ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

Сборник трудов
1964 г. Института математики СО АН СССР Выпуск I4

ОБ ОШИБКАХ КЛАССИФИКАЦИИ ОБРАЗОВ ПРИ НЕРАВНЫХ МАТРИЦАХ КОВАРИАЦИИ

Г.С. Лбов

В данной статье задача опознавания речевых образов рассматривается как статистическая проблема. Каждому образу ставится в соответствие генеральная совокупность, т.е.закон распределения некоторого случайного вектора. Задача состоит в том, чтобы на основе реализации случайного вектора отнести его к одному из нескольких законов распределения. Для решения этой задачи необходимо определить оптимальную решающую функцию, обеспечивающую минимум вероятности неправильной классификации, а также величину этой вероятности.

В этой работе решение указанных задач рассматривается только для нормального распределения случайного вектора. Случай многомерных нормальных совокупностей с разными векторами средних, но равными матрицами ковариации, подробно рассмотрен в работе [I]. Если же генеральные совокупности имеют разные матрицы ковариации, то нахождение вероятности неправильной классификации значительно усложняется. Тогда для определения указанной вероятности был использован метод моделирования [2], который, однако, требует больших затрат машинного времени,особенно при увеличении размерности случайного вектора.

Однако с увеличением размерности случайного вектора для распределения разности квадратичных форм, появляющейся при решении задачи, можно использовать асимптотическое выражение.

Это позволяет избежать неэкономичного моделирования.

С целью введения необходимых обозначений и сравнения методов решения задачи рассмотрим сначала случай равных матриц ковариаций.

§ I. Случай равных матриц ковариаций

Пусть $\pi_1,\dots,\pi_m$ – m генеральных совокупностей с законами распределения $N(\mu^{(i)},\lambda)$ где $\mu^{(i)}=(\mu_1^{(i)},\dots,\mu_n^{(i)})$ – вектор среднего значения i -ой генеральной совокупности ($i=1,\dots,m$), а λ – ковариационная матрица каждой совокупности; $x=(x_1,x_2,\dots,x_n)$ – случайный вектор.

Тогда i -ая плотность распределения вероятностей будет

$$p_i(x)=\frac{1}{(2\pi)^{n/2}|\lambda|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu^{(i)})'\lambda^{-1}(x-\mu^{(i)})}$$

Пространством результатов наблюдений Y служит n -мерное векторное пространство. Разбиваем все пространство Y на m попарно пересекающихся областей $Y_1,\dots,Y_m$. Если наблюдение попадает в область Y_i, то мы говорим, что оно произведено над π_i. При этом вероятность ошибочной классификации наблюдения, отнесенного к генеральной совокупности π_j, в то время как оно принадлежит к π_i, равна

$$P(j/i,Y)=\int_{Y_j}p_i(x)dx.$$

Если априорные вероятности $q_1,\dots,q_m$ того, что выборка произведена из соответствующей генеральной совокупности, известны, то общая вероятность ошибочной классификации равна

$$\mathcal{P}=\sum_{i=1}^{m}q_i\sum_{\substack{j=1\\ j\neq i}}^{m}\mathcal{P}(j/i,Y)=\sum_{i=1}^{m}q_i\mathcal{P}_i, \qquad (\mathrm{I})$$

где $\mathcal{P}_i$ – вероятность ошибочной классификации, когда выборка производится из π_i.

Области $Y_1,\dots,Y_m$ должны быть выбраны так, чтобы $\mathcal{P}$ было минимальным. Метод Байеса обеспечивает минимум $\mathcal{P}$ при данных $q_1,\dots,q_m$. Согласно этому методу область Y_i определяется как совокупность точек x, удовлетворяющих условию

$$\frac{p_i(x)}{p_j(x)}\geqslant\frac{q_j}{q_i}, \quad \text{для } j=1,\dots m;\ j\neq i \qquad (2)$$

В случае равных матриц ковариации условие (2) записывается так:

$$u_{ij}(x)=\ln\frac{p_i(x)}{p_j(x)}=x'\lambda^{-1}(\mu^{(i)}-\mu^{(j)})+$$

$$+\frac{1}{2}(\mu^{(i)}+\mu^{(j)})'\lambda^{-1}(\mu^{(i)}-\mu^{(j)})\geqslant\ln\frac{q_j}{q_i}=c_{ij}.$$

Равенство $u_{ij}=c_{ij}$ определяет оптимальную решающую функцию для генеральных совокупностей π_i и π_j. Заметим, что случайная величина $u_{ij}(x)$ имеет нормальное распределение, так как является линейной функцией от x.

