Опубликовано а журнале Математические машины и системы. – 1998. – № 1.– С. 57-64.

УДК 62-50

Л.С. ФАЙНЗИЛЬБЕРГ

ОЦЕНКА ПОЛЕЗНОСТИ ПРИЗНАКОВ ПРИ РЕШЕНИИ ЗАДАЧ ДИАГНОСТИКИ В СТАТИСТИЧЕСКОЙ ПОСТАНОВКЕ

Введение. При построении компьютерных систем медицинской и технической диагностики актуальной является проблема выбора диагностических признаков. Исследованию этой проблемы посвящены многочисленные публикации, в частности, работы [1-11]. Традиционно выбор диагностических признаков производится в два этапа. На первом этапе конструируется исходное пространство признаков $x^{(N)} = (x_1, \dots, x_N)$, однако эта процедура чаще всего не поддается формализации: конструктор включает в первоначальный набор признаков любой измеряемый параметр, который по его мнению имеет некоторую диагностическую ценность для распознавания классов состояний объекта $V_1, \dots, V_M$. При этом обычно предположение о диагностической ценности признака делается лишь на основании опыта и интуиции конструктора либо эксперта в предметной области.

На второй этапе решается задача селекции признаков, которая сводится к минимизации имеющегося набора признаков. Согласно [1] основная цель этого этапа состоит в конструировании на основе исходного множества признаков подмножества из N_0 признаков ($N_0 < N$) без уменьшения достоверности диагностики. Такая цель может быть достигнута только на основе использования некоторого *формального* критерия, позволяющего оценить полезность любого признака.

На первый взгляд может показаться, что проблема формальной оценки полезности признаков не столь актуальна: можно попытаться провести обучение распознаванию классов в пространстве всех имеющихся признаков и получить приемлемое решающее правило. Однако на самом деле это не так, поскольку необоснованное увеличение размерности пространства признаков может ухудшить качество обучения, особенно при малом объеме обучающей выборки наблюдений [12]. К тому же, использование в решающем правиле бесполезных признаков, измерение которых в ряде случаев сопряжено с большими затратами, приводит к необоснованному увеличению стоимости диагностической системы.

В связи с этим возникает естественный вопрос: каким образом оценить полезность любого предварительно выбранного признака при решении конкретной прикладной задачи диагностики? Настоящая статья посвящена исследованию этого вопроса.

Критерии диагностической ценности признака

Рассмотрим различные подходы к определению ценности признаков при решение задачи диагностики в статистической постановке. В соответствии с этой постановкой будем считать классы $V_1, \dots, V_M$ случайными событиями с априорными вероятностями $P(V_k)$, $\sum_{k=1}^{M} P(V_k) = 1$, а совокупность признаков $x^{(N)} = (x_1, \dots, x_N)$ - случайным вектором с условными распределениями в классах $p(x^{(N)} | V_k)$, $k = 1,\dots,M$. При этом множества

$$X_k^{(N)} = \{ x^{(N)} : p(x^{(N)} / V_k) > 0 \}, \qquad k = 1,\dots,M \tag{1}$$

образуют собой собственные области классов в пространстве признаков $x_1, \dots, x_N$.

При таких исходных предпосылках качество системы диагностики в общем случае может быть оценено математическим ожиданием потерь от ошибочной диагностики (средним риском) [13]. Для заданной совокупности признаков средний риск зависит от решающего правила (алгоритма диагностики) и определяется выражением[1]

$$R(\delta) = \sum_{x^{(N)} \in X^{(N)}} \sum_{k=1}^{M} L(k, j = \delta(x^{(N)})) P(V_k) p(x^{(N)} / V_k) , \tag{2}$$

где $k = 1,\dots,M$ - номера классов, формируемых для текущего значения вектора $x^{(N)}$ алгоритмом диагностики $\delta(.)$, $L(k, j)$ - потеря (штраф) при отнесении вектора, порожденного классом V_k, к классу V_j, а $X^{(N)} = \bigcup_{k=1}^{M} X_k^{(N)}$ - множества возможных значений $x^{(N)}$.

