$, $r_{21}^{(i)}$ – соответственно центры и радиусы доверительных интервалов условных вероятностей (3.46), (3.47) ошибок первого и второго рода i -го диагностический теста ($i = 1, 2$), которые, согласно (3.50), (3.51), зависят исключительно от доверительной вероятности β и частот ошибок (3.48), допущенных соответствующим диагностический тестом на экспериментальных выборках известного объема, а θ – константа, которая в соответствии с (3.52) с учетом (3.5) определяется значениями L_{11} , L_{22} , L_{12} , L_{21} и P .

Доказательство. Поскольку по определению доверительного интервала имеем $P_{12}^{(i)} \in \langle P_{12}^{c(i)}, r_{12}^{(i)} \rangle$ и $P_{21}^{(i)} \in \langle P_{21}^{c(i)}, r_{21}^{(i)} \rangle$, то с доверительной вероятностью β выполняются условия

$$P_{21}^{(1)} < P_{21}^{c(1)} + r_{21}^{(1)},$$

$$P_{21}^{(2)} > P_{21}^{c(2)} - r_{21}^{(2)},$$

т.е.

$$P_{21}^{(1)} - P_{21}^{(2)} < P_{21}^{c(1)} + r_{21}^{(1)} - P_{21}^{c(2)} + r_{21}^{(2)}. \quad (3.66)$$

Аналогично легко показать, что с доверительной вероятностью β выполняется условие

$$P_{12}^{(2)} - P_{12}^{(1)} > P_{12}^{c(2)} - r_{12}^{(2)} - P_{12}^{c(1)} - r_{12}^{(1)}. \quad (3.67)$$

Следовательно, если справедливо (3.64), то на основании (3.66), (3.67), заведомо выполняется условие

$$\theta(P_{21}^{(1)} - P_{21}^{(2)}) < P_{12}^{(2)} - P_{12}^{(1)}. \quad (3.68)$$

Подстановка (3.52) в (3.68) с учетом (3.5) дает

$$P(V_1)(P_{21}^{(1)} - P_{21}^{(2)})(L_{21} - L_{11}) < (1 - P(V_1))(P_{12}^{(2)} - P_{12}^{(1)})(L_{12} - L_{22}).$$

Отсюда после элементарных преобразований имеем

$$L_{11}P(V_1)(1 - P_{21}^{(1)}) + L_{22}(1 - P(V_1))(1 - P_{12}^{(1)}) +$$

$$\begin{aligned}
& + L_{21}P(V_1)P_{21}^{(1)} + L_{12}(1 - P(V_1))P_{12}^{(1)} < \\
& < L_{11}P(V_1)(1 - P_{21}^{(2)}) + L_{22}(1 - P(V_1))(1 - P_{12}^{(2)}) + \\
& + L_{21}P(V_1)P_{21}^{(2)} + L_{12}(1 - P(V_1))P_{12}^{(2)},
\end{aligned}$$

А это как раз и означает, что с доверительной вероятностью β выполняется строгое неравенство $R_1 < R_2$.

Аналогичным образом можно показать, что выполнение условия (3.65) с доверительной вероятностью β влечет выполнение строгого неравенства $R_1 > R_2$. Теорема 3.7 доказана.

Понятно, что если не выполняется ни одно из условий (3.64), (3.65), невозможно определить, какой из диагностических тестов является более эффективным. Поэтому представляет интерес определение условий, налагаемых на объем экспериментальной выборки, по которой должны быть оценены частоты ошибок диагностических тестов для того, чтобы с заданной доверительной вероятностью β сравнить их эффективность.

Прежде чем доказать теорему, определяющую эти условия, докажем еще одну вспомогательную лемму.

Лемма 3.2. Функция

$$\varphi_1(n) = P_{X_1}^c(n) - r_{X_1}(n) - P_{X_2}^c(n) - r_{X_2}(n), \quad (3.69)$$

строго монотонно возрастает с ростом n , а функция

$$\varphi_2(n) = P_{X_2}^c(n) + r_{X_2}(n) - P_{X_1}^c(n) + r_{X_1}(n), \quad (3.70)$$

строго монотонно убывает с ростом n .

Доказательство. Рассмотрим функцию $\varphi_1(n) = f_1(n) - f_2(n)$ от непрерывного аргумента n , где

$$f_1(n) = P_{X_1}^c(n) + r_{X_1}(n), \quad f_2(n) = P_{X_2}^c(n) - r_{X_2}(n).$$

Поскольку, согласно лемме 3.1,

$$f_1' < 0, \text{ а } f_2' > 0,$$

то

$$\varphi_1'(n) = f_1'(n) - f_2'(n) < 0.$$

Следовательно, функция $\varphi_1(n)$ строго монотонно возрастает с ростом n .

Аналогично можно показать, что функция $\varphi_2(n)$ строго монотонно убывает при любых непрерывных $n > 0$. Очевидно, что утверждения леммы 3.2 остаются справедливыми и при целочисленных (дискретных) значениях $n > 0$. Лемма 3.2 доказана.

Теорема 3.8. Для любой доверительной вероятности β и фиксированном значении θ существует такие числа $n_1^0 > 0$ и $n_2^0 > 0$, что после оценки частот (3.48) ошибок диагностический тестов по выборкам объемом $n_1 > n_1^0$ и $n_2 > n_2^0$ можно с вероятностью β определить, какой из диагностический тестов эффективнее с точки зрения определения 3.3.

Доказательство. Пусть по выборкам, содержащим n_1^* и n_2^* наблюдений, оценены частоты ошибок первого и второго рода каждого из диагностический тестов, однако для заданных значений θ и β совместно выполняются условия

$$\theta(P_{21}^{c(1)} + r_{21}^{(1)} - P_{21}^{c(2)} + r_{21}^{(2)}) > P_{12}^{c(2)} - r_{12}^{(2)} - P_{12}^{c(1)} - r_{12}^{(1)}, \quad (3.71)$$

$$P_{12}^{c(2)} + r_{12}^{(2)} - P_{12}^{c(1)} + r_{12}^{(1)} > \theta(P_{21}^{c(1)} - r_{21}^{(1)} - P_{21}^{c(2)} - r_{21}^{(2)}), \quad (3.72)$$

В этом случае ни одно из условий теоремы 3.7 не выполняется и невозможно с вероятностью β определить, какой из диагностический тестов более полезен.

Поскольку $\theta > 0$ и, согласно лемме 3.2, функции, фигурирующие в левой части (3.71), (3.72), строго монотонно убывают при увеличении числа наблюдений n_1 и n_2 , а функции, фигурирующие в правой части (3.71), (3.72) строго монотонно возрастают при увеличении числа наблюдений n_1 и n_2 , то всегда найдутся такие значения $n_1^0 > n_1^*$ и $n_2^0 > n_2^*$, что одно из строгих неравенств (3.71) или (3.72) нарушится. Следовательно, при числе наблюдений $n_1 > n_1^0$ и $n_2 > n_2^0$ будет выполняться одно из условий теоремы 3.7. Теорема 3.8 доказана.

Для иллюстрации на рисунке 3.13 приведены графики функций, представляющие собой зависимости от числа наблюдений n левой (кривая 1) и правой (кривая 2) частей неравенства (3.64), а также левой (кривая 3) и правой (кривая 4) частей неравенства (3.65).

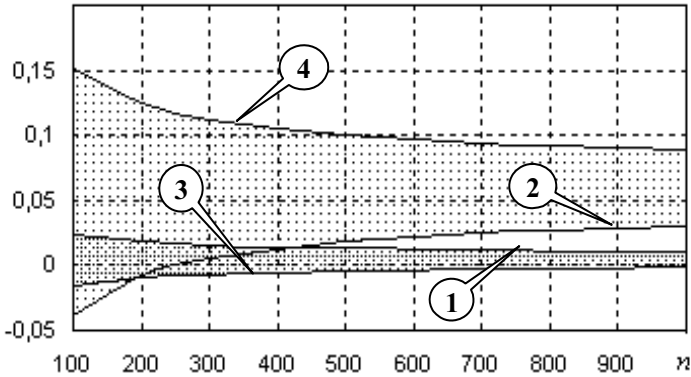


Рис. 3.13. Области, в которых возможно сравнения эффективности двух диагностических тестов

Указанные зависимости построены для значения доверительной вероятности $\beta = 0,99$ при частотах ошибок диагностических тестов $P_{12}^{*(1)} = 0,01$, $P_{12}^{*(2)} = 0,07$, $P_{21}^{*(1)} = 0,04$, $P_{21}^{*(2)} = 0,02$, априорной вероятности $P(V_1) = 0,1$ и значениях элементов матрицы потерь $L_{11} = L_{22} = 0$, $L_{21} = 2$, $L_{12} = 1$, т.е. для случая $\omega = 2$, когда $\theta \approx 0,22$.

Легко видеть, что в рассматриваемом случае однозначное сравнение эффективностей диагностических тестов возможно лишь в том случае, когда частоты их ошибок оценены более чем по $n > 430$ наблюдениям.

Заметим, что величина θ , фигурирующая в условиях (3.64), (3.65) теоремы 3.7, является константой лишь в том случае, когда известны точные значения элементов платежной матрицы (3.2) и априорные вероятности состояния объекта. Общий случай, в котором сделан переход от точечных значений L_{ij} и P к их интервальным аналогам $\langle L_{ij}^c, r_{ij} \rangle$ и $\langle P^c, r_p \rangle$ в форме центр-радиус, рассмотрен в работе [21].

ГЛАВА 4

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ РЕЗУЛЬТАТОВ

4.1. Анализ диагностических признаков термограммы

Одним из перспективных методов быстрого определения содержания углерода C в жидкой стали признан метод, основанный на статистической зависимости между C и температурой начала кристаллизации (температурой ликвидуса) T_L [36].

Величину T_L определяют как температуру, при которой в процессе кристаллизации расплава на кривой охлаждения (термограмме) за счет выделения скрытой теплоты кристаллизации появляется своеобразный фрагмент в виде горизонтальной или наклонной площадки¹ (рис. 4.1). Между тем оказывается, что в ряде случаев подобный аномальный участок появляется не при T_L , а при некоторой другой температуре.

До тех пор, пока не будет зарегистрирована сама термограмма, нельзя определенно знать, появится ли характерный фрагмент при температуре T_L или при другой температуре.

Как видно из сравнения реальных измерений (рис. 4.1), термограммы, на которых фрагмент появлялся при температуре T_L (термический эффект) и при другой температуре (псевдотермический эффект) внешне настолько сходны, что сразу еще не ясно, по каким признакам можно осуществлять распознавание термограмм с термическим (класс V_1) и псевдотермическим (класс V_2).

¹ На рисунке 4.1 ось температур T направлена вправо, а ось времени t – вверх.

Поэтому понадобилась заняться построением исходного пространства признаков термограммы, эффективных для распознавания классов V_1 и V_2 .

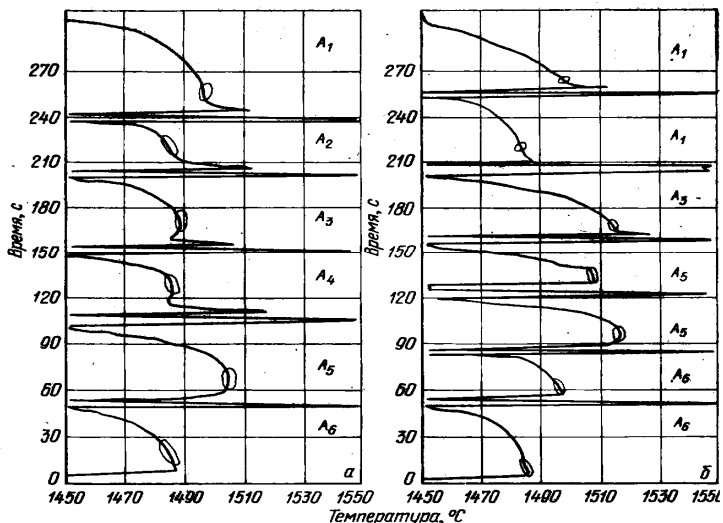


Рис. 4.1. Термограммы, на которых фрагмент появлялся при температуре T_L (слева) и при другой температуре (справа)

В результате выяснилось, что такими признаками в принципе могут быть тип A термограммы, продолжительность τ фрагмента, время ϑ от начала измерения до момента появления фрагмента, величина перегрева ΔT и переохлаждения δT металла относительно температуры T^* , при которой появляется площадка, и даже сама температура T^* (рис. 4.2).

