Проблемы управления и информатики.–2003. – № 2. – С.100-111.

Условия полезности диагностических тестов с позиции теории статистических решений

Л.С. Файнзильберг

Введение. Одним из центральных вопросов распознавания образов и диагностики состояния объектов различной физической природы является выбор эффективной системы признаков [1-4]. Существуют различные подходы к оценке диагностической ценности признаков. Так, согласно, [5,6] признак считается информативным, если он снижает априорную неопределенность по Шеннону относительно распознаваемых классов. Легко показать, что такой критерий эквивалентен критерию релевантности [7], основанному на различии условных распределений признака в классах. По сути дела именно такой подход получил широкую известность для оценки эффективности диагностических тестов на основе ROC-анализа (*receiver operating characteristic curve*) [8-11].

В то же время в работах [12,13] было показано, что один лишь факт различия условных распределений признака в классах является необходимым, но не достаточным условием его полезности с точки зрения уменьшения средней вероятности ошибочных решений.

Между тем известно [14], что средняя вероятность ошибочных решений является лишь частным случаем среднего риска, и не учитывает соотношения потерь от ошибок различного рода. Поскольку для задач медицинской и технической диагностики потери от ошибок пропуска цели и ложной тревоги чаще всего не равнозначны, актуальным является получение условий, гарантирующих полезность как отдельных признаков, так и диагностического теста в целом с точки зрения уменьшения среднего риска.

В настоящей работе приведено исследование данного вопроса.

Постановка задачи. Пусть некоторый объект Z находится в одном из двух возможных состояний V_1 , V_2 с априорными вероятностями $P(V_1)$ и $P(V_2)$ = 1 - $P(V_1)$. Имеется алгоритм (диагностический тест), который на основании измерения доступной информации принимает решения

$$\delta = \begin{cases} 1, & \text{\textit{если принято решение}}\ Z \in V_1; \\ 2, & \text{\textit{если принято решение}}\ Z \in V_2. \end{cases} \qquad (1)$$

Предполагается, что помимо правильных решений, алгоритм может ошибаться, т.е. допускать как ошибки пропуска цели, так и ложной тревоги, или, что то же самое, ошибки первого и второго рода [15]:

$$\begin{aligned} E_1 &= (Z \in V_1) \wedge (\delta = 2) \\ E_2 &= (Z \in V_2) \wedge (\delta = 1) \end{aligned}.$$

Оценки вероятностей $P(E_1)$ и $P(E_2)$ могут быть получены путем проверки («экзамена») теста на некоторой представительной выборке наблюдений, для которых известны истинные состояния Z.

Будем, как это принято в теории статистических решений [16], характеризовать возможные потери платежной матрицей вида

$$L = \begin{Vmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{Vmatrix} \tag{2}$$

в которой L_{11} и L_{22} – потери, связанные с правильными решениями, а L_{12} и L_{21} – потери, связанные с ошибками пропуска цели и ложной тревоги. Тогда математическое ожидание потерь (средний риск) определяет взвешенная сумма указанных потерь с учетом вероятностей их появления:

$$R = \sum_{k=1}^{2} \sum_{j=1}^{2} L_{kj} P(V_k, \delta = j) \ . \tag{3}$$

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.

Определение 1. Диагностический тест полезен, если средний риск R принимаемых на его основе решений, меньше риска R_0 априорных решений, принимаемых по вероятностям $P(V_1)$ и $P(V_2)$:

$$R < R_0. \tag{4}$$

Ставится задача получить условия, гарантирующие выполнение строгого неравенства (4).

Условия полезности диагностического теста. Введем обозначение

$$\omega = \frac{L_{12} - L_{11}}{L_{21} - L_{22}}, \tag{5}$$

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. Заметим, что в частном случае, когда $L_{11} = L_{22} = 0$, величина ω просто определяет соотношение потерь от ошибок пропуска цели и ложной тревоги.

Утверждение 1. Диагностический тест полезен в смысле (4), в том и только в том случае, когда

$$[(1 - P(V_1)]P(E_2) < \omega P(V_1)[1 - P(E_1)] \tag{6}$$

при $P(V_1)(1+\omega) < 1$ либо

$$[1 - P(V_1)][1 - P(E_2] > \omega P(V_1)P(E_1) \tag{7}$$

при $P(V_1)(1+\omega) > 1$.

