А. С. ФАЙНЗИЛЬБЕРГ

К ВОПРОСУ О БЕЗОШИБОЧНОМ РАСПОЗНАВАНИИ ДВУХ КЛАССОВ ПО СОВОКУПНОСТИ ПЕРЕСЕКАЮЩИХСЯ ПРИЗНАКОВ

УДК 62—50

Пусть каждый признак x_n $(1 \leqslant n \leqslant N)$ описания $x^{(N)} = (x_1, \ldots, x_N)$ имеет пересекающиеся условные распределения $p(x_n \mid V_1)$ и $p(x_n \mid V_2)$ в классах V_1 и V_2, иначе говоря, пусть множества $X_{1n} = \{x_n : p(x_n \mid V_1) > 0\}$ и множества $X_{2n} = \{x_n : p(x_n \mid V_2) > 0\}$ заведомо пересекаются:

$$X_{1n} \cap X_{2n} \neq \varnothing \text{ для всех } n = 1, \ldots, N. \quad (1)$$

Ясно, что безошибочное распознавание по каждому отдельному такому пересекающемуся в классах признаку x_n невозможно. В то же время при $N \geqslant 2$, вообще говоря, возможны случаи, когда описание $x^{(N)}$ позволяет провести безошибочное распознавание, несмотря на ограничения (1) [1]. В связи с этим возникает естественный вопрос: каким условиям должны удовлетворять пересекающиеся признаки, чтобы можно было рассчитывать на возможность безошибочного распознавания по их совокупности?

Согласно [2] при ограничениях (1) описание $x^{(N)}$ позволяет получить вероятность ошибочных решений $P_N(e) = 0$, если признаки этого описания связаны функциональной зависимостью в классах. Между тем оказывается, что для обеспечения $P_N(e) = 0$ требование наличия функциональной связи между признаками является чересчур жестким. Более того, ниже будет показано, что даже если каждый из N признаков описания $x^{(N)}$ имеет одинаковые маргинальные распределения в классах, т. е. если

$$p(x_n \mid V_1) \equiv p(x_n \mid V_2) \text{ для всех } n = 1, \ldots, N, \quad (2)$$

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

$$p(x^{(N)}) \equiv \prod_{n=1}^{N} p(x_n), \quad (3)$$

то и в этом, казалось бы совсем «безнадежном», случае все же существует возможность безошибочного распознавания.

Определим необходимые условия безошибочного распознавания по совокупности пересекающихся признаков. Но прежде чем непосредственно переходить к решению этой задачи, установим некоторые довольно любопытные свойства признаков, которые будут полезны в дальнейших рассуждениях.

Пусть x_n и x_m $(1 \leqslant n \leqslant N,\ 1 \leqslant m \leqslant N,\ n \neq m)$ — любые два признака описания $x^{(N)} = (x_1, \ldots, x_N)$. Следуя [3, 4], будем различать *безусловную* статистическую независимость x_n и x_m, когда

$$p(x_n, x_m) \equiv p(x_n)\, p(x_m), \quad (\text{4}$$

и *условную* статистическую независимость эти признаков в k-м классе, когда

$$p(x_n, x_m \mid V_k) \equiv p(x_n \mid V_k)\, p(x_m \mid V_k). \quad (\text{5}$$

Из работы [3] известно, что соотношения (4) (5) не могут выполняться совместно, если признак x_n и x_m имеют различные распределения в класса т. е. если[1]

$$p(x_n \mid V_1) \not\equiv p(x_n \mid V_2), \quad p(x_m \mid V_1) \not\equiv p(x_m \mid V_2). \quad (\text{6}$$

Тем не менее из этого результата в той форме, которой он был сформулирован в [3], еще не ясно обязательно ли из условий (4) и (6) вытекает ста тистическая зависимость x_n и x_m сразу в обоих кла сах. Поэтому понадобилось провести дополнитель ные исследования, позволившие несколько усилит формулировку этого свойства. Попутно был обн ружен целый ряд других, не менее любопытны свойств, формулировке и доказательству которы предпошлем следующее вспомогательное утверж дение.

