ЭМПИРИЧЕСКОЕ ПРЕДСКАЗАНИЕ И РАСПОЗНАВАНИЕ ОБРАЗОВ
(Вычислительные системы)
1980 год Выпуск 83

УДК 681.3:621.391

О СВОЙСТВАХ ИНВАРИАНТНОСТИ И СОСТОЯТЕЛЬНОСТИ АЛГОРИТМОВ РАСПОЗНАВАНИЯ, ОСНОВАННЫХ НА ЛОГИЧЕСКИХ ДЕРЕВЬЯХ

А.Н.Манохин

I. Постановка задачи и основные понятия

Пусть для каждого объекта a из некоторой генеральной совокупности определены (но не обязательно известны) значения признаков $X_1,\dots,X_n$ и целевого признака X_{n+1}, принимающего два значения: 1 либо 2. Предполагается, что X_i $(i=1,\dots,n,n+1)$ - случайные величины, которые имеют совместное распределение $P(\bar{y})$ в R^{n+1}.

ОПРЕДЕЛЕНИЕ. Отображение F, ставящее в соответствие каждому вектору $\bar{x} = (x_1,\dots,x_n) \in R^n$ значение $i \in \{1,2\}$, называется р е ш а ю щ и м п р а в и л о м.

Произвольный объект a определяет векторы:

$\bar{x}=(X_1(a),\dots,X_n(a))$ и $\bar{y}=(X_1(a),\dots,X_n(a),X_{n+1}(a))$.

Поэтому любому объекту a ставится в соответствие значение $F(\bar{x})$. Мерой качества решающего правила F будем считать вероятность ошибочной классификации

$$P_0(F,P) = \int g(\bar{y})dP,$$

где

$$g(\bar{y}) = \begin{cases} 0, \text{ если } x_{n+1} = F(x_1,\dots,x_n); \\ 1, \text{ если } x_{n+1} \neq F(x_1,\dots,x_n). \end{cases} \qquad (1)$$

Оптимальное байесовское правило F_0 определяется соотношением:

$$F_0(\bar{x}) = \begin{cases} 1, \text{ если } P(X_{n+1}=1/\bar{x}) \geq 1/2; \\ 2, \text{ если } P(X_{n+1}=1/\bar{x}) < 1/2, \end{cases} \qquad (2)$$

где $P(X_{n+1}=1/\bar{x})$ - условная вероятность. Если предполагать, что условные распределения $X_1,\dots,X_n$ при $X_0=1$ и $X_0=2$ имеют плотности $f_1(\bar{x})$ и $f_2(\bar{x})$, то F_0 запишется в виде:

$$F_0(\bar{x}) = \begin{cases} 1, \text{ если } \dfrac{P(X_{n+1}=1)\, f_1(\bar{x})}{P(X_{n+1}=2)\, f_2(\bar{x})} \geq 1; \\ 2 - \text{в противном случае.} \end{cases}$$

В задачах распознавания P неизвестно, но имеется случайная, независимая выборка $\bar{y}_1,\dots,\bar{y}_n$, полученная в соответствии с P.

ОПРЕДЕЛЕНИЕ. Отображение D, ставящее в соответствие любому набору $\tilde{X} = \{\bar{y}_1,\dots,\bar{y}_n\}$ решающее правило F из некоторого множества S, называется алгоритмом обучения.

Часто представляют $\tilde{X}$ в виде матрицы размерности $n \times m$. В настоящее время нет общепринятых или сколько-нибудь распространенных принципов сравнения алгоритмов обучения. Этим, видимо, и объясняется их чрезмерное изобилие.

2. Свойства инвариантности алгоритмов обучения

Признаки $X_1,\dots,X_n$ могут иметь различную природу. Например, если а - человек и X_1 - его рост, а X_2 - его профессия, выраженная числовым кодом, то ясно, что эти характеристики существенно различны. Предлагается формализовать интуитивные представления о различной природе признаков в рамках теории измерений [1], для этого с каждым признаком X_i свяжем группу допустимых преобразований $E_i = \{\varphi\}$, φ - отображение $R \to R$. Последняя и будет определять тип шкалы признака. Примеры различных шкал можно найти в [1, 2]. Выбирая для каждого X_i свое $\varphi_i \in E_i$, получим набор $\bar{\varphi} = \{\varphi_1, \dots, \varphi_n, \varphi_{n+1}\}$.