Тип A как признак связан с представлениями о ее форме. Следуя данным работы [26] будем различать шесть типов термограмм, качественное описание которых дано в таблице 4.1. Все шесть типов термограмм приведены на рис.4.1.

Поскольку тип термограммы не может быть предсказан до измерения, будем считать, что это признак случайным образом принимает

одно из своих «значений» A_1, A_2, \dots, A_6 в соответствии с неким распределением вероятностей $P(A_1), P(A_2), \dots, P(A_6)$.

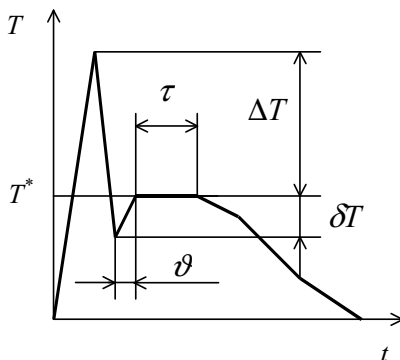


Рис. 4.2. Термограмма кристаллизации жидкой стали и ее признаки

Таблица 4.1.

Типы термограмм кристаллизации жидкой стали

Тип термограммы	Качественное описание термограммы		
	перегрев над T^*	переохлаждение ниже T^*	форма фрагмента
A_1	есть	нет	горизонтальный
A_2	есть	нет	наклонный
A_3	есть	есть	горизонтальный
A_4	есть	есть	наклонный
A_5	нет	нет	горизонтальный
A_6	нет	нет	наклонный

Как показывают наблюдения, для каждого типа термограммы признаки $\tau, \vartheta, \Delta T, \delta T$, и T^* принимают самые разные значения независимо от того, появляется ли фрагмент при температуре ликвидуса или при другой температуре. Поэтому будем считать эти признаки непрерывными случайными величинами, для которых объективно существуют условные распределения $p(\cdot | V_1)$ и $p(\cdot | V_2)$.

Анализ литературных источников позволил получить целый ряд важных априорных сведений. Например, согласно данным работы [26], вероятность того, что фрагмент вызван термическим эффектом, зависит от типа термограммы, причем

$$P(V_1 | A_1) > P(V_1 | A_2) > \dots > P(V_1 | A_6), \quad (4.1.)$$

однако всегда

$$P(V_1 | A) > P(V_2 | A) \quad \text{для всех } A = A_1, A_2, \dots, A_6. \quad (4.2)$$

Условие (4.2) исключает возможность распознавания термического V_1 и псевдотермического V_2 эффектов только на основании измерения признака A .

Установлено также, что вероятность события V_2 не зависит от того, при какой температуре T^* появляется фрагмент, и какое время ϑ прошло от начала измерения до момента появления фрагмента. Следовательно, можно положить, что

$$P(V_2 | T^*) \equiv P(V_2). \quad (4.3)$$

$$P(V_2 | \vartheta) \equiv P(V_2). \quad (4.4)$$

На основании (4.2) - (4.4) распознавание классов V_1 и V_2 по отдельно взятым признакам A , T^* и ϑ невозможно, т.е.

$$P_A(e) = P_{T^*}(e) = P_\vartheta(e) = \min\{P(V_1), P(V_2)\}. \quad (4.5)$$

Опираясь на физические представления о процессах, порождающих термограмму, можно получить ряд дополнительных сведений о рассматриваемых признаках, в частности, о статистической зависимости между некоторыми из них.

Так из работы М.А.Анфиногеновой [1], в которой предложена идеализированная модель процесса кристаллизации металла в пробнице, следует, что с увеличением температуры ликвидуса T_L увеличивается продолжительность τ фрагмента, появляющегося при этой температуре. Там же установлено, что перегрев ΔT оказывает существенное влияние на время ϑ , тогда как на величину τ это влияние значительно меньше.

Для исследования этих зависимостей нами была проведена статистическая обработка реальных термограмм [55]. В результате установлено, что величины τ и T_L действительно связаны статистической зависимостью, которая описывается уравнением регрессии

$$\hat{M}\{T_L | \tau\} = 1508,7 + 0,8155\tau, \quad (4.6)$$

где $\hat{M}\{T_L | \tau\}$ – оценка условного математического ожидания. Хотя эмпирический коэффициент корреляции составил $r = 0,386$, но проверка значимости r по методике, описанной в [34], показала, что с надежностью вывода $P > 0,99$ гипотеза о некоррелированности T_L и τ должна быть отвергнута.

Проведена также оценка безусловной $\hat{D}\{T_L\}$ и условной $\hat{D}\{T_L | \tau\}$ дисперсий величины T_L . Оказалось, что $\hat{D}\{T_L\} = 144,4 \text{ град}^2$, тогда как при изменении τ от 0 до 28 с условная дисперсия $\hat{D}\{T_L | \tau\}$ находилась в пределах

$$25,6 \text{ град}^2 \leq \hat{D}\{T_L | \tau\} \leq 93,9 \text{ град}^2. \quad (4.7)$$

Оценка расхождений $\hat{D}\{T_L\}$ и $\hat{D}\{T_L | \tau\}$ по методу Бартлетта [37] с надежностью вывода $P > 0,99$ подтвердила гипотезу о том, что

$$\hat{D}\{T_L\} > \hat{D}\{T_L | \tau\} \quad \forall \tau,$$

или, в эквивалентной форме записи,

$$\hat{D}\{T^* | V_1\} > \hat{D}\{T^* | \tau, V_1\} \quad \forall \tau. \quad (4.8)$$

Наличие корреляционной связи между T^* и τ в тех случаях, когда фрагмент был вызван псевдотермическим эффектом, не обнаружено. И хотя следующее предположение требует более веских аргументов, мы все же допустим, что для класса V_2 величины T^* и τ статистически независимы в общем смысле этого понятия, т.е.

$$p(T^* | \tau, V_2) \equiv p(T^* | V_2). \quad (4.9)$$

Наблюдения показали, что при больших значениях τ продолжительности фрагмента чаще всего он вызван термическим эффектом

(класс V_1), а при малых – псевдотермическим эффектом (класс V_2). Поэтому можно предположить, что

$$P(V_1 | \tau) > P(V_1 | \tau) \text{ при больших значениях } \tau,$$

$$P(V_1 | \tau) < P(V_1 | \tau) \text{ при малых значениях } \tau.$$

Отсюда следует, что признак τ уже сам себе полезен. Однако ввиду того, что при одних и тех же τ может наблюдаться и класс V_1 , и класс V_2 , т.е. τ имеет пересекающиеся распределения в классах, этот признак не позволяет провести безошибочное распознавание.

Следовательно

$$0 < P_\tau(e) < \min\{P(V_1), P(V_2)\}. \quad (4.10)$$

Поскольку в соответствии с [1] время ϑ зависит *только* от величины ΔT , то вполне правомерна гипотеза о том, что случайная величина ϑ в обоих классах статистически не зависит от T^* и τ , т.е.

$$p(\vartheta | \tau, T^*, V_k) \equiv P(\vartheta | V_k), \quad k = 1, 2. \quad (4.11)$$

Из работы [1] непосредственно также следует, что при любой температуре ликвидуса T_L продолжительность τ фрагмента, появляющегося при этой температуре, не зависит от величины ΔT . Считая, что это в равной степени относится и для фрагментов, появляющихся при другой температуре, введем следующую гипотезу

$$p(\tau | T^*, \Delta T, V_k) \equiv P(\tau | T^*, V_k), \quad k = 1, 2. \quad (4.12)$$

Естественно также предположить, что сама величина ΔT , которая определяется лишь начальной температурой расплава, не зависит от того, при какой температуре T^* появится фрагмент, а значит

$$p(\Delta T | T^*, V_k) \equiv P(\Delta T | V_k), \quad k = 1, 2. \quad (4.13)$$

В то же время, как следует из работы [1] от величины перегрева ΔT зависит и время снятия перегрева ϑ , т.е.

$$p(\vartheta | \Delta T, V_k) \neq p(\vartheta | V_k), \quad k = 1, 2. \quad (4.14)$$

Поскольку в соответствии с (4.10) признак τ сам по себе уже является полезным, а условные распределения $p(\tau | V_k)$, $k = 1, 2$ – непрерывные функции, заданные на односвязных множествах (по крайней мере, исходя из физических представлений, нет оснований допустить противное), то в соответствии с леммой 2.2 найдется такое значение τ , при котором

$$\frac{p(\tau | V_1)}{p(\tau | V_2)} = \frac{P(V_2)}{P(V_1)}. \quad (4.15)$$

Полагая, что уравнение (4.15) имеет единственное решение $\tau = \tau_0 = \text{const}$, приходим к достаточно простому пороговому правилу:

$$\left. \begin{array}{l} \text{решаем в пользу } V_1, \text{ если } \tau \geq \tau_0, \\ \text{решаем в пользу } V_2, \text{ если } \tau < \tau_0 \end{array} \right\}. \quad (4.16)$$

В отличие от признака τ , признаки T^* и ϑ , как следует из (4.5) сами по себе являются бесполезными. Более того, как следует из (4.3), (4.4), T^* и ϑ имеют одинаковые распределения в классах

$$p(T^* | V_1) \equiv p(T^* | V_2), \quad (4.17)$$

$$p(\vartheta | V_1) \equiv p(\vartheta | V_2), \quad (4.18)$$

а значит они не только бесполезны, но неинформативны сами по себе.

Тем не менее, опираясь на приведенные выше результаты наших исследований, можно показать, что использование признака T^* совместно с признаком τ оказывается полезным.

Действительно, ввиду наличия корреляционной связи (4.6) между T^* и τ в классе V_1 имеем

$$p(T^* | \tau, V_1) \neq p(T^* | V_1). \quad (4.19)$$

Кроме того, из условия (4.8) следует

$$D\{T^* | \tau, V_1\} \neq D\{T^* | V_1\} \quad \forall \tau. \quad (4.20)$$

Поскольку

$$D\{T^* | V_1\} \triangleq \int_0^{+\infty} [T^* - \int_0^{+\infty} T^* p(T^* | V_1) dT^*]^2 p(T^* | V_1) dT^*, \quad (4.21)$$

а

$$D\{T^* | \tau, V_1\} \triangleq \int_0^{+\infty} [T^* - \int_0^{+\infty} T^* p(T^* | \tau, V_1) dT^*]^2 p(T^* | \tau, V_1) dT^*, \quad (4.22),$$

то из условия (4.20) с учетом (4.21) и (4.22) немедленно следует, что соотношение (4.19) справедливо для любого τ , т.е.

$$p(T^* | \tau, V_1) \neq p(T^* | V_1) \quad \forall \tau. \quad (4.23)$$

В самом деле, если допустить противное, т.е. $\exists \tau = \tau^*$, что

$$p(T^* | \tau^*, V_1) = p(T^* | V_1),$$

то на основании (4.21), (4.22) получим

$$D\{T^* | \tau^*, V_1\} \neq D\{T^* | V_1\},$$

что противоречит (4.20).