Доказательство. Поскольку по определению $P(V_k,\delta{=}j) = P(V_k)P(\delta = j/V_k)$, а $P(\delta = 2/V_1) = P(E_1)$ и $P(\delta = 1/V_2) = P(E_2)$, причем $P(\delta = 1/V_1) = 1 - P(E_1)$ и $P(\delta = 2/V_2) = 1 - P(E_2)$ то, согласно (3), при известных $P(V_1)$, $P(E_1)$, $P(E_2)$ и L средний риск решений (1) можно представить в эквивалентной форме записи

$$\begin{aligned} R = L_{11}P(V_1)[1 - P(E_1)] + L_{22}[1 - P(V_1)][1 - P(E_2)] + \\ + L_{12}P(V_1)P(E_1) + L_{21}[1 - P(V_1)]P(E_2) \end{aligned}, \tag{8}$$

С другой стороны, если 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 $L_{22}[1- P(V_1)] + L_{12} P(V_1)$.

Принимая во внимание обозначение (5) заключаем, что если выполняется условие

$$P(V_1)(1+\omega) < 1, \tag{9}$$

то с точки зрения минимума априорного риска лучше всегда принимать решение $Z \in V_2$, а если выполняется условие

$$P(V_1)(1+\omega) > 1 \tag{10}$$

то лучшими оказываются решения $Z \in V_1$.

Следовательно априорный риск R_0, фигурирующий в правой части (4), можно представить в виде

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

Подстановка (8) и (11) в (4) после простых преобразований приводит к условиям (6),(7). Утверждение доказано.

Из Утверждения 1 следует, что если не выполняются условия (6) или (7), то с точки зрения уменьшения априорного риска диагностический тест абсолютно бесполезен. Этот факт позволяет сформулировать еще два важных утверждения.

Утверждение 2. Пусть выполняется условие $P(V_1)(1+\omega) < 1.$ Тогда при *сколь угодно малой* вероятности ошибки пропуска цели диагностический тест заведомо бесполезен с точки зрения уменьшения априорного риска, если вероятность ошибки ложной тревоги удовлетворяет условию

$$\boldsymbol{P(E_2) \geq \frac{\omega P(V_1)}{1-P(V_1)}}. \tag{12}$$

Доказательство. Принимая во внимание, что $P(V_1) \neq 1$, представим условие (6), гарантирующее полезность теста для области $\boldsymbol{P(V_1)(1+\omega) < 1}$, в эквивалентной форме записи

$$P(E_2) < \frac{\omega P(V_1)[1-P(E_1)]}{1-P(V_1)}. \tag{13}$$

Поскольку всегда $P(E_1) \geq 0$, то усиление (13) путем подстановки $P(E_1) = 0$ приводит к оценке сверху допустимого значения вероятности ошибки ложной тревоги полезного теста

$$\sup P(E_2) = \frac{\omega P(V_1)}{1 - P(V_1)} \tag{14}$$

для области $P(V_1)(1+\omega) < 1$.Отсюда непосредственно следует справедливость утверждения.

Утверждение 3. Пусть выполняется условие $P(V_1)(1+\omega) > 1$. Тогда при любой *сколь угодно малой* вероятности ошибки ложной тревоги диагностический тест заведомо бесполезен с точки зрения уменьшения априорного риска, если вероятность ошибки пропуска цели удовлетворяет условию

$$P(E_1) \geq \frac{1 - P(V_1)}{\omega P(V_1)}. \tag{15}$$

Доказательство. Поскольку $P(V_1) \neq 0$, то условие (7), гарантирующее полезность теста для области $P(V_1)(1+\omega) > 1$ можно представить так

$$P(E_1) < \frac{[1 - P(V_1)[1 - P(E_2)]}{\omega P(V_1)}. \tag{16}$$

Если учесть, что всегда $P(E_2) \geq 0$, то усиление (16) путем подстановки $P(E_2) = 0$ приводит к оценке сверху допустимого значения вероятности ошибки пропуска цели полезного теста

$$\sup P(E_1) = \frac{1 - P(V_1)}{\omega P(V_1)} \tag{17}$$

для области $P(V_1)(1+\omega) > 1$. Утверждение доказано.