Утверждение 1. Пусть в N-мерном пространств R^N случайных векторов $x^{(N)} = (x_1, \ldots, x_N)$ с зада ным на нем многомерным распределением $p(x^{(N}$ выделено множество

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

включающее те и только те значения $x^{(N)} \in R^N$, дл которых $p(x^{(N)})$ отлично от нуля. Тогда, для то чтобы компоненты $x_1, \ldots, x_N$ векторов $x^{(N)}$ был статистически независимы, необходимо, чтобы мн жество $X^{(N)}$ представляло собой прямое произвед ние всех своих проекций $X_1, \ldots, X_N$ соответствен на множества значений $x_1, \ldots, x_N$, т. е. чтобы

$$X^{(N)} = X_1 \times \ldots \times X_N. \quad (\text{8}$$

Доказательство. Если $x_1, \ldots, x_N$ статистическ независимы, то непременно

$$p(x^{(N)}) \triangleq p(x_1, \ldots x_N) \equiv p(x_1) \ldots p(x_N), \quad (\text{9}$$

[1] Здесь и далее запись вида $p_1(x) \not\equiv p_2(x)$ будем пони мать так: существует такое непустое подмножество X_0 зн чений x, что $p_1(x) \neq p_2(x)$ $\forall\, x \in X_0$. При этом имеется в вид что x может быть как дискретной, так и непрерывной вел чиной, для которой объективно существует обобщенна вероятностная плотность $p(x)$ [5].

где $p(x_n)$ — одномерное распределение n-й ($n = 1, \ldots, N$) 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 (7) имеем

$$p(x_n) > 0 \text{ для всех } x_n \in X_n,\ n = 1, \ldots, N. \qquad (10)$$

На основании (9) и (10) заключаем, что

$$p(x^{(N)}) > 0 \text{ для всех } x^{(N)} \in X_1 \times \ldots \times X_N. \qquad (11)$$

Поскольку же всегда

$$X^{(N)} \subseteq X_1 \times \ldots \times X_N, \qquad (12)$$

а по определению (7) множество $X^{(N)}$ включает те и только те значения $x^{(N)}$, для которых $p(x^{(N)}) > 0$, то из (11) с учетом (12) как раз и следует соотношение (8).

Утверждение 1 доказано.

Условие утверждения 1, по сути — простое обобщение известного положения, состоящего в том, что в случае статистической независимости двух непрерывных случайных величин, совместное распределение которых отлично от нуля на односвязном ограниченном множестве, это множество представляет собой прямоугольник со сторонами, параллельными координатным осям [2]. Следует обратить внимание, что множество $X^{(N)}$, удовлетворяющее условию (8), может быть не только односвязным, но и многосвязным, ограниченным или же неограниченным и даже совпадать со всем пространством R^N, а компоненты векторов $x^{(N)}$ могут быть как непрерывными, так и дискретными величинами.

Перейдем теперь к рассмотрению свойств признаков.

Свойство 1. Если x_n и x_m безусловно независимы в смысле (4) и, кроме того, каждый из этих признаков имеет различные в смысле (6) распределения в классах, то, по крайней мере, в одном из классов эти признаки будут статистически зависимыми.

Доказательство. Воспользовавшись формулой полной вероятности, выразим распределения, фигурирующие в (4), через соответствующие распределения в классах:

$$p(x_n, x_m) \equiv P(V_1)\, p(x_n, x_m \mid V_1) + P(V_2)\, p(x_n, x_m \mid V_2), \qquad (13)$$

$$p(x_n) \equiv P(V_1)\, p(x_n \mid V_1) + P(V_2)\, p(x_n \mid V_2), \qquad (14)$$

$$p(x_m) \equiv P(V_1)\, p(x_m \mid V_1) + P(V_2)\, p(x_m \mid V_2), \qquad (15)$$

где $P(V_1)$ и $P(V_2)$ — априорные вероятности классов V_1 и V_2.