В частном случае, когда штраф равен нулю при правильной диагностике и единице при любой неправильной, то, как известно, риск (2) сводится к средней вероятности ошибочных решений. При этом, если проводить диагностику на основании правила максимума апостериорных вероятностей

$$P(V_k / x^{(N)}) = P(V_k)\, p(x^{(N)} / V_k) \left[\sum_{i=1}^{M} P(V_i)\, p(x^{(N)} / V_i)\right]^{-1} \tag{3}$$

ISSN 1028-9763. Математические машины и системы, 1998, № 1

т.е. если

$$\delta\ (x^{(N)}) = \arg\max_{1\le k\le M} P(V_k / x^{(N)}),$$

то средняя вероятность ошибочных решений будет минимальна и равна величине

$$P_N(e) = \sum_{x^{(N)}\in X^{(N)}} p(x^{(N)})\ [1-\ \max_k P(V_k / x^{(N)})]. \qquad (4)$$

Пусть x_n , $n = 1,...,\ N$ - один из имеющихся признаков с условными распределениями в классах $p(x_n / V_k)$, k = 1,...,M. При этом предполагается, что множества $X_{nk} = \{x_n : p(x_n / V_k) > 0\}$, k = 1,...,M заведомо пересекаются, поскольку в противном случае уже по одному такому признаку может быть проведено безошибочное распознавание и статистическая постановка задачи теряет смысл.

В работе [2] дано следующее определение.

Определение 1. Признак x_n называется *нерелевантным*, если апостериорная вероятность k-го класса сохраняется неизменной при всех возможных значений x_n, т. е.

$$P(V_k / x_n) = P(V_k) \qquad \forall x_n \in X_n, \qquad (5)$$

где $X_n = \bigcup_{k=1}^{M} X_{nk}$.

Из работы [6], в которых задача распознавания рассматривается с позиций теории информации, следует такое определение.

Определение 2. Признак x_n *неинформативен*, если количество информации, содержащееся в множестве X_n его возможных значений относительно множества классов $V = \{V_1,...,V_M\}$ равно нулю, т.е. если

$$H(V) - H(V / X_n) = 0, \qquad (6)$$

где $H(V) = -\sum_{k=1}^{M} P(V_k)\log_2 P(V_k)$ - начальная (априорная) энтропии множества V, а

а $H(V\ /\ X) = -\sum_{x_n\in X_n} p(x_n) \sum_{k=1}^{M} P(V_k / x_n)\log_2 P(V_k / x_n)$ - средняя условная энтропия множества V.

Введем еще одно определение, непосредственно связанное со средней вероятностью ошибочных решений.

Определение 3. Признак x будем называть *бесполезным*, если средняя вероятность ошибочных решений

$$P_n(e) = \sum_{x_n\in X_n} p(x_n)\ [1 - \max_k P(V_k / x_n)], \qquad (7)$$

принимаемых по правилу максимуму апостериорных вероятностей $P(V_k | x_n)$, равна вероятности ошибочного решения

$$P_0(e) = 1 - \max_k (P(V_k)\ , \qquad (8)$$

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

$$P_n(e) = P_0(e). \qquad (9)$$

Если признак не удовлетворяет какому-либо из этих определений, то будем его называть соответственно релевантным, информативным или полезным признаком.

Ограничимся далее рассмотрением случая двух классов (M=2), к которому сводится, например, диагностика штатной и нештатной (аварийной) ситуации в объекте.

В соответствии с [2] при M=2 признак x_n будет нерелевантным в том и только в том случае, когда он имеет одинаковые условные распределения в классах:

$$p(x_n \,|V_1) \equiv p\,(x_n \,|V_2). \tag{10}$$

Отсюда легко показать, что Определения 1 и 2, несмотря на различие их формулировок, абсолютно эквивалентны: признак нерелевантен, когда он неинформативен и наоборот.

Совершенно очевидно также, что при выполнении условия (10) признак оказывается бесполезным в смысле Определения 3, т.е. всякий нерелевантный (а значит и неинформативный) признак является заведомо бесполезным.

