В Ы Ч И С Л И Т Е Л Ь Н Ы Е С И С Т Е М Ы
1972 год Выпуск 50

УДК 621.391:512.2

СТАТИСТИЧЕСКИЕ ОЦЕНКИ КАЧЕСТВА РАСПОЗНАЮЩИХ СИСТЕМ

А.Г. Францув, А.А. Кельзон, Б.В. Козловский, Б.П. Марков

Введение

Распознающие системы (классификаторы), предназначенные для отнесения объекта (ситуации) на основе наблюдений над ним к одному из конечного набора классов (альтернатив), детально рассматриваются в [1 - 7] и ряде других работ. Показателем качества классификатора обычно служит некоторая функция потерь, выбираемая в зависимости от предъявляемых требований. Так, например, байесов критерий минимизирует математическое ожидание потерь (средний риск), минимаксный критерий минимизирует максимум ожидаемых потерь (условный риск) и т.п.

Вычисление рисков для конкретного классификатора не представляет затруднений, если задано решающее правило классификатора и известна вероятностная структура входных воздействий в пространстве наблюдения. Однако при решении практических задач априорная информация недостаточна или чаще отсутствует вообще. Вычисление рисков классическими методами в таких условиях невозможно. В связи с этим на практике для оценки качества классификаторов широко используют относительные частоты ошибок на случайной независимой выборке (часто называемой "контрольной" или "экзаменационной"). Однако ввиду наличия статистического разброса между выборками и неучета имеющихся априорных сведений такими оценками, смысловое содержание и надежность последних для выборок ограниченного объема нуждаются в уточнении.

Ниже предлагается альтернативный подход, позволяющий строить различные статистические оценки качества классификаторов в условиях априорной неопределенности. Этот подход связан с построением апостериорных распределений условных вероятностей ошибочной классификации. Вид распределений, параметры которых вычисляются по результатам классификации случайной независимой выборки, при определенной форме представления имеющихся априорных данных оказывается не зависящим от характера пространства наблюдений, статистической структуры входных воздействий и типа классификатора. Данное обстоятельство позволяет получить простые выражения для расчета различных оценок качества классификаторов и построить оценки статистически надежного выигрыша по условным вероятностям ошибочных решений, обеспечиваемого при выборе одного из двух классификаторов.

Апостериорные распределения условных вероятностей ошибочной классификации

Всякий классификатор порождает конечное разбиение $\{\mathcal{A}_k\}$, $k = 1,2,\ldots,z$, некоторого измеримого пространства $(\Omega, \mathcal{A})$, где $\mathcal{A}$ - борелевское поле подмножеств множества Ω. Заметим, что пространство Ω не обязательно должно совпадать с пространством $\mathcal{X}$ наблюдаемых случайных векторных величин $X = (x_1, x_2, \ldots, x_n) \in \mathcal{X}$.

Пусть E - множество возможных состояний классификатора (предполагается, что классификатор может в процессе работы изменять свое состояние, например, в результате адаптации или самообучения). Сопоставим классификатору отображение $T:(\mathcal{X}, e) \to \Omega$, где $e \in E$. Пусть классификатор относит объект по наблюдению $x \in \mathcal{X}$ к i-му классу, если

$$\omega = Tx \in \Omega_i\,; \quad i = 1,\ldots,H,$$

где

$$\Omega_i = \bigcup_{k=t_{i-1}+1}^{t_i} A_k, \quad t_0=0, \quad t_H = r;$$

$$\bigcup_{i=1}^{H} \Omega_i = \Omega; \quad A_k \cap A_m = \varnothing; \quad k \neq m.$$

Ошибка классификации возникает в том случае, когда объект, принадлежащий i - му классу, порождает наблюдение $x \in \mathcal{X}$ такое, что $Tx \in \Omega_j$; i, j = I,2,..., H; $i \neq j$. Обозначим через $Q = \| q_{ik} \|$, $i=1,\dots,H; k=1,\dots,r$, матрицу,неизвестных условных вероятностей отображений наблюдений объектов $X \in \mathcal{X}$ в области A_k.

В дальнейшем будем употреблять сокращения: п.в. – для плотностей вероятностей и у.в. – для условных вероятностей.