Подстановка выражений для $p(x_n, x_m)$, $p(x_n)$ и $p(x_m)$ из (13) — (15) в (4) дает

$$P(V_1)\, p(x_n, x_m \mid V_1) + P(V_2)\, p(x_n, x_m \mid V_2) \equiv P(V_1)[P(V_1)\, p(x_n \mid V_1)\, p(x_m \mid V_1) + P(V_2)\, p(x_n \mid V_1) \times p(x_m \mid V_2)] + P(V_2)[P(V_1)\, p(x_n \mid V_2)\, p(x_m \mid V_1) + P(V_2)\, p(x_n \mid V_2)\, p(x_m \mid V_2)]. \qquad (16)$$

[2] См., например, ([6], с. 204—205).

На основании условий (6) можно заключить, что

$$X_n^0 \neq \varnothing, \quad X_m^0 \neq \varnothing, \qquad (17)$$

где

$$X_n^0 = \{x_n : p(x_n \mid V_1) - p(x_n \mid V_2) \neq 0\}, \qquad (18)$$

$$X_m^0 = \{x_m : p(x_m \mid V_1) - p(x_m \mid V_2) \neq 0\}. \qquad (19)$$

Поскольку же в силу (4) признаки x_n и x_m статистически независимы, то, согласно утверждению 1, множество $X^{(2)}$, на котором $p(x_n, x_m) > 0$, удовлетворяет условию

$$X^{(2)} = X_n \times X_m. \qquad (20)$$

Из условия (20) с учетом (17) и того, что $X_n^0 \subset X_n$ и $X_m^0 \subset X_m$, следует $X_0^{(2)} \neq \varnothing$, где $X_0^{(2)} = \{x^{(2)} \equiv (x_n, x_m) : x_n \in X_n^0,\ x_m \in X_m^0\}$, а это означает, что непременно имеет место соотношение

$$[p(x_n \mid V_1) - p(x_n \mid V_2)]\,[p(x_m \mid V_1) - p(x_m \mid V_2)] \not\equiv 0. \qquad (21)$$

Соотношение (21) в силу того, что $P(V_1) + P(V_2) = 1$, после несложных преобразований может быть представлено в виде

$$P(V_1)[P(V_1)\, p(x_n \mid V_1)\, p(x_m \mid V_1) + P(V_2)\, p(x_n \mid V_1) \times p(x_m \mid V_1)] + P(V_2)[P(V_1)\, p(x_n \mid V_2)\, p(x_m \mid V_2) + P(V_2)\, p(x_n \mid V_2)\, p(x_m \mid V_2)] \not\equiv P(V_1)[P(V_1)\, p(x_n \mid V_1) \times p(x_m \mid V_1) + P(V_2)\, p(x_n \mid V_1)\, p(x_m \mid V_2)] + P(V_2) \times [P(V_1)\, p(x_n \mid V_2)\, p(x_m \mid V_1) + P(V_2)\, p(x_n \mid V_2) \times p(x_m \mid V_2)]. \qquad (22)$$

Как видно, в правых частях (16) и (22) фигурирует одно и то же выражение. Отсюда можно заключить, что левые части (16) и (22) связаны соотношением

$$P(V_1)\, p(x_n, x_m \mid V_1) + P(V_2)\, p(x_n, x_m \mid V_2) \not\equiv P(V_1)[P(V_1)\, p(x_n \mid V_1)\, p(x_m \mid V_1) + P(V_2)\, p(x_n \mid V_1) \times p(x_m \mid V_1)] + P(V_2)[P(V_1)\, p(x_n \mid V_2)\, p(x_m \mid V_2) + P(V_2)\, p(x_n \mid V_2)\, p(x_m \mid V_2)],$$

которое после очевидных преобразований можно записать так:

$$P(V_1)\, p(x_n, x_m \mid V_1) + P(V_2)\, p(x_n, x_m \mid V_2) \not\equiv P(V_1)\, p(x_n \mid V_1)\, p(x_m \mid V_1) + P(V_2)\, p(x_n \mid V_2) \times p(x_m \mid V_2). \qquad (23)$$

Предположим, что в одном из классов, например в классе V_1, признаки x_n и x_m статистически независимы. В этом случае

$$p(x_n, x_m \mid V_1) \equiv p(x_n \mid V_1)\, p(x_m \mid V_1). \qquad (24)$$

При таком предположении из (23) с учетом (24) сразу же следует

$$p(x_n, x_m | V_2) \not\equiv p(x_n | V_2) p(x_m | V_2). \qquad (25)$$

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

Разумеется, соотношение (23) может выполняться и в том случае, когда совместно с соотношением (25) выполняется и соотношение $p(x_n, x_m | V_2) \not\equiv \not\equiv p(x_n | V_2) p(x_m | V_2)$, т. е. когда и в одном, и в другом классе признаки x_n и x_m статистически зависимы[3].