Для иллюстрации на рис.1 и 2 показаны границы областей (12) и (15) при $P(V_1) < 0.1$ для различных значениях ω.

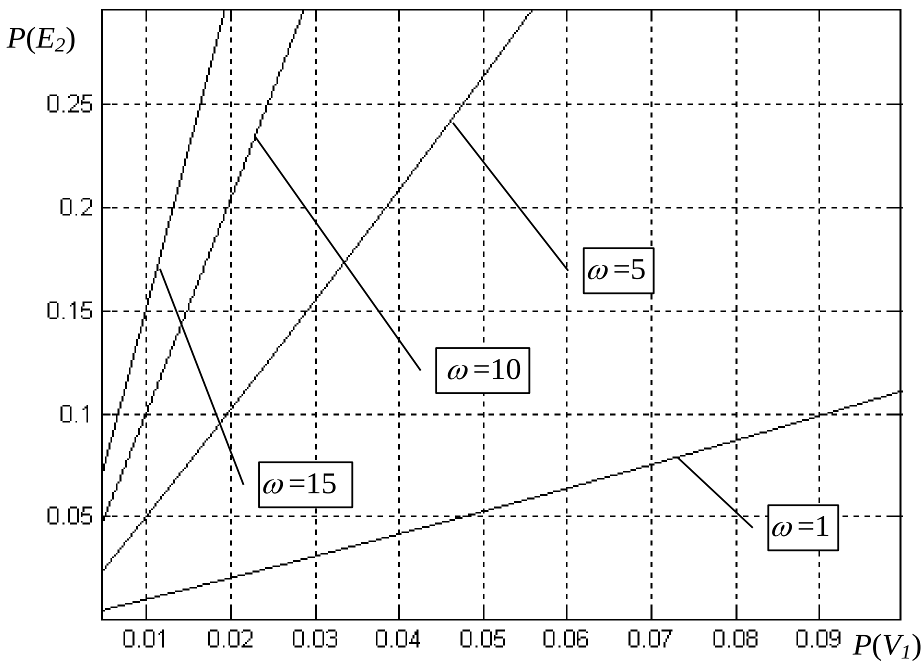

Рис.1

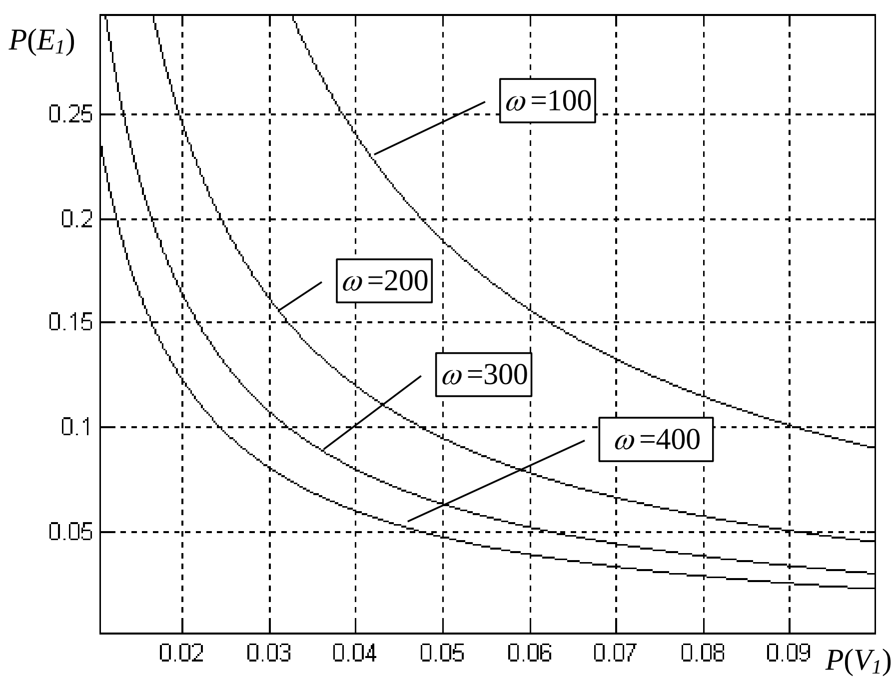

Рис.2

Условия полезности совокупности диагностических признаков. Любой алгоритм диагностики основан на измерении некоторой совокупности признаков $x_1,..., x_N$, образующих N-мерное описание

