Вестник
МОСКОВСКОГО УНИВЕРСИТЕТА

№ 6 — 1973

УДК 621.391.2

В. А. БУРОВ, О. В. ДМИТРИЕВ

КЛАССИФИКАЦИЯ ПО СЛАБЫМ ПРИЗНАКАМ МЕТОДОМ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Рассматривается задача классификации выборки к одному из нескольких классов. Предполагается, что функция правдоподобия выборки зависит от информативных параметров, характеризующих каждый класс, а также от общих параметров обстановки. При условии, что все информативные параметры малы, получен достаточно простой алгоритм классификации, основанный на методе максимального правдоподобия.

Постановка задачи

Предположим, что выборка, представленная в виде вектора v, может принадлежать одному из классов: $H_0, H_1, ..., H_M$, причем функция распределения выборки известна и зависит от $M+1$ вектора неизвестных параметров $\lambda, \varepsilon_1, ..., \varepsilon_M$.

Будем также считать, что функция правдоподобия выборки $w(v/\lambda, \varepsilon_1, ..., \varepsilon_M)$ в том случае, когда $v \in H_0$, может быть представлена в виде

$$w(v/\lambda, 0, \ldots, 0) = G_0(\lambda) \tag{1}$$

и зависит от n неизвестных параметров $\lambda_1, \lambda_2, ..., \lambda_n$.

Если же выборка относится к классу H_j, то функция правдоподобия

$$w(v/\lambda, 0, \ldots, \varepsilon_j, \ldots, 0) = G_j(\lambda, \varepsilon_j) \tag{2}$$

будет дополнительно зависеть от m_j неизвестных параметров $\varepsilon_{j1}, \varepsilon_{j2}, ...$ $..., \varepsilon_{jm}$.

Таким образом, каждый из классов H_j $(j=1, 2, ..., M)$ характеризуется информативными параметрами ε_{jk} $(k=1, 2, ..., m_j)$ и, кроме того, имеются общие для всех классов параметры λ_i $(i=1, 2, ..., n)$, которые можно назвать параметрами обстановки [1].

Задача классификации в такой постановке может решаться как задача проверки гипотез совместно с оценкой параметров, соответствующих этим гипотезам [1, 2]. При этом байесовское решение определяется выбором функции потерь и априорными вероятностями для каждой из гипотез. Однако для проверки гипотез относительно функций распределения, зависящих от неизвестных параметров, можно использовать метод максимального правдоподобия, который хотя и не основан на строгих соображениях оптимальности, но часто позволяет получать достаточно простые результаты, близкие к оптимальным при больших выборках [3], и не требует той полноты априорной информации, которая необходима для получения байесовских решений.

Правило классификации по методу максимального правдоподобия удобно сформулировать в терминах выигрышей: если принимается неправильное решение, то выигрыш полагается равным нулю, а в случае верной классификации выборки v классу H_j выигрыш равен $a_j>0$ $(j=0, 1, ..., M)$. При этом правило классификации состоит в выборе наибольшей из величин:

$$a_0 \max_{\lambda} G_0(\lambda); \quad a_j \max_{\lambda, \varepsilon_j} G_j(\lambda, \varepsilon_j) \quad (j=1, 2, \ldots, M).$$

На возможность использования такой процедуры указано в [4]. Кроме того, к задачам подобного типа может быть применен теоретико-игровой подход [5]. В частном случае, когда имеется всего два класса H_0 и H_1, правило классификации сводится к рассмотрению отношения максимумов функций правдоподобия:

$$\frac{a_1 \max\limits_{\lambda, \varepsilon_1} G_1(\lambda, \varepsilon_1)}{a_0 \max\limits_{\lambda} G_0(\lambda)} \gtreqless 1. \tag{3}$$

В [3] показано, что критерий (3) обладает свойством асимптотической оптимальности.

В общем случае классификация выборки также может производиться с помощью отношений типа (3). Обозначим вектор, составленный из оценок максимального правдоподобия парметров $\lambda_1, \lambda_2, ..., \lambda_n$, полученных в предположении $v \in H_0$ с помощью $G_0(\lambda)$, через $\widehat{\lambda}_0$, а вектора оценок максимального правдоподобия, полученных с помощью $G_j(\lambda, \varepsilon_j$, через $\widehat{\lambda}_j$, $\widehat{\varepsilon}_j$. Переходя к логарифмам функций правдоподобия, вместо отношений типа (3) будем рассматривать разности:

$$L_j = \ln G_j(\widehat{\lambda}_j, \widehat{\varepsilon}_j) - \ln G_0(\widehat{\lambda}_0) - \ln \frac{a_0}{a_j}, \tag{4}$$
$$j = 1, 2, \ldots, M.$$

Если все $L_j \leqslant 0$, то выборка принадлежит классу H_0; если же некоторые из $L_j > 0$, то выборка относится к тому классу, для которого L_j принимает максимальное значение.