Свойство 1 доказано.

Свойство 2. Пусть x_n и x_m безусловно независимы в смысле (4), и по крайней мере один из этих признаков имеет одинаковые распределения в классах. Тогда если x_n и x_m независимы в одном классе, то они непременно будут независимыми и в другом классе, и наоборот, если x_n и x_m 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_m, т. е.

$$p(x_m | V_1) \equiv p(x_m | V_2). \qquad (26)$$

Тогда, повторяя выкладки, использованные при доказательстве свойства 1, из условия (4) с учетом (13)—(15) приходим к соотношению (16), которое в силу (26) теперь уже может быть преобразовано к виду

$$\begin{aligned} P(V_1) p(x_n, x_m | V_1) + P(V_2) p(x_n, x_m | V_2) &\equiv \\ \equiv P(V_1) p(x_n | V_1) p(x_m | V_1) &+ \\ + P(V_2) p(x_n | V_2) p(x_m | V_2) & \end{aligned} \qquad (27)$$

или, что то же самое,

$$\begin{aligned} P(V_1)[p(x_n, x_m | V_1) - p(x_n | V_1) p(x_m | V_1)] &\equiv \\ \equiv P(V_2)[p(x_n, x_m | V_2) - p(x_n | V_2) p(x_m | V_2)]. & \end{aligned} \qquad (28)$$

Предположим, что в одном из классов, например в классе V_1, признаки x_n и x_m статистически независимы, т. е. что $p(x_n, x_m | V_1) \equiv p(x_n | V_1) p(x_m | V_1)$. Тогда в силу (28) сразу же следует $p(x_n, x_m | V_2) \equiv \equiv p(x_n | V_2) p(x_m | V_2)$, т. е. x_n и x_m оказываются независимыми и в классе V_2.

Если же предположить, что в одном из классов, например в 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 в силу (28) $p(x_n, x_m | V_2) \not\equiv p(x_n | V_2) p(x_m | V_2)$. А это означает, что x_n и x_m оказываются статистически зависимыми и в другом классе.

[3] Именно такая трактовка свойства 1, по сути, и дана была в работе [3]. Между тем, как следует из приведенного доказательства, наличие статистической зависимости между x_n и x_m *сразу в обоих* классах вовсе не является обязательным: при выполнении условий (4) и (6) условие (5) может иметь место при $k=1$ или же при $k=2$.

Тем самым завершается доказательство свойства 2.

Заметим, что в соответствии со свойством 2 безусловно независимые признаки x_n и x_m могут оказаться условно зависимыми в классах (причем непременно *сразу в обоих*), даже если оба эти признака имеют одинаковые распределения в классах, т. е.

$$p(x_n | V_1) \equiv p(x_n | V_2), \quad p(x_m | V_1) \equiv p(x_m | V_2). \quad (29)$$

При этом имеет место следующее свойство.

Свойство 3. Пусть признаки x_n и x_m безусловно независимы в смысле (4) и оба имеют одинаковые распределения в классах, т. е. пусть выполняется условие (29). Тогда если x_n и x_m условно зависимы в классах, то

$$p(x_m | x_n, V_1) \not\equiv p(x_m | x_n, V_2), \qquad (30)$$

иначе говоря, зависимость между этими признаками непременно различна в классах.