$x^{(N)} = x_1,..., x_N$ распознаваемых классов V_1 и V_2. В простейшем случае для диагностики используется только один признак (N = 1). Тогда, если известны вероятности ошибок пропуска цели и ложной тревоги, допускаемые этим признаком, априорная вероятность $P(V_1)$ и значение величины ω, то полезность отдельного признака можно оценить на основе полученных выше условий.

Для оценки полезности признаков в совокупности (N > 1) введем следующее определение.

Определение 2. Диагностический признак $x_n (1 \le n \le N)$ полезен в совокупности с остальными N-1 признаками, если его исключение из описания $x^{(N)} = x_1,..., x_N$ приводит к увеличению среднего риска, т.е.

$$R_{N-1} > R_N \ , \tag{18}$$

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 (18) для бинарных признаков (симптомов), характерных для задач медицинской диагностики [17]. 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 условиям

$$P(x_1^+ / V_1) = 1 , \quad 0 < P(x_1^+ / V_2) < 1. \tag{19}$$

Поскольку из (19) следует, что $P(x_1^- / V_1) = 0$, то при значении признака $x_1 = x_1^-$ (отсутствии симптома) следует однозначно принимать решение в пользу V_2. Ясно, что в этом случае вероятность ошибочного отнесения больного к здоровым (вероятность ошибки пропуска цели) составит $P(E_1) = 0$.

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

$$P(x_1^+ / V_2) \geq \frac{\omega P(V_1)}{1 - P(V_1)}. \tag{20}$$

Легко видно, что для фиксированных значений $P(V_1)$ и ω в области $P(V_1)(1+\omega) < 1$ условие (20) не противоречит второму из ограничений (19). Если при $x_1 = x_1^+$ принимать решение в пользу класса V_1, другими словами считать Z больным при наличии у него симптома, то вероятность ошибочного отнесения здорового к больным (вероятность ошибки ложной тревоги) определяет $P(x_1^+ / V_2)$. Однако, если $P(E_2) = P(x_1^+ / V_2)$, то при выполнении условия (20) в соответствии с Утверждением 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 при $x_1 = x_1^+$. Риск решении $Z \in V_1$ 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 решении $Z \in V_2$ равен

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

Если учесть, что $P(x_1^+ / V_1) = 1$, то с учетом обозначения (5) разность $\Delta R = R_2 - R_1$ определяется соотношением

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

из которого в силу (20) и того, что $L_{21} > L_{22}$ следует, что $R_2 \leq R_1$. Но это означает то, что и при втором значении признака $x_1 = x_1^+$ (наличии симптома) правомерно 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 имеется два таких бесполезных признака. Возникает естественный вопрос: в каких случаях их совокупность может быть полезна с точки зрения Определения 3?

В работе [18] показано, что совокупность неинформативных в отдельности признаков может быть полезна, если они статистически связаны в классах. Покажем теперь, что совокупность бесполезных в отдельности признаков-симптомов может быть полезна и в том случае, когда такая связь отсутствует.

Этот факт в точных формулировках определяет следующее утверждение.

Утверждение 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

$$P(x_i^+ / V_1) = 1, \quad i = 1,2\,, \tag{22}$$

Пусть, кроме того, признаки статистически независимы в обоих классах, т.е.

$$P(x_1, x_2 / V_k) \equiv P(x_1 / V_k) P(x_2 / V_k), \quad k = 1,2\,. \tag{23}$$

Тогда для любых $P(V_1)$ и ω, удовлетворяющих условию $P(V_1)(1+\omega) < 1$, существуют такие значения вероятностей $P(x_i^+ / V_2), \quad i = 1,2$, при которых каждый из этих признаков бесполезен сам по себе, но их совокупность полезна.

