ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ
1971 г. вып. 44

УДК 621.391:519.2

ФОРМИРОВАНИЕ РЕШАЮЩИХ ПРИЗНАКОВ

В.И.Котюков

В данной работе рассмотрена задача последовательного формирования линейных решающих признаков распознавания.

§ 1. Постановка задачи и общее описание метода

Структурная схема элементарного распознающего автомата может быть представлена так: $X \to D \to S$, где $X = \{X_1, \ldots, X_p\}$ – исходная система параметров, в p-мерном пространстве которых необходимо построить решающее правило D, $S = \{S_1, \ldots, S_m\}$ – алфавит объектов распознавания [1]. Часто возникает задача построения автомата следующей структуры: $X \to Y \to D \to S$, где $Y = \{Y_1, \ldots, Y_n\}$ – новая, более информативная система признаков, сформированная на базе исходной системы X. $Y = F(X)$, где F – оператор некоторого преобразования.

Очевидно, что любое преобразование F целесообразно производить лишь когда выполняется соотношение: $H_f + H_y < H_x$, где H_f – "сложность" преобразования F, а H_y и H_x – "сложность" решающего правила D в пространствах Y и X соответственно. Большое распространение получили алгоритмы формирования, основанные на статистическом методе главных компонент [2,3,4]. За новые признаки здесь берутся векторы, соответствующие собственным числам $\{\lambda_1, \ldots, \lambda_p\}$ общей ковариационной матрицы A_0. Эти алгоритмы обладают рядом недостатков, которые связаны с тем, что A_0 рассматривается как матрица единой совокупности, т.е. при этом плохо учитываются особенности распределения каждого класса в пространстве X. Неясным остается и вопрос о соотношения $(H_f + H_y)$ и H_x.

Здесь предлагается иной алгоритм формирования. Система новых признаков $\{y_j\}$ также ищется в классе линейных преобразований F. Класс D фиксируется заранее – решающие правила должны быть ортогональны вновь образуемым признакам y. Метод представляет собой последовательную, сходящуюся на конечной выборке итерационную процедуру, на каждом шаге которой ищется такой признак $y_j = \sum\limits_{i=1} \alpha_i x_i$, распознавание по которому ещё нераспознанной обучающей выборки было бы наилучшим. Иными словами, на j-ом шаге алгоритма на признак y_j проецируется множество объектов L_{j-1}, оставшееся нераспознанным на $(j-I)$-м шаге, и по признаку y_j определяются зоны "распознавания" и зоны "пересечения" классов.

Таким образом, на каждом шаге алгоритма мы сталкиваемся с однотипной задачей минимизации множества L_j объектов, попадающих в зоны "пересечения" по признаку y_j. Нетрудно видеть, что каждый такой признак y_j эквивалентен в общем случае $2(m-I)$ параллельным гиперплоскостям (m – число классов), т.е. осуществляет нелинейное разбиение пространства X. Это и создает предпосылки для выполнения соотношения $H_f + H_y \leq H_x$.

В дальнейшем для простоты рассмотрения мы ограничимся случаем двух классов $\{A, B\}$, ибо проблему многих классов можно тем или иным способом свести к решению целого ряда задач дихотомии. Далее будем рассматривать также лишь типовую задачу определения признака y, распознавание по которому было бы наилучшим.

§ 2. Классы линейно разделимы

Обучающие выборки классов A и B обозначим LA и LB соответственно. Введем некоторые определения.

Классы A и B называются __линейно-разделимыми__, если существует такой вектор $y_0 = \sum\limits_{i=1}^{p} \alpha_i x_i$, для которого выполняются соотношения:

$$\left.\begin{aligned} y_0(x) \leq q_0\,, \quad x \in LB \\ y_0(x) \geq q_0\,, \quad x \in LA \end{aligned}\right\} . \qquad (I)$$

Вектор y_0 при этом называется __линейно разделяющим признаком__.

Условие (I) эквивалентно существованию __разделяющей гиперплоскости__ $Z_0 = \sum\limits_{i=1}^{p} \alpha_i x_i + \alpha_{p+1}$, ортогональной вектору y_0 в точке q_0.