В силу (4.23) с учетом (4.9) и (4.17) окончательно имеем

$$p(T^* | \tau, V_1) \neq p(T^* | \tau, V_2) \quad \forall \tau. \quad (4.24)$$

Будем попрежнему считать, что условные распределения $p(\tau | V_k)$, $k = 1, 2$ – непрерывные функции, а множества $\{\tau\}$ возможных значений τ – односвязны. По-видимому, такого рода допущения всегда можно сделать при решении задач распознавания сигналов, порожденных физическими процессами (в данном случае – теплофизическим), если только нет веских оснований опровергнуть такие допущения.

Условия (4.23) и (4.24) с учетом указанных выше допущений о распределениях $p(\tau | V_k)$, $k = 1, 2$ и множествах $\{\tau\}$ позволяют на основании теоремы 2.4 утверждать, что признак T^* является полезным в совокупности (τ, T^*) , т.е.

$$P_{T, \tau^*}(e) < P_\tau(e). \quad (4.24)$$

При такой комбинации признаков решающее правило имеет вид

$$\left. \begin{array}{l} \text{решаем в пользу } V_1, \text{ если } \tau \geq \tau_0(T^*), \\ \text{решаем в пользу } V_2, \text{ если } \tau < \tau_0(T^*) \end{array} \right\}, \quad (4.25)$$

причем оптимальная величина $\tau_0(T^*)$, при которой будет достигнут минимум средней вероятности ошибки, является решением уравнения [55]

$$\frac{p(\tau | T^*, V_1)}{p(\tau | V_2)} = \frac{P(V_2)}{P(V_1)}. \quad (4.26)$$

Если к тому же учесть, что согласно (4.17) признак T^* имеет одинаковые распределения в классах и по предположению (4.9) статистически не зависит от τ в классе V_2 , то по теореме 1.2 совокупность (τ, T^*) еще не позволяет провести безошибочное распознавание. Тогда с учетом (4.10) и (4.24) можем записать

$$0 < P_{T^*, \tau}(e) < P_\tau(e) < \min\{P(V_1), P(V_2)\}. \quad (4.27)$$

А это значит, что принципиально может идти речь о полезности какого-либо из оставшихся признаков совместно с совокупностью (τ, T^*) . Исследуем такую возможность с учетом имеющихся априорных сведений.

Поскольку признак ϑ по предположению (4.11) статистически независим от совокупности (τ, T^*) в обоих классах и, к тому же, согласно (4.18) имеет одинаковые распределения в классах, можно утверждать, что добавление такого признака к совокупности (τ, T^*) не может уменьшить вероятность ошибки, т.е.

$$P_{T^*, \tau, \vartheta}(e) = P_{T^*, \tau}(e).$$

В отличие от признаков ϑ и T^* , тип термограммы A имеет различные условные распределения в классах, т.е.

$$P(A | V_1) \neq P(A | V_2). \quad (4.28)$$

Справедливость соотношения (4.28) легко показать с использованием формулы Байеса методом от противного, опираясь на условие (4.1).

Введем в рассмотрение два укрупненных типа термограммы, а именно, тип термограммы с участком перегрева

$$A' = A_1 \cup A_2 \cup A_3 \cup A_4 \quad (4.29)$$

и тип термограммы без участка перегрева

$$A'' = A_5 \cup A_6, \quad (4.30)$$

игнорируя наличие и отсутствие других более тонких различий типов, описанных в таблице 4.1.

При этом будем считать, что

$$A = \begin{cases} A', & \text{если } \Delta T \geq \Delta T_0, \\ A'', & \text{если } \Delta T < \Delta T_0, \end{cases} \quad (4.31)$$

где $\Delta T_0 = \text{const}$ – некоторый малый порог.

Обратимся к соотношению (4.12), которое с учетом (4.13) после простых преобразований сводится к виду

$$P(\Delta T | \tau, T^*, V_k) \equiv P(\Delta T | V_k), \quad k = 1, 2. \quad (4.32)$$

Интегрируя левую и правую части (4.32) по ΔT в пределах от ΔT_0 до $+\infty$ получим

$$\int_{\Delta T_0}^{+\infty} p(\Delta T | \tau, T^*, V_k) d\Delta T \equiv \int_{\Delta T_0}^{+\infty} p(\Delta T | V_k) d\Delta T. \quad (4.33)$$

Из (4.33) с учетом (4.31) немедленно следует

$$P(A' | \tau, T^*, V_k) = P(A' | V_k), \quad k = 1, 2. \quad (4.34)$$

Аналогичным образом, интегрируя левую и правую часть соотношения (4.32), но уже в пределах от 0 до ΔT_0 , можно показать, что

$$p(A'' | \tau, T^*, V_k) = p(A'' | V_k), \quad k = 1, 2. \quad (4.35)$$

Если теперь рассматривать тип термограммы A как дискретный признак с двумя градациями (4.29) и (4.30) и принять во внимание соотношения (4.34) и (4.35), то можно заключить, что этот признак статистически не зависит от совокупности (τ, T^*) в обоих классах, т.е.

$$p(A | \tau, T^*, V_k) \equiv p(A | V_k), \quad k = 1, 2. \quad (4.36)$$

На основании соотношений (4.28) и (4.36) по теореме 2.3 заключаем, что тип термограммы A – полезный признак в совокупности с признаками τ и T^* , а значит

$$P_{A, T^*, \tau}(e) < P_{T^*, \tau}(e) < P_\tau(e) < \min\{P(V_1), P(V_2)\}. \quad (4.37)$$

В данном случае решающее правило имеет вид

$$\left. \begin{array}{l} \text{решаем в пользу } V_1, \text{ если } \tau \geq \tau_0(A, T^*), \\ \text{решаем в пользу } V_2, \text{ если } \tau < \tau_0(A, T^*) \end{array} \right\}, \quad (4.38)$$

причем оптимальная величина $\tau_0(A, T^*)$, при которой будет достигнут минимум средней вероятности ошибки, является решением уравнения [55]

$$\frac{p(\tau | T^*, V_1)}{p(\tau | V_2)} = \frac{P(V_2 | A)}{P(V_1 | A)}. \quad (4.39)$$

Построенные выше решающие правила обеспечивают возможность распознавания термического и псевдотермического эффектов, если на термограмме появляется только один фрагмент. В то же время на практике наблюдаются и такие случаи, когда на одной термограмме образуются $G > 1$ фрагментов, но только один из них может быть вызван термическим эффектом.

Для построения решающего правила в таком более общем случае введем в рассмотрение следующую систему $G + 1$ гипотез:

$$W_0 = \bigcap_{j=1}^G V_{2j}, \quad W_m = \bigcap_{j=1}^G B_{mj}, \quad m = 1, \dots, G. \quad (4.40)$$

Здесь

$$B_{mj} = \begin{cases} V_{1j}, & \text{если } j = m, \\ V_{2j}, & \text{если } j \neq m, \end{cases} \quad (4.41)$$

где V_{1j} и V_{2j} – элементарные случайные события, состоящие в том, что появление j -го фрагмента вызвано соответственно термическим и псевдотермическим эффектами.

Для того чтобы не усложнять изложение излишними деталями, ограничимся здесь построением схемы принятия решений только по признакам τ_1, \dots, τ_G , представляющим собой продолжительности по времени соответствующих фрагментов.

Построение решающего правила основано на следующих предпосылках, вытекающих из особенностей процесса порождения термограмм [53]:

1. Продолжительности фрагментов считаются независимыми случайными величинами как при гипотезе W_0 , так и при любой из гипотез W_m , $m = 1, \dots, G$, т.е.

$$p(\tau_1, \dots, \tau_G | W_0) \equiv \prod_{j=1}^G p(\tau_j | W_0), \quad (4.42)$$

$$p(\tau_1, \dots, \tau_G | W_m) \equiv \prod_{j=1}^G p(\tau_j | W_m), \quad m = 1, \dots, G. \quad (4.43)$$

2. Условные плотности распределения продолжительностей фрагментов, вызванных термическим эффектом, полагаются одинаковыми при всех W_m , т.е.

$$p(\tau_1 | W_1) = p(\tau_2 | W_2) = \dots = p(\tau_G | W_G) \quad (4.44)$$

для любых $\tau_1 = \tau_2 = \dots = \tau_G$.

3. Условные плотности распределения продолжительностей фрагментов, вызванных псевдотермическим эффектом, полагаются одинаковыми

$$p(\tau_j | W_0) = p(\tau_j | W_m) = \dots = p(\tau_j | W_l) \quad (4.45)$$

для всех $j, m, l = 1, 2, \dots, G$, $j \neq m$, $m \neq l$ и

$$p(\tau_m | W_0) = p(\tau_m | W_l) = \dots = p(\tau_l | W_m) \quad (4.46)$$

при любых $\tau_m = \tau_l$, $m, l = 1, 2, \dots, G$, $m \neq l$.

Здесь W_l – гипотеза, состоящая в том, что l -й фрагмент вызван термическим эффектом, а остальные $G-1$ фрагментов – псевдотермическим эффектом.

4. Априорные вероятности всех гипотез W_m , $m = 1, \dots, G$ считаются одинаковыми, причем

$$P(W_m) = \frac{P(V_1)}{G}. \quad (4.47)$$

Принимая во внимание (4.42), (4.43) запишем условие равенства апостериорных вероятностей $P(W_m | \tau_1, \tau_2, \dots, \tau_G)$ и $P(W_0 | \tau_1, \tau_2, \dots, \tau_G)$ в виде

$$\frac{p(\tau_1 | W_m)p(\tau_2 | W_m) \cdots p(\tau_G | W_m)}{p(\tau_1 | W_0)p(\tau_2 | W_0) \cdots p(\tau_G | W_0)} = \frac{P(W_0)}{P(W_m)}. \quad (4.48)$$

В силу соотношений (4.45) и (4.47) из (4.48) получим

$$\frac{p(\tau_m | W_m)}{p(\tau_m | W_0)} = \frac{G[1 - P(V_1)]}{P(V_1)}. \quad (4.49)$$

Уравнение (4.49) определяет границу между областями решений в пользу гипотезы W_0 и любой из гипотез W_m , $m = 1, \dots, G$.

Запишем теперь условие равенство апостериорных вероятностей $P(W_m | \tau_1, \tau_2, \dots, \tau_G)$ и $P(W_l | \tau_1, \tau_2, \dots, \tau_G)$. С учетом (4.42), (4.43) и (4.47) это условие можно записать так:

$$\frac{p(\tau_1 | W_m)p(\tau_2 | W_m) \cdots p(\tau_G | W_m)}{p(\tau_1 | W_l)p(\tau_2 | W_l) \cdots p(\tau_G | W_l)} = 1. \quad (4.50)$$

На основании (4.45) из (4.50) имеем

$$\frac{p(\tau_m | W_m)}{p(\tau_m | W_l)} = \frac{p(\tau_l | W_l)}{p(\tau_l | W_m)}, \quad l, m = 1, \dots, G, \quad l \neq m. \quad (4.51)$$

В силу соотношений (4.44) и (4.46) уравнение (4.51), определяющее границу между областями решений относительно гипотез W_l и W_m , переходит в тождество при

$$\tau_m = \tau_l. \quad (4.52)$$

Если исходить из того, что отношения условных распределений продолжительностей фрагментов, фигурирующие в уравнениях (4.49) и (4.51), возрастают по мере увеличения τ_m и τ_l , то решение (4.52) будет единственным решением уравнения (4.51). Единственное решение будет иметь в этом случае и уравнение (4.49).

Тогда принятие решения по максимуму апостериорных вероятностей

$$\max \left\{ \max_m P(W_m | \tau_1, \dots, \tau_G), P(W_0 | \tau_1, \dots, \tau_G) \right\}$$

снова приводит к достаточно простому решающему правилу

$$\text{решаем в пользу } W_m, m = 1, \dots, G, \text{ если } \tau_m = \max_{\tau_j \geq \tau_0(G)} \tau_j,$$

$$\text{решаем в пользу } W_0, \text{ если } \tau_j < \tau_0(G) \quad \forall j.$$

Здесь $\tau_0(G)$ – порог по продолжительности, удовлетворяющий уравнению (4.49), который увеличивается с увеличением числа G фрагментов, одновременно зарегистрированных на термограмме.