Доказательство. Воспользуемся полученным при доказательстве свойства 2 соотношением (27), которое в силу того, что $p(x_n | V_1) \equiv p(x_n | V_2)$, теперь уже может быть представлено так:

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

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 (29) имеем

$$P(V_1) p(x_m | x_n, V_1) + P(V_2) p(x_m | x_n, V_2) \equiv p(x_m | V_k),$$
$$k = 1, 2. \qquad (31)$$

Предположим, что сформулированное свойство неверно, т. е.

$$p(x_m | x_n, V_1) \equiv p(x_m | x_n, V_2). \qquad (32)$$

При таком предположении из (31) с учетом (32) получим $p(x_m | x_n, V_k) \equiv p(x_m | x_n)$, $k = 1, 2$. Но это означало бы, что признаки x_n и x_m статистически независимы в классах, что противоречит условию свойства.

Полученное противоречие подтверждает справедливость свойства 3.

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

Свойство 4. Если x_n и x_m безусловно независимы в смысле (4) и, кроме того, условно независимы в обоих классах, т. е. условие (5) выполняется при $k = 1$ и при $k = 2$, то *по крайней мере* один из этих признаков имеет одинаковые распределения в классах, т. е.

$$p(x_n | V_1) \equiv p(x_n | V_2) \quad \text{и/или} \quad p(x_m | V_1) \equiv p(x_m | V_2). \qquad (33)$$

Свойство 5. Если x_n и x_m условно независимы в обоих классах, т. е. условие (5) имеет место при

$k=1$ и при $k=2$, и, кроме того, оба эти признака имеют различные в смысле (6) распределения в классах, то x_n и x_m безусловно зависимы, т. е.

$$p(x_n, x_m) \neq p(x_n)\,p(x_m). \qquad (34)$$

Свойство 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, другими словами, условие (5) выполняется *только* при $k=1$ или же при $k=2$, то x_n и x_m безусловно зависимы в смысле (34).

Свойство 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, т. е. условие (5) выполняется при $k=1$ и при $k=2$, то эти признаки безусловно независимы в смысле (4).

Свойство 8. Если x_n и x_m условно независимы в смысле (5) в обоих классах, но безусловно зависимы в смысле (34), то *каждый* из этих признаков непременно имеет различные распределения в классах, иначе говоря, имеют место соотношения (6).

Понятно, что все приведенные выше свойства справедливы и в том случае, когда вместо двух отдельных признаков — x_n и x_m — рассматриваются любые две совокупности признаков, выделенных из описания $x^{(N)}=(x_1,\dots,x_N)$. В частности, эти свойства можно распространить на тот случай, когда рассматривается один из признаков x_n описания $x^{(N)}$ и сокращенное описание $x^{(N-1)}$, которое не содержит этот признак x_n.

Заметим также, что сформулированные свойства устанавливают взаимосвязи именно тех индивидуальных и совместных вероятностных характеристик признаков, которые фигурируют в достаточных условиях полезности признаков, полученных в работе [7].

Докажем теперь утверждения, которые определяют необходимые условия безошибочного распознавания по описанию $x^{(N)}$ при ограничениях (1).

Утверждение 2. Пусть каждый из N признаков ($N \geqslant 2$) описания $x^{(N)}=(x_1,\dots,x_N)$ имеет пересекающиеся распределения в классах, т. е. пусть выполняются условия (1). Тогда для безошибочного распознавания классов V_1 и V_2 по такому описанию необходимо, чтобы *по крайней мере* в одном из них признаки были статистически зависимыми, т.е.

$$p(x^{(N)} \mid V_k) \neq \prod_{n=1}^{N} p(x_n \mid V_k), \quad k=1 \text{ и/или } k=2. \qquad (35)$$

Доказательство. Рассмотрим множества

$$S_1 = X_{11} \times \dots \times X_{1N}, \qquad (36)$$

$$S_2 = X_{21} \times \dots \times X_{2N}, \qquad (37)$$

образованные прямым произведением всех N множеств X_{1n} и X_{2n} соответственно.