В то же время в общем случае обратное утверждение неверно: как будет показано ниже при определенных условиях релевантный (а значит информативный) признак, который имеет различные распределения в классах

$$p\,(x_n \,|V_1) \neq p\,(x_n \,|V_2), \tag{11}$$

может оказаться бесполезным в смысле (9), несмотря на то, что в этом случае непременно при различных значениях $x_n \in X_n$ апостериорные вероятности классов изменяются.

Условия бесполезности информативного признака

Для того, чтобы показать возможность того, что информативный (релевантный) признак может быть бесполезным, нам понадобится следующая лемма.

Лемма. Пусть $P(V_1) \neq P(V_2)$. Тогда признак x является бесполезным в смысле (9) в том и только в том случае, когда

$$[P\,(V_1) - P(V_2)]\,[P\,(V_1 / x_n) - P\,(V_2 / x_n)] > 0 \quad \forall \;\; x_n \in X_n, \tag{12}$$

т. е. для каждого возможного значения $x_n \in X_n$ знак разности апостериорных вероятностей

$$\Delta\,(x_n) = P\,(V_1 / x_n) - P\,(V_2 /x_n) \tag{13}$$

совпадает со знаком разности априорных вероятностей

$$\Delta_0 = P\,(V_1) - P\,(V_2)\,. \tag{14}$$

Доказательство. Предположим, что неравенство (12) выполняется, например, что $P\,(V_1) > P\,(V_2)$ и $P(V_1 / x_n) > P(V_2 / x_n)$ для всех возможных $x_n \in X_n$. Это значит, что $\min\,\{P\,(V_1),\ P\,(V_2)\} = P\,(V_2)$ и $\min\,\{P\,(V_1 | \,x_n),\ P\,(V_2 | \,x_n)\} = P(V_2 | \,x_n)$ для всех возможных $x_n \in X_n$. Следовательно, если (12) справедливо, то, в соответствии с выражением (7) при $M = 2$ имеем

$$P_n\,(e) = \sum_{x_n \in X_n} p\,(x_n)\,P\,(V_2 | \,x_n) = P\,(V_2) = \min\{P\,(V_1),\,P\,(V_2)\} = P_0\,(e).$$

Предположим теперь, что (12) не выполняется, например, что $P\,(V_1) > P\,(V_2)$, но существует подмножество $X^* \in X_n$ значений x_n таких, что $P\,(V_1 \,|\, x_n) < P\,(V_2 | \,x_n)$ для всех $x_n \in X^*$. В этом случае в соответствии с (7) имеем

$$P_n\,(e) = \sum_{x_n \in X_n \setminus X^*} p\,(x_n)\,P\,(V_2 / \,x_n) + \sum_{x_n \in X^*} p\,(x_n)\,P\,(V_1 /x_n). \tag{15}$$

Заменяя в правой части (15) $P\,(V_1 / x_n)$ на $P\,(V_2 / x_n)$ с учетом того, что $P(V_1/x_n) < P(V_2/x_n)$ $\forall\, x_n \in X^*$ заключаем, что если $X^* \neq \varnothing$, то

$$P_n\,(e) < \sum_{x_n \in X_n} p\,(x_n)\,P\,(V_2 | \,x_n),$$

т. е. $P_n\,(e) < P_0\,(e)$. Лемма доказана.

Как видно условие леммы совершенно не предполагает выполнение равенства (3). Следовательно условие (12) может выполняться не только в том тривиальном случае, когда признак x_n одинаково распределен в классах, но и в общем случае, когда при возможных изменениях x_n апостериорные вероятности $P\ (V_1\,|x_n)$ и $P\ (V_2\,|x_n)$ изменяются, но сохраняется знак разности (13). Заметим, что такая ситуация возможна только в том случае, когда различны априорные вероятности классов, т. е. $P\ (V_1) \neq P\ (V_2)$.

Рассмотрим отдельно два случая, представляющие интерес для задач медицинской и технической диагностики.