ОПРЕДЕЛЕНИЕ. Алгоритм обучения D называется инвариантным относительно допустимых преобразований шкал, если для каждой тройки $(\tilde{X}, \bar{x}, \bar{\varphi})$ справедливо равенство $[D(\tilde{X})](\bar{x}) = [D(\bar{\varphi}(\tilde{X}))](\bar{\varphi}(\bar{x}))$, где $\bar{\varphi}(\tilde{X}) = \{\varphi_j(x_{ij})\}$, $\bar{\varphi}(\bar{x}) = (\varphi_1(x_1), \varphi_2(x_2), \dots, \varphi_n(x_n))$.

Аналогично можно было бы в точных терминах определить инвариантность относительно наименований объектов и признаков.

Далее, будем считать корректными те алгоритмы, которые обладают свойством инвариантности относительно сформулированных выше преобразований. Именно в этом точном смысле предлагается понимать

выражения вида: "Алгоритм может обрабатывать разнотипную информацию".

Мы рассматривали упрощенную модель алгоритма, когда D ставит в соответствие обучающей матрице $\tilde{X}$ единственное решающее правило. Как правило, когда алгоритмы строятся на основе оптимизации некоторого критерия Z, то D определяет F неоднозначно, т.е. ставит в соответствие $\tilde{X}$ некоторое подмножество $D(\tilde{X}) = S_0$ множества S. Это происходит, например, тогда, когда экстремум Z достигается на нескольких F из S.

Введенные выше понятия инвариантности легко распространяются и на этот случай. Пусть $D(\tilde{X},\bar{x}) = \{ k$ таких, что существует $F \in D(\tilde{X})$ и $k = F(\bar{x}) \}$. Алгоритм D будем считать инвариантным, если выполняется $D(\tilde{X},\bar{x}) = D(\bar{\varphi}(\tilde{X}), \bar{\varphi}(\bar{x}))$ для любой тройки $\tilde{X},\bar{x},\bar{\varphi}$.

3. Алгоритмы, основанные на логических деревьях

Этот класс алгоритмов рассматривался в [2]. Будем называть элементарными высказывания C вида $X(a) = x$, $X(a) \neq x$ для признаков в шкале наименований и $X(a) \leq x$, $X(a) > x$ в шкалах порядка, отношений, интервалов. Здесь X – признак, x – его конкретное значение (порог). Высказывания $X(a) \neq x$ и $x(a) > x$ являются дополнительными к $X(a) = x$, $X(a) \leq x$. Дополнительное к C высказывание обозначается символом $\bar{C}$. Далее рассмотрим конъюнкции на элементарных высказываниях $B = C_1 \wedge \ldots \wedge C_l$. На наборе таких конъюнкций $A = \{B_1, \ldots, B_t\}$ можно различным образом задавать решающие правила. Соответствующий класс называют классом логических решающих правил. Если в качестве элементарных высказываний рассматривать высказывания вида $\alpha_1 X_{i_1}(a) + \ldots + \alpha_p X_{i_p}(a) + \alpha_0 \leq 0$ для $X_{i_1}, \ldots, X_{i_p}$ в шкалах отношений и интервалов, то получим класс линейно-логических правил. Решающие правила в виде логических деревьев являются одним из конкретных видов логических решающих правил. В каждой вершине дерева A стоит элементарное высказывание (рис.I), причем если в вершине с номером $2k+1$ стоит C, то в вершине с номером $2k+2$ стоит $\bar{C}$, для каждой k-й вершины определена пара чисел a_k, b_k – количество объектов первого и второго образов, попавших в вершину. Если в каждой конечной вершине дерева A задано решение I либо 2, то получим решающее правило F_A в виде логического дерева.