Алгоритм классификации при наличии малых параметров

Оценки максимального правдоподобия параметров обстановки λ, вообще говоря, существенно зависят от того, какая из функций правдоподобия $G_j(\lambda, \varepsilon_j)$ использовалась для их получения. Однако если

$$\left| \frac{\ln G_j(\lambda, \varepsilon_j) - \ln G_0(\lambda)}{\ln G_0(\lambda)} \right| \ll 1; \quad j = 1, 2, \ldots, M, \tag{5}$$

все параметры ε_{jk} $(j=1, 2, ..., M;\ k=1, 2, ..., m_j)$ можно считать «малыми», и вектор достаточно оценить в предположении $v \in H_0$. Алгоритм классификации при этом значительно упрощается. Если объем выборки v велик, то оценки максимального правдоподобия близки к истинным значениям параметров, и в этом случае можно считать, что условия, аналогичные (5), выполнены и для оценок $\widehat{\lambda}_j$, $\widehat{\varepsilon}_j$ и $\widehat{\lambda}_0$.

Предположим, что функция правдоподобия выборки вне зависимости от того, какому классу $H_0, H_1, \ldots, H_M$ она принадлежит, является дважды дифференцируемой по всем входящим в нее параметрам. Оценки максимального правдоподобия, входящие в (4), удовлетворяют системам уравнений

$$\frac{\partial}{\partial \lambda_i} \ln G_0(\hat{\lambda}_0) = 0 \qquad i = 1, 2, \ldots, n, \tag{6}$$

$$\frac{\partial}{\partial \lambda_i} \ln G_j(\hat{\lambda}_j, \hat{\varepsilon}_j) = 0 \quad i = 1, 2, \ldots, n,$$

$$\frac{\partial}{\partial \varepsilon_{jk}} \ln G_j(\hat{\lambda}_j, \hat{\varepsilon}_j) = 0 \quad k = 1, 2, \ldots, m_j. \tag{7}$$

Кроме того, с учетом (1) и (2):

$$\frac{\partial}{\partial \lambda_i} \ln G_j(\hat{\lambda}_0, 0) = \frac{\partial}{\partial \lambda_i} \ln G_0(\hat{\lambda}_0) = 0,$$

$$\frac{\partial^2}{\partial \lambda_i \partial \lambda_k} \ln G_j(\hat{\lambda}_0, 0) = \frac{\partial^2}{\partial \lambda_i \partial \lambda_k} \ln G_0(\hat{\lambda}_0). \tag{8}$$

Введем матрицы

$$B = \left\| -\frac{\partial^2}{\partial \lambda_i \partial \lambda_k} \ln G_0(\hat{\lambda}_0) \right\| \quad i, k = 1, 2, \ldots, n, \tag{9}$$

$$P_j = \left\| -\frac{\partial^2}{\partial \lambda_i \partial \varepsilon_{jk}} \ln G_j(\hat{\lambda}_0, 0) \right\| \quad \begin{matrix} i = 1, 2, \ldots, n, \\ k = 1, 2, \ldots, m_j, \end{matrix} \tag{10}$$

$$D_j = \left\| -\frac{\partial^2}{\partial \varepsilon_{ji} \partial \varepsilon_{jk}} \ln G_j(\hat{\lambda}_0, 0) \right\| \quad i, k = 1, 2, \ldots, m_j. \tag{11}$$

При высокой апостериорной точности с учетом (5), а также (6)—(8) запишем

$$\ln G_j(\hat{\lambda}_j, \hat{\varepsilon}_j) \simeq \ln G_0(\hat{\lambda}_0) + \sum_{k=1}^{m_j} \hat{\varepsilon}_{jk} \frac{\partial}{\partial \varepsilon_{jk}} \ln G_j(\hat{\lambda}_0, 0) -$$

$$- \frac{1}{2} \hat{\varepsilon}'_j D_j \hat{\varepsilon}_j - \frac{1}{2} (\hat{\lambda}_j - \hat{\lambda}_0)' B (\hat{\lambda}_j - \lambda_0) - (\hat{\lambda}_j - \hat{\lambda}_0)' P_j \hat{\varepsilon}_j, \tag{12}$$

где $\hat{\varepsilon}_{jk}$ — оценка максимального правдоподобия k-го компонента вектора ε_j, $\hat{\varepsilon}'_j$ — вектор полученный транспонированием $\hat{\varepsilon}_j$, а $(\hat{\lambda}_j - \hat{\lambda}_0)$ — вектор, равный разности $\hat{\lambda}_j$ и $\hat{\lambda}_0$.

