Опубиковано в журнале «Автоматика». – 1985. – № 5. – С. 78-82.

УДК 658.012.011

ОБ ОЦЕНКЕ ПОЛЕЗНОСТИ ПРИЗНАКОВ ПРИ СТАТИСТИЧЕСКОМ РАСПОЗНАВАНИИ ДВУХ КЛАССОВ

Л.С. Файнзильберг, Г.А. Шкляр

При решении задач распознавания довольно часто возникает необходимость в оценке полезности признаков $x_1,...,x_N$, составляющих описание $x=(x_1,...,x_N)$ неких образов (классов) V_1, V_2 [1-5]. Применительно к задаче распознавания в статистической постановке, как известно, наиболее подходящей оценкой полезности любого признака x_n $(1 \le n \le N)$ может служить изменение средней вероятности ошибочных решений

$$P(e) = P(V_1)\int\limits_{\Omega_2} p(x/V_1)dx + P(V_2)\int\limits_{\Omega_1} p(x/V_2)dx \qquad (1)$$

при переходе от исходного ($N-$ мерного) описания $x^{(N)}$ к сокращенному ($(N-1)$ – мерному) описанию $x^{(N-1)}$, не содержащему этот признак x_n. Здесь $P(V_m), m=1,2$ – априорные вероятности классов, $p(x/V_m)$ – условные распределения признаков в классах, а $\Omega_1 = \{x: P(V_1/x) > P(V_2/x)\}$ и $\Omega_2 = \{x: P(V_2/x) > P(V_1/x)\}$ – области оптимальных байесовых решений. При этом следует различать полезность признака x_n самого по себе и его полезность в описании.

Определение 1. Признак x_n полезен сам по себе, если

$$P_1(e) < P_0(e), \qquad (2)$$

где $P_1(e)$ – вероятность ошибочных решений, принимаемых по максимуму апостериорных вероятностей $P(V_1/x_n), P(V_2/x_n)$, а

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

– априорная вероятность ошибки.

Определение 2. Признак x_n полезен в описании $x^{(N)} = (x_1, \ldots, x_N)$ при $N \ge 2$, если для вероятностей ошибочных решений, принимаемых по исходному и сокращенному описаниям, справедливо строгое неравенство

$$P_N(e) < P_{N-1}(e). \qquad (3)$$

Следует обратить внимание, что о полезности признака x_n в описании $x^{(N)}$ в смысле (3) можно говорить лишь в случае, когда заведомо

$$P_N(e) < P_0(e), \qquad P_{N-1}(e) \neq 0, \qquad (4)$$

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

Разумеется, в том случае, когда имеются исчерпывающие сведения об априорных вероятностях $P(V_m)$ и многомерных условных распределениях $p(x^{(N)}/V_m)$, то в конце концов можно прямо проверить выполнение неравенства (3), непосредственно определив $P_N(e)$ и $P_{N-1}(e)$. Но дело в том, что случай, когда имеется столь полная априорная информация, является скорее редким исключением, чем правилом. Обычно же при решении прикладных задач распознавания имеются лишь ограниченные априорные сведения, которые и хотелось бы использовать для оценки полезности признаков.

В настоящей статье поставлена задача получения условий, гарантирующих выполнение строгого неравенства (3) на основе ограниченных априорных сведений о вероятностных характеристиках признаков. Формулировке утверждения, определяющего эти условия, предпошлем следующие леммы.

Лемма 1. Для того, чтобы при исключении из описания $x^{(N)} = (x_{1,\ldots,}x_N)$ признака x_n вероятность $P(e)$ сохранялась неизменной, т.е. чтобы $P_N(e) = P_{N-1}(e)$, необходимо и достаточно, чтобы $\Omega_m^{(N)} = (\Omega_m^{(N-1)} \times X_n) \cap X^{(N)})$, $m = 1,2$, где $\Omega_m^{(N)}, \Omega_m^{(N-1)}$- области оптимальных решений соответственно в N–мерном и $(N-1)$ –мерном пространствах признаков, а

$$X^{(N)} = \left\{x^{(N)} : p)x^{(N)} \neq 0 \right\}, X_n = \left\{x_n : p(x_n) \neq 0 \right\}$$

– соответственно множества возможных значений $x^{(N)}$ *и* x_n.

Доказательство леммы 1 ввиду ограниченного объема статьи не приводится.

Лемма 2. Пусть $\xi = (\xi_{1,\ldots,} \xi_N)$ $-N-$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 $\theta_1 \in \Theta, \theta_2 \in \Theta$

$$p_{\xi}(x,\theta_2) \equiv p_{\xi}(x-\delta,\theta_1), \qquad (5)$$

где $\delta \equiv (\delta_{1,} \ldots, \delta_N) = \delta(\Delta\theta)$ – 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

$$M_{\theta_2}\xi = M_{\theta_1}\xi + \delta, \qquad (6)$$

б) центральные смешанные моменты величин $\xi_1, \ldots, \xi_N$ всех порядков не зависят от параметра θ, т.е. для любых натуральных чисел $k_1, \ldots, k_N$ выполняется равенство

$$G_1(\xi) = G_2(\xi), \qquad (7)$$

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

На основании сформулированных лемм может быть доказано следующее утверждение, определяющее достаточные условия полезности признака x_n в описании $x^{(N)}$.