Опишем предлагающийся здесь метод "__крайних точек__", позволяющий находить признак y_0. Метод проиллюстрирован на рис. I

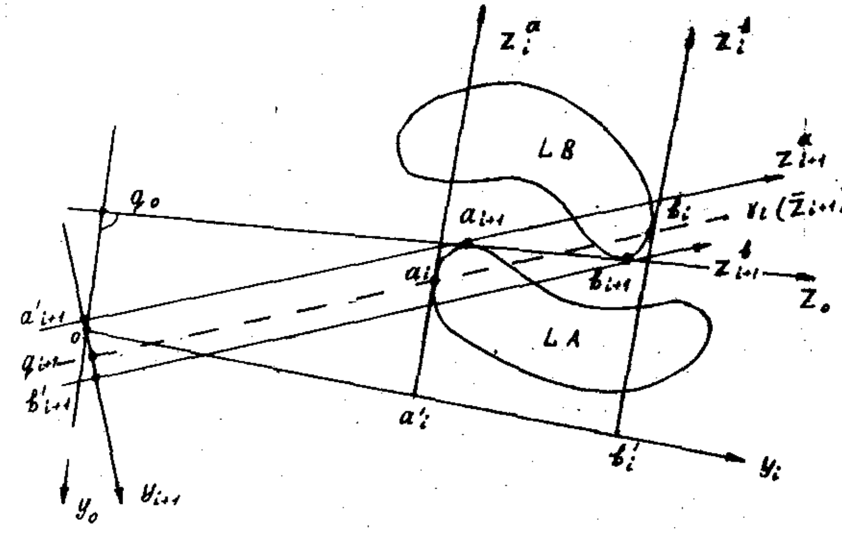

Рис. I

Допустим, мы имеем некоторый признак y_i. Проектируем на y_i выборки LA и LB и получаем максимум три зоны. Возможен, например, такой вариант:

$$\left.\begin{aligned} y_i(x) < a_i'\,, \quad & x \in LB_1 \\ a_i' \leq y_i(x) < b_i'\,, \quad & x \in \{LA_2 \cup LB_2\} \\ y_i(x) > b_i'\,, \quad & x \in LA_1 \end{aligned}\right\} , \qquad (2)$$

где $LA = LA_1 \cup LA_2$ и $LB = LB_1 \cup LB_2$.

Зона (a_i', b_i') называется зоной "пересечения" классов. Множество точек $L_i = \{LA_2 \cup LB_2\}$ называется множеством "пересечения" классов. Возьмем крайние точки множеств LA и LB (но не LA_2 и LB_2) по их проекциям на y_i — a_i и b_i соответственно. При каком-либо ином чем (2) расположении проекций: LA и LB на признаке y_i поступают аналогичным образом.

Крайние точки a_i и b_i соединяются прямой γ_i. Следующей итерацией метода "крайних точек" является вектор y_{i+1}, ортогональный прямой Y_i. По вектору y_{i+1} также находятся свои крайние точки (a_{i+1}, b_{i+1}), по которым делается следующая итерация y_{i+2} аналогичным образом, и т.д.

ТЕОРЕМА I. *Если классы A и B линейно разделимы, то метод "крайних точек" является итерационной процедурой, сходящейся за конечное число шагов к линейно разделяющему признаку y_0.*

ДОКАЗАТЕЛЬСТВО. Если $a_i' = b_i'$, то имеем $y_i = y_0$, и процесс на этом заканчивается. Если имеем соотношение $y_i(x) = a_i' = b_i'$, $x \in \{LA \cup LB\}$, то за y_{i+1} берется любой вектор, отличный от y_i. Как известно, y_i эквивалентен двум параллельным гиперплоскостям $y_i = \{Z_i^a, Z_i^b\}$. Допустим, имеем:

$$\left.\begin{array}{l} Z_i^a(a_i) = Z_i^b(b_i) = 0 \\ Z_i^a(x) < 0, \ x \in LB_1 \\ Z_i^b(x) > 0, \ x \in LA_1 \end{array}\right\} \qquad (3)$$

$$L_i = \{Z_i^a(x) \geq 0\} \cap \{Z_i^b(x) \leq 0\}, \ x \in \{LA \cup LB\}. \qquad (4)$$

По методу "крайних точек" получим следующую итерацию:

$$y_{i+1} = \{Z_{i+1}^a, Z_{i+1}^b\}, \text{ где } Z_{i+1}^a(a_{i+1}) - Z_{i+1}^b(b_{i+1}) = 0.$$

Пусть $y_{i+1} \neq y_0$, т.е. $a_{i+1}' \neq b_{i+1}'$.