Поскольку условия (1) выполняются для всех $n=1,\dots,N$, то множества S_1 и S_2, образованные в соответствии с (36) и (37), непременно пересекаются:

$$S_1 \cap S_2 \neq \varnothing. \qquad (38)$$

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

$$p(x^{(N)} \mid V_1) \equiv \prod_{n=1}^{N} p(x_n \mid V_1). \qquad (39)$$

Тогда в силу утверждения 1 собственная область этого класса $X_1^{(N)}=\{x^{(N)}: p(x^{(N)} \mid V_1) > 0\}$ должна быть образована прямым произведением всех своих проекций на множества значений $x_1,\dots,x_N$, иначе говоря, прямым произведением $X_{11}\times\dots\times X_{1N}$. Принимая во внимание (36), заключаем, что в этом случае $S_1 \setminus X_1^{(N)} = \varnothing$, откуда на основании (38) имеем

$$X_1^{(N)} \cap S_2 \neq \varnothing. \qquad (40)$$

Вполне очевидно, что для безошибочного распознавания классов V_1 и V_2 по описанию $x^{(N)}$ необходимо, чтобы множество $X_1^{(N)}$ не пересекалось с множеством $X_2^{(N)}=\{x^{(N)}: p(x^{(N)} \mid V_2) > 0\}$, представляющим собственную область второго класса в пространстве описаний $x^{(N)}$, т. е. чтобы

$$X_1^{(N)} \cap X_2^{(N)} = \varnothing. \qquad (41)$$

Легко увидеть, что условия (40) и (41) могут выполняться совместно в том и только лишь в том случае, когда $S_2 \setminus X_2^{(N)} \neq \varnothing$. А это в силу утверждения 1 с учетом (37) означает, что в классе V_2 признаки $x_1,\dots,x_N$ должны уже быть статистически зависимыми.

Тем самым утверждение 2 доказано.

Утверждение 3. Пусть каждый из N признаков ($N \geqslant 2$) описания $x^{(N)}=(x_1,\dots,x_N)$ имеет одинаковые распределения в классах, т. е. пусть выполняются условия (2). Тогда для безошибочного распознавания классов V_1 и V_2 по такому описанию $x^{(N)}$ необходимо, чтобы *сразу в обоих* классах признаки были статистически зависимыми, т. е.

$$p(x^{(N)} \mid V_k) \neq \prod_{n=1}^{N} p(x_n \mid V_k), \quad k=1 \text{ и } k=2. \qquad (42)$$

Доказательство. Прежде всего заметим, что условия (2) могут выполняться только в том случае, когда

$$X_{1n} \setminus X_{2n} = \varnothing \text{ для всех } n=1,\dots,N, \qquad (43)$$

т. е. только тогда, когда множества X_{1n} и X_{2n}, на которых отличны от нуля распределения $p(x_n \mid V_1)$ и $p(x_n \mid V_2)$, совпадают при всех $n=1,\dots,N$. Действительно, если хотя бы для одного из признаков эти множества 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 (2).

Принимая во внимание (43), на основании (36) и (37) заключаем:

$$S_1 \setminus S_2 = \varnothing. \qquad (44)$$

Предположим, что в одном из классов, например в классе V_1, признаки $x_1, \ldots, x_N$ статистически независимы, т. е. что выполняется условие (39). Тогда в силу утверждения 1 с учетом (36) имеем $S_1 \setminus X_1^{(N)} = \varnothing$. В этом случае условие (44) переходит к виду

$$X_1^{(N)} \setminus S_2 = \varnothing. \qquad (45)$$

Поскольку в соответствии с определением множества S_2 непременно $X_2^{(N)} \subseteq S_2$ и, разумеется, $X_2^{(N)} \neq \varnothing$, то из условия (45) следует $X_1^{(N)} \cap X_2^{(N)} \neq \varnothing$. Но это означает, что безошибочное распознавание классов V_1 и V_2 по описанию $x^{(N)}$ невозможно. Следовательно, в условиях данного утверждения безошибочное распознавание становится невозможным, если хотя бы в одном из классов признаки $x_1, \ldots, x_N$ статистически независимы.

Утверждение 3 доказано.