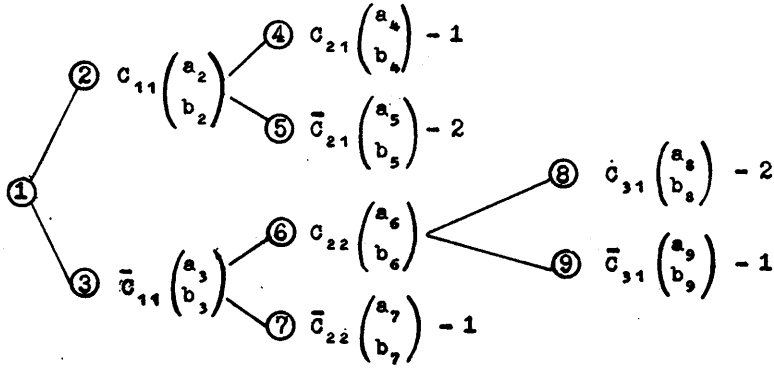

Рис. I

Формальное описание дерева приводится в [2]. Если перенумеровать конечные вершины индексами от I до t, то каждой конечной вершине будут соответствовать конъюнкция B_i (которая для дерева называется ветвью) и пара чисел n_i, m_i – количество объектов первого и второго образов соответственно, попавших в конечную вершину с номером i (что эквивалентно истинности ветви B_i). Характеристика t(A) – количество ветвей дерева – определяет его сложность. Определим класс S_i как множество всех деревьев с количеством ветвей, не большим чем i. Таким образом, задана вложенная последовательность $S_1 \subset S_2 \subset S_3 \ldots$.

Пусть $\nu(F)$ – частота ошибок для F. Задавая функцию $Z(\nu, t, n, m)$, мы определяем тем самым алгоритм обучения, который из множества S – всех решающих правил в виде деревьев – выбирает F, минимизирующее Z. Можно сформулировать некоторые разумные минимальные требования к Z.

I. При фиксированных ν, t и n имеет место $\lim_{m \to \infty} Z(\nu, t, n, m) = \nu$, или, на содержательном уровне, с ростом m вклад сложности t в Z уменьшается, и выбор определяется частотой ошибки ν.

2. Функция Z монотонно возрастает по t (при фиксированных n, m и ν) и по ν (при фиксированных n, m и t). Содержательная трактовка этого пункта очевидна.

Очевидно, что формулировка алгоритма обучения в столь общем виде мало что дает для практической реализации, но вполне доста –

точна для доказательства инвариантности любого алгоритма такого вида.

УТВЕРЖДЕНИЕ I. Для алгоритма D, определяемого функцией Z, справедливо соотношение $D(\tilde{X},\bar{x}) = D(\bar{\varphi}(\tilde{X}), \bar{\varphi}(\bar{x}))$.

ДОКАЗАТЕЛЬСТВО. Пусть $k \in D(\tilde{X})$. Тогда существуют $F_1 \in D(\tilde{X})$, заданное на дереве A_1, и a такие, что $\bar{x} = (X_1(a), \ldots, X_n(a))$ и $Z(\nu(F_1), t(A_1), n, m) \le Z(\nu(F), t(A), n, m)$ для любого F на A. Зададим решающее правило $F^{\bar{\varphi}}$ следующим образом. Каждому C поставим в соответствие C' по правилу: $X_j = x$, $X_j \ne x$, $X_j \le x$, $X_j > x$ переходят в $X_j = \varphi_j(x)$, $X_j \ne \varphi_j(x)$, $X_j \le \varphi_j(x)$, $X_j > \varphi_j(x)$ соответственно. Из определения C' очевидно, что если $\bar{x}$, описывающий a, таков, что на $\bar{x}$ выполняется C, то на $\bar{\varphi}(\bar{x})$ выполняется C', и наоборот. Далее, B' получается из B заменой каждого входящего C на C'. Набор $A^{\bar{\varphi}}$ состоит из $\{B'_1, \ldots, B'_t\}$ и определяет решающее правило $F^{\bar{\varphi}}$ так, что $F^{\bar{\varphi}}(B'_i) = F(B_i)$. Тогда очевидно, что $\nu(F)$ на $\tilde{X}$ совпадает с $\nu(F^{\bar{\varphi}})$ на $\bar{\varphi}(\tilde{X})$ и $t(A) = t(A^{\bar{\varphi}})$. Справедливо неравенство

$$Z(\nu(F_1^{\bar{\varphi}}), t(A_1^{\bar{\varphi}}), n, m) \le Z(\nu(F), t(A), n, m)$$

для любых F на A и $F_1^{\bar{\varphi}}(\bar{\varphi}(\bar{x})) = k$. Следовательно, $k \in D(\bar{\varphi}(\tilde{X}), \bar{\varphi}(\bar{x}))$. Обратное включение доказывается аналогично.