Пусть имеется случайная независимая выборка, состоящая из $\sum_{i=1}^{H} N_i$ наблюдений, где N_i - число наблюдений из i - го класса. Рассмотрим матрицу

$$L_0 = \| l_{ik} \|, \quad i=1,\dots,H; \quad k=1,\dots,r,$$

где l_{ik} - число объектов i - го класса, попадающих в область A_k при отображении T. Рассмотрим далее векторы – столбцы матриц L_0 и Q:

$$l_i = \| l_{ik} \|_{k=\overline{1,r}}; \quad q_i = \| q_{ik} \|_{k=\overline{1,r}};$$

$$\sum_{k=1}^{r} l_{ik} = N_i; \quad \sum_{k=1}^{r} q_{ik} = 1.$$

Апостериорная п.в. $f(q_i \mid l_i)$ по теореме Байеса определится как

$$f(q_i \mid l_i) = \frac{P(l_i \mid q_i) f_a(q_i)}{P_a(l_i)}, \tag{I}$$

где $P(l_i|q_i)$ - у.в. случайного вектора l_i при известном векторе q_i;

$f_a(q_i)$ - априорная r - мерная п.в. неизвестного вектора q_i;

$P_a(l_i)$ –априорная вероятность случайного вектора l_i.

У.в. $P(l_i/q_i)$ определяется мультиномиальным законом

$$P(l_i/q_i) = c_1(l_i) q_{i1}^{l_{i1}} q_{i2}^{l_{i2}} \dots q_{ir}^{l_{ir}}; \tag{2}$$

$$0 < q_{ik} \leq 1; \quad k=1,\dots,r; \quad i=1,\dots,H,$$

где

$$c_1(l_i) = \frac{N_i!}{l_{i1}! \dots l_{ir}!}.$$

Семейство априорных распределений

$$F(q) \; (f_a(q_i) \in F(q). \; i=1,\dots,H,$$

должно удовлетворять следующим требованиям:

I) значения параметров q_{ik_r}, $i=1,\dots,H; k=1,\dots,r$, должны лежать в интервале [0,I], а $\sum_{k=1}^{r} q_{ik}=1$;

2) семейство F должно быть достаточно "богатым" в том смысле, чтобы существовал элемент $f \in F$, соответствующий имеющейся априорной информации;

3) семейство F должно быть достаточно легко интерпретируемо с тем, чтобы можно было проверить соответствие выбранного элемента F априорным данным.

Поставленным требованиям для априорных распределений векторов $q_i, i=1,\dots,H$, удовлетворяет семейство ($r-1$) – мерных распределений Дирихле $D(\xi_{i1}, \xi_{i2}, \dots, \xi_{i,r-1}; \xi_{ir})$ с параметрами $\xi_{ik} > 0$, $k=1,\dots,r$:

$$f_a(q_i) = \begin{cases} c_2(\xi_i) q_{i1}^{\xi_{i1}-1} \dots q_{ir}^{\xi_{ir}-1} & \text{при} \quad 0 < q_{ik} \leq 1, \\ 0 & \text{– в остальных случаях}, \end{cases} \tag{3}$$

где

$$c_2(\xi_i) = \frac{\Gamma(\xi_{i1} + \dots + \xi_{ir})}{\Gamma(\xi_{i1}) \dots \Gamma(\xi_{ir})}$$

Распределение (3) удобно и в том смысле, что оно является "воспроизводящим" распределением [6], [8], сопряженным с полиномиальным, что облегчает проведение требуемых преобразований. Заметим также, что распределение Дирихле играет большую роль в теории выборочных распределений порядковых статистик [9],позволяющей строить непараметрические оценки вероятностей попа –

дания наблюдений в области специального вида (блоки) в пространстве $\mathcal{X}$ (для непрерывных распределений наблюдений). С помощью упомянутой теории в [10], [11] и других работах показывается, что вероятности попадания наблюдений в систему специальным образом построенных параллелепипедов подчиняются распределению Дирихле.*)

Априорное распределение вероятностей $P_a(\ell_i)$ вектора ℓ_i можно найти из условий нормировки выражения (1)

$$\int f(q_i/\ell_i)dq_i = \frac{1}{P_a(\ell_i)}\int f_a(q_i)P(\ell_i/q_i)dq_i = 1,$$

откуда с учетом (2) и (3) получим

$$P_a(\ell_i) = \int f_a(q_i)P(\ell_i/q_i)dq_i = C_1(\ell_i)C_2(\xi_i)\int_{S_{i,z-1}} q_{i1}^{\ell_{i1}+\xi_{i1}-1}\cdots q_{i,z-1}^{\ell_{i,z-1}+\xi_{i,z-1}-1} \times \left(1-\sum_{k=1}^{z-1} q_{ik}\right) dq_{i1}\, dq_{i2}\ldots dq_{i,z-1}, \qquad (4)$$

где интегрирование проводится по симплексу

$$S_{i,z-1} = \left\{(q_{i1},\ldots,q_{i,z-1});\ q_{ik}\geq 0,\ k=1,\ldots,z-1;\ \sum_{k=1}^{z-1} q_{ik}\leq 1\right\}$$