Следует обратить внимание, что для выполнения необходимых условий безошибочного распознавания, о которых шла речь в утверждениях 2 и 3, вовсе не обязательно наличие *безусловной* статистической зависимости между признаками $x_1, \ldots, x_N$. В самом деле, если каждый из этих признаков имеет различные распределения в классах, а сами признаки безусловно независимы (в смысле (3)) то, согласно свойству 1, по крайней мере в одном из классов эти признаки непременно будут условно зависимыми. А это как раз в силу утверждения 2 и является необходимым условием безошибочного распознавания. Точно так же, даже если выполняются условия (2), то, согласно свойству 2, безусловно независимые в смысле (3) признаки могут оказаться условно зависимыми сразу в обоих классах, причем в силу свойства 3 эта зависимость непременно различна в классах. Следовательно, и в этом случае может выполняться необходимое условие безошибочного распознавания, сформулированное в утверждении 3.

Поскольку примеры довольно часто оказываются более убедительными, чем формальные выкладки, проиллюстрируем изложенное следующим наглядным примером.

Пример. Пусть $P(V_1) = P(V_2) = 0{,}5$, $N = 2$, т. е. описание содержит два признака — x_1 и x_2, каждый из которых имеет соответственно четыре градации: $x_1^1, x_1^2, x_1^3, x_1^4$ и $x_2^1, x_2^2, x_2^3, x_2^4$.

Пусть условные распределения этих признаков— $p(x_1, x_2 \mid V_1)$ $p(x_1 \mid V_1)$, $p(x_2 \mid V_1)$ и $p(x_1, x_2 \mid V_2)$, $p(x_1 \mid V_2)$, $p(x_2 \mid V_2)$ — заданы соответственно табл. 1 и 2[4]. Как видно из этих таблиц, оба признака имеют одинаковые маргинальные распределения в классах. Действительно,

$$p(x_1^1 \mid V_1) = p(x_1^1 \mid V_2) = 0{,}4,$$
$$p(x_1^2 \mid V_1) = p(x_1^2 \mid V_2) = 0{,}1,$$
$$p(x_1^3 \mid V_1) = p(x_1^3 \mid V_2) = 0{,}4,$$
$$p(x_1^4 \mid V_1) = p(x_1^4 \mid V_2) = 0{,}1.$$

Точно так же

$$p(x_2^1 \mid V_1) = p(x_2^1 \mid V_2) = 0{,}3125,$$
$$p(x_2^2 \mid V_1) = p(x_2^2 \mid V_2) = 0{,}1875,$$
$$p(x_2^3 \mid V_1) = p(x_2^3 \mid V_2) = 0{,}375,$$
$$p(x_2^4 \mid V_1) = p(x_2^4 \mid V_2) = 0{,}125.$$

[4] Таблицы распределений составлены по форме, предлагаемой в [8]. В клетке, расположенной на пересечении столбца x_1^i и строки x_2^j (в данном случае $i, j = 1, \ldots, 4$) указана вероятность $p(x_1^i, x_2^j \mid V_k)$. В правой колонке приводятся вероятности $p(x_2^j \mid V_k)$, полученные суммированием всех значений $p(x_1^i, x_2^j \mid V_k)$ в соответствующей строке, а в нижней строке— вероятности $p(x_1^i \mid V_k)$, полученные суммированием всех $p(x_1^i, x_2^j \mid V_k)$ в соответствующей колонке.

Таблица 1

x_2 \ x_1	x_1^1	x_1^2	x_1^3	x_1^4	$p(x_2^j \mid V_1)$
x_2^4	0	0	0,1	0,025	0,125
x_2^3	0	0	0,3	0,075	0,375
x_2^2	0,15	0,0375	0	0	0,1875
x_2^1	0,25	0,0625	0	0	0,3125
$p(x_1^i \mid V_1)$	0,4	0,1	0,4	0,1	

Таблица 2

x_2 \ x_1	x_1^1	x_1^2	x_1^3	x_1^4	$p(x_2^j \mid V_2)$
x_2^4	0,1	0,025	0	0	0,125
x_2^3	0,3	0,075	0	0	0,375
x_2^2	0	0	0,15	0,0375	0,1875
x_2^1	0	0	0,25	0,0625	0,3125
$p(x_1^i \mid V_2)$	0,4	0,1	0,4	0,1	