4. Статистические свойства алгоритмов, основанных на логических деревьях

Мы определили S_i как множество решающих правил на деревьях с количеством ветвей, не большим i. Пусть алгоритм обучения выбирает в фиксированном классе S_i правило F_i^*, минимизирующее частоту ошибок $\nu(F)$. Используя оценки вероятности равномерного уклонения частоты от вероятности по классу событий [3, гл.X], можно получить оценки для $P\{P_0(F_i^*, P) - P_0(F_i^0, P) > 2\varepsilon\}$ (заметим, что величина $P_0(F_i^*,P) - P_0(F_i^0,P)$ всегда положительна), где F_i^0 – правило, на котором достигается $\inf_{P \in S_i} P_0(F,P)$; $P_0(F,P)$ – вероятность ошибки.

Как и в [3], свяжем с S_i класс событий $\tilde{S}_i$, поставив в соответствие каждому решающему правилу F событие $C_F = \{\bar{y}$ таких, что $F(\bar{x}) \ne x_{n+1}\}$ (напомним, что $\bar{y} = x_1, \ldots, x_n, x_{n+1}$; $\bar{x} = x_1, \ldots, x_n$).

Если для $\tilde{S}_i$ определить функцию роста $m^{\tilde{S}_i}(n)$ (см. определение в [3, гл. У, § 6,7]), то будут справедливы оценки [3, гл.Х, §6]:

$$P\{P_0(F_i^*,P) - P_0(F_i^0,P) > 2\varepsilon\} \le 6m^{\tilde{S}}(2n)\exp\left(-\frac{\varepsilon^2(n-1)}{4}\right)$$

и (3)

$$P\{|P_0(F_i^*,P) - \nu(F_i^*)| > \varepsilon\} \le 6m^{\tilde{S}}(2n)\exp\left(-\frac{\varepsilon^2(n-1)}{4}\right).$$

Эти неравенства дают доверительные интервалы для отклонения полученной вероятности ошибки от минимальной для F из S_i и от полученной минимальной частоты ошибки.

Получены оценки для $m^{\tilde{S}_i}$ и емкости $r(\tilde{S}_i)$ (определение см. в [3, гл.У, §6,7]) для частных случаев. В общем случае точное вычисление этих характеристик приводит к сложной комбинаторной задаче. Но довольно просто можно показать, что $r(S_i)$ конечна для любого S_i. А отсюда сразу следует [3, гл.Х, §6], что $P_0(F_i^*, P)$ сходится по вероятности к $P_0(F_i^0, P)$ при любом фиксированном i. Это утверждение устанавливает асимптотические свойства алгоритма для фиксированного S_i.

Теперь рассмотрим асимптотические свойства алгоритмов, которые выбирают решающее правило из упорядоченной по вложению последовательности классов решающих правил $S_1 \subset S_2 \subset S_3 \dots$.

ЛЕММА I. Для любого распределения P

$$\lim_{i\to\infty} \inf_{F\in S_i} P_0(F,P) = P_0(F_0,P),$$

где F_0 – байесовское решающее правило.

ДОКАЗАТЕЛЬСТВО. На σ-алгебре борелевских множеств в R^m распределение P определяет две вероятностные меры μ_1 и μ_2, которые соответствуют условным распределениям в R^m при $X_0 = 1$, $X_0 = 2$. Байесовское решающее правило определяет множество $O \subset R^m$, на котором $F(x) = 1$. Это множество измеримо относительно σ-алгебры борелевских множеств. Для вероятностной меры, в силу ее регулярности, существует открытое множество O_1 такое, что $\mu((O_1 \setminus O) \cup (O \setminus O_1)) \le \varepsilon$. Но, как известно, любое открытое множество в R^m можно представить в виде счетного объединения полуоткрытых прямоугольников

$$\bigcup_{i=1}^{\infty} \Pi_i, \quad \Pi_i = \{\bar{x}\in R^m;\ b_i < x_i \le a_i;\ i = 1,\dots,m\}.$$