В [9] показано, что $\int$ в (4) равен

$$\int_{S_{i,z-1}} = \frac{\Gamma(\ell_{i1}+\xi_{i1})\ldots\Gamma(\ell_{iz}+\xi_{iz})}{\Gamma\left[\sum_{k=1}^{z}(\ell_{ik}+\xi_{ik})\right]} = C_3(\ell_i+\xi_i). \qquad (5)$$

Подставляя (5) в (4), получим

$$P_a(\ell_i) = C_1(\ell_i)C_2(\xi_i)C_3(\ell_i+\xi_i). \qquad (6)$$

Подстановка в (1) выражений (2), (3) и (6) и проведение сокращений показывают, что апостериорное распределение $f(q_i/\ell_i)$

*) В.А. Герман обратил внимание авторов на то, что вероятность попадания в систему параллелепипедов более общего вида подчиняется распределению Дирихле, параметры которого определяются по числу выборочных объектов и числу параллелепипедов.

является $(z-1)$ – мерным распределением Дирихле*)

$$D(\ell_{i1}+\xi_{i1},\ldots,\ell_{i,z-1}+\xi_{i,z-1};\ \ell_{iz}+\xi_{iz}).$$

Рассмотрим множество $\Omega_j(\Omega)$, подмножества которого A_k; $k=(t_{j-1}+1),\ldots,t_j$ занумеруем для удобства порядковыми числами от 1 до S_j. Маргинальное распределение векторной случайной величины $q_{i1},q_{i2},\ldots,q_{iS_j}$, $S_j<z$, является S_j-мерным распределением Дирихле [9].

$$D\left(\ell_{i1}+\xi_{i1},\ldots,\ell_{iS_j}+\xi_{iS_j};\ \sum_{k=S_j+1}^{z}(\ell_{ik}+\xi_{ik})\right).$$

Тогда случайная величина $W_{ij}=\sum_{k=1}^{S_j} q_{ik}$ – п.в. отображения наблюдения над объектом, принадлежащим к i-му классу, в область $\Omega_j\subset\Omega$; $i,j=1,\ldots,H$, имеет бета-распределение [9]

$$Be\left(\sum_{k=1}^{S_j}(\ell_{ik}+\xi_{ik}),\ \sum_{k=S_j+1}^{z}(\ell_{ik}+\xi_{ik})\right). \qquad (7)$$

Обозначив через $\ell_i(\Omega_j)=\sum_{k=1}^{S_j}\ell_{ik}$ число объектов i-го класса из выборки объема N_i, отнесенных при классификации к j-му классу,

$N_i=\sum_{j=1}^{H}\ell_i(\Omega_j)$, и через $\xi_i(\Omega_j)=\sum_{k=1}^{S_j}\xi_{ik}$, $N_{ia}=\sum_{j=1}^{H}\xi_i(\Omega_j)$ –

– параметры априорного распределения, а также положив

$$u_{ij}=\ell_i(\Omega_j)+\xi_i(\Omega_j);\quad i,j=1,\ldots,H, \qquad (8)$$

и

$$N_i' = N_i+N_{ia},\quad i=1,\ldots,H, \qquad (9)$$

получим окончательное выражение для апостериорной п.в. $f(W_{ij}/u_{ij},N_i')$ классификации объекта из i-го класса как объекта j-го класса в виде бета-распределения:

$$f(W_{ij}/u_{ij},N_i')=\frac{\Gamma(N_i')}{\Gamma(u_{ij})\Gamma(N_i'-u_{ij})}W_{ij}^{u_{ij}-1}(1-W_{ij})^{N_i'-u_{ij}-1} \qquad (10)$$

*) Тот же результат может быть получен из (3), исходя из свойств воспроизводящих распределений.

Если в формуле (I0) положить $i=j$, то получим выражение для апостериорной п.в. правильной классификации, если $i\neq j$ – для апостериорной п.в. ошибочной классификации, заключающейся в отнесении к j-му классу объекта из i-го класса.

Заметим, что при выводе выражения (I0) не использовались какие-либо ограничивающие предположения относительно характера пространства наблюдений, вида условных распределений $F_i(x)$, $i=1,\dots,H$, или структуры классификатора. Таким образом, апостериорный закон распределения вероятностей правильной и ошибочной классификации является универсальным. Очевидно, что в случаях полного априорного знания или асимптотического стремления объема выборки к бесконечности бета-распределение (I0) вырождается в дельта-функцию $\delta(W_{ij}-W^o_{ij})$, где W^o_{ij} – истинное значение параметра W.