Случай 1. Признак-симптом, принимающий два значения

Пусть x_n - дискретный признак (симптом), принимающий только два возможных значения $x_n = x^+$ и $x_n = x^-$ с условными вероятностями $p(x^+|V_k)$, $p(x^-|V_k)$, k=1,2. Очевидно, что в этом случае $p(x^-|V_k) = 1- p(x^+|V_k)$.

Теорема 1. Если априорные вероятности классов неравны, т.е. $P(V_1)\ ^1P(V_2)$, то для любой заданной условной вероятности $p\ (x^+|V_1)$ всегда может быть указана условная вероятность $p(x^+|V_2)$, такая, что

$$p(x^+|V_1) \neq p(x^+|V_2) \tag{16}$$

и при этом выполняется равенство (9), а значит x_n будет информативным, но бесполезным признаком.

Доказательство. Предположим вначале, что $P(V_2) < P(V_1)$. Введем в рассмотрение множество W возможных значений $p(x^+|V_1)$ и $p(x^+|V_2)$, которые удовлетворяют следующим неравенствам

$$p(x^+/V_1) > l_0\, p(x^+/V_2)\ , \tag{17}$$

$$p(x^+/V_1) < l_0\, p(x^+/V_2) + (1- l_0), \tag{18}$$

где $l_0 = P(V_2)/P(V_1)$ - отношение априорных вероятностей классов.

Из (17), (18) легко видно, что если $P(V_2) < P(V_1)$, то $\Omega \neq \varnothing$, поскольку в этом случае $l_0 < 1$. Следовательно для заданной вероятности $p(x^+|V_1)$ всегда можно указать соответствующую вероятность $p(x^+/V_2)$, которая удовлетворит совместно (16) - (18).

Тогда, принимая во внимание, что $p(x_n) = P(V_1)p(x_n/V_1) + P(V_2)p(x_n/V_2)$ при $x_n = x^+$ и $x_n = x^-$, а $p(x^-/V_k) = 1 - p(x^+/V_k)$ для k = 1,2 из (17) и (18) по формуле Байеса следует, что $p(V_1/x^+) > p(V_2/x^+)$ и $p(V_1/x^-) > p(V_2/x^-)$. Но это свидетельствует о том, что выполняется условие (12), а значит в соответствии с Леммой выполняется равенство (9).

Доказательство Теоремы 1 для случая $P(V_2) > P(V_1)$ выполняется аналогичным образом, но в этом случае вместо (17), (18) рассматриваются неравенства

$$p(x^+/V_1) < l_0\, p(x^+/V_2), \tag{19}$$

$$p(x^+/V_1) > l_0\, p(x^+/V_2) + (1-l_0), \tag{20}$$

которые определяют множество $\Omega \neq \varnothing$ при $l_0 > 1$.

Следствие 1. Если априорные вероятности классов равны, т.е. $P(V_1) = P(V_2)$, а признак информативен, т.е. $p(x^+|V_1) \neq p(x^+|V_2)$, то он заведомо полезен. Этот факт непосредственно вытекает из (17), (18) или (19), (20), поскольку при l_0 = 1 непременно $\Omega = \varnothing$.

На рис. 1 показаны границы областей W бесполезности информативного признака при различных значениях l_0 . Как видно область Ω уменьшается по мере приближения l_0 к единице, и $\Omega = \varnothing$, когда l_0 = 1.

Пример. Пусть $P(V_1) = 0.8$; $P(V_1) = 0.2$. Тогда $l_0 = 0.25$.

Предположим, что $p(x^+/V_1) = 0.7$. Выберем соответствующее значение $p(x^+/V_2) = 0.2$, при котором точка $P = (0.2,\ 0.7)$ попадает в область Ω , показанную на рис.1, а). Легко видно, что признак x_n имеет различные распределения в классах: $p(x^+/V_1) \neq p(x^+ V_2)$, а значит x_n релевантный и информативный признак.