Согласно условию $u_{ij}\geqslant c_{ij}$, определяющему область Y_i, вероятность ошибочной классификации $\mathcal{P}_i$ определяется как:

$$\mathcal{P}_i=\int_{-\infty}^{c_{im}}\dots\int_{-\infty}^{c_{i2}}\int_{-\infty}^{c_{i1}}p_i(u_{i1},\dots,u_{i,i-1},u_{i,i+1},\dots u_{im})du_{i1},\dots du_{i,i-1},$$

$$du_{i,i+1},\dots du_{im},$$

где p_i – плотность распределения вектора $u_i=(u_{i1},\dots,u_{i,i-1},u_{i,i+1},\dots,u_{im})$ для генеральной совокупности π_i. В случае генеральной совокупности π_i вектор u_i имеет закон распределения $N[M_i(u_i),\Delta^{(i)}]$, где j -й – элемент вектора средних $M_i(u_i)$ имеет вид

$$M_i(u_{ij})=\frac{1}{2}(\mu^{(i)}-\mu^{(j)})'\lambda^{-1}(\mu^{(i)}-\mu^{(j)})=\frac{1}{2}\alpha_{ij},$$

а ковариация между u_{ip} и u_{ik}, где $p,k=1,\dots,m;\ p,k\neq i$,

$$\Delta_{pk}^{(i)}=(\mu^{(i)}-\mu^{(k)})'\lambda^{-1}(\mu^{(i)}-\mu^{(p)}).$$

Величина α_{ij} называется "дивергенцией" между $N(\mu^{(i)},\lambda)$ и $N(\mu^{(j)},\lambda)$

§ 2. Случай неравных матриц ковариаций

Пусть $p_i(x)$ – нормальная плотность с вектором средних $\mu^{(i)}$ и ковариационной матрицей $\lambda^{(i)}(i=1,\dots,m)$. В этом случае условие (2), определяющее совокупность точек x, составляющих область Y_i, имеет вид:

$$\xi_{ij}(x)=\ln\frac{p_i(x)}{p_j(x)}=\ln\sqrt{\frac{|\lambda^{(j)}|}{|\lambda^{(i)}|}}+\frac{1}{2}(Q_j^{-1}-Q_i^{-1})\geqslant\ln\frac{q_j}{q_i},$$

где $Q_i^{-1}=(x-\mu^{(i)})'(\lambda^{(i)})^{-1}(x-\mu^{(i)})$ - положительно определенная квадратичная форма, $j=1,\dots,m;\ j\neq i$. Последнее неравенство можно записать:

$$\xi_{ij}(x)=Q_j^{-1}-Q_i^{-1}\geqslant 2\ln\frac{q_j}{q_i}+2\ln\sqrt{\frac{|\lambda^{(i)}|}{|\lambda^{(j)}|}}=C_{ij}.$$

Функция $\xi_{ij}(x)$, являющаяся нелинейной функцией от x, не подчиняется нормальному закону распределения. Поэтому вероятность неправильной классификации нельзя определить так просто, как это было рассмотрено в § I для случая равных матриц ковариаций.

В случае неравных матриц ковариаций "дивергенция" (см. [2]) между $N(\mu^{(i)},\lambda^{(i)})$ и $N(\mu^{(j)},\lambda^{(j)})$ имеет вид

$$\alpha_{ij}=\frac{1}{2}tr(\lambda^{(i)}-\lambda^{(j)})\left[(\lambda^{(j)})^{-1}-(\lambda^{(i)})^{-1}\right]+$$

$$+\frac{1}{2}tr\left[(\lambda^{(i)})^{-1}+(\lambda^{(j)})^{-1}\right](\mu^{(i)}-\mu^{(j)})(\mu^{(i)}-\mu^{(j)})'. \quad (3)$$

Используя метод моделирования, авторы работы [2], находят по значению "дивергенции" α_{ij} вероятность неправильной классификации.

Чтобы сократить машинное время мы вмест^ моделирования будем применять метод, основанный на том, что при увеличении размерности случайного вектора закон распределения случайной величины $\xi_{ij}(x)$ стремится к нормальному [4].

Для оценки скорости сходимости распределения этой величины к нормальному можно на ЭВМ поставить следующий эксперимент: определить, в какой степени эмпирическое распределение $\xi_{ij}(x)$, полученное с помощью моделирования выражения (8), соответствует нормальному закону распределения при возрастании n. Итак, при достаточно большом n будем считать, что если x имеет закон распределения $N(\mu^{(i)},\lambda^{(i)})$, то вектор $\xi_i=(\xi_{i1},\dots,\xi_{i\,i-1},\xi_{i\,i+1},\dots,\xi_{im})$ имеет закон распределения $N(M_i(\xi_i),\Delta^{(i)})$.