Доказательство. Предположим, что

$$P(x_i^+ / V_2) \ge \rho, \quad i = 1,2\,, \tag{24}$$

где

$$\rho = \frac{\omega P(V_1)}{1 - P(V_1)} \quad . \tag{25}$$

Как было показано ранее в этом случае рассматриваемые признаки сами по себе бесполезны. Рассмотрим условные риски возможных решений при следующей комбинации значений признаков $(x_1 = x_1^+) \wedge (x_2 = x_2^+)$. Принимая во внимание (23) для такой комбинации риск решения $Z \in V_1$ равен

$$R_1(x_1^+, x_2^+) = L_{11} P(V_1) P(x_1^+ / V_1) P(x_2^+ / V_1) + L_{21}[1 - P(V_1)] P(x_1^+ / V_2) P(x_2^+ / V_2)\,, \tag{26}$$

а риск решения $Z \in V_2$ равен

$$R_2(x_1^+, x_2^+) = L_{22}[1 - P(V_1)]P(x_1^+ / V_2)P(x_2^+ / V_2) + L_{12}P(V_1)P(x_1^+ / V_1)P(x_2^+ / V_1), \quad (27)$$

Поскольку при $P(V_1)(1+\omega) < 1$ минимизация априорного риска приводит к решению $Z \in V_2$, потребуем, чтобы для рассматриваемой комбинации $(x_1 = x_1^+) \wedge (x_2 = x_2^+)$ с точки зрения минимизации условного риска оптимальным было противоположное решение, т.е. решение $Z \in V_1$. Другими словами, потребуем, чтобы выполнялось строгое неравенство $R_1(x_1^+, x_2^+) < R_2(x_1^+, x_2^+)$.

На основании (26),(27) с учетом обозначения (25) последнее неравенство можно представить в виде

$$P(x_1^+ / V_2)P(x_2^+ / V_2) < \rho. \quad (28)$$

С другой стороны, в соответствии с (24)

$$P(x_1^+ / V_2)P(x_2^+ / V_2) \geq \rho^2. \quad (29)$$

Поскольку при $P(V_1) \neq 0$ для области $P(V_1)(1+\omega) < 1$ 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 (24), при которых условия (28),(29) будут выполняться совместно. Утверждение доказано.

Для иллюстрации на рис. 3 в координатах ρ и $\chi = P(x_1^+ / V_2)P(x_2^+ / V_2)$ показана область Ω значений, удовлетворяющих (28),(29).

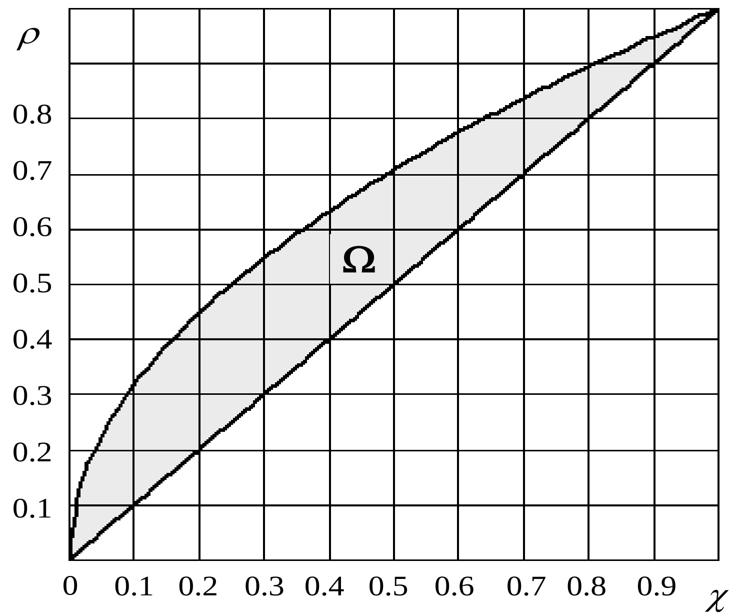

Рис. 3

Модельный пример. 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 статистически независимых в классах бинарных признака, условные распределения которых представлены в таблицах 1 и 2.

