УДК 621.391:519.2

# ОПТИМИЗАЦИЯ КРИТЕРИЯ ФИШЕРА-УИЛКСА И СОКРАЩЕНИЕ ИСХОДНОЙ СИСТЕМЫ ОПИСАНИЯ В ЗАДАЧАХ РАСПОЗНАВАНИЯ ОБРАЗОВ

В.И. Котюков

В данной работе предложен метод, позволяющий достаточно точно определять дискриминантные функции в пространстве большой размерности. Кроме того, описан алгоритм минимизации исходной системы признаков (параметров). Рассмотрена также задача распознавания одного класса; введен дискриминантный критерий, учитывающий ее специфику.

## § I. Постановка задачи дискриминантного анализа

Допустим, в $\rho$ - мерном исходном пространстве признаков $X=\{X_1,\dots,X_\rho\}$ задана обучающая выборка из $M$ совокупностей (классов). Необходимо построить $n$ таких векторов

$$\{y_1,\dots,y_n\} \quad (y_i=\alpha_{1i}X_1+\alpha_{2i}X_2+\dots+\alpha_{\rho i}X_\rho),$$

каждый из которых максимизировал бы дискриминантный критерий Фишера-Уилкса [I] и был бы ортогонален остальным. Напомним, что данный критерий по вектору $y$ равен

$$F=\frac{\sum\limits_{\ell=1}^{M}(\bar{y}_\ell-\bar{y}_o)^2}{\sum\limits_{\ell=1}^{M}\bar{\sigma}_\ell}, \tag{I}$$



где $\bar{y}_o$, $\bar{y}_\ell$ и $\bar{\sigma}_\ell$ есть соответствующие оценки общего математического ожидания (м.о.), м.о. и дисперсии $\ell$-го класса, полученные по выборке, спроектированной на вектор $y$. "Весовые" множители $N_\ell$ (объем выборки) при $(\bar{y}_\ell-\bar{y}_o)^2$ [I] для простоты рассмотрения здесь опущены. В работах [I]-[3] предложены различные методы решения данной задачи, требующие громоздких вычислений и, что самое главное, обладающие малой точностью при достаточно большой размерности исходного пространства ($\rho>20$) [4]. Ниже будет предложен иной метод, не обладающий вышеупомянутыми недостатками.

## § 2. Описание метода

Допустим, что на основе обучающей выборки получены матрицы $Q$ и $A=\sum\limits_{\ell=1}^{M}A_\ell$, каждая размером $\rho\times\rho$, где $Q$ есть оценка матрицы ковариаций м.о. классов относительно общего м.о., а $A_\ell$ - оценка матрицы собственных ковариаций класса $\ell$.

Образуем матрицу $H=Q-\kappa A$, где $\kappa$ - есть некоторое положительное число.

Нетрудно видеть, что вектор $y=\alpha'X$, соответствующий максимальному собственному числу $\lambda_{max}$ матрицы $H$, будет максимизировать критерий

$$f=\sum_{\ell=1}^{M}(\bar{y}_\ell-\bar{y}_o)^2-\kappa\sum_{\ell=1}^{M}\bar{\sigma}_\ell \tag{2}$$

Такой вектор будем обозначать $y_{max}(H)\equiv\lambda_{max}(H)$. Заметим, что в вычислительной математике [4] существуют простые и точные методы определения вектора $y$, соответствующего максимальному по модулю собственному числу матрицы. Поэтому необходимо экстремизировать не матрицу $H$, а матрицу $H^*=Q-\kappa A+aE$, где $E$ - единичная матрица, а положительное число $a$ всегда можно взять таким, чтобы $H^*$ была положительно определенной [4].