Предложенный метод распознавания термических и псевдотермических эффектов и цифровое устройство для его реализации запатентовано в ряде индустриально развитых стран, в том числе в США [76].

4.2. Анализ диагностических признаков электрокардиограммы

Электрокардиограмма (ЭКГ) представляет собой запись сигнала, несущего информацию об изменениях во времени суммарного электрического потенциала, который возникает в сердечной мышце за счет движением ионов через мышечную мембрану [60]. Диагностика заболеваний сердца по ЭКГ основана на морфологическом анализе амплитудно-временных параметров зубца P , комплекса QRS , сегментов $RS - T$, $Q - T$ и зубца T (рис. 4.3).

Например, основным электрокардиологическим признаком ишемической болезни сердца (ИБС) является депрессия $-\Delta_{S-T}$ (смещение вниз) или элевация $+\Delta_{S-T}$ (подъем) сегмента $S - T$ относительно изоэлектрической линии на величину, превышающую некоторый порог Δ_{S-T}^0 [29, с. 207].

В то же время известно, что этот традиционный диагностический признаки не обеспечивают требуемую достоверность результатов диагностики. Согласно [69], ЭКГ покоя, оцениваемая по общепринятым критериям, остается нормальной приблизительно у 50 % больных с хронической формой ИБС.

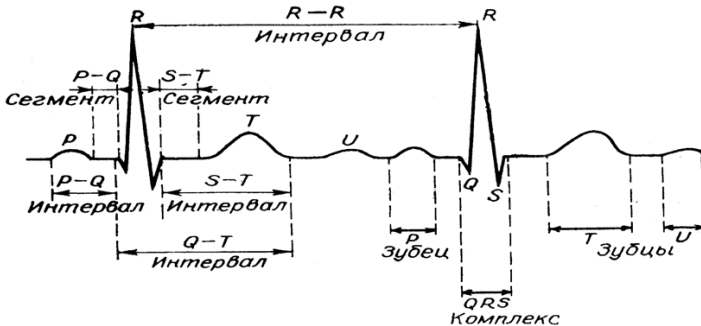


Рис. 4.3. Фрагменты нормальной ЭКГ в одном отведении

Для повышения чувствительности и специфичности ЭКГ диагностики нами предложен новый подход [43, 45, 71, 73]. Основная идея метода состоит в том, что в каждой точке дискретно заданной ЭКГ $y(t)$, $t = 1, 2, \dots$ численными методами оценивается первая производная $\dot{y}(t)$ и вся последующая обработка сигнала осуществляется на фазовой плоскости в координатах $y(t) - \dot{y}(t)$ (рис. 4.4).

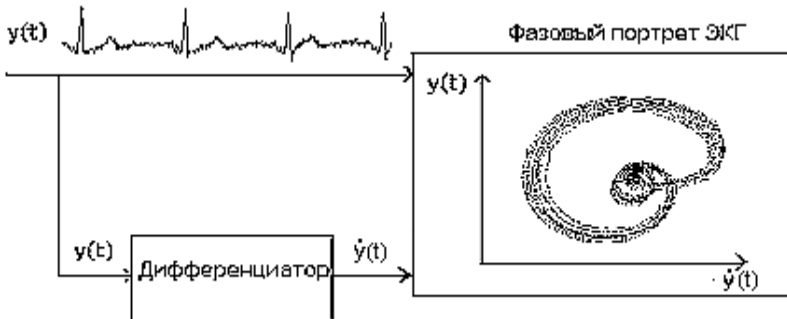


Рис. 4.4. Основная идея метода

Рассмотрим кратко математические методы, положенные в основу данной технологии [40, 46, 77].

Предположим, что наблюдаемая ЭКГ $y(t) = \Phi[\hat{y}(t), \zeta(t)]$ представляет собой результат искажения некоторой периодической функции $\hat{y}(t)$ случайным возмущением $\zeta(t)$, где $\Phi(\cdot)$ – неизвестная функция.

Пусть $y_0(t)$ – часть ненаблюдаемой функции $\hat{y}(t)$ на одном из ее периодов T_0 . Задача состоит в том, чтобы на основе обработки наблюдаемой ЭКГ оценить «идеальный» кардиоцикл $y_0(t)$ [75].

Функция $y_0(t)$ представляет собой последовательность K_F фрагментов $y_0^{(i)}(t), i = 1, \dots, K_F$ – зубцов P, Q, R, S, T и элементов изоэлектрической линии, которые характеризуют стадии возбуждения отдельных участков сердечной мышцы:

$$y_0(t) = \begin{cases} y_0^{(1)}(t), & \text{если } 0 \leq t < t_0^{(1)}, \\ \dots & \\ y_0^{(K_F)}(t), & \text{если } t_0^{(K_F-1)} \leq t < t_0^{(K_F)} = T_0 \end{cases}. \quad (4.53)$$

Будем считать, что каждый i -й фрагмент m -го цикла наблюдаемой ЭКГ представляет собой результат операторного преобразования над соответствующим фрагментом «идеального» цикла $y_0(t)$:

$$y_m^{(i)}(t) = a_m y_0^{(i)}\left(\frac{t - \tau_m^{(i)}}{b_m^{(i)}}\right), \quad m = 1, 2, \dots, \quad i = 1, \dots, K_F, \quad (4.54)$$

где $a_m, b_m^{(i)}$ – случайные искажения по амплитуде и времени, а $\tau_m^{(i)}$ – сдвиг по времени. Для обеспечения непрерывности порождаемого сигнала предполагается, что $y_0(0) = 0$. Последнее требование всегда можно обеспечить, выполнив предварительно операцию центрирования $\tilde{y}_0(t) = y_0(t) - y_0(0)$.

Пусть в пределах каждого m -го цикла параметр a_m принимает фиксированное значение:

$$a_m = 1 + \xi_m, \quad m = 1, 2, \dots, \quad (4.55)$$

где ξ_m - последовательность реализаций независимых случайных величин, которые с нулевым математическим ожиданием $M\{\xi\}=0$ распределены на интервале $[-\varepsilon_0, \varepsilon_0]$, ограниченном фиксированным числом $\varepsilon_0 \in [0, 1)$.

Предполагается также, что параметр $b_m^{(i)}$ принимает фиксированное значение в процессе порождения каждого отдельного i -го фрагмента m -го цикла, т.е.

$$b_m^{(i)} = 1 + \delta_m^{(i)}, \quad m = 1, 2, \dots, \quad i = 1, \dots, K_F, \quad (4.56)$$

где $\delta_m^{(i)}$ - последовательность реализаций независимых случайных величин, которые с нулевым математическим ожиданием $M\{\delta\}=0$ распределены на интервалах $[-\Delta_0^{(i)}, \Delta_0^{(i)}]$, ограниченными фиксированными числами $\Delta_0^{(i)} \in [0, 1)$.

При таких предположениях продолжительность i -го фрагмента m -го цикла сигнала $y(t)$ связана с продолжительностью $\Delta T_0^{(i)} = t_0^{(i)} - t_0^{(i-1)}$ соответствующего фрагмента «идеального» цикла соотношением

$$\Delta T_m^{(i)} = (t_0^{(i)} - t_0^{(i-1)})(1 + \delta_m^{(i)}) .$$

Следовательно, общая продолжительность m -го цикла порождаемого сигнала $y(t)$ определяется выражением

$$T_m = T_0 + \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)}) \delta_m^{(i)},$$

началу m -го цикла соответствует момент времени

$$t_m = (m-1)T_0 + \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)}) \delta_j^{(i)},$$

а началу i -го фрагмента m -го цикла – момент времени

$$t_m^{(i-1)} = (m-1)T_0 + \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)}) \delta_j^{(i)} + \sum_{l=1}^{i-1} (t_0^{(l)} - t_0^{(l-1)})(1 + \delta_m^{(l)}). \quad (4.57)$$

Применим к i -му фрагменту «идеального» цикла $y_0(t)$ операторное преобразование (4.54), положив параметр сдвига

$$\tau_m^{(i)} = t_m^{(i-1)} - (1 + \delta_m^{(i)})t_0^{(i-1)}.$$

Тогда из (4.54) с учетом соотношений (4.55)-(4.57) следует, что процесс порождения i -го фрагмента на m -м цикле можно представить в виде

$$y_m^{(i)}(t) = (1 + \xi_m) y_0^{(i)}(\theta), \quad m = 1, 2, \dots, \quad i = 1, \dots, K_F, \quad (4.58)$$

где нелинейная величина θ , имеющая размерность времени, определяется соотношением

$$\theta = \{t - (m-1)T_0 + (1 + \delta_m^{(i)})t_0^{(i-1)} - \\ - \sum_{j=1}^{m-1} \sum_{i=1}^{K_F} (t_0^{(i)} - t_0^{(i-1)})\delta_j^{(i)} - \sum_{l=1}^{i-1} (t_0^{(l)} - t_0^{(l-1)})(1 + \delta_m^{(l)})\} (1 + \delta_m^{(i)})^{-1}, \quad (4.59)$$

Примечательно, что модель (4.58), (4.59) является стохастическим обобщением моделей почти периодической и периодической функций и последовательно сводится к ним при условии, что $\delta_j^{(i)} = 0 \quad \forall j = 1, \dots, m, i = 1, \dots, K_F$ и $\varepsilon_0 = 0$ [42].

Модель (4.58), (4.59) позволила обосновать эффективность метода восстановления $y_0(t)$ по последовательности искаженных реализаций [41], суть которого состоит в следующем.

Пусть $y_1(t), \dots, y_M(t)$ – циклы наблюдаемой ЭКГ, каждый из которых представлен дискретными значениями $y_m[k] \equiv y_m(t_k)$, $m = 1, \dots, M$, зафиксированными в равноотстоящие моменты времени $t_k \equiv k\Delta$, $k = 1, \dots, K_m$, где Δ – шаг квантования по времени.

Используя алгоритм численного дифференцирования, оценим производные $\dot{y}_m[k]$ в каждой k -й точке m -го цикла и перейдем от исходной последовательностей скалярных величин $y_m[1], \dots, y_m[K_m]$ к последовательностям векторов

$$\begin{aligned} Q_1 &= \{y_m^*[k], \dot{y}_m^*[k], \tau_m[k], k = 1, \dots, K_1\}, \dots, \\ \dots, Q_M &= \{y_m^*[k], \dot{y}_m^*[k], \tau_m[k], k = K_{M-1} + 1, \dots, K_M\}, \end{aligned} \quad (4.60)$$

компонентами которых служат нормированные величины

$$y_m^*[k] = \frac{y_m[k] - \min_{1 \leq k \leq K_m} y_m[k]}{\max_{1 \leq k \leq K_m} y_m[k] - \min_{1 \leq k \leq K_m} y_m[k]} \in [0, 1], \quad k = 1, \dots, K_m,$$

$$\dot{y}_m^*[k] = \frac{\dot{y}_m[k] - \min_{1 \leq k \leq K_m} \dot{y}_m[k]}{\max_{1 \leq k \leq K_m} \dot{y}_m[k] - \min_{1 \leq k \leq K_m} \dot{y}_m[k]} \in [0, 1], \quad k = 1, \dots, K_m,$$

а также нормированная величина

$$\tau_m[k] = \frac{k-1}{K_m-1} \in [0, 1], \quad k = 1, \dots, K_m,$$

имеющая смысл относительного времени появления k -й точки на m -й фазовой траектории.