Таблица 1

-	$x_2 = x_2^-$	$x_2 = x_2^+$	$P(x_1/V_1)$
$x_1 = x_1^-$	0	0	0
$x_1 = x_1^+$	0	1	1
$P(x_2/V_1)$	0	1	-

Таблица 2

-	$x_2 = x_2^-$	$x_2 = x_2^+$	$P(x_1/V_2)$
$x_1 = x_1^-$	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	-

Легко видно, что для обоих признаков выполняются условия (22),(24), а значит эти признаки сами по себе бесполезны. В этом легко убедиться, если сравнить условные риски возможных решений при различных значениях этих признаков. Так, например, для признака x_1 решение $Z \in V_1$ приводит к риску $R_1(x_1^+) = 0.288$ при $x_1 = x_1^+$ и риску $R_1(x_1^-) = 0.672$ при $x_1 = x_1^-$, 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 риска должны приниматься решения в пользу класса V_2, которые при заданных $P(V_1) = 0.04$ и $\omega = 4$ совпадают с априорными решениями. Аналогичная ситуация имеет место и для признака x_2.

Вместе с тем, в условиях данного примера $\chi = P(x_1^+ / V_2) P(x_2^+ / V_2) = 0.09$, а значит для $\rho \approx 0.167$ выполняется условие $\chi < \rho \le \sqrt{\chi}$ и совокупность признаков оказывается полезной. 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.

Частный случай. Хорошо известно [14], что если матрица потерь (2) «антидиогональна», т.е. $L_{11} = L_{22} = 0$, а $L_{12} = L_{21} = 1$, то средний риск равен средней вероятности ошибки, а оптимальные решения, обеспечивающие минимум этой вероятности, должны приниматься по правилу максимума апостериорных вероятностей.

Если при этом допустить, что в условиях Утверждения 4 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

$$\frac{P(V_1)}{1 - P(V_1)} \le P_2^+ < \sqrt{\frac{P(V_1)}{1 - P(V_1)}} \qquad (30)$$

такие признаки будут бесполезны сами по себе, но полезны в совокупности.

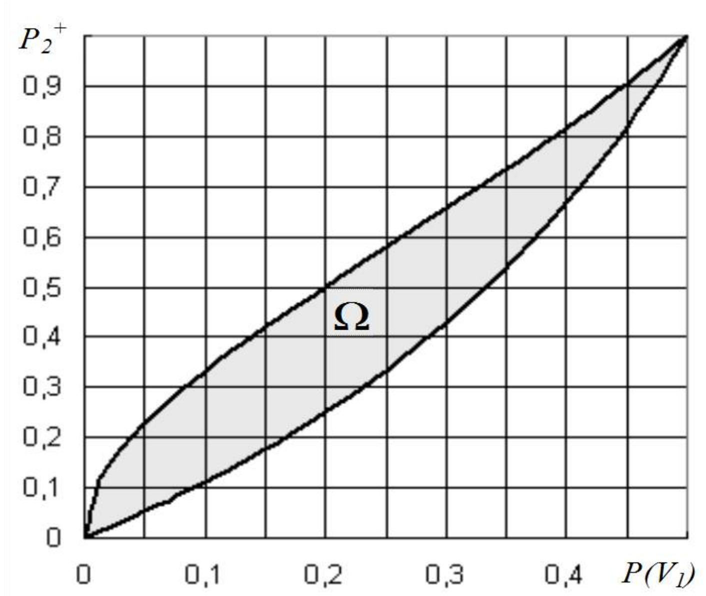

Рис.4

На рис. 4 в координатах $P_2^+ - P(V_1)$ показана область Ω значений, удовлетворяющих условию (30). Так, например, если $P(V_1) = 0.1$ и некоторый симптом присутствует для всех объектов первого класса ($P_1^+ = 1$) и для 20% объектов второго класса ($P_2^+ = 0.2$), то, как видно из рис. 4, точка с такими координатами попадаем в область Ω .

Непосредственной проверкой по формуле Байеса можно убедиться в том, что $P(V_1 / x^+) \approx 0.357$, в то время как $P(V_2 / x^+) \approx 0.642$. Если к тому же учесть, что $P(V_1 / x^-) = 0$, а $P(V_2 / x^-) = 1$, то становится ясно, что при любом значении признака оптимальным оказывается решение $Z \in V_2$, совпадающее с решением, принимаемых по максимуму априорных вероятностей $P(V_1)$ и $P(V_2) = 1 - P(V_1)$ распознаваемых классов (рис. 5 а). Значит с точки зрения средней вероятности ошибки такой признак-симптом абсолютно бесполезен.