Нетрудно доказать, что при $\alpha'\alpha=1$ имеем:

$$y_{max}(H)=y_{max}(H^*) \quad \text{и} \quad f_{max}=\lambda_{max}(H)=\lambda_{max}(H^*)-a$$

Введем обозначения:

$$\sum_{l=1}^{M}(\bar{y}_l-\bar{y}_0)^2=\mu, \quad \sum_{l=1}^{M}\bar{\sigma}_l=\sigma \quad \text{и} \quad F_{max}=\frac{\mu_0}{\sigma_0}$$

ЛЕММА. Существует такое число $k=\bar{k}$, при котором имеем $F=F_{max}$ для вектора $y_{max}(H)$, где $H=Q-\bar{k}A$.

ДОКАЗАТЕЛЬСТВО. Допустим, имеются два вектора $y_0$ и $y_1$, где $y_0 \equiv (\mu_0/\sigma_0)$ и $y_1 \equiv (\mu_1/\sigma_1)$. Следовательно, имеем $\mu_0/\sigma_0 \geq \mu_1/\sigma_1$. Покажем, что при $k=\bar{k}=\mu_0/\sigma_0$ выполняется соотношение:

$$\mu_0-\bar{k}\sigma_0 \geq \mu_1-\bar{k}\sigma_1. \quad (3)$$

Раскроем (3), подставив вместо $\bar{k}$ величину $\mu_0/\sigma_0$:

$$\mu_0\sigma_0-\mu_0\sigma_0 \geq \mu_1\sigma_0-\mu_0\sigma_1. \quad (4)$$

Из (4) следует:

$$\mu_0\sigma_1 \geq \mu_1\sigma_0.$$

Откуда получаем $\mu_0/\sigma_0 \geq \mu_1/\sigma_1$, что справедливо в силу исходного предположения.

Таким образом, мы видим, что при $\bar{k}=\mu_0/\sigma_0$ справедливо (3), что означает $y_0=y_{max}(Q-\bar{k}A)$. Лемма доказана.

Будем обозначать через $f_{max}(k)$ величину $f_{max}$, которая соответствует $y_{max}(Q-kA)$.

ТЕОРЕМА. Функция $f_{max}(k)$ является убывающей при возрастании величины аргумента $k$ и при $k=\bar{k}$ имеет место $f_{max}(\bar{k})=0$.

ДОКАЗАТЕЛЬСТВО. Пусть при различных $k_1$ и $k_2$, где $k_1<k_2$, мы получили:

$$f_{max}(k_1)=\mu_1-k_1\sigma_1 \quad \text{и} \quad f_{max}(k_2)=\mu_2-k_2\sigma_2.$$

Так как $k_1<k_2$, то имеем

$$\mu_2-k_2\sigma_2<\mu_2-k_1\sigma_2 \quad (5)$$

В силу того, что $f_{max}(k_1)$ есть результат максимизации при $k_1$, справедливо:

$$\mu_2-k_1\sigma_2 \leq \mu_1-k_1\sigma_1.$$

Следовательно, выполняется соотношение:

$$\mu_2-k_2\sigma_2<\mu_1-k_1\sigma_1,$$



что доказывает справедливость первого утверждения теоремы.

Исходя из леммы, мы при $k=\bar{k}$ получаем:

$$f_{max}(\bar{k})=\mu-\bar{k}\sigma=\mu_0-\bar{k}\sigma_0=\mu_0-(\mu_0/\sigma_0)\sigma_0=0.$$

Таким образом, теорема полностью доказана.

Следовательно, меняя величину $k$ и на каждом шаге экстремизируя матрицу $H$, а также пользуясь известными методами нахождения единственного корня монотонной функции одного аргумента ($f_{max}(k)$), мы можем определить вектор $y_1$, для которого имеем $F=F_{max}$.

На основе подобной методики нетрудно получить и вектор $y_2$, для которого $F=F_{max}$ при условии $y_1\cdot y_2=0$, и т.д. Если при этом пользоваться, например, методом "ортогонализации" ([4], стр. 523), то критерием оптимальности вектора $y_2$ будет по-прежнему условие $f_{max}(\bar{k})=0$.