Очевидно, существует такое k , что

$$O_\varepsilon = \bigcup_{i=1}^{k} \Pi_i, \quad \mu((O_1 \setminus O_2) \cup (O_2 \setminus O_1)) \le \varepsilon .$$

Теперь выделим из всех признаков подмножество признаков $X_{i_1}, \ldots, X_{i_p}$, замеренных в шкале наименований. Мера в соответствующем подпространстве R^p дискретна. Тогда в R^p можно выделить конечное множество точек $Z = \{\bar{z}_1, \ldots, \bar{z}_l\}$ такое, что $\mu(Z) \ge 1-\varepsilon$. В пространстве R^n поставим в соответствие множеству Z множество $Z' = \{\bar{x}$ таких, что проекция $\bar{x}$ в R^p совпадает с одним из $\bar{z}_i$; $i = 1, \ldots, l\}$. Пусть $O_3 = Z' \cap O_2$. Тогда $\mu((O_3 \setminus O_2) \cup (O_2 \setminus O_3)) \le \varepsilon$. Из приведенных неравенств легко следует

$$\mu((O \setminus O_3) \cup (O_3 \setminus O)) \le \varepsilon. \qquad (4)$$

Вероятность $P_0(F,P) = P(X_0 = 1) \times \mu_1(R^n \setminus O) + P(X_0 = 2) \times \mu_2(O)$. В силу неравенства (4), существует множество O', устроенное так же, как и O_3, т.е. состоящее из конечного числа непересекающихся множеств таких, что их проекция в R^p – точка, а в дополнение к R^p – полуоткрытый прямоугольник, и для O' выполняются неравенства:

$$\mu_1((O \setminus O') \cup (O' \setminus O)) \le \varepsilon,$$

$$\mu_2((O \setminus O') \cup (O' \setminus O)) \le \varepsilon .$$

Тогда для правила F такого, что $F(\bar{x}) = 1$ при $\bar{x} \in O'$ и $F(\bar{x}) = 2$ при $\bar{x} \in R^n \setminus O'$, выполняется неравенство $0 \le P_0(F,P) - P_0(F_0,P) \le \varepsilon$. Отсюда для произвольного ε существует i , для которого

$$0 \le \inf_{F \in S_i} (F,P) - P_0(F_0,P) \le \varepsilon .$$

Доказательство закончено.

Из неравенства (3) легко получить

$$P\{P_0(F_i,P) - P_0(F_i^0,P) > 2\varepsilon(n,i,\eta)\} \le \eta ,$$

где $$\varepsilon(n,i,\eta) = \sqrt{-\frac{4}{n-1} \ln \frac{\eta}{6 n^{S_i}(2n)}} .$$

ЛЕММА 2 Существуют последовательности $i(n)$ и $\eta(n)$ такие, что $\eta(n)$ не возрастает, $i(n)$ не убывает,

$$\lim_{n\to\infty} \eta(n) = 0, \quad \lim_{n\to\infty} i(n) = \infty \quad \text{и} \quad \lim_{n\to\infty} \varepsilon(n, i(n), \eta(n)) = 0 .$$