В то же время, если два подобных признака статистически независимы в классах, то их совместное использование оказывается полезным: при обнаружении обоих симптомов оптимальным является решение в пользу класса V_1, поскольку апостериорная вероятность $P(V_1 / x_1^+, x_2^+) \approx 0.74$, тогда как $P(V_2 / x_1^+, x_2^+) \approx 0.26$ (рис. 5 б).

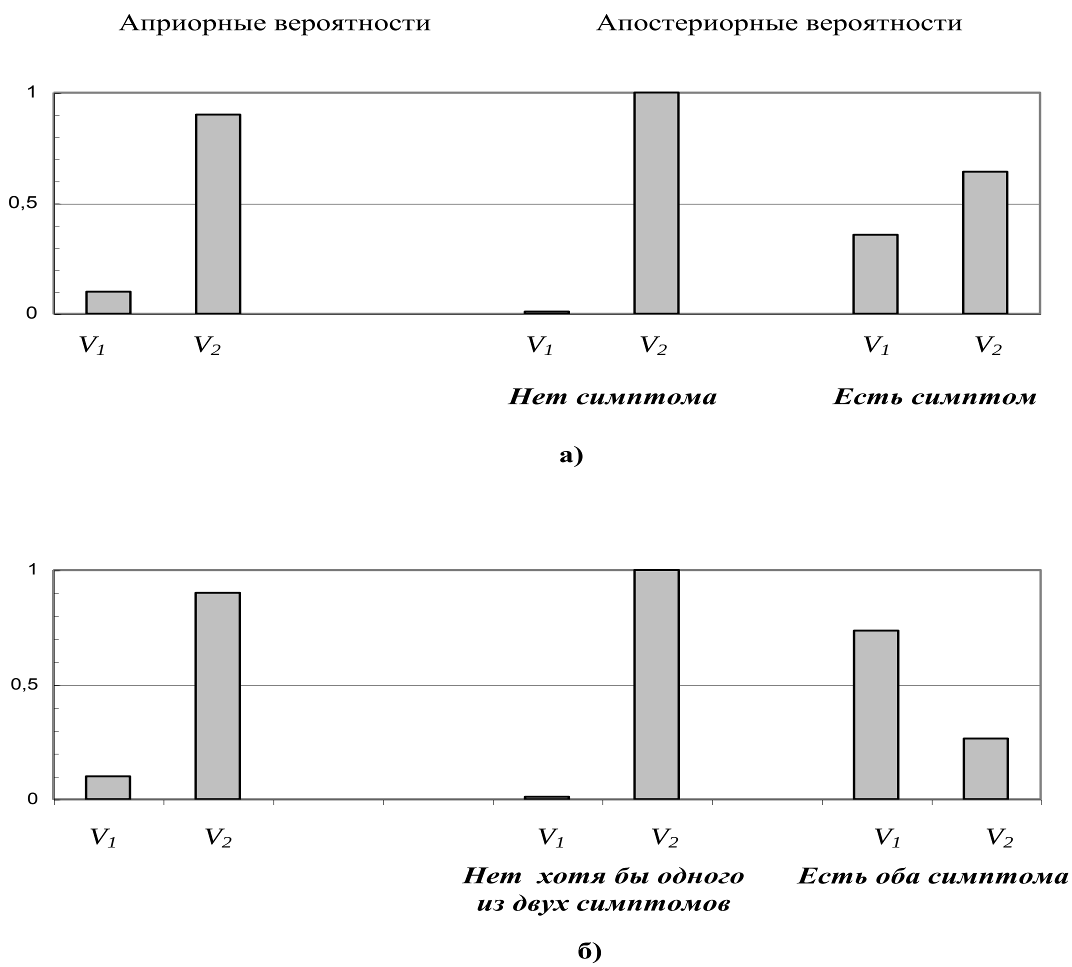

Рис. 5.

Полученные результаты еще раз напоминают о том, насколько осмотрительно нужно подходить к вопросу об исключения «неэффективных» диагностических признаков в задачах медицинской и технической диагностики.