Заметим, что в области $\{Z_i^a(x) < 0\}$ гиперплоскость Z_{i+1}^a не может пересекать выпуклую оболочку, натянутую на LB, а в области $\{Z_i^b(x) > 0\}$ Z_{i+1}^b не пересекает оболочку LA, иначе множества LA и LB не были бы линейно отделимыми. Через точку q_{i+1} проведем гиперплоскость $\bar{Z}_{i+1}$, ортогональную y_{i+1}.

Имеем
$$(\bar{Z}_{i+1}(a_{i+1}) \leq 0 \wedge \bar{Z}_{i+1}(b_{i+1}) \geq 0).$$

В силу всего вышесказанного получаем:

$$\left.\begin{array}{ll} (Z_{i+1}^a(x) < 0 \wedge x \in LB), & x \in \{Z_i^a(x) < 0\} \\ (Z_{i+1}^b(x) > 0 \wedge x \in LA), & x \in \{Z_i^b(x) > 0\} \end{array}\right\}, \qquad (5)$$

$$L_{i+1} = \{LA \cup LB\} \setminus \{\{Z_{i+1}^a(x) < 0\} \cup \{Z_{i+1}^b(x) > 0\}\}. \qquad (6)$$

Из (4),(5) и (6) следует:

$$L_{i+1} \subseteq L_i. \qquad (7)$$

Заметим, что если бы на i-м шаге мы имели $LB_1 = LA_1 = \emptyset$, то уже

на $(i+1)$-м шаге получили бы $(LB_1 \neq \emptyset \wedge LA_1 \neq \emptyset)$. Будем полагать, что шкалы $\{X_j\}$ непрерывны, и через V_i будем обозначать суммарный объем выпуклых оболочек LA и LB, заключенный между Z_i^a и Z_i^b.

Тогда из-за того, что $\{Z_i^a, Z_i^b\}$ и $\{Z_{i+1}^a, Z_{i+1}^b\}$ не параллельны и в силу (7), имеем:

$$V_{i+1} < V_i. \qquad (8)$$

Условие разделимости (I) эквивалентно следующему условию:

$$V_0 = 0. \qquad (9)$$

В силу (8) и того, что множества LA и LB конечны, мы на каком-то K-м шаге нашего процесса получим $V_K = 0$, то есть $y_K = y_0$, что и требовалось доказать. Заметим, что положение вектора y_i было произвольным.

ТЕОРЕМА 2. *Если классы LA и LB не являются линейно разделимыми, то существует такой l-й шаг метода "крайних точек", на котором не будет выполняться соотношение $L_i \subseteq L_{i-1}$.*

ДОКАЗАТЕЛЬСТВО. Допустим, что всегда выполняется лишь соотношение $L_i \subseteq L_{i-1}$. Тогда в силу того, что гиперплоскости $\{Z_i^a, Z_i^b\}$ и $\{Z_{i-1}^a, Z_{i-1}^b\}$ не параллельны, будем иметь $V_i < V_{i-1}$, т.е. будем иметь сходимость к разделяющему признаку y_0, что невозможно по условию теоремы. Таким образом, теорема доказана.

Заметим, что если классы A и B линейно разделимы с выполнением условий

$$\left.\begin{array}{ll} y_0(x) \leq q_0, & x \in LB \\ y_0(x) > q_0, & x \in LA \end{array}\right\}, \qquad (10)$$

то, сделав после сходимости метода "крайних точек" достаточно малый шаг Δ в направлении градиента сходимости, получим выполнение соотношений (10).

Равенство $a_i' = b_i'$ является критерием останова алгоритма в случае линейно разделимых классов, а невыполнение соотношения $L_i \subseteq L_{i-1}$ – в случае линейно не разделимых классов.

§ 3. Классы линейно не разделимы

В этом случае мы должны уметь определять признак $y_{опт}$, по которому распознавание классов было бы максимально.