Вычислим $M \times M$ матрицу $D = [R_H(Q_i, Q_j)]$ хаусдорфовых расстояний [35]

$$R_H(Q_i, Q_j) = \max \left\{ \max_{z_j \in Q_j} \min_{z_i \in Q_i} \rho(z_i, z_j), \max_{z_i \in Q_i} \min_{z_j \in Q_j} \rho(z_i, z_j) \right\}, \quad (4.61)$$

между парами Q_i и Q_j последовательностей (4.60), где

$$z_m = y_m^*[k], \dot{y}_m^*[k], \tau_i[k], \quad m = i, j, \quad k = 1, \dots, K_M,$$

а

$$\rho(z_i, z_j) = \|z_i - z_j\|$$

– евклидово расстояние между точками $z_i \in Q_i$ и $z_j \in Q_j$.

Последовательность

$$Q_0 = \{z_0[k] = (y_0^*[k], \dot{y}_0^*[k], \tau_0[k])\}, k = 1, \dots, K_0$$

нормированных векторов такая, что

$$Q_0 = \arg \min_{1 \leq j \leq M} \sum_{i=1}^M R_H(Q_i, Q_j) \quad (4.62)$$

определит одну из фазовых траекторий, которую назовем опорным циклом.

Будем использовать Q_0 в качестве первого приближения ненаблюдаемой траектории «идеального» кардиоцикла, а для уточнения оценки $\hat{y}_0(\cdot)$ усредним точки наблюдаемых траекторий в окрестности соответствующих точек опорного цикла Q_0 .

С этой целью вычислим последовательность векторов $\hat{z}[k], k = 1, \dots, K_0$ по формуле

$$\hat{z}[k] = \frac{z_0[k] + \sum_{m=1}^M z_{mk}^{\min}}{M+1}, \quad (4.63)$$

где

$$z_{mk}^{\min} = \arg \min_l \rho_\tau(z_0[k], z_m[l]), \quad l = 1, \dots, K_m \quad (4.64)$$

– вектор, принадлежащий m -той траектории (не являющейся опорной), который находится на минимальном евклидовом расстоянии

$$\rho_\tau(z_0[\cdot], z_m[\cdot]) = \sqrt{(y_0^*[\cdot] - y_m^*[\cdot])^2 + (\dot{y}_0^*[\cdot] - \dot{y}_m^*[\cdot])^2 + (\tau_0[\cdot] - \tau_m[\cdot])^2} \quad (4.65)$$

от вектора $z_0[k] \in Q_0$, а K_0 – число точек на опорной траектории.

Последовательность векторов $\hat{z}[k] = (\hat{y}^*[k], \hat{y}^*[k]), k = 1, \dots, K_0$, вычисленная согласно (4.63), определяет усредненную фазовую траекторию, а соответствующая последовательность значений $\hat{y}[1], \dots, \hat{y}[K_0]$ дает оценку $y_0(t)$ во временной области (рис. 4.5).

Для формирования последовательности (4.60) необходимо предварительно разбить наблюдаемый сигнал $y(t)$ на отдельные сердечные циклы $y_1(t), \dots, y_M(t)$. Коль скоро мы отказались от обработки сигнала во временной области, привлекательно было бы и эту задачу

решить в фазовом пространстве. Покажем, что такая возможность действительно существует.

Вычислим квадраты расстояний

$$r_0^2[k] = (y^*[k] - y_0^*)^2 + (\dot{y}^*[k] - \dot{y}_0^*)^2, \quad k = 1, \dots, K \quad (4.66)$$

между нормированными фазовыми координатами каждой k -й точки и некоторой фиксированной точкой $k_0 \in [1, K]$.

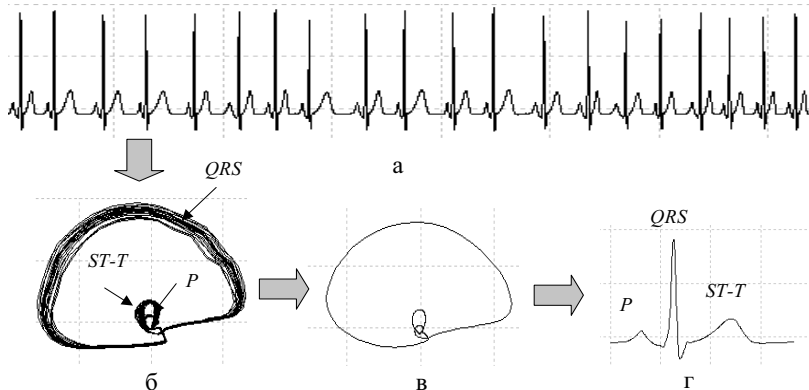


Рис. 4.5. Последовательность этапов обработки ЭКГ: а — исходная ЭКГ ; б — фазовая траектория ЭКГ; в — усредненная фазовая траектория ; г — эталонный цикл ЭКГ во временной области

Поскольку комплекс QRS отличают от других фрагментов ЭКГ высокие значения амплитуды y и производной \dot{y} , то в моменты появления QRS - комплексов на профиле расстояний $r_0^2[k]$ будут появляться выраженные экстремумы даже в тех случаях, когда наблюдаемый сигнал имеет сложную форму, в частности, высокий зубец T .

Поэтому, если выделить подмножества Ω_m точек $k \in [1, K]$, таких, что

$$\rho_0^2[k] \geq \rho_{QRS} \quad \forall k \in \Omega_m,$$

где ρ_{QRS} — некоторый порог, и определить номера точек $k_m, m = 1, \dots, M$, удовлетворяющие условию

$$k_m = \arg \max_{k \in \Omega_m} \rho_0^2[k], \quad (4.67)$$

то полученная последовательность номеров $k_m, m = 1, \dots, M$ как раз и определит границы отдельных сердечных циклов ЭКГ.

«Оптимальные» значения ρ_{QRS} и k_0 , обеспечивающие необходимый компромисс между ошибками пропуска цели и ложной тревоги, можно определить экспериментальным путем. Например, значения k_0 для конкретной обрабатываемой реализации можно найти методом полного перебора, используя критерий

$$k_0 = \arg \max_{1 \leq k_j \leq K} \left\{ \frac{\max_{1 \leq k \leq K} r[k, k_j]}{\frac{1}{K} \sum_{k=1}^K r[k, k_j]} \right\}. \quad (4.68)$$

В этом случае фрагменты функции $r_0^2[k]$, соответствующие QRS -комплексам, становятся более выраженными по сравнению с остальными участками области определения.

Процедуру выделения QRS -комплексов можно сделать еще более эффективной, если вместо (4.66) использовать функцию

$$\tilde{r}_0^2[k] = (\tilde{y}^*[k] - \tilde{y}_0^*)^2 + (\tilde{y}'^*[k] - \tilde{y}'_0^*)^2, \quad k = W, \dots, K-1-W, \quad (4.69)$$

в которой фигурируют модифицированные значения сигнала

$$\tilde{y}[k] = y[k] - \frac{1}{2W} \sum_{i=-W}^W y[k-i], \quad k = W, \dots, K-1-W,$$

где W – окно сглаживания. Поскольку продолжительность QRS -комплекса существенно меньше продолжительности зубцов P и T , такой прием позволяет еще больше усилить «контраст» искомых фрагментов на профиле квадратов расстояний (4.69).

На реальных ЭКГ могут появляться экстрасистолы и артефакты. Понятно, что такие «нетипичные» циклы не должны использоваться в процедуре усреднения (4.63)-(4.65).

Рассмотрим один из возможных приемов селекции «нетипичных» циклов, предполагая лишь, что число таких циклов $M_N \ll M$ (в противном случае определение «нетипичный» цикл теряет смысл). При

таком предположении опорный цикл Q_0 , найденный согласно (4.62), правомерно считать «типичным», что позволяет обнаружить «нетипичные» циклы по расстоянию до Q_0 .

С этой целью упорядочим по возрастанию элементы строки матрицы $D = [R_H(Q_i, Q_j)]$, соответствующей опорному циклу Q_0 . Полученная последовательность значений

$$\mathfrak{X} = R_H(Q_0, Q_\mu), \mu \in [1, M]$$

дает наглядное представление о вариабельности формы траекторий наблюдаемых сигналов по отношению к Q_0 [41].

Если на сигнале появляются «нетипичные» циклы, форма которых существенно отличается от опорного, то последовательность \mathfrak{X} содержит один или несколько выраженных скачков. Определив положение первого из таких скачков, можно найти пороговое значение \mathfrak{X}_0 и тем самым выделить подмножество траекторий, подлежащих усреднению:

$$\Omega_0 = \{Q_\mu : R_H(Q_0, Q_\mu) < \mathfrak{X}_0\}.$$

Для расширения исходного пространства признаков предложено использовать по четыре дополнительных параметра, характеризующих форму ЭКГ в фазовом пространстве в каждом из 12 отведений, а именно

- параметры $\sigma[i], i = 1, \dots, 12$ рассеивания точек фазовых траекторий;
- углы $\alpha[i], i = 1, \dots, 12$ ориентации усредненной фазовой траектории;
- параметры $\beta_P[i], i = 1, \dots, 12$ симметрии относительно оси $\dot{y} = 0$ фрагментов фазовых траекторий, соответствующих зубцам P ;
- параметры $\beta_T[i], i = 1, \dots, 12$ симметрии относительно оси $\dot{y} = 0$ фрагментов фазовых траекторий, соответствующих зубцам T (рис.4.5).

Приемлемую оценку параметров σ_j ($j = 1, \dots, 12$) можно получить по элементам матрицы $D = [R_H(Q_i, Q_j)]$ хаусдорфовых расстояний (4.61), используя формулу

$$\sigma = \frac{1}{M-1} \sum_m^{M-1} R_H(Q_0, Q_m).$$

Угол ориентации α определяется по направлению прямой, проходящей через точку «самопересечения» внешнего витка траектории, соответствующему QRS комплексу, и точку z_c с фазовыми координатами

$$y_c = \frac{1}{K_0} \sum_{k=1}^{K_0} \hat{y}^*[k]; \quad \dot{y}_c = \frac{1}{K_0} \sum_{k=1}^{K_0} \dot{\hat{y}}^*[k],$$

соответствующей центру «тяжести» усредненной фазовой траектории.

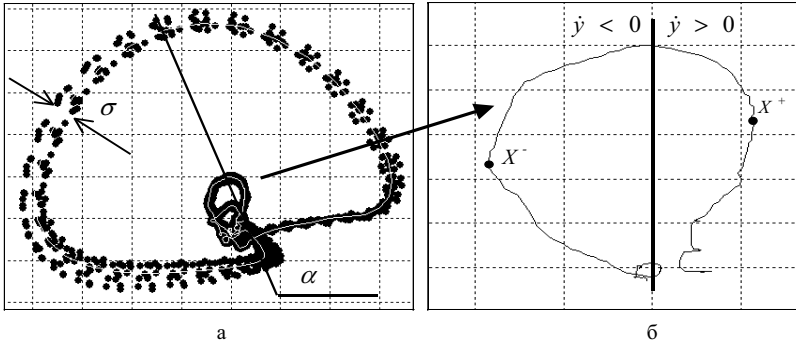


Рис. 4.5. Дополнительные признаки ЭКГ в фазовом пространстве: а – точки фазовых траекторий, б – фрагмент усредненной фазовой траектории, соответствующий зубцу T .

Параметры β_p и β_T определяются отношением абсолютных значений $y_{X^+}^*$ и $y_{X^-}^*$ в точках X^+ и X^- соответствующего фрагмента усредненной фазовой траектории (рис. 4.5).

Для повышения надежности определялись средние значения параметров σ , β_p , β_S во всех 12 отведениях, а вместо абсолютного значения угла α ориентации усредненной фазовой траектории использовалось его отклонение $\Delta\alpha_i$ ($i = 2, \dots, 12$) по отношению к первому стандартному отведению.