Определив математическое ожидание $M_i(\xi_i)$ и матрицу ковариации $\Delta^{(i)}$ вектора ξ_i для $i=1,\dots,m$, вероятность ошибочной классификации $\mathcal{P}$ находим точно так же, как и в случае равных матриц ковариаций. Ясно, что чем выше размерность вектора x, тем ближе закон распределения к нормальному и тем меньше ошибка в определении $\mathcal{P}$.

§ 3. Определение $M_i(\xi_i)$ и $\Delta^{(i)}$

Вектор $M_i(\xi_i)$ легко определяется; j-ая компонента этого вектора

$$M_i(\xi_{ij})=\sum_{g=1}^{n}\sum_{h=1}^{n}(\lambda^{(j)})^{-1}_{gh}\left[(\lambda^{(i)})_{gh}+(\mu_g^{(i)}-\mu_g^{(j)})(\mu_h^{(i)}-\mu_h^{(j)})\right]-n.$$

Однако при вычислении $\Delta^{(i)}$ возникают некоторые трудности. Рассмотрим процедуру вычисления элемента $\Delta^{(i)}_{pk}(p,k=1,\dots,m;\ p,k\neq i)$.

$$\Delta^{(i)}_{pk}=M_i\left[\xi_{ip}-M_i(\xi_{ip})\right]\left[\xi_{ik}-M_i(\xi_{ik})\right]=$$

$$=M_i(\xi_{ip}\cdot\xi_{ik})-M_i(\xi_{ip})\cdot M_i(\xi_{ik})$$

$$M_i(\xi_{ip}\cdot\xi_{ik})=M_i(Q_p^{-1}\cdot Q_k^{-1})-M_i(Q_i^{-1}Q_k^{-1})-M_i(Q_p^{-1}Q_i^{-1})+M_i(Q_i^{-1})^2 \quad (4)$$

Вычислим первое слагаемое из (4).

$$M_i(Q_p^{-1}Q_k^{-1})=M_i\left[\sum_{g=1}^{n}\sum_{h=1}^{n}(\lambda^{(p)})^{-1}_{gh}\cdot(x_g-\mu_g^{(p)})\cdot(x_h-\mu_h^{(p)})\times\right.$$

$$\left.\times\sum_{g=1}^{n}\sum_{h=1}^{n}(\lambda^{(k)})^{-1}_{gh}(x_g-\mu_g^{(k)})(x_h-\mu_h^{(k)})\right]=$$

$$=\sum_{g=1}^{n}\sum_{h=1}^{n}\sum_{\ell=1}^{n}\sum_{d=1}^{n}(\lambda^{(p)})^{-1}_{gh}(\lambda^{(k)})^{-1}_{\ell d}M_i\left[(x_g-\mu_g^{(p)})(x_h-\mu_h^{(p)})\times\right.$$

$$\left.\times(x_\ell-\mu_\ell^{(k)})(x_d-\mu_d^{(k)})\right].$$

Определение $M_i\left[(x_g-\mu_g^{(p)})(x_h-\mu_h^{(p)})(x_\ell-\mu_\ell^{(k)})(x_d-\mu_d^{(k)})\right]$ сводится к громоздкому вычислению четырехкратного интеграла. Для нахождения $M_i(Q_p^{-1}Q_k^{-1})$ необходимо вычислить n^4 подобных интегралов. Аналогичные вычисления требуются для определения других слагаемых выражения (4).

Эти трудности в определении $\Delta^{(i)}$ можно избежать, если преобразовать координаты так, чтобы от функции $\xi_{ij}(x)$ зависимых компонент вектора x перейти к $\xi_{ij}(V)$, функции независимых компонент случайного вектора V.

Сначала приведем независимо друг от друга квадратичные формы $Q_i^{-1}(x)$ и $Q_j^{-1}(x)$ к каноническому виду. Матрицы преобразования обозначим соответственно F_i и F_j. Так как Q_i^{-1} и Q_j^{-1} являются положительно определенными квадратичными формами, то они в каноническом виде будут представлены суммой n квадратов от своих переменных, т.е.

$$Q_i^{-1}(V)=V'V=\sum_{g=1}^{n} V_g^2$$

$$Q_j^{-1}(W)=W'W=\sum_{g=1}^{n} W_g^2,$$

тогда
$$\xi_{ij}(V,W)=W'W-V'V. \tag{5}$$

Новые и старые координаты связаны следующим образом:

$$V=F_i(x-\mu^{(i)}), \tag{6}$$

$$W=F_j(x-\mu^{(j)}). \tag{7}$$

Заметим, что если x имеет закон распределения $N(\mu^{(i)},\lambda^{(i)})$, то V имеет закон распределения $N(0,E)$. Матрица E является единичной матрицей ковариации.

Таким образом, в случае генеральной совокупности π_i функцию ξ_{ij} необходимо выражать только через компоненты вектора V, которые в этом случае являются независимыми нормальными переменными с нулевыми математическими ожиданиями и единичными дисперсиями.