Условимся в качестве первого приближения метода "крайних "точек" брать наилучший (в смысле распознавания) из следующих двух векторов:

а) вектор Y, соответствующий наибольшему собственному числу матрицы $\| C_{ij} \|$ $(i,j=1,\dots,p)$, где

$$C_{ij} = \frac{(\bar\mu_i^a - \bar\mu_i^b)(\bar\mu_j^a - \bar\mu_j^b)2 - (\bar\sigma_{ij}^a + \bar\sigma_{ij}^b)}{(\bar\sigma_{jj}^a + \bar\sigma_{jj}^b) + (\bar\sigma_{ii}^a + \bar\sigma_{ii}^b)},$$

$\bar\mu_i$ и $\bar\sigma_{ij}$ – оценки математического ожидания и дисперсии классов. Нетрудно доказать, что вектор Y будет при этом максимизировать критерий Фишера: $\left(\frac{(\bar\mu^a - \bar\mu^b)^2}{\sigma^a + \sigma^b}\right)$;

б) вектор Y, соответствующий наибольшему собственному числу общей матрицы ковариаций.

Распознавание различных гипотетических односвязных двумерных распределений показывает, что вектор $Y_{опт}$, как правило, близок к вектору Y, полученному на основе вот такого модифицированного метода "крайних точек".

Для получения вектора $Y=\alpha X = \sum_{i=1}^{p} \alpha_i X_i$, приближающего вектор $Y_{опт}$, можно использовать и итерационную процедуру метода стохастической аппроксимации:

$$\alpha^{n+1} = \alpha^n - \Gamma[n+1]\,\nabla F(\alpha^n),$$

где $\bar\nabla F(\alpha^n)$ – оценка градиента "качества" (число "нераспознанных" объектов) по признаку $Y^n = \alpha^n X$.

Последовательность матриц Γ числовых значений должна удовлетворять определенным требованиям сходимости [5].

§ 4. Информативность параметров

После того, как в результате обучения по вышеописанному методу будет получено необходимое число последовательных признаков распознавания $\{Y_1,\dots,Y_n\}$, мы можем оценить информативность каждого из p исходных параметров $\{X_i\}$.

Пусть информативность признака Y_j будет определяться так:

$$g(Y_j) = \frac{N[j-1]-N[j]}{N_0}, \qquad (11)$$

где $N[j]$ – число объектов из $(LA \cup LB)$, оставшихся "нераспознанными" при обучении после "применения" признака Y_j, а N_0 – исходный объем выборки.

Поскольку $Y_j = \sum_{i=1}^{p} \alpha_{ji} X_i$, то мы имеем возможность по величине модуля α_{ji} судить о "вкладе" параметра X_i в признак Y_j. Таким образом, информативность каждого параметра X_i можно оценить так:

$$J(X_i) = \sum_{j=1}^{n} g(Y_j)\cdot|\alpha_{ji}|. \qquad (12)$$

На основании величин $\{J(X_i)\}$ мы можем отобрать некоторую подсистему наиболее информативных исходных параметров $X' \subset X$.

§ 5. Доверительная вероятность правильной классификации

Часто бывает важно после процесса обучения, не прибегая к контрольным испытаниям, определить доверительную вероятность надежной классификации по полученному решающему правилу $D = \{Y_1,\dots,Y_n\}$.

Вероятность правильной классификации можно определить так:

$$p(e) = \sum_{i=1}^{m} q_i p_i(e), \qquad (13)$$

где q_i – априорная вероятность i-го класса, а $p_i(e)$ – вероятность его правильной классификации. $\sum_{i=1}^{m} q_i = 1$.

Имеем:

$$p_i(e) = \sum_{\ell=1}^{n} p_i(Y_\ell)\cdot p_i(e, Y_\ell), \qquad (14)$$

где $p_i(Y_\ell)$ – вероятность того, что реализация i-го класса будет распознаваться (попадать в зону "распознавания" i-го класса) по признаку Y_ℓ, а $p_i(e,Y_\ell)$ – вероятность её правильной классификации при этом. Пусть для $i=1,\dots,m$ выполняется $\sum_{\ell=1}^{n} p_i(Y_\ell)=1$. $N_0=\sum_{i=1}^{m} N_i$, где N_i – объем обучающей выборки i-го класса. По обучающей выборке мы можем найти следующие оценки:

$$\bar p(e) = \sum_{i=1}^{m} q_i \bar p_i(e)$$

и

$$\bar p_i(e) = \sum_{\ell=1}^{n} \bar p_i(Y_\ell)\bar p_i(e,Y_\ell). \qquad (15)$$

Оценка $\bar p_i(Y_\ell)$ вычисляется так:

$$\bar p_i(Y_\ell) = \frac{N_i[\ell-1]-N_i[\ell]}{N_i}, \qquad (16)$$

где $N_i[\ell]$ – объем обучающей выборки i-го класса, оставшийся нераспознанным по признаку Y_ℓ. Для простоты рассмотрения далее будем полагать $N_i[n]=0$.