Утверждение. Пусть при ограничениях (4) выполняются следующие условия:

а) признак x_n имеет одинаковые распределения в классах, т.е.

$$p(x_n / V_1) \equiv p(x_n / V_2); \qquad (8)$$

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

$$p(x^{(N)} / V_1) \equiv p(x^{(N-1)} / V_1)\, p(x_n / V_1), \qquad (9)$$

$$M\{x^{(N)} / V_2\} \neq M\{x^{(N-1)} / V_2\} M\{x_n / V_2\}, \qquad (10)$$

причем эта связь такова, что корреляционные моменты

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

при всех $k_1 = 0,1,\ldots$ не зависят от x_n;

в) условные распределения $p(x^{(N)}/V_m), m=1,2$ являются непрерывными функциями, производные от которых по любому направлению отличны от нуля за исключением, быть может, изолированных точек, а собственные области классов $X_m^{(N)} = \left\{x^{(N)} : p(x^{(N)}/V_m) \neq 0 \right\}$ представляют собой односвязные множества.

При выполнении этих условий x_n заведомо полезный признак в описании $x^{(N)}$ в смысле строгого неравенства (3).

Заметим, что в силу условия (8) признак x_n сам по себе бесполезен в смысле определения (1). Вместе с тем такой признак оказывается полезным в описании, если выполняются все остальные условия сформулированного утверждения.

Следует также обратить внимание, что проверка условий, фигурирующих в утверждении, вовсе не предполагает знание самих законов распределений признаков в классах. Как показано в работе [5], подобного рода ограниченные сведения о распределениях в ряде случаев могут быть получены на основе известных физических представлений о распознаваемых классах. В тех же случаях, когда правомерна гипотеза о нормальном законе распределения признаков в классах, разумно использовать следствие, вытекающее из сформулированного утверждения.

Следствие. Пусть описание $x^{(N)} = (x_1, \ldots x_N)$ имеет невырожденное многомерное нормальное распределение в классах V_1 *и* V_2. Тогда, если: а) признак x_n $(1 \leq n \leq N)$ 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.

$$M\{x_n/V_1\} = M\{x_n/V_2\}, \tag{11}$$

$$D\{x_n/V_1\} = D\{x_n/V_2\}, \tag{12}$$

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

Заключение. В статье показано, что при наличии статистической зависимости между признаками существуют условия, при которых бесполезный сам по себе признак оказывается полезным в совокупности с другими признаками. Это еще раз свидетельствует о том, насколько важно в каждом конкретном случае исследовать статистическую связь между признаками, прежде чем решать вопрос об их полезности в описании. При этом оказывается, что в случае нормальных законов распределений оценка полезности в описании признака, который сам по себе бесполезен, сводится к проверке достаточно простых условий.

П Р И Л О Ж Е Н И Е

Доказательство леммы 2. Необходимость доказывается непосредственной проверкой. Для доказательства достаточности применим формулу бинома Ньютона и, проведя ряд преобразований, представим выражение для G_j в виде

$$G_j(\xi) = \sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} a^{(j)}_{i_1 \dots i_N}(\xi) \int \cdot \underset{R^N}{\cdot} \cdot \int x_1^{i_1} \dots x_N^{i_N} \, p_\xi(x, \theta_j) dx,$$

где

$$a^{(j)}_{i_1 \dots i_N}(\xi) = (-1)^L \begin{pmatrix} i_1 \\ k_1 \end{pmatrix} \dots \begin{pmatrix} i_N \\ k_N \end{pmatrix} (M_{\theta_j} \xi_1)^{k_1 - i_1} \dots (M_{\theta_j} \xi_N)^{k_N - i_N} ,$$

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

$\begin{pmatrix} i_l \\ k_l \end{pmatrix} = \dfrac{k_l!}{i_l!(k_l - i_l)!}$, $l = \overline{1, N}$ – число сочетаний из k_l по i_l.

Из условия (6) следует, что $a^{(2)}_{i_1 \dots i_N}(\xi) = a^{(1)}_{i_1 \dots i_N}(\xi + \delta)$. 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 (7) примет вид:

$$\sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} a^{(2)}_{i_1 \dots i_N}(\xi) \int \cdot \underset{R^N}{\cdot} \cdot \int x_1^{i_1} \dots x_N^{i_N} \, p_\xi(x, \theta_2) dx =$$

$$= \sum_{i_1=0}^{k_1} \dots \sum_{i_N=0}^{k_N} a^{(2)}_{i_1 \dots i_N}(\xi) \int \cdot \underset{R^N}{\cdot} \cdot \int x_1^{i_1} \dots x_N^{i_N} \, p_\xi(x - \delta, \theta_1) dx.$$

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

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

или в эквивалентной форме записи

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

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 $x_1^{k_1} \dots x_N^{k_N}$. Отсюда, в силу полноты системы полиномов вида $x_1^{k_1} \dots x_N^{k_N}$ в пространстве непрерывных функций [6] следует, что $g(x) \equiv 0$.