В то же время по формуле Байеса мы имеем $p(V_1/x^+) = 0.933$; $p(V_2/x^+) = 0.067$; $p(V_1/x^-) = 0.6$; $p(V_1/x^-) = 0.4$. Значит $p(V_1/x^+) > p(V_2/x^+)$ и $p(V_1/x^-) > p(V_2/x^-)$, т. е. при любом значении признака решение принимается в пользу класса V_1. Поэтому $P_n(e) = P_0(e) = 0.2$, а значит этот признак бесполезен в смысле Определения 3.

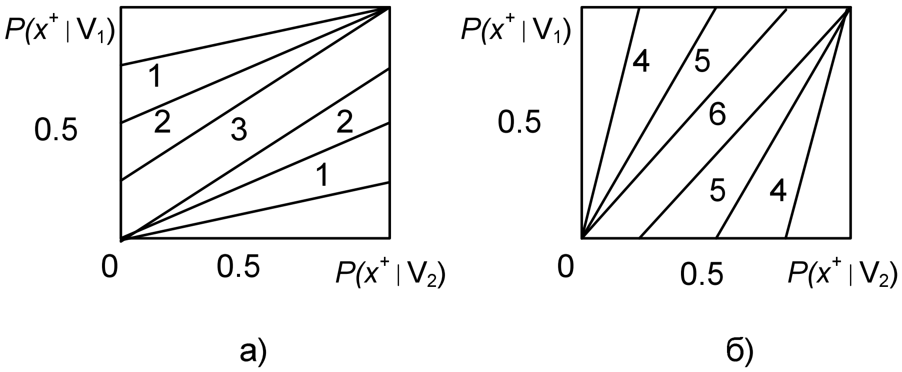

Рис.1. Области Ω бесполезности информативного признака
1 - $\lambda_0 = 0.25$; 2 - $\lambda_0 = 0.5$; 3 - $\lambda_0 = 0.75$;
4 - $\lambda_0 = 4.0$; 5 - $\lambda_0 = 2.0$; 6 - $\lambda_0 = 1.33$.

На основании полученных результатов может быть сформулирована следующая теорема, которая определяет необходимые и достаточные условия полезности диагностического признака-симптома.

Теорема 2. Любой дискретный признак-симптом, принимающий два возможных значения, будет полезен в смысле Определения 3 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 $\lambda_0 > 1$.

Случай 2. Непрерывный признак, нормально распределенный в классах

Теорема 3. Пусть выполняется условие $P(V_1) > P(V_2)$, а признак x_n имеет нормальные распределения в классах

$$p(x_n/V_1) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[\frac{-(x_n-a_1)^2}{2\sigma_1^2}\right], \tag{21}$$

$$p(x_n/V_2) = \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left[\frac{-(x_n-a_2)^2}{2\sigma_2^2}\right], \tag{22}$$

с математическими ожиданиями a_1, a_2 и различными дисперсиями σ_1^2, σ_2^2.

Тогда, если выполняются условия

$$\sigma_1 > \sigma_2, \tag{23}$$

$$\lambda_0 < \frac{\sigma_2}{\sigma_1}\exp\left[-\frac{(a_1-a_2)^2}{2(\sigma_1^2-\sigma_2^2)}\right], \tag{24}$$

где $\lambda_0 = P(V_2)/P(V_1)$, то x_n - релевантный (информативный), но бесполезный признак.

Доказательство. Прежде всего заметим, что в силу (23) условие (10) не выполняется, а значит x_n - релевантный (информативный) признак.

Поскольку $P(V_1) > P(V_2)$, то в соответствии с условием (14) Леммы признак x_n будет бесполезным в том и только в том случае, когда

$$p(x_n/V_1) > \lambda_0 p(x_n/V_2), \quad \forall x_n \in X_n. \tag{25}$$

Подстановка распределений из (21), (22) в (25) дает

$$\exp\left[\frac{(x_n-a_2)^2}{2\sigma_2^2}-\frac{(x_n-a_1)^2}{2\sigma_1^2}\right] > \lambda_0\frac{\sigma_1}{\sigma_2}, \quad \forall x_n \in X_n, \tag{26}$$