В результате, вместо 48 первичных признаков использовалось 14 признаков, характеризующих все 12 отведений ЭКГ в фазовом пространстве, в том числе

σ^I – интегральный показатель рассеивания фазовых траекторий;

β_P^I – интегральный параметр симметрии P -зубцов;

β_T^I – интегральный параметр симметрии T -зубцов;

$\alpha_{II}, \alpha_{III}, \alpha_{AVR}, \alpha_{AVL}, \dots, \alpha_{V5}, \alpha_{V6}$ – относительные углы ориентации усредненных траекторий соответствующих отведений.

Результаты статистической обработки данных подробно описаны в работе [39]. Поэтому дадим лишь краткую оценку наиболее интересным результатам.

Установлено, что с надежностью вывода $P > 0,99$ у больных с ревматоидным артритом по сравнению с контрольной группой здоровых добровольцев наблюдались неслучайные изменения средних значений признаков ЭКГ в фазовом пространстве, в том числе:

– уменьшение интегрального показателя σ^I рассеивания фазовых траекторий (0,06 против 0,11 соответственно);

– увеличение относительного угла ориентации $\Delta\alpha_{V1}$ усредненной фазовой траектории в первом грудном отведении (165,2 против 123,3 соответственно);

– увеличение относительного угла ориентации $\Delta\alpha_{V3}$ усредненной фазовой траектории в третьем грудном отведении (101,9 против 63,8 соответственно);

– уменьшение относительного угла ориентации $\Delta\alpha_{V5}$ усредненной фазовой траектории в третьем грудном отведении (97,3 против 184,9 соответственно).

Исследования также показали, что совокупность признаков ЭКГ во временной области и в фазовом пространстве позволяет распознавать представителей трех исследуемых групп – больных ревматоидным артритом (РА), больных ИБС и контрольной группа, включающей здоровых добровольцев (КГ).

Пороговое правило, осуществляющее такое распознавание, имеет вид

РА, если $R_1 > 0,25$;

ИБС, если $R_1 < 0,25$ и $R_2 > 1,5$;

КГ, если $R_1 < 0,25$ и $R_2 < 1,5$,

где

$$R_1 = -26,08 \cdot \sigma^I + 3,5318 \cdot \beta_P^I + 8,3044 \cdot \beta_T^I - 0,0047 \cdot \Delta\alpha_3 + \\ + 0,0252 \cdot \Delta\alpha_{V1} + 0,00177 \cdot \Delta\alpha_{V2} + 0,0094 \cdot \Delta\alpha_{V4} + 0,0009 \cdot \Delta\alpha_{V6} - \\ - 66,297 \cdot F_V + 0,0210 \cdot P_\Sigma - 12,5815 ;$$

$$R_2 = -7,33449 \cdot \sigma^I + 6,0665 \cdot \beta_P^I - 1,42512 \cdot \beta_T^I - \\ - 0,00422 \cdot \Delta\alpha_3 - 0,00707 \cdot \Delta\alpha_{V1} + 0,00076 \cdot \Delta\alpha_{V2} + 0,00382 \cdot \Delta\alpha_{V4} + \\ + 0,00775 \cdot \Delta\alpha_{V6} + 19,62837 \cdot F_V - 0,00832 \cdot P_\Sigma - 1,09178 .$$

Здесь

σ^I – интегральный показатель расеивания фазовых траекторий;

β_P^I, β_T^I – интегральные показатели симметрии фрагментов усредненных фазовых траекторий, соответствующих зубцам P и T ;

$\Delta\alpha_3, \Delta\alpha_{V1}, \Delta\alpha_{V2}, \Delta\alpha_{V4}, \Delta\alpha_{V6}$ – относительные углы ориентации усредненных фазовых траекторий в третьем стандартном, первом, втором, четвертом и шестом отведениях;

F_V – фронтальный вектор волны P ;

P_Σ – признак, характеризующий суммарную «площадь» волны P во всех 12 отведениях.

При экзамене решающего правила на выборке, состоящей из 44 наблюдений, не участвующих в обучении, была допущена только 1 ошибка: здоровый испытуемый был отнесен к группе РА (рис. 4.6).

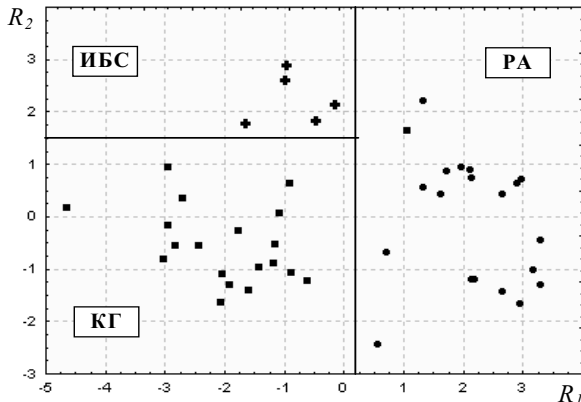


Рис. 4.6. Результаты экзамена решающего правила

Дальнейшие исследования показали, что признак β_T , который характеризует симметрию фрагмента фазовой траектории, соответствующей периоду реполяризации, уже сам по себе является полезным и позволяет обнаружить начальные признаки ИБС под действием физических и эмоциональных нагрузок.

Для оценки полезности β_T проводилась статистическая обработка клинических данных, полученных в отделении ишемических болезней сердца Национального научного центра «Институт кардиологии им. акад. Н.Д. Стражеско» АМН Украины (Киев), а также в четырех клиниках Германии – кардиологической клинике университета Дуйсбург-Эссен (Essen University Hospital), католическом госпитале «Филлипусстифт» (Katholical Hospital "Phillpusstift", Essen), Центре сердца земли Северный Рейн-Вестфалия (Heart and Diabetes Center of North Rhein-Weastfalia, Bad-Oeynhausen), Германском центре сердца (German Heart Center, Berlin)².

Клинический материал составлял 441 записей ЭКГ верифицированных больных ИБС и 387 записей ЭКГ здоровых добровольцев, включенных в контрольную группу.

Статистическая обработка данных показала, что среднее значение параметра β_T существенно различалось в группе ИБС и контрольной группе и составило $0,956 \pm 0,43$ против $0,665 \pm 0,12$ соответственно [55].

Проверка полученного результата по t -критерию Стьюдента подтвердила, что с высокой вероятностью ($P > 0,99$) гипотеза о случайном различии средних может быть отброшена.

Согласно данным медицинской статистики [33], частота (преваленс) ИБС среди лиц старше 40 лет в крупных городах Европы составляет 4–6 %.

На рис. 4.7 показано окно программного модуля, реализующего методику усиленного ROC-анализа, описанную в разделе 3.3. Серым цветом выделена область допустимых значений чувствительности и специфичности теста, построенная согласно условию (3.33) при $P_{ИБС} = 6\%$ и соотношении потерь ложноотрицательных и ложноположительных ошибок $\omega_{ИБС} = 5$.

² Методическое руководство экспериментами осуществлял канд.мед.наук И.А.Чайковский

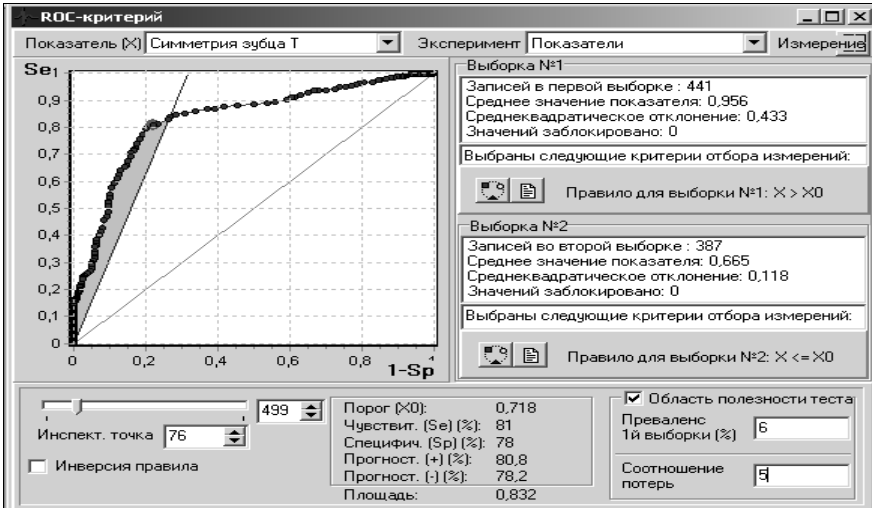


Рис. 4.7. Окно программного модуля усиленного ROC-анализа

Установлено, что принятие решений по пороговому правилу

$$\begin{aligned} \text{ИБС, если } \beta_T > \beta_0 \\ \text{Норма, если } \beta_T < \beta_0' \end{aligned} \quad (4.70)$$

где $\beta_0 \approx 0,72$, обеспечивает чувствительность $S_E = 81\%$ и специфичность $S_p = 78\%$. Легко видеть (рис. 4.7), что тест с такими операционными характеристиками попадает в область гарантированно полезных тестов.

Следует подчеркнуть, что в группу больных были включены лишь пациенты, у которых традиционный ЭКГ анализ не выявил каких либо отклонений от нормы, хотя диагноз ИБС был установлен по результатам коронароангиографии. Поэтому предложенное диагностическое правило, подтвердившее сравнительно высокие показатели чувствительности и специфичности на таком «сложном» клиническом материале, можно считать вполне приемлемым для решения задач скрининга ИБС.

Заметим также, что согласно условиям (3.43) тест, основанный на правиле (4.70), остается полезным в достаточно широком диапазоне допустимых значений соотношения потерь: $4,2 \leq \omega_{ИБС} \leq 64,1$.

Диагностическая ценность признака β_T подтверждена также в экспериментах на животных. Основная цель эксперимента состояла в наблюдении за изменением признака β_T в условиях искусственной ишемии и реперфузии. Эксперименты проводились в Институте физиологии им. А.А. Богомольца НАН Украины совместно с группой сотрудников института под руководством академика А.А.Мойбенко [64].

Эксперименты еще раз подтвердили гипотезу о полезности измерения β_T для выявления начальных признаков ИБС: как только животному проводилась манипуляция, вызывающая искусственную ишемию, значения показателя β_T увеличивались, постепенно приближаясь к порогу «опасных» значений (рис. 4.8).

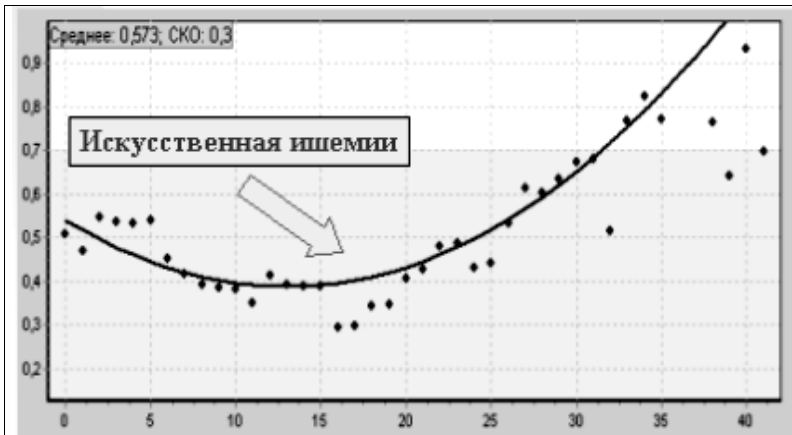


Рис. 4.8. Результаты эксперимента на беспородной собаке (искусственная ишемия)

Адекватная динамика изменения показателя β_T наблюдалась также в процессе эксперимента с велоэргометрической пробой (рис. 4.9), в котором участвовал здоровый доброволец. Легко видеть, что значения β_T увеличивались по мере увеличения нагрузки, а затем снижались в период отдыха (реституции).