Но последнее соотношение как раз и эквивалентно выполнению равенства (5). Лемма доказана.

Доказательство утверждения. Поскольку для $P_{N-1}(e)$ всегда справедлива оценка $P_{N-1}(e) \le P_0(e)$, то в силу первого из ограничений (4) при $P_{N-1}(e) = P_0(e)$ выполнение неравенства (3) следует немедленно. Поэтому остается показать справедливость сформулированного утверждения лишь для случая, когда $P_{N-1}(e) < P_0(e)$. Легко показать, что в этом случае в силу второго из ограничений (4) и условия в) существует граница между областями решений в пространстве $X^{(N-1)}$, т.е. $\exists X_0^{(N-1)} \subset X^{(N-1)}$ такое, что

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

Принимая во внимание условия (8), (9) представим выражение для отношения правдоподобия $\lambda(x^{(N)}) \overset{\Delta}{=} \dfrac{p(x^{(N)}/V_1)}{p(x^{(N)}/V_2)}$ в точке $x_0^{(N-1)} \in X_0^{(N-1)}$ так:

$$\lambda(x_0^{(N-1)}, x_n) = \lambda(x_0^{(N-1)}) \frac{p(x_0^{(N-1)}/V_2)}{p(x_0^{(N-1)}/x_n, V_2)} \quad . \tag{13}$$

На основании условия б) в силу леммы 2 имеем

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

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

.

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

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

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 1 с учетом (13) имеем: $p(x_0^{(N-1)}/V_2) = p(x_0^{(N-1)}/x_n, V_2)$ $\forall x_n \in X_n$. А это значит, что $grad\ p(x_0^{(N-1)} + \gamma\delta^{(N-1)}/x_n^{(1)}, V_2)\,\delta^{(N-1)} = 0$. 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 сформулированного утверждения.

Доказательство следствия. Если

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

где $\mu_m^{(N)} = (M\{x_1 | V_m\}, \ldots, M\{x_N | V_m\})$ – вектор математических ожиданий, $D_m^{(N)} = \{M(x_i | V_m)(x_i - M\{x_j V_m\}), i, j = 1, N\}$ – 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 при выполнении условий (11), (12) будет иметь место условие (8).

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:

$$p(x^{(N)} / V_1) = (2\pi)^{-\frac{N-1}{2}} (2\pi)^{-\frac{1}{2}} \left| D_1^{(N-1)} \right|^{-\frac{1}{2}} (D\{x_n / V_1\})^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (x^{(N-1)} - \right.$$
$$\left. - \mu_1^{(N-1)})^T (D_1^{(N-1)})^{-1} (x^{(N-1)} - \mu_1^{(N-1)}) - \frac{(x_n - M\{x_n / V_1\})^2}{2D\{x_n / V_1\}} \right\} , \tag{14}$$

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. Поскольку правая часть (14) представляет собой произведение $p(x^{(N-1)} | V_1) p(x_n | V_1)$, то значит имеет место условие (9).

Легко видно, что условное распределение $p(x^{(N-1)} | x_n, V_m)$ представляет собой нормальное распределение с вектором математических ожиданий

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

и ковариационной матрицей $S_m^{(N-1)} = D_m^{(N-1)} - D_m^n (D_m^n)^T / D \times \{x_n / V_m\}$, где D_m^n- вектор, представляющий собой $n-й$ столбец матрицы $D_m^{(N-1)}$ без диагонального элемента. Так как 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 (10).

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 же согласно [7] 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) утверждения. Если к тому же учесть, что для нормальных распределений условие в) утверждения очевидно, то в силу этого утверждения признак x_n полезен в описании $x^{(N)}$.

1. *Івахненко О.Г.* Критерій числа розв'язуваних суперечок для вибору корисних ознак для розпізнаючих і передбачаючих систем.– Автоматика, 1966.– № 4.– С.26-40.
2. *Горелик А.Л.* О выборе пространства признаков системы распознавания объектов и явлений.– Кибернетика, 1980.– № 6.– С.72-75.
3. *Васильев В.И.* Конструирование пространства в процессе обучения распознаванию образов. – Автоматика, 1982.– № 5.– С.18-27.
4. *Барабаш Ю.Л.* Минимизация описаний в задаче автоматического распознавания образов.– Изв. АН СССР. Техническая кибернетика, 1964.– № 3.– С.14-23.
5. *Файнзильберг Л.С.* Применение методов статистического распознавания в термографическом анализе состава металла. – Кибернетика, 1978.– № 6.– С.159-162.
6. *Партасарати К.* Введение в теорию вероятностей и теорию меры.– М.: Мир, 1983.– 336 с.
7. *Гихман И.И., Скороход А.В.* Введение в теорию случайных процессов.– М.: Наука, 1977. – 568 с.