Заключение. В статье показано, что диагностический тест (признак) полезен с точки зрения уменьшения априорного риска (Определение 1) в том и только в том случае, когда вероятности ошибок пропуска цели и ложной тревоги удовлетворяют условиям (6),(7) Утверждения 1.

Получены оценки сверху граничных значений вероятностей ошибок пропуска цели и ложной тревоги, которые в соответствии с Утверждениями 2 и 3 определяются априорной вероятностью класса и величиной ω, характеризующей соотношение потерь от ошибочных и правильных результатов.

Показано, что с точки зрения Определения 2 совокупность бесполезных в отдельности признаков может быть полезна даже в том случае, когда признаки статистически независимы в обоих классах (Утверждение 4).

Полученные результаты могут быть использованы для обоснованного конструирования пространства признаков в задачах медицинской и технической диагностики.

1. *Pudil P., Ferri F.J., Novovicova J., Kittler J.* Floating Search Methods for Feature selection with Non monotonic Criterion Function. - Proc. of the 12-th Int. Conf. on Pattern Recognition. - vol.2, Jerusalem. -1994. - P.279-283.
2. *Kittler J.* Feature Selection and Extraction. - Handbook of Pattern Recognition and Image Processing. - 1986.- 450 p.
3. *Барабаш Ю.Л.* Учет свойств признаков при распознавании. - Изв. АН СССР. Техн. кибернетика.- 1965. - № 5. - С.85-92.
4. *Fainzilberg L.S.* Interconnection Between Feature Properties and Probability of Error in Statistical Recognition of Two Classes. - Proc. of the 12-th Int. Conf. on Pattern Recognition. - vol.2. - Jerusalem. -1994. - P.544-546.
5. *Lewis P.M.* The Characteristic Selection Problem in Recognition System. - IRE Trans. Inform. Theory. - vol. 8. – N 2. - 1962. - P.171-178.
6. *Биргер И.А.* Определение диагностической ценности признаков. - Кибернетика. - 1968. - № 3. - С. 80-85.
7. *Ben-Bassat M.* Irrelevant Features in Pattern Recognition. - IEEE Trans. Comput. - vol.C-27. - 1978. – N 8. - pp.749-766.
8. *McNeil B.J., Hanley J.A.* Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Decision Making.- 1984.- N 4.- P. 137-150
9. *Gray R., Begg C.B., Greenes R.A.* Construction of receiver operating characteristic curves when disease verification is subject to selection bias. Med Decision Making.- 1984.- N 4.- P. 151-164.
10. *Pan X., Metz C.E.* The "proper" binormal model: parametric ROC curve estimation with degenerate data. Acad. Radiology. - 1997.- N 4.- P. 380-389.
11. *Metz C.E.* Fundamental ROC analysis. Progress in Medical Physics and Psychophysics. In: Handbook of Medical Imaging, Vol. 1: (R Van Metter, J Beutel and H Kundel, eds.). Bellingham, WA; SPIE Press.- 2000.- P .754-769.
12. *Файнзильберг Л.С.* Оценка полезности признаков при решении задач диагностики в статистической постановке .- Математические машины и системы, 1998, № 1.- С. 57-64.
13. *Fainzilberg L.S.* Why Relevant Features May Be Unuseful in Statistical Recognition of Two Classes // Proc. of the 13-th International Conference on Pattern Recognition (ICPR'96).- Viena (Austria). - 1996. - P. 730-734.

14. *Ковалевский В.А.* Методы оптимальных решений в распознавании изображений. -М. - Наука. -1976. - 328 с.
15. *Васильев В.И.* Распознающие системы: Справочник. – К.: Наук. думка, 1983.- 422 с.
16. *Горелик А.Л., Скрипник В.А.* Методы распознавания. – М.: Высшая школа, 1977. – 220 с.
17. *Власов В.В.* Эффективность диагностических исследований. – М.: Медицина, 1988.- 256 с.
18. *Файнзильберг Л.С.* К вопросу о безошибочном распознавании двух классов по совокупности пересекающихся признаков. - Кибернетика, 1982, № 4.- С.104-109.