Логарифмируя левую и правую часть (26) получим соотношение

$$\frac{(x_n - a_2)^2}{2\sigma_2^2} - \frac{(x_n - a_1)^2}{2\sigma_1^2} \quad > \quad \ln[\lambda_0 \frac{\sigma_1}{\sigma_2}], \ \forall x_n \in X_n, \tag{27}$$

которое после очевидных преобразования можно представить в виде

$$[\sigma_1^2 - \sigma_2^2]\, x_n^2 + [2(a_1\sigma_2^2 - a_2\sigma_1^2)]\, x_n +$$
$$+ a_2^2\sigma_1^2 - a_1^2\sigma_2^2 - 2\sigma_1^2\sigma_2^2 \ln(l_0\, \sigma_1/\sigma_2)] > 0, \quad \forall x_n \in X_n. \tag{28}$$

Таким образом в рассматриваемом случае условие бесполезности признака x_n сводится к условию положительности квадратного трехчлена, фигурирующего в левой части (28), при любом $x_n \in X_n$. Поскольку по условию теоремы $\sigma_1 \neq \sigma_2$, то это возможно в том и только в том случае, когда

$$s_1^2 - \sigma_2^2 > 0 \tag{29}$$

и, кроме того

$$4(a_1\sigma_2^2 - a_2\sigma_1^2)^2 - 4(\sigma_1^2 - \sigma_2^2)\,[a_2^2\sigma_1^2 - a_1^2\sigma_2^2 - 2\sigma_1^2\sigma_2^2 \ln(l_0\, \sigma_1/\sigma_2)] < 0. \tag{30}$$

Из (29) немедленно следует (23), а условие (30) после очевидных упрощений сводится к неравенству

$$\ln(\lambda_0 \frac{\sigma_2}{\sigma_1}) < -\frac{(a_1 - a_2)^2}{2(\sigma_1^2 - \sigma_2^2)}, \tag{31}$$

которое, с учетом того, что exp (.) > 0, эквивалентно условию (24). Теорема доказана.

Следствие 2. При выполнении условия $P(V_1) > P(V_2)$ признак, имеющий нормальные распределения в классах с одинаковыми математическими ожиданиями и среднеквадратическими отклонениями $\sigma_1 > \sigma_2$, будет релевантным (информативным), но бесполезным, если отношение среднеквадратических отклонений превышает отношение соответствующих априорных вероятностей классов, т.е.

$$\sigma_1/\sigma_2 > l_0.$$

Следствие прямо вытекает из (24), если положить $a_1 = a_2$.

Заметим также, что при $\sigma_1 = \sigma_2$ квадратный трехчлен (28) переходит в линейную функцию x_n, которая не может быть положительна для любого $x_n \in X_n$, если $a_1 = a_2$. Отсюда немедлено вытекает такое следствие.

Следствие 3. Всякий информативный признак, имеющий нормальные распределения в классах с одинаковыми дисперсиями, заведомо полезен.

Информационный критерий полезности признака

Как было показано выше информативность признака является лишь необходимым, но не достаточным условием его полезности в смысле Определения 3.

С другой стороны легко убедиться в том, что в рассматриваемом случае диагностики двух классов при любом фиксированном значении $x_n \in X_n$ условная вероятность ошибки

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

однозначно определяет частную условную энтропию

$$h(V/x_n) = -P(e/x_n)\log P(e/x_n) - [1 - P(e/x_n)]\log[1 - P(e/x_n)].$$

В то же время известно [14], что не существует однозначной связи между средней вероятностью ошибки

$$P_n(e) = \sum_{x_n \in X_n} p(x_n) P(e/x_n)$$

и средней условной энтропией

$$H(V/X_n) = -\sum_{x_n \in X_n} p(x_n)\, h(V/x_n).$$

В связи с этим возникает естественный вопрос : можно ли вообще судить о полезности признака с точки зрения уменьшения вероятности ошибочной диагностики по количеству информации, содержащемуся в множестве его возможных значений?

Ответ на этот вопрос дает следующая теорема.