ДОКАЗАТЕЛЬСТВО. Пусть $i(1) = 1$, $\eta(1) = \xi_1$, где $0 < \xi_1 < 1$. Зададим последовательности: $\{\xi_i\}$, $0 < \xi_i < 1$, $\{\varepsilon_j\}$, $0 < \varepsilon_j < 1$, которые не возрастают и стремятся к нулю. Далее, $\eta(n)$ и $i(n)$ определим рекуррентным соотношением: если $i(n) = k$, $\eta(n) = \eta$, то при $\varepsilon(n+1, k+1, \xi_{k+1}) \geq \varepsilon_{k+1}$ полагаем $i(n+1) = i(n) = k$; $\eta(n+1) = \eta(n) = \eta$, а при $\varepsilon(n+1, k+1, \xi_{k+1}) < \varepsilon_{k+1}$ полагаем $i(n+1) = k+1$, $\eta(n+1) = \xi_{k+1}$. Поскольку $\lim_{n\to\infty} \varepsilon(n, k+1, \xi_{k+1}) = 0$ при фиксированном k, то всегда найдется n такое, что $i(n+1) = k+1$, следовательно, $i(n)$ не убывает и $\lim_{n\to\infty} i(n) = \infty$. Далее, $\eta(n)$ монотонно не возрастает, и $\lim_{n\to\infty} \eta(n) = 0$ в силу того, что она принимает значения последовательности $\{\xi_i\}$, причем для любого K существует N такое, что для $n \geq N$ $\eta(n) = \xi_k$ и $k \geq K$. Далее, очевидно, что верно $\lim_{n\to\infty} \varepsilon(n, i(n), \eta(n)) = 0$, так как для любого k при некотором n выполняется $\varepsilon(n, i(n), \eta(n)) < \varepsilon_k$. Доказательство закончено.

Пусть алгоритм ставит в соответствие обучающей матрице $\tilde{X}_n$ (n – объем выборки) решающее правило $F^*_{i(n)}$, минимизирующее частоту ошибки в классе $S_{i(n)}$, где $\{i(n)\}$ – заданная последовательность.

УТВЕРЖДЕНИЕ 2. Существует $\{i(n)\}$ такая, что $P_0(F^*_{i(n)}, P)$ сходится к $P_0(F_0, P)$ по вероятности, т.е. алгоритм является состоятельным.

ДОКАЗАТЕЛЬСТВО. Выберем $\{i(n)\}$, удовлетворяющую условиям леммы 2. Тогда для любого ε существует N_1 такое, что $\varepsilon(n, i(n), \eta(n)) < \varepsilon/4$ при $n \geq N_1$, следовательно,

$$P\{P_0(F^*_{i(n)}, P) - \inf_{F\in S_{i(n)}} P_0(F, P) \geq \varepsilon/2\} \leq \eta(n) .$$

По лемме 1, существует N_2 такое, что при $n \geq N_2$ выполняется

$$\inf_{F\in S_{i(n)}} P_0(F, P) - P_0(F_0, P) \leq \varepsilon/2 .$$

Тогда при $n \geq N = \max(N_1, N_2)$ имеет место

$$P\{P_0(F^*_{i(n)}, P) - P_0(F_0, P) > \varepsilon\} \leq \eta(n).$$

Следовательно, при любом ε справедливо

$$\lim_{n\to\infty} P\{|P_0(F^*_{i(n)}, P) - P_0(F_0, P)| > \varepsilon\} = 0.$$

Доказательство закончено.

ЗАМЕЧАНИЕ. Мы не только доказали существование $\{i(n)\}$, но указали в лемме 2 конструктивную процедуру построения.

5. Связь с известными методами

В статистическом подходе к решению задачи распознавания можно выделить два направления.

В первом фиксируется некоторый класс решающих правил, и в нем выбирается правило, минимизирующее некоторый критерий (обычно это частота ошибок на обучающей выборке $\nu(F)$). Зачастую исследователь в явном или неявном виде работает с последовательностью вложенных классов решающих правил $S_1 \subset S_2 \subset \ldots$. Ясно, что минимальная частота ошибки $\nu(F^*_i)$ в классе S_i не возрастает с ростом i. Но с расширением класса решающих правил у нас уменьшается уверенность в том, что $P_0(F^*_i, P)$ незначительно отклоняется от $\nu(F^*_i)$. В [3] предлагается формальная схема, в которой задача ставится в точных терминах и предлагается критерий выбора правила F^*. Выше логические деревья рассматривались именно в этой схеме.

Второе направление заключается в построении оценок $\hat{f}_1(\bar{x})$ и $\hat{f}_2(\bar{x})$ для плотностей $f_1(\bar{x})$ и $f_2(\bar{x})$. Предполагается, что для условных распределений существуют соответствующие плотности. Оценки $\hat{f}_1(\bar{x})$ и $\hat{f}_2(\bar{x})$ подставляются в (2) на место $f_1(\bar{x})$ и $f_2(\bar{x})$, таким образом задается решающее правило. Во втором направлении можно выделить две ветви: первая - параметрическая, когда $f_1(\bar{x})$ и $f_2(\bar{x})$ известны с точностью до параметров. Тогда по обучающей выборке оценивают эти параметры и получают соответствующие оценки для плотностей. Вторая ветвь - непараметрическая, когда используются непараметрические оценки для $f_1(\bar{x})$, $f_2(\bar{x})$ в (2).