Используя (8), преобразуем второй член в правой части (12):

$$\sum_{k=1}^{m_j} \hat{\varepsilon}_{jk} \frac{\partial}{\partial \varepsilon_{jk}} \ln G_j(\hat{\lambda}_0, 0) \simeq \hat{\varepsilon}'_j D \hat{\varepsilon}_j + (\hat{\lambda}_j - \hat{\lambda}_0)' P_j \hat{\varepsilon}_j. \tag{13}$$

Тогда

$$\ln G_j(\hat{\lambda}_j, \hat{\varepsilon}_j) \simeq \ln G_0(\hat{\lambda}_0) + \frac{1}{2} \hat{\varepsilon}'_j D_j \hat{\varepsilon}_j - \frac{1}{2} (\hat{\lambda}_j - \hat{\lambda}_0)' B (\hat{\lambda}_j - \hat{\lambda}_0). \tag{14}$$

Используя $\lambda = \hat{\lambda}_0$ и $\varepsilon_j = 0$ в качестве приближенного решения системы (7), можно получить вектор $(\hat{\lambda}_j - \hat{\lambda}_0)$, выраженный через $\hat{\varepsilon}_j$:

$$(\widehat{\lambda}_j - \lambda_0) = -B^{-1} P_j \widehat{\varepsilon}_j. \quad (15)$$

Подставив (15) в (14), получим

$$L_j = \frac{1}{2} \widehat{\varepsilon}_j' T_j \widehat{\varepsilon}_j - \ln \frac{a_0}{a_j} \text{ (см. (4))}, \quad (16)$$

где матрица

$$T_j = D_j - P_j' B^{-1} P_j. \quad (17)$$

Таким образом, алгоритм классификации при наличии малых параметров сводится к получению оценок параметров обстановки $\widehat{\lambda}_0$ в предположении $v \in H_0$, а также к нахождению оценок всех векторов $\widehat{\varepsilon}_j$ $(j=1, 2, ..., M)$, которые затем используются для формирования квадратичных форм (16). Выбор максимума среди L_j позволяет отнести выборку к определенному классу, основываясь лишь на оценках информативных параметров, характеризующих этот класс.

Частные случаи. Интерпретация полученных результатов

Полная матрица, составленная из вторых производных $\ln G_j(\lambda, \varepsilon_j)$, взятых со знаком минус в точке $(\lambda=\lambda_0;\ \varepsilon_j=0)$, будет блочной:

$$R_j = \begin{pmatrix} B & P_j \\ P_j' & D_j \end{pmatrix}. \quad (18)$$

Матрица, обратная R_j, имеет вид

$$R_j^{-1} = \begin{pmatrix} B^{-1} + (B^{-1} P_j) T_j^{-1} (B^{-1} P_j)' & -(B^{-1} P_j) T_j^{-1} \\ -T_j^{-1} (B^{-1} P_j) & T_j^{-1} \end{pmatrix}, \quad (19)$$

где T_j определена (17). Так как элементы R_j (18) при высокой апостериорной точности могут служить оценками элементов информационной матрицы Фишера, то диагональные элементы матрицы T_j^{-1} как видно из (19), являются оценками дисперсий оценок $\widehat{\varepsilon}_{jk}$ $(k=1, 2, ..., m_j)$ в точке $(\varepsilon_j=0;\ \lambda=\lambda_0)$. Кроме того, матрица T_j (не совпадающая с D_j^{-1} *см* (18)) является ядром квадратичной формы (16).

В частном случае, когда каждый из векторов ε_j имеет всего один компонент (например, ε_{j1}), матрица P_j вырождается в вектор, а матрица T_j — в число t_j. При этом (16) будет

$$L_j = \frac{1}{2} t_j \widehat{\varepsilon}_{j_1}^2 - \ln \frac{a_0}{a_j}, \quad (20)$$

причем величины t_j^{-1} являются оценками дисперсий оценок $\widehat{\varepsilon}_{j1}$ при $\varepsilon_{j1}=0$.

Если имеется всего два возможных класса H_0 и H_1, то вместо правила классификации (3) получим

$$\widehat{\varepsilon}_1' T_1 \widehat{\varepsilon}_1 \gtrless 2 \ln \frac{a_0}{a_1}. \quad (21)$$

Если же, кроме того, класс H_1 характеризуется единственным малым информативным параметром ε, то придем к правилу классификации в виде

$$\widehat{\varepsilon}^2 \gtrless t^{-1}\, 2 \ln \frac{a_0}{a_1}, \tag{22}$$

т. е. квадрат оценки малого параметра сравнивается с порогом, пропорциональным оценке дисперсии $t^{-1} = \widehat{\sigma}_\varepsilon^2$, взятой в точке ($\varepsilon=0$; $\lambda=\widehat{\lambda}_0$).

Соотношения (20) и (21) могут быть применены, в частности, для решения задачи порогового обнаружения сигналов (в том числе и случайных) с неизвестными параметрами, принимаемых на фоне помех, также имеющих неизвестные параметры.

В заключение авторы выражают благодарность проф. В. А. Красильникову и доц. В. И. Шмальгаузену за обсуждение работы и ряд полезных замечаний.

ЛИТЕРАТУРА

1. Тартаковский Г. П., Репин В. Г. «Радиотехника», **26**, № 4, 1971.
2. Левин Б. Р., Шинаков Ю. С. «Радиотехника», **26**, № 2, 1971.
3. Wald A. Trans Am. Math. Soc., **54**, 426—482, 1943.
4. Леман Э. Проверка статистических гипотез. М., 1964.
5. Терпугов А. Ф. «Радиотехника и электроника», № 1, 1964.

Поступила в редакцию
22.2 1972 г.

Кафедра
акустики