Теорема 4. Любой информативный признак является заведомо полезным, если количество информации, содержащееся в множестве его возможных значений превышает информационный барьер, равный

$$I_0 = \log_2(1+\lambda_0) - \lambda_0(1+\lambda_0)^{-1}\log_2\lambda_0 - 2\min\{(1+\lambda_0)^{-1}, \lambda_0(1+\lambda_0)^{-1}\}, \quad (32)$$

т.е. если

$$H(V) - H(V|X_n) > I_0, \quad (33)$$

где $\lambda_0 = P(V_2)/P(V_1)$ - отношение априорных вероятностей классов.

Доказательство. Для доказательства воспользуемся известным из [14] выражением для точной нижней границы средней условной энтропии $H(V|X_n)$ при заданной средней вероятности P(e), которое при числе классов M=2 может быть представлено в виде

$$\inf H(V/X_n) = 2\,P(e). \quad (34)$$

Предположим, что соотношение (33) выполняется. Тогда поскольку $P_0(e) = \min(P(V_1), P(V_2)\}$, то с учетом (32) это соотношение в конечном счете можно представить так

$$H(V/X) < 2\,P_0(e). \quad (35)$$

Покажем, что в этом случае $P(e) < P_0(e)$. Для этого предположим противное, т.е. что $P(e) = P_0(e)$ (случай $P(e) > P_0(e)$ вообще невозможен). При таком предположении в соответствии с (34) имеем $\inf H(V/X) = 2P_0(e)$, но это противоречит (35). Полученное противоречие подтверждает справедливость сформулированной теоремы.

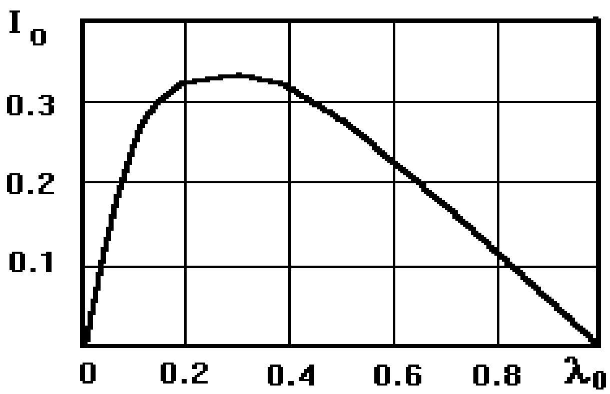

Рис. 2. Зависимость между информационным порогом I_0 и λ_0.

Для иллюстрации на рис.2 показана зависимость информационного барьера I_0 от отношения априорных вероятностей классов λ_0 при $0 \le \lambda_0 \le 1$. Если учесть, что в соответствии с (32) $I_0(\lambda_0) = I_0(1/\lambda_0)$, то этой же зависимостью можно пользоваться при $\lambda_0 > 1$, поскольку в этом случае $0 \le (\lambda_0)^{-1} \le 1$.

Следствие 4. Если априорные вероятности классов равны, т.е. $P(V_1) = P(V_2)$, то всякий информативный признак является полезным.

Следствие прямо вытекает из (33) с учетом (32) в силу того, что I_0 =0 при λ_0=1.

Примечательно, что такой же результат эквивалентности критериев информативности и полезности признака при $P(V_1) = P(V_2)$ был получен выше другим путем.

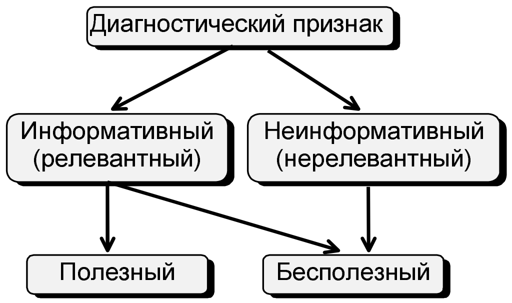

Рис.3. Взаимосвязь критериев информативности (релевантности) и полезности признака при