Пусть $P_1(\bar{x})$ и $P_2(\bar{x})$ - условные функции распределений в E^n при $X_0 = 1, X_0 = 2$ соответственно. Для функции распределения по выборке может быть построена оценка, например, выборочная функция распределения. Оценки функции распределения изучались раньше, чем оценки для плотности, и для них известен ряд классических свойств

(например, выборочная функция $\hat{P}(x)$ сходится к истинной функции распределения $P(x)$ по вероятности равномерно по x). Почему же в литературе не рассматриваются правила, полученные использованием в (I) оценок $\hat{P}_1(\bar{x})$ и $\hat{P}_2(\bar{x})$ для $P_1(\bar{x})$ и $P_2(\bar{x})$? Причина заключается в том, что близость $\hat{P}_i(\bar{x})$ к $P_i(\bar{x})$ ни в среднеквадратичной, ни в равномерной метрике не влечет близости $P_0(F^*,\bar{P})$ к $P_0(F_0, P)$, где F^* - правило, построенное по оценкам.

Алгоритмы, основанные на классе логических деревьев, можно рассматривать в рамках второго направления, но строить оценки не для $f_i(\bar{x})$ или $P_i(\bar{x})$, а для соответствующих вероятностных мер μ_i в R^n. Мы строим конечную систему непересекающихся событий и каждому событию ставим в соответствие меру $\hat{\mu}_i(A_j)$, $i = 1,2$; $j = 1,\ldots,t$. Результаты, полученные в работе, позволяют утверждать, что решающее правило, построенное на основе таких оценок, приближается к искомому.

Предложенный подход является естественным продолжением идей, на которых основаны оценки плотности типа гистограмм и полиграмм [4]. Характерные же особенности предложенного подхода заключаются в следующем.

I. При фиксированном объеме выборки часто оказывается невозможным удовлетворительно восстановить плотности в R^n. Но, учитывая специфику задачи распознавания, можно предполагать, что существует подмножество в R^n, в котором меры μ_i приближаются достаточно удовлетворительно для построения решающего правила.

2. Плотность вообще может не существовать. Однако и в этом случае можно говорить о приближении вероятностных мер μ_1 и μ_2.

3. Необходимо рассматривать и случай разнотипных признаков. Заметим, что дискретность распределения признака еще не определяет его типа.

В теории распознавания образов известны работы [5,6], где для формулировки решающих правил используется язык, близкий к предложенному. В этих работах не рассматривается статистическая постановка и признаки могут быть лишь бинарными, тем не менее основные идеи описания языка в них заложены и предложенный подход можно считать развитием этих идей.

Литература

I. СУППЕС П., ЗИНЕС Дж. Основы теории измерений.-В кн.: Психологические измерения.-М.: Мир, 1967, с.9-110.

2. МАНОХИН А.Н. Методы распознавания образов, основанные на логических решащих функциях. -В кн.: Эмпирическое предсказание и распознавание образов. (Вычислительные системы, вып. 67.) Новосибирск, 1976, с. 42-53.

3. ВАПНИК В.Н., ЧЕРВОНЕНКИС А.Я. Теория распознавания образов.- М.: Наука, 1974. - 414 с.

4. ТАРАСЕНКО Ф.П. Непараметрическая статистика.-Томск: 1976. - 292 с.

5. ЖУРАВЛЁВ Ю.И., ДМИТРИЕВ А.Н., КРЕНДЕЛЕВ Ф.П. О математических принципах классификации предметов и явлений. -В кн.: Дискретный анализ, Новосибирск, 1966, вып.7, с. 3-17.

6. ХАНТ Э., МАРТИН Дж., СТОУН Ф. Моделирование процесса формирования понятий на вычислительной машине. - М.: Мир, 1970.-301с.

Поступила в ред.-изд.отд.
17 апреля 1980 года