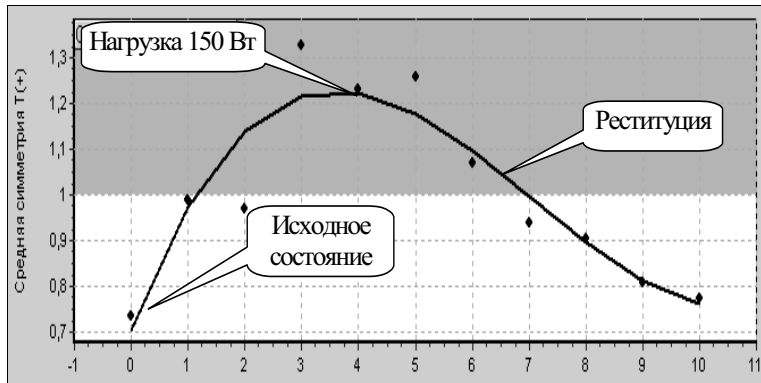


Рис. 4.9. Результаты эксперимента с велоэргометрической пробой

Интересные результаты удалось получить после статистической обработки экспериментальных данных, полученных в инфарктном отделении городской клинической больницы г. Киева. Установлено [63], что изменение показателя β_T симметрии волны T в фазовом пространстве отражает динамику протекания острого коронарного синдрома на госпитальном этапе. Уменьшение симметрии волны T может служить маркером благоприятного исхода лечения. В то же время отсутствие динамики или увеличение симметрии волны T позволяет отнести больных к группе высокого риска развития неблагоприятного исхода острого коронарного синдрома.

На кафедре внутренних болезней № 1 Медицинского института Украинской ассоциации народной медицины были обследованы амбулаторные больные с гипертонической болезнью. Статистическая обработка результатов показала [62], что симметризация зубца T является более ранним признаком электрофизиологических изменений в миокарде, возникших у больных с высоким артериальным давлением, чем традиционные критерии. Обнаружены также достоверные корреляционные связи между β_T и индексом напряжения, который отражает активность механизмов симпатической регуляции, а также между β_T и уровнем диастолического давления.

На основе проведенных исследований разработано устройство ФАЗАГРАФTM, ориентированное на массовые профилактические об-

следования в амбулаторных и домашних условиях для выявления начальных признаков ИБС [48, 51, 56, 61]. Устройство состоит из миниатюрного датчика (рис. 4.10, слева) с пальцевыми электродами и компьютерной программы, реализующей описанные выше алгоритмы.

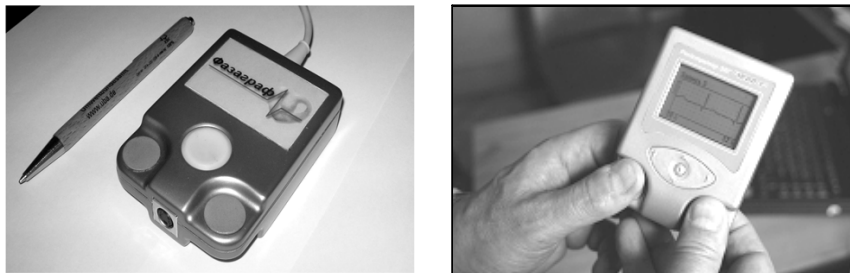


Рис. 4.10. Микроэлектронные регистраторы ЭКГ
(слева – ФАЗАГРАФ™, справа – ИКАР)

На основании проведенной технической и медицинской экспертиз изделие ФАЗАГРАФ™ получило Свидетельство о государственной регистрации № 8398/2008 от 3 декабря 2008 г. и разрешение на его использование в медицинской практике, а также Сертификат № UA-MI/1-2558-2009 от 14 октября 2009 г., в соответствии с которым прибор зарегистрирован в Государственном реестре средств измерительной техники Украины за № У2446-09.

Дальнейшее развитие технологии – полностью автономное микропроцессорное устройство ИКАР (рис. 4.10, справа), которое имеет внутренний экран для отображения регистрируемой ЭКГ и результатов ее обработки.

Рассмотренные в монографии результаты нашли применение и при решении других задач медицинской диагностики, в частности, для оценки полезности тестов, основанных на анализе магнитокардиограмм [8]

СПИСОК ЛИТЕРАТУРЫ

1. *Анфиногенова М.В.* Математическое моделирование и исследование процессов теплообмена и кристаллизации металла в датчиках автоматического контроля состава стали // Автореферат канд. диссертации.– М.:МИСИС,1976.– 26 с.
2. *Барабаш Ю.Л., Б.В. Варский, Зиновьев В.Т.* и др. Вопросы статистической теории распознавания.– М.: Советское радио,1967.– 224 с.
3. *Барабаш Ю.Л.* Минимизация описаний в задаче автоматического распознавания образов // Изв. АН СССР. Техническая кибернетика.– 1964.– № 3.– С.14–23.
4. *Барабаш Ю.Л.* Учет свойств признаков при распознавании // Изв. АН СССР. Техническая кибернетика.– 1965.– № 5.– С.85–92.
5. *Биргер И.А.* Определение диагностической ценности признаков. – Кибернетика.– 1968.– № 3.– С. 80 –85.
6. *Биргер И.А.* Техническая диагностика.– М: Машиностроение. 1978.– 240 с.
7. *Блох Э.Л.* К вопросу о минимальном описании // Радиотехника.– 1960.–Том. 15.– № 2.– с.10-14.
8. *Васецкий Ю.М., Файнзильберг Л.С. Чайковский И.А.* Методы анализа структуры распределения тока в проводящей среде применительно к магнитокардиографии // Электронное моделирование, 2004.– № 3.– С. 95-115.
9. *Васильев В.И.* Распознающие системы. – К.: Наукова думка, 1983. – 422 с.
10. *Васильев В.И., Овсянникова Ф.П.* Оптимизация пространства в процессе обучения распознаванию образов // Автоматика. –1985. – № 3.– С. 8–15.
11. *Вентцель Е.С.* Теория вероятностей. – М.: Наука, 1969.– 575 с.
12. *Власов В.В.* Введение в доказательную медицину. – М.: Медиа Сфера, 2001.– 392 с.

13. *Власов В.В.* Эффективность диагностических исследований.– М.: Медицина, 1988.– 256 с.
14. *Гихман И.И., Скороход А.В.* Введение в теорию случайных процессов.– М: Наука. –1977. – 568 с.
15. *Головкин Б.А.* Машинное распознавание и линейное программирование.– М.: Сов. радио, 1973.– 100 с.
16. *Горелик А.Л.* О выборе пространства признаков системы распознавания объектов и явлений // Кибернетика.–1976.– № 3.– С.35-39.
17. *Горелик А.Л., Скрипкин В.А.* Методы распознавания. – М.: Высшая школа, 1977.– 222 с.
18. *Горелик А.Л., Этштейн С.С.* Об условиях аддитивности информации в задачах распознавания объектов и явлений // Кибернетика.– 1983.– № 6.– С. 85-88.
19. *Гульчак М.Г.* Об информативности совокупности признаков // Автоматика и телемеханика.– 1970.– № 6.– С. 161-163.
20. *Житецкий Л.С., Файнзильберг Л.С.* Об информационном подходе к оценке полезности признаков при статистическом распознавании образов // Изв. АН СССР. Техническая кибернетика.–1983.– № 4. – С.120-126.
21. *Жуковская О.А., Файнзильберг Л.С.* Формальная оценка квалификации эксперта на основе байесовской модели и методов интервального анализа // Проблемы управления и информатики.– 2005.– № 3.– С. 103-115.
22. *Загоруйко Н.В.* Методы распознавания и их применение.– М.: Советское радио.– 1972.– 208 с.
23. *Иваненко В.И.* Неопределенность с стохастических системах управления // Автоматика и телемеханика.– 1983.– № 4.– С. 50-57.
24. *Киселев Н.В., Сечкин В.А.* Техническая диагностика методами нелинейного преобразования.– Л.: Энергия, 1980.– 112 с.
25. *Ковалевский В.А.* Методы оптимальных решений в распознавании изображений. – М.: Наука, 1976. – 328 с.

26. *Кочо В.С., Рудой П.С., Сукачев О.И.* Скоростной метод определения содержания углерода в пробе мартеновской стали, серия 14, инф. № 3. – М.: Черметинформация, 1970.– С.3-8.
27. *Ластед Л.* Введение в проблему принятия решений в медицине: Пер. с англ.– М.: Мир, 1971.– 282 с.
28. *Лунц А. Л., Брашловский В.Л.* Об оценке признаков, получаемых в статистических решающих правилах // Изв. АН СССР. Техническая кибернетика.– 1967.– № 3.– С.99-110.
29. *Мурашко В.В., Струтынский А.В.* Электрокардиография. – М.: Медицина.– 1991.– 288 с.
30. *Неймарк Ю.И., Баталова З.С., Образцова Н.Д.* К вопросу о выборе признаков при распознавании образов // Изв. АН СССР. Техническая кибернетика.– 1970.– № 1.– С. 105-112.
31. *Партасари К.* Введение в теорию вероятностей и теорию меры.– М.: Мир.– 1983.– 336 с.
32. *Распознавание образов и медицинская диагностика* / Под ред. Ю.М. Неймарка.– М.: Наука, 1972.– 328 с.
33. *Руководство по внутренним болезням.* Издание 2-е, переработанное / Под ред. Н.А.Жукова.–Омск: Омская Государственная медицинская академия, 1998.– 512 с.
34. *Румицкий Л.З.* Математическая обработка результатов экспериментов.– М.: Наука, 1971.– 192 с.
35. *Скворцов В.А.* Примеры метрических пространств.– М.: МЦНМО, 2002.– 24 с.
36. *Скурихин В.И., Житецкий Л.С., Файнзильберг Л.С.* Цифровой анализатор “Углерод” // Управляющие системы и машины.– 1978.– № 2.– С. 99-105.
37. *Смирнов Н.В., Дунин-Барковский И.В.* Курс теории вероятностей и математической статистики для технических приложений.– М.: Наука, 1965.– 511 с.
38. *Турбович И.Т., Гитис В.Г.* Дополнительная минимизация числа первичных признаков при использовании информации о параметрах опознающей машины // Опознающие машины.– М.: Наука.– 1968.– С.67-76.

39. *Файнзильберг, А.Ф. Клубова, Стаднюк Л.А.* и др. Новый метод анализа ЭКГ больных ревматоидным артритом // Украинський ревматологічний журнал.– 2001.– № 2.– С. 48-51.
40. *Файнзильберг Л.С.* Диагностика состояния объектов по фазовым траекториям наблюдаемых сигналов с локально сосредоточенными признаками // Проблемы управления и информатики.– 2004.– № 2.– С.56-67.
41. *Файнзильберг Л.С.* Компьютерный анализ и интерпретация электрокардиограмм в фазовом пространстве // Системні дослідження та інформаційні технології.– 2004.– № 1.– С.32-46.
42. *Файнзильберг Л.С.* Восстановление эталона циклических сигналов на основе использования хаусдорфовой метрики в фазовом пространстве // Кибернетика и системный анализ.–2003.– № 3.– С. 20-28.
43. *Файнзильберг Л.С., Потапова Т.П.* Спосіб інтегральної оцінки функціонального стану серцево-судинної системи людини // Патент України № 24516, МКИ А 61 В 5/024.– Заявлено 21.05.97, опубл. 21.07.98.– Бюл. № 5,1998.– 4 с.
44. *Файнзильберг Л.С.* Условия полезности диагностических тестов с позиции теории статистических решений // Проблемы управления и информатики.– 2003.– № 2.– С. 100-111.
45. *Файнзильберг Л.С.* Спосіб інтегральної оцінки поточного стану серцево-судинної системи людини // Патент України № 24517, МКИ А61 В 5/024.– Заявлено 21.05.97, опубл. 30.10.98.– Бюл. № 5, 1998.– 4 с.
46. *Файнзильберг Л.С.* Информационный технологии обработки сигналов сложной формы. Теория и практика.– Киев.: Наукова думка, 2008.– 333 с.
47. *Файнзильберг Л.С.* К вопросу о полезности диагностических методов в задачах скрининга // Управляющие системы и машины.– 2002.– № 6.– С. 10-17.
48. *Файнзильберг Л.С.* Информационная технология для диагностики функционального состояния оператора // Управляющие системы и машины.– 1998.– № 4.– С. 40-45.