$P(V_1) \neq P(V_2)$

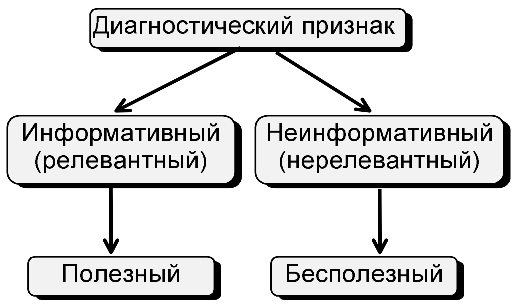

Рис.4. Взаимосвязь критериев информативности (релевантности) и полезности признака при

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

Заключение. В статье показано, что информативность признака является необходимым, но не достаточным условием его полезности с точки зрения уменьшения вероятности ошибочной диагностики двух классов. Следовательно в общем случае, когда $P(V_1) \neq P(V_2)$ всякий полезный в соответствии с Определением 3 признак является релевантным и информативным, но не наоборот (см. рис. 3).

Показано также, что для гарантированной полезности признака при $P(V_1) \neq P(V_2)$ достаточно, чтобы количество информации, содержащееся в множестве возможных значений информативного признака, превышало информационный барьер, который согласно (32) определяется лишь отношением априорных вероятностей классов.

На примерах дискретного признака с двумя возможными градациями (признака-симптома) и признака, нормально распределенного в классах продемонстрирована возможность существования информативного, но бесполезного признака.

В то же время оказалось, что если диагностируемые классы равновероятны, то критерии информативности (релевантности) признака и его полезности являются эквивалентными (рис. 4).

Полученные результаты нашли применение при разработке компьютерных систем диагностики в технических и медицинских приложениях [15,16].

СПИСОК ЛИТЕРАТУРЫ

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. - pp.279-283.
2. Ben-Bassat M. Irrelevant Features in Pattern Recognition. - IEEE Trans. Comput. - vol.C-27. - 1978. - N8. - pp.749-766.
3. Kittler J. Feature Selection and Extraction. - Handbook of Pattern Recognition and Image Processing. - 1986.- 450p.
4. Narendra P.M., Fukunaga K.A. Branch and Bound Algorithm for Feature Subset Selection. - IEEE Trans. Comp. - vol.C-26. - Sept. 1977.- pp.917-922.
5. Siedlecki W., Sklansky J. - On Automatic Feature Selection. - Intern. Journal of Pattern Recognition and Artificial Intelligence. - vol.2. - 1988. - pp.197-220.
6. Lewis P.M. The Characteristic Selection Problem in Recognition System. - IRE Trans. Inform. Theory. - vol.8. - N2. - 1962. - pp.171-178.
7. Tou J.T., Heydorn R.P. Some Approaches to Optimum Feature Selection. - In Comp. and Inform. Sci. Acad. Press. - N4. - vol.11. - 1967. - pp.57-89.
8. Stearns S.D. On Selecting Features for Pattern Classifiers. - Proc. Of the Third Int. Conf. on Pattern recognition. - Coronado. - CA. - 1976. - pp.71-75.
9. Барабаш Ю.Л. Учет свойств признаков при распознавании.- Изв. АН СССР. Техн. кибернетики.- 1965. - № 5. - С.85-92.
10. Биргер И.А. Определение диагностической ценности признаков. - Кибернетика. - 1968. -№ 3. - С. 80-85.
11. 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. - pp.544-546.
12. Загоруйко Н.Г. Методы распознавания и их применение.- М.- Сов. Радио.- 1972.- 208 с.
13. Гимельфарб Г.Л. Риск распознавания. -Энциклопедия кибернетики.- т.2. 1974.- С.297-298.
14. В.А. Ковалевский, Методы оптимальных решений в распознавании изображений. -М. - Наука. -1976. - 328 с.
15. Fainzilberg L.S. Method and Device for Discriminating Thermal Effect of Phase Transformation of Metals and Alloys in the Process of their Cooling.- Патент (США). - N4198679. - 1980. - 78 с.
16. Fainzilberg L.S., Potapova T.P. Computer Analysis and Recognition of Cognitive Phase Space Electro-Cardio Graphics Image. -Proc. of the 6-th Int. Conf. on Computer Analysis of Images and Patterns.- Prague.- 1995. - pp. 668-673.