Последовательные методы получения ортогональных векторов $\{y_1,\dots,y_n\}$, как известно [4], дают хорошую точность лишь при $n \leq 6 \div 8$. Но этого вполне достаточно для решения практических задач распознавания.

Если же пользоваться последовательной решающей процедурой [5], то на каждом следующем этапе нужно получать вектор $y_i$, который бы наилучшим образом "распознавал" то, что не "распознал" (не "отделил") предыдущий вектор $y_{i-1}$ (матрицы $Q$ и $A$ для нахождения вектора $y_i$ определяются по выборке, оставшейся "нераспознанной" по вектору $y_{i-1}$).

## § 3. Минимизация исходной системы признаков

Необходимо из исходной системы $p$ признаков $\{X_i\}$ выбрать такие $m$ признаков $\{X_j\}$ ($m<p$), $m$-мерное пространство которых обладало бы наилучшими дискриминантными возможностями.

Допустим, в исходном $p$-мерном пространстве $X$ найдены $n$ наилучших ортогональных дискриминантных векторов $y_l = \alpha_{(l)}X$, $l=1,\dots,n$. Согласно изложенной выше методике, каждый такой вектор $y_{(l)}$ есть "наилучший" собственный вектор определенной матрицы $H_{(l)}$. Матрицы $H_{(l)}=Q-\bar{k}_l A$ отличаются друг от друга лишь величиной $\bar{k}_l$, где $\bar{k}_l$ есть значение функционала (I) для вектора $y_l$. Величина $n$ выбирается такой, чтобы дискриминантные способности $n$ - мерного пространства $Y$, равные $\bar{k}=\bar{k}_1+\ldots+\bar{k}_n$, "исчерпывали" бы основную долю дискриминантных способностей пространства $X$. В теоретически оптимальном случае имеем $n=p$. Обычно $n\leq 5$. Напомним, что вектор $\alpha_{(e)}=(\alpha_{1(e)},\ldots,\alpha_{p(e)})$ в выражении $Y_e = \alpha_{(e)} X$ есть результат максимизации величины $\alpha' H_{(e)} \alpha$ при соответствующих условиях ортогональности.

$$(\alpha' H_{(e)} \alpha)_{max} = \alpha'_{(e)} H_{(e)} \alpha_{(e)} = 0;\ l=1,\ldots,n.$$

Пусть $\{h_{ij(e)}\}$ есть элементы матрицы $H_{(e)}$, $i,j = 1,\ldots,p$.

Образуем функционал:

$$B = \bar{k}_1 (\alpha'_{(1)} H_{(1)} \alpha_{(1)}) + \ldots + \bar{k}_n (\alpha'_{(n)} H_{(n)} \alpha_{(n)}). \tag{6}$$

Введем обозначение

$$\beta_{ij} = \bar{k}_1 \alpha_{i(1)} \alpha_{j(1)} h_{ij(1)} + \ldots + \bar{k}_n \alpha_{i(n)} \alpha_{j(n)} h_{ij(n)}.$$

Тогда будем иметь:

$$B = \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_{ij} \tag{7}$$

Заметим, что аддитивность функционала (6) справедлива в силу ортогональности векторов $\{Y_1,\ldots,Y_n\}$. Поскольку, как уже говорилось, система $\{Y_1,\ldots,Y_n\}$ находилась из условия максимизации величины $B$, исходную задачу минимизации $\{X_i\}$ мы можем заменить следующей задачей. Необходимо найти такую подсистему из $m$ исходных признаков $\{X_j\}$ $(m<p)$, которая вносила бы наибольший "вклад" в величину $B$ для полученного пространства $Y$. Это возможно в силу того, что каждый элемент $h_{ij(e)}$ матрицы $H_{(e)}$ определяется лишь исходными признаками $X_i$ и $X_j$.