Индекс "0" будет обозначать какую-либо допустимую величину. Пусть $p_0(e)$ – допустимая вероятность правильной классификации. Тогда $P\{\bar{p}(e) \geq p_0(e)\} = \gamma$, где γ – доверительная вероятность такого события.

$$\left.\begin{aligned} p_0(e) &= \sum_{i=1}^{m} q_i p_{io}(e), \\ p_{io}(e) &= \sum_{l=1}^{n} p_i(Y_e) p_{io}(e, Y_e) \end{aligned}\right\} . \qquad (17)$$

Пусть $p_{io}(e) = p_{io}(e, Y_e) = p_{io}(e, Y_c)$, где $e, c = 1, \ldots, n$. Исходя из (17), будем определять:

$$\left.\begin{aligned} \gamma &= \sum_{i=1}^{m} q_i \gamma_i , \\ \gamma_i &= \sum_{l=1}^{n} \bar{p}_i(Y_e) \cdot \beta_i(Y_e) \cdot \gamma_i(Y_e) \end{aligned}\right\} , \qquad (18)$$

где $\beta_i(Y_e)$ и $\gamma_i(Y_e)$ – доверительные вероятности

$$\left.\begin{aligned} &P\{\bar{p}_i(e, Y_e) \geq p_{io}(e, Y_e)\} = \gamma_i(Y_e), \\ &P\{\bar{p}_i(Y_e) \geq p_i(Y_e)\} = \beta_i(Y_e) \end{aligned}\right\} . \qquad (19)$$

<u>Определение $\beta_i(Y_e)$</u>. Если имеем $\bar{p}_i(Y_e) > p_i(Y_e)$, то это только, как будет видно из дальнейшего рассмотрения, увеличит $\bar{p}_i(e)$. Полагаем, что $p_i(Y_e) = \bar{p}_i(Y_e)$; $\bar{p}_i(Y_e)$ определяется по (16). Величина $\bar{p}_i(Y_e)$ есть частота попадания реализации i-го класса в зону "распознавания" по признаку Y_e, и поэтому для определения $\beta_i(Y_e)$ мы можем воспользоваться биномиальным законом:

$$\beta_i(Y_e) = \sum_{\nu = N_i[e-1] - N_i[e]}^{N_i} C_{N_i}^{\nu} (p_i(Y_e))^{\nu} (1 - p_i(Y_e))^{N_i - \nu} . \qquad (20)$$

<u>Определение $\gamma_i(Y_e)$</u>. Так как по признаку Y_e мы определяем зону i-го класса (или "толерантные пределы" класса) крайними членами вариационного ряда проекции выборки i-го класса на Y_e, то для произвольного "односвязного" закона распределения этого ряда [6] мы имеем:

$$\gamma_i(Y_e) = 1 - N_i[e-1] p_{io}(e, Y_e)^{N_i[e-1]-1} + (N_i[e-1] - 1) p_{io}(e, Y_e)^{N_i[e-1]} . \qquad (21)$$

Назовем наше решающее правило <u>последовательным решающим правилом</u> (ПРП). Тогда из формул, определяющих γ , непосредственно следует такая теорема.

ТЕОРЕМА 3. Доверительная вероятность γ правильной классификации по ПРП с надежностью не менее $p_0(e)$ при фиксированном N_0 (а, стало быть, и наоборот – N_0 при фиксированном γ) не зависит (в явном виде) от размерности исходного пространства X.

Попытаемся связать "сложность" ПРП (число признаков $\{Y_e\}$) с величиной γ .

Пусть при решении двух различных задач мы для каждой получили своё ПРП: $D_1 = \{Y_1^1, \ldots, Y_n^1\}$ и $D_2 = \{Y_1^2, \ldots, Y_k^2\}$, где $n < k$.