Для этого выразим вектор W через V, используя (6) и (7),

$$W=F_jF_i^{-1}V+F_j(\mu^{(i)}-\mu^{(j)}).$$

Введем обозначения

$$R^{(ij)}=F_jF_i^{-1} \quad \text{и} \quad \delta^{(ij)}=F_j(\mu^{(i)}-\mu^{(j)}),$$

тогда (5) представится в виде:

$$\xi_{ij}(V)=(R^{(ij)}V+\delta^{(ij)})'(R^{(ij)}V+\delta^{(ij)})-V'V$$

или

$$\xi_{ij}(V_1,\dots,V_n)=\sum_{d=1}^{n}\Big(\sum_{g=1}^{n}R_{dg}^{(ij)}V_g+\delta_d^{(ij)}\Big)^2-\sum_{g=1}^{n}V_g^2. \tag{8}$$

Используя выражение (8), определим $M_i(\xi_i)$ и матрицу

ковариации $\Delta^{(i)}$. При этом j -ый элемент вектора $M_i(\xi_i)$ определяется так:

$$M_i(\xi_{ij})=\sum_{d=1}^{n}\sum_{g=1}^{n}(R_{gd}^{(ij)})^2+\sum_{g=1}^{n}(\delta_g^{(ij)})^2-n \quad (j=1,\dots,m;\ j\neq i)$$

а pk -ый элемент матрицы $\Delta^{(i)}$:

$$\Delta_{pk}^{(i)}=M_i(\xi_{ip}\cdot\xi_{ik})-M_i(\xi_{ip})\cdot M_i(\xi_{ik})$$

для $p,k=1,\dots,m;\ p,k\neq i$.

Математическое ожидание $M_i(\xi_{ip}\cdot\xi_{ik})$ имеет вид:

$$M_i(\xi_{ip}\xi_{ik})=\sum_{g=1}^{n}(\delta_g^{(ip)})^2\cdot\sum_{g=1}^{n}(\delta_g^{(ik)})+\sum_{g=1}^{n}(\delta_g^{(ip)})^2\cdot\Big[\sum_{d=1}^{n}\sum_{g=1}^{n}(R_{gd}^{(ik)})^2-n\Big]+$$

$$+\sum_{g=1}^{n}(\delta_g^{(ik)})^2\Big[\sum_{d=1}^{n}\sum_{g=1}^{n}(R_{gd}^{(ip)})^2-n\Big]+\sum_{e=1}^{n-1}\Big\{\Big[\sum_{g=1}^{n}(R_{ge}^{(ip)})^2-1\Big]\times$$

$$\times\Big[\sum_{d=e+1}^{n}\sum_{g=1}^{n}(R_{gd}^{(ik)})^2-n\Big]+\Big[\sum_{g=1}^{n}(R_{ge}^{(ik)})^2-1\Big]\cdot\Big[\sum_{d=e+1}^{n}\sum_{g=1}^{n}(R_{gd}^{(ip)})^2-n\Big]\Big\}+$$

$$+3\sum_{d=1}^{n}\Big\{\Big[\sum_{g=1}^{n}(R_{gd}^{(ip)})^2-1\Big]\Big[\sum_{g=1}^{n}(R_{gd}^{(ik)})^2-n\Big]\Big\}+$$

$$+4\Big\{\sum_{d=1}^{n}\Big(\sum_{g=1}^{n}R_{gd}^{(ip)}\delta_g^{(ip)}\sum_{g=1}^{n}R_{gd}^{(ik)}\delta_g^{(ik)}\Big)+$$

$$+\sum_{e=1}^{n-1}\Big[\sum_{d=e+1}^{n}\Big(\sum_{g=1}^{n}R_{ge}^{(ip)}R_{gd}^{(ip)}\sum_{g=1}^{n}R_{ge}^{(ik)}R_{gd}^{(ik)}\Big)\Big]\Big\}.$$

В заключение отметим, что рассмотренный метод нахождения вероятности ошибочной классификации $\mathcal{P}$ целесообразен только при большой размерности вектора x ; в противном случае, данный метод будет давать бо́льшую ошибку в определении этой величины, чем метод моделирования.

Литература

I. Т. Андерсон. Введение в многомерный статистический анализ. М., ФМ., 1963 г.

2. T. Marill and D.M.Green. On the effectiveness of receptors in recognitions systems . IRE Trans. on information theory, v.IT-9, Yanuary, 1963, pp. 11 - 17.

3. S. Kullback, Information theory and statistics, John Wiley and Sons, Inc., New York, N Y, 1959.

4. М.К. Камалов. Распределение квадратичных форм, Издательство АН УзССР, Ташкент, 1958.