Введем неизвестные переменные $z_{ij}=\{0,1\}$, $i=1,\ldots,p$; $j=1,\ldots,(p+1)$. Если $z_{ij}=1$, то слагаемое $\beta_{ij}$ в (7) учитывается; в противном случае - нет. Фиктивный $(p+1)$ - й столбец в матрице $\{z_{ij}\}$ введен для контроля величины $m$ - числа отбираемых признаков $\{X_j\}$. Тогда наша задача сводится к максимизации целевой функции

$$B^* = \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_{ij} z_{ij} \tag{8}$$

при ограничениях на неизвестные:



$$\sum_{j=1}^{p} z_{ij} - m z_{i(p+1)} = 0, \quad i=1,\ldots,p; \tag{9}$$

$$\sum_{i=1}^{p} z_{ij} - m z_{jj} = 0, \quad j=1,\ldots,p; \tag{10}$$

$$\sum_{i=1}^{p} z_{i(p+1)} = m. \tag{11}$$

Задача $\langle$ (8), (9), (10), (11) $\rangle$ есть задача линейного программирования обобщенно-транспортного типа, имеющая целочисленное решение [6]. Если $z_{ij}=1$, то $X_i$ и $X_j$ входят в $m$ - мерную подсистему $\{X_j\}$, в противном случае - нет.

Данный алгоритм минимизации исходной системы признаков применим и в некоторых задачах аппроксимационного анализа при использовании метода главных компонент или разложения Карунена-Лоэва [1].

## § 4. Распознавание одного класса

В ряде случаев возникает такая задача. Необходимо с заданной надежностью классифицировать реализации классов $C$ и $\bar{C}$, где класс $\bar{C}$ означает определенное множество объектов, не входящих в $C$. В этом случае критерий Фишера-Уилкса не адекватен уже смыслу задачи. Не совсем адекватен и аппроксимационный критерий (метод главных компонент) [1], обычно используемый в данном случае.

В настоящей работе для этой цели предлагается следующий дискриминантный критерий:

$$\tilde{F} = \bar{\sigma}_{\bar{C}}(C) / \bar{\sigma}_C, \tag{12}$$

где $\bar{\sigma}_C$ - оценка дисперсии (разброса) реализаций класса $C$ относительно своего м.о., а $\bar{\sigma}_{\bar{C}}(C)$ - оценка дисперсии (разброса) реализаций класса $\bar{C}$ относительно м.о. класса $C$. Пусть матрица $Q$ - оценка матрицы ковариаций (разброса) класса $\bar{C}$ относительно м.о. класса $C$, а $A$ - оценка собственных ковариаций класса $C$. Оптимизация критерия $\tilde{F}$ соответствует максимальному сжатию "области" класса $C$ при условии максимального исключения из этой "области" реализаций класса $\bar{C}$.

На основе матриц $Q$ и $A$ и процедуры, описанной во 2-м параграфе, можно получить систему векторов $\{\varphi_1, \dots, \varphi_n\}$, максимизирующих критерий $\tilde{F}$.

Заметим, что если $Q$ – единичная матрица, то данный критерий вырождается в аппроксимационный.

Автор признателен Г.С. Лбову за обсуждение работы.

## Литература


1. УИЛКС С. Математическая статистика. М., "Наука", 1967.
2. СЕБАСТИАН Г.С. Процессы принятия решений при распознавании образов. Пер. с англ. Киев, "Техника", 1965, 151 стр.
3. Опознание образов. Под ред. Турбовича И.Т. М., "Наука", 1968.
4. УИЛКИНСОН Дж.Х. Алгебраическая проблема собственных значений. М., "Наука", 1970.
5. КОТЮКОВ В.И. Формирование решающих признаков. -"Вычислительные системы". Новосибирск, 1972, вып. 44.
6. ГОЛЬШТЕЙН Е.Г., ЮДИН Д.Б. Задачи линейного программирования транспортного типа. М., "Наука", 1969.


Поступила в ред.-изд.отд.
5. I. 1972 г.