Таблица 3

x_2 \ x_1	x_1^1	x_1^2	x_1^3	x_1^4	$p(x_2^j)$
x_2^4	0,05	0,0125	0,05	0,0125	0,125
x_2^3	0,15	0,0375	0,15	0,0375	0,375
x_2^2	0,075	0,01875	0,075	0,01875	0,1875
x_2^1	0,125	0,03125	0,125	0,03125	0,3125
$p(x_1^i)$	0,4	0,1	0,4	0,1	

Поэтому по каждому в отдельности из этих признаков невозможно провести не только безошибочное распознавание, но и распознавание с вероятностью ошибочных решений, меньшей, чем $P_0(e) = \min\{P(V_1),\ P(V_2)\} = 0{,}5$. Другими словами, каждый из этих признаков сам по себе неинформативен.

В то же время, как видно из табл. 1 и 2, совокупность признаков x_1 и x_2 позволяет провести безошибочное распознавание классов V_1 и V_2 по схеме:

$$\text{решаем в пользу } V_1, \text{ если } x^{(2)} = (x_1, x_2) \in \Omega_1,$$

$$\text{решаем в пользу } V_2, \text{ если } x^{(2)} = (x_1, x_2) \in \Omega_2,$$

где Ω_1 и Ω_2 — области решений, представляющие собой соответственно множества векторов (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) и (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).

Для того чтобы убедиться в безусловной статистической независимости признаков x_1 и x_2, перейдем по формулам полной вероятности (13)—(15), положив $n = 1$ и $m = 2$, от заданных табл. 1 и 2 условных распределений признаков x_1 и x_2 в классах к безусловным распределениям $p(x_1,\ x_2)$, $p(x_1)$, $p(x_2)$. Результаты расчета сведены в табл. 3.

Из приведенных в табл. 3 данных видно, что $p(x_1^1)\,p(x_2^1) = 0{,}4 \cdot 0{,}3125 = 0{,}125$ и $p(x_1^1, x_2^1) = 0{,}125$, т. е. $p(x_1^1, x_2^1) = p(x_1^1)\,p(x_2^1)$. Аналогичным образом можно убедиться в том, что $p(x_1^i, x_2^j) = p(x_1^i)\,p(x_2^j)$ для всех $i, j = 1, \ldots, 4$. Следовательно, $p(x_1, x_2) \equiv p(x_1)\,p(x_2)$, а значит, признаки $x_1 x_2$ безусловно независимы.

Рассмотренный пример наглядно иллюстрирует тот факт, что совокупность одних только неинформативных самих по себе признаков, удовлетворяющих условиям (2), принципиально позволяет провести безошибочное распознавание не только при функциональной связи между признаками в классах, как полагали авторы работы ([2], с. 95), но и при наличии статистической связи между признаками в классах, в том числе и в случае, если эти признаки сами по себе безусловно независимы в смысле (3). Это лишний раз показывает, насколько важно в каждом конкретном случае исследовать именно условную связь в классах, а не безусловную связь между признаками, прежде чем решать вопрос об их полезности в описании.

СПИСОК ЛИТЕРАТУРЫ

1. Харкевич А. А. О выборе признаков при машинном опознавании.— Изв. АН СССР. Техн. кибернетика, 1963, № 2, с. 3—10.
2. Вопросы статистической теории распознавания / Барабаш Ю. Л., Варский Б. В., Зиновьев В. Т. и др.—М.: Сов. радио, 1967.— 400 с.
3. Барабаш Ю. Л. Учет свойств признаков при распознавании.— Изв. АН СССР. Техн. кибернетика, 1965, № 5, с. 85—93.
4. Биргер И. А. Техническая диагностика.— М.: Машиностроение, 1978.— 240 с.
5. Ковалевский В. А. Методы оптимальных решений в распознавании изображений.— М. : Наука, 1976.— 328 с.
6. Минский М., Пейперт С. Персептроны.— М. : Мир, 1971.— 261 с.
7. Файнзильберг Л. С. Применение методов статистического распознавания в термографическом анализе состава металла.— Кибернетика, 1978, № 6, с. 133—136.
8. Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений.— М. : Наука, 1965.— 511 с.

Поступила в редакцию 7.V 1981