Условия задач 1 и 2 будем называть <u>одинаковыми</u>, если выполняются следующие соотношения:

$$\left.\begin{aligned} &m_1 = m_2 = m ; \\ &N_i^1 = N_i^2 = N_i ; \quad i = 1, \ldots, m ; \\ &q_i^1 = q_i^2 = q_i ; \quad i = 1, \ldots, m ; \\ &p_{io}(e, Y_e^1) = p_{io}(e, Y_e^2) = p_{io}(e, Y_e) < 1 ; \\ &i = 1, \ldots, m ; \quad e = 1, \ldots, n . \end{aligned}\right\} . \qquad (22)$$

Будем говорить, что D_1 <u>мажорирует</u> D_2 , если выполняются соотношения:

$$\left.\begin{aligned} &g(Y_e^1) \geq g(Y_e^2), \qquad e = 1, \ldots, n, \\ \text{и} \quad &\max\{g(Y_a^2)\}_{n > a \geq k} \leq \min\{g(Y_e^1)\}_{1 \geq e \geq n} \end{aligned}\right\} . \qquad (23)$$

Сформулируем такую теорему.

ТЕОРЕМА 4. Если ПРП $D_1 = \{Y_1^1, \ldots, Y_n^1\}$ и $D_2 = \{Y_1^2, \ldots, Y_k^2\}$, где $n < k$ являются соответственно решениями двух задач, имеющих одинаковые условия, и при этом D_1 мажорирует D_2 , а также выполняются соотношения $\{q_i = q_j, \ i, j = 1, \ldots, m$ и $p_{io}(e, Y_e) = p_{io}(e, Y_c)$, $e, c = 1, \ldots, n(k)\}$, то справедливо неравенство $\gamma_1 > \gamma_2$, где γ_1 и γ_2 – доверительные вероятности надежной классификации по D_1 и D_2 соответственно.

ДОКАЗАТЕЛЬСТВО. В силу условий теоремы имеем:

$$\begin{aligned} \gamma_j &= q_i \sum_{i=1}^{m} \sum_{l=1}^{n,k} \bar{p}_i(Y_e^j) \cdot \beta_i(Y_e^j) \cdot \gamma_i(Y_e^j) = \\ &= q_i \sum_{l=1}^{n,k} \sum_{i=1}^{m} \bar{p}_i(Y_e^j) \cdot \beta_i(Y_e^j) \cdot \gamma_i(Y_e^j) ; \quad j = 1, 2 . \end{aligned} \qquad (24)$$

Исходя из (I6) и (22), имеем:

$$\sum_{i=1}^{m} \sum_{\ell=1}^{n} \bar{p}_i (\mathcal{Y}_\ell^1) = \sum_{i=1}^{m} \sum_{\ell=1}^{k} \bar{p}_i (\mathcal{Y}_\ell^2). \tag{25}$$

Учитывая условие мажорируемости, получаем:

$$\sum_{i=1}^{m} \bar{p}_i (\mathcal{Y}_\ell^1) \geq \sum_{i=1}^{m} \bar{p}_i (\mathcal{Y}_\ell^2). \tag{26}$$

Величина $\sum_{i=1}^{m} \bar{p}_i (\mathcal{Y}_\ell)$ линейно зависит от величины ($N[\ell-1] - N[\ell]$) и не зависит при этом от самой величины $N[\ell]$. Функция же $\gamma(N) = 1 + (N-1)\rho^N - N\rho^{N-1}$ является выпуклой [6], то есть сходимость $\lim_{N\to 0} \gamma[N] = 0$ увеличивается при уменьшении значений N.

Величина $\sum_{i=1}^{m} \beta_i (\mathcal{Y}_\ell)$ равномерно уменьшается при уменьшении $\sum_{i=1}^{m} \bar{p}_i (\mathcal{Y}_\ell)$.

Для обеих задач имеем $N[\ell-1] > N[\ell]$. Признак $\mathcal{Y}_\ell$ индуцирует величину $N[\ell]$, соответствующую величине $\beta(\mathcal{Y}_\ell)$. Однако в формуле (24) величина $\beta(\mathcal{Y}_\ell)$ встречается в ℓ-м члене ряда, в то время как $N[\ell]$ – в (ℓ+ I)-м члене ряда (в определении $\gamma(\mathcal{Y}_{\ell+1})$). Исходя из этого факта и из характера изменения $\gamma(\mathcal{Y}_\ell)$, а также учитывая (25) и (24), нетрудно вывести соотношение $\gamma_1 > \gamma_2$, которое и требовалось доказать.

Таким образом, более простое ПРП при прочих равных условиях и некоторых не сильных ограничениях "лучше" более сложного ПРП.

Пусть N_o^* – необходимый объем обучающей выборки для достижения величины γ при данной $\rho_o(\ell)$. Поскольку при фиксированном значении $\rho_o(\ell)$ величины γ и N_o^* связаны между собой однозначно, то, учитывая результат теоремы 4, можно сформулировать следующее утверждение.