49. *Файнзильберг Л.С.* Оценка полезности признаков при решении задач диагностики в статистической постановке // Математические машины и системы.– 1998.– № 1.– С. 57-64.
50. *Файнзильберг Л.С., Шкляр Г.А.* Об оценке полезности признаков при статистическом распознавании двух классов // Автоматика.– 1985.– № 5.– С.78-82.
51. *Файнзильберг Л.С.* Компьютерная система Фазаграф™ для экспресс диагностики состояния сердечно-сосудистой системы в клинических и домашних условиях // Материалы Четверной международной научной школы «Наука и инновации-2009».– Йошкар-Ола: МарГУ, 2009.– С. 279-286.
52. *Файнзильберг Л.С., Жук Т.Н.* Гарантированная оценка эффективности диагностических тестов на основе усиленного ROC-анализа // Управляющие системы и машины.– 2009.– № 5.– С.3-13.
53. *Файнзильберг Л.С.* Применение методов статистического распознавания в термографическом анализе состава металла // Кибернетика.– 1978.– № 6.– С.133-136.
54. *Файнзильберг Л.С.* К вопросу о безошибочном распознавании двух классов по совокупности пересекающихся признаков // Кибернетика. –1982. – № 4. – С. 104-109.
55. *Файнзильберг Л.С.* Новый подход к термическому анализу жидкой стали на углерод // Известия вузов. Черная металлургия.–1980.– № 6.– С. 113-119.
56. *Файнзильберг Л.С.* Новая информационная технология обработки ЭКГ для выявления ишемической болезни сердца при массовых обследованиях населения // Управляющие системы и машины. – 2005.– № 3.– С. 63-71.
57. *Фихтенгольц Г.М.* Курс дифференциального и интегрального исчисления.– Том 1.– М.: Гос.изд-во физ-мат литературы, 1958.– 607 с.
58. *Фукунага К.* Введение в статистическую теорию распознавания образов.– М.: Наука, 1979.– 368 с.
59. *Харкевич А.А.* О выборе признаков при машинном опознавании // Изв. АН СССР. Техническая кибернетики.–1963.– № 2.– С.3.-10.

60. *Циммерман Ф.* Клиническая электрокардиография. – М.: Биком. – 1998. – 448 с.
61. *Чайковский И.А., Файнзильберг Л.С.* Медицинские аспекты применения устройства ФАЗАГРАФ в клинической практике и в домашних условиях.– Киев: МНУЦ ИТИС НАН Украины, 2009.– 74 с.
62. *Чабан Т.І., Чайковский І.А., Файнзильберг Л.С.* и др. Можливості аналізу електрокардіограми у фазовому просторі та варіабельності ритму серця в амбулаторних пацієнтів із гіпертонічною хворобою // Український медичний часопис. – 2009. – № 2 (70) – III-IV 2009.– С. 126-128.
63. *Чайковский И.А., Батушкин В.В., Файнзильберг Л.С.* и др. Эффективность оценки течения острого коронарного синдрома по данным анализа первого отведения ЭКГ на фазовой плоскости // Журнал Академії медичних наук.– 2007.– Том 13.– № 1.– С. 104-113.
64. *Чайковський І.А., Неццерет О.П., Файнзильберг Л.С.* и др. Дослідження функції серця при ішемії міокарда за допомогою нового методу обробки електрокардіограми // Фізіологічний журнал. – 2008.– Том 54.– № 6.– С.42-48.
65. *Шлезингер М.И.* Самообучение распознаванию образов – теоретические вопросы и алгоритмы // Труды IV Всесоюзной школы-семинара “Автоматическое распознавание слуховых образов” (АРСО-IV) –Киев.–Институт кибернетики.– С. 42-48.
66. *Шлезингер М., Главач В.* Десять лекций по статистическому и структурному распознаванию.– К.: Наукова думка, 2004.– 545 с.
67. *Barber C., Dobkin D., Huhdanpaa H.* The quickhull algorithm for convex hull. // Technical Report GCG-53, University of Minnesota.– 1993.– 185 P.
68. *Ben-Bassat M.* Irrelevant features in pattern recognition // IEEE Trans. Comput.– 1978.– Vol. C-27.– No 8.– P. 749-766.
69. *Connolly DC., Elveback LR., Oxman HA.* Coronary heart disease in residents of Rochester, Minnesota: Prognostic value of the resting electrocardiogram at the time of initial diagnosis of angina pectoris // Mayo Clinic Proceedings.–1984.– Vol. 59.– P. 247-250.

70. *Egan J.P.* Signal detection theory and ROC analysis.– New York: Academic Press.– 1975.– 386 P.
71. *Fainzilberg L.S.* Heart Functional State Diagnostic Using Pattern Recognition of Phase Space ECG-Images.– Proc. of the 6th European Congress on intelligent techniques and soft computing (EUFIT '98).– Aachen (Germany).– 1998.– No: B- 27.– Vol. 3.– P. 1878-1882.
72. *Fainzilberg L.S.* Interconnection between Features Properties and Probability of Error in Statistical Recognition of Two Classes // Proceeding of the 12th International Conference on Pattern Recognition (ICPR'94).– Vol. 2.– Jerusalem (Israel).–IEEE Computer Society Press (LA, CA).– 1994.– P. 544-546.
73. *Fainzilberg L.S., Potapova T.P.* Computer Analysis and Recognition of Cognitive Phase Space Electro-Cardio Graphic Image // Proc. of the 6th Intern. Conference on computer analysis of images and patterns (CAIP'95). –Prague (Czech. Republic).– 1995.– P. 668-673.
74. *Fainzilberg L.S.* Why Relevant Features May Be Useless in Statistical Recognition of Two Classes // Proceeding of the 13th International Conference on Pattern Recognition (ICPR'96).– Viena (Austria).– 1996.–P. 730-734.
75. *Fainzilberg L.S.* ECG Averaging Based on Hausdorff Metric // International Journal of Biomagnetism.– 2003.– Vol. 5.– № 1.– P.236-237.
76. *Fainzilberg L.S.* Methods and device for discriminating thermal effect of phase transformation of metals and alloys in the process of their cooling // Парент СІІІА № 4198679, МКІ G 01N.– Опыбл. 15.04.80.
77. *Fainzilberg L.S.* Nowa metoda interpretacji zapisu EKG w balaniach skринingowych oraz w opiece domowej // Zdrowie publiczne.– 2005.– Vol. 115.– Number 4.– P.458-464.
78. *Fawcett T.* ROC Graphs: Notes and Practical Considerations for Researchers // Technical Report MS 1143.– 2004.– Kluwer Academic Publishers.– 38 P.
79. *Fawcett T.* Using Rule Sets to Maximize ROC Performance // Proceedings of the IEEE International Conference on Data Mining (ICDM-2001).– 2001.– IEEE Computer Society.– Los Alamitos, CA.– P. 131-138.

-
80. *Ferri C., Hernández-Orallo J., Salido M.A.* Volume Under the ROC Surface for Multiclass Problems. Exact Computation and Evaluation of Approximations // Technical Report DSIC. –2003.– Univ. Politèc. València.– P. 36–43.
 81. *Flach P., Wu S.* Repairing concavities in ROC curves. // Proc. 2003 UK Workshop on Computational Intelligence.– London, 2003.– P. 38 - 44.
 82. *Green D. M., Swets J. A.* Signal detection theory and psychophysics. – New York: John Wiley and Sons Inc.–1966.– 328 P.
 83. *Hand D. J., Till R. J.* A simple generalization of the area under the ROC curve to multiple class classification problems // Machine Learning.– 2001.– No. 45(2).– P. 171-186.
 84. *Hanley J. A., McNeil B. J.* The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve // Radiology.– 1982.– No. 143.– P. 29–36.
 85. *Kittler J.* Feature selection and extraction // Handbook of Pattern Recognition and Image Processing.– 1986.– 450p.
 86. *Lewis P.M.* The characteristic selection problem in recognition system // IRE Trans. Inform. Theory.– 1962.– Vol.8.– No2.– P. 171-178.
 87. *Metz C.E.* Fundamental ROC analysis. Progress in medical physics and psychophysics // Handbook of Medical Imaging (Vol. 1).– Bellingham: WA: SPIE Press, 2000.– P. 754-769.
 88. *Narendra P.M., Fukunaga K.A.* Branch and bound algorithm for feature subset selection // IEEE Trans. Comp.– 1977.– Vol.– C-26.–No 9.– P. 917-922.
 89. *Provost F., Fawcett T.* Robust Classification for Imprecise Environments. // Machine Learning.– 2001.– Vol. 42.– No. 3.– P. 203-231.
 90. *Siedlecki W., Sklansky J.* On automatic feature selection // Int. Journal of Pattern Recognition and Artificial Intelligence.– 1988.– Vol.2.– P. 197-220.
 91. *Schlisterman E.F., Perkins N.J., Liu A., Bondell H.* Optimal Cut-point and Its Corresponding Youden Index to Discriminate Individuals Using Pooled Blood Samples // Epidemiology.– 2005.– Vol. 16.– P. 73–81.

92. *Spackman, K. A.* (1989). Signal detection theory: Valuable tools for evaluating inductive learning. Proceedings of the Sixth International Workshop on Machine Learning.–1989.– San Mateo, CA.– P. 160 - 163.
93. *Swets J.A.* Measuring the accuracy of diagnostic systems // Science. – 1988.– No. 240. – P. 1285-1292.
94. *Srikanth T., Napper S.A., Gu H.* Bottom-up approach to uniform feature extraction in time and frequency domains for single-lead ECG signal // International Journal of Bioelectromagnetism.– 2002.– Vol. 4. – P. 21-29.
95. *Stearns S.D.* On selecting features for pattern classifiers. – Proceeding of the Third Int. Conference on Pattern recognition (Coronado, CA).– 1976.– P.71-75.
96. *Tou J.T., Heydorn R.P.* Some approaches to optimum feature selection // Computers and Information Science Academy Press.– 1967.– Vol. 11.– No. 4.– P. 57-89.
97. *Van den Hout W. B.* The area under an ROC curve with limited information. Medical Decision Making.– 2003.– No. 23.– P. 160-166.
98. *Youden W. J.* Index for rating diagnostic tests // Cancer.– 1950.– No. 3. – P. 32–35.
99. *Zweig M.H, Campbell G.* Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine // Clinical Chemistry.– 1993.– Vol. 39.– No. 4.– P. 561–577.

Монография

ФАЙНЗИЛЬБЕРГ Леонид Соломонович

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ
ОЦЕНКИ ПОЛЕЗНОСТИ ДИАГНОСТИЧЕСКИХ
ПРИЗНАКОВ**

Подписано в печать 02.03.2010 г.

Формат 60x84/16.

Усл. печ. л. 4,41. Уч.-изд. л. 6,84.

Тираж 500 экз.

Издатель и изготовитель:

Издательство «Освита Украины»

04213, г. Киев, ул. Героев Днепра, 63, к. 40

Свидетельство о внесении в Государственный реестр издателей

ДК №1957 от 23.04.2009 г.

Тел./факс (044) 411–43–97; 237–59–92

E-mail: osvita2005@ukr.net, www.rambook.ru

Издательство «Освита Украины» приглашает авторов к сотрудничеству по выпуску изданий, касающихся вопросов управления, модернизации, инновационных процессов, технологий, методических и методологических аспектов образования и учебного процесса в высших учебных заведениях.

Предоставляем все виды издательских и полиграфических услуг.