ТЕОРЕМА 5. Если имеются два различные ПРП $\mathcal{D}_1 = \{\mathcal{Y}_1^1, \ldots, \mathcal{Y}_n^1\}$ и $\mathcal{D}_2 = \{\mathcal{Y}_1^2, \ldots, \mathcal{Y}_k^2\}$, где $n<k$ и выполняются все условия теоремы 4 (за исключением $N_i^1 = N_i^2$, $i=1,\ldots,m$), то для достижения равенства $\gamma_1 = \gamma_2$ необходимо, чтобы выполнилось соотношение $N_{o1}^* < N_{o2}^*$,где N_{oj}^* – необходимый объем обучающей выборки для j-й задачи.

То есть, более "сложное" ПРП требует большую величину N_o^*.

Оптимальным ПРП (ОПРП) будем называть такое ПРП $\mathcal{D} = \{\mathcal{Y}_1, \ldots, \mathcal{Y}_n\}$, у которого $\mathcal{Y}_\ell = \mathcal{Y}_{\ell\, onm}$, $\ell = 1,\ldots,n$.

Совершенно очевидно, что увеличение размерности исходного пространства X может привести только лишь к "упрощению" (уменьшению числа признаков $\mathcal{Y}_\ell$) ОПРП, но не к его "усложнению". Из этого факта и из теоремы 5 следует справедливость следующей теоремы.

ТЕОРЕМА 6. Увеличение размерности исходного пространства X при наличии ОПРП и соблюдении условий теоремы 4 (за исключением $N_i^1 = N_i^2$, $i=1,\ldots,m$) при фиксированном γ может привести лишь к уменьшению (но никак не к увеличению) величины N_o^*.

Определим теперь "момент" окончания построения ПРП.

Часто бывает нежелательно строить ПРП до полного разделения классов в пространстве X (это относится и к требованию определения зон "распознавания" по любому $\mathcal{Y}_\ell$ именно крайними членами вариационного ряда).

Пусть γ_o – минимально допустимое значение γ при данном $\rho_o(\ell)$ и N_o .

Если мы на обучающей выборке получаем $\bar{\rho}(\ell) \geq \rho_o(\ell)$ при $\gamma \geq \gamma_o$, то дальнейшее "дробление" нежелательно, так как при этом мы получим лишь меньшее значение γ .

Допустим, мы имеем два ПРП-$\mathcal{D}_1$ и $\mathcal{D}_2$ ($\mathcal{D}_2$ является "продолжением" $\mathcal{D}_1$), для которых по обучающей выборке получили следующие значения соответственно: $\{\gamma_1, \bar{\rho}^1(\ell)\}$ и $\{\gamma_2, \bar{\rho}^2(\ell)\}$, где $\{\gamma_1 \geq \gamma_o, \bar{\rho}^1(\ell) < \rho_o(\ell), \gamma_2 < \gamma_o, \bar{\rho}^2(\ell) \geq \rho_o(\ell)\}$.

Тогда предпочесть следует то $\mathcal{D}_j$, для которого

$$\min \{(\gamma_o - \gamma_2)^a, (\rho_o(\ell) - \bar{\rho}^1(\ell))^b\}. \tag{27}$$

При отсутствии какой-либо априорной информации можно полагать $a = b = 1$.

В заключение автор выражает благодарность Н.Г.Загоруйко и Г.С.Лбову за обсуждение работы и критические замечания, а также Т.Н.Гади за помощь при составлении программы для ЭВМ.

Литература

1. Загоруйко Н.Г. Современное состояние проблемы распознавания образов. "Вычислительные системы". Новосибирск, вып. 28, 1967.

2. Сборник "Опознание образов" под ред. И.Т.Турбовича, М., Наука, 1968.

3. Tou, Heydorn. Some approaches to optimum feature extraction. Computer and Information Sciences-2, New - York, 1967.

4. S. Watanabe. Evaluation and selection of variable in pattern recognition. Computer and Information Sciences-2, N-Y, 1967.

5. Цыпкин Я.З. Адаптация и обучение в автоматических системах. М., Наука, 1968.

6. Дунин-Барковский И.В., Смирнов Н.В. Теория вероятностей и математическая статистика в технике. М., Гостехиздат, 1955.

Поступила в редакцию
21.I.1971 г.