Если априорная информация о распределении параметра W_{ij} отсутствует вовсе, то можно принять постулат Байеса [I2] о равновероятных априорных гипотезах. В общем случае это соответствует равноплотному распределению на симплексе

$$S_{i,r-1}=\left\{(q_{i1},\dots,q_{i,r-1});\ q_{ik}\geq 0;\ k=1,\dots,r-1;\ \sum_{k=1}^{r-1} q_{ik}\leq 1\right\},$$

т.е. $(r-1)$-мерному распределению Дирихле D (I,...,I;I). В этом случае из (7) видно, что случайная величина W_{ij} имеет бета-распределение

$$Be\left(\sum_{k=1}^{S_j} l_{ik}+S_j,\ \sum_{k=S_j+1}^{r} l_{ik}+(r-S_j)\right)=$$

$$=Be(l_i/\Omega_j)+S_j,\ N_i+r-l_i(\Omega_j)-S_j), \qquad (I0')$$

т.е. в параметрах (8) и (9) распределения (I0) $\xi_i(\Omega_j)$ заменяется на S_j, а N_{iQ} – на r.

Распределения (I0) – (I0') могут быть использованы для построения статистических оценок качества классификаторов.

Апостериорный риск

Пусть для некоторого классификатора определены матрица

$$F=\|f(W_{ij}/u_{ij},N_i')\|=\|f_{ij}(w)\|$$

апостериорных п.в. правильной и ошибочной классификации (первые из них образуют диагональные элементы матрицы F), матрица потерь $C(W)=\|C_{ij}(W)\|$ и вектор $P(i)$ априорных вероятностей классов $i,j=1,\dots,H$. Тогда апостериорный риск $R(C,F)$ определяется выражением

$$R(C,F)=\sum_{i,j=1}^{H}\int_0^1 P(i)C_{ij}(w)f_{ij}(w)dw. \qquad (II)$$

Важным частным случаем является апостериорный риск при линейной функции потерь $C_{ij}=a_{ij}w$:

$$R(C,F)=\sum_{i,j=1}^{H}P(i)a_{ij}\int_0^1 w f_{ij}(w)dw \qquad (I2)$$

Учитывая, что $\int w f_{ij}(w)dw=M_w(w)$ и вычисляя математическое ожидание $M^o_w(w)$ по апостериорному распределению (I0), получим из (I2)

$$R(C,F)=\sum_{i,j=1}^{H}\frac{P(i)a_{ij}u_{ij}}{N_i'}=\sum_{i=1}^{H}\frac{P(i)}{N_i'}\sum_{i=1}^{H}a_{ij}u_{ij}. \qquad (I3)$$

Если считать потери при правильных решениях нулевыми и положить $a_{ij}=a_i$ при всех j, т.е.

$$a_{ij}=\begin{cases}0 & \text{при } i=j,\\ a_i & \text{при } i\neq j,\end{cases}$$

то получим из (I3)

$$R(C,F)=\sum_{i=1}^{H}P(i)a_i\left(1-\frac{u_i}{N_i'}\right), \qquad (I4)$$

где

$$u_i=u_{ii}=N_i'-\sum_{j\neq i}u_{ij}.$$

Минимальная оценка $R(L,F)$ получается из (I4) при

$$\xi_i(\Omega_i)=S_i,\quad N_{iQ}=r;\ i,j=1,\dots,H;$$

$$\inf_F R(L,F)=\sum_{i=1}^{H}P(i)a_i\left(1-\frac{l_i(\Omega_i)+S_i}{N_i+r}\right). \qquad (I5)$$

Оптимальный выбор классификатора из группы классификаторов

Как известно [I3], оптимальный (байесов) выбор решающего правила заключается в сравнении апостериорных рисков для рассматриваемых решающих правил и выборе правила по минимуму апостериорного риска. Каждый классификатор можно отождествить с соответствующим решающим правилом и применить байесов выбор по одному из выражений предыдущего раздела для апостериорного риска. Можно показать, что частными случаями такого общего подхода являются выбор классификатора по критериям идеального наблюдателя и Неймана-Пирсона. Критерий идеального наблюдателя означает выбор классификатора, реализующего минимум вероятности ошибочных решений. Этому условию соответствует апостериорный риск

$$R_{uH} = 1 - \sum_{i=1}^{H} P(i)\frac{u_i}{N_i'} . \qquad (I6)$$

Нетрудно также показать, что выбор классификатора по критерию Неймана-Пирсона сводится к фиксации величины $\frac{u_1}{N_1'} = const$ для сравниваемых классификаторов и выбору классификатора, минимизирующего величину

$$R_{HП} = 1 - \frac{u_2}{N_2'}, \qquad (I7)$$

Оптимальные точечные оценки параметра W

В ряде случаев требуется построить оптимальную оценку параметра W. Построим оптимальную (байесову) оценку $\hat{W}^*$ параметра W, то есть определим оценку $\hat{W}$, дающую минимум среднего риска, когда для каждой пары (W, $\hat{W}$) заданы потери C (W, $\hat{W}$). При квадратичной функции потерь C (W, $\hat{W}$) $= \lambda$ (W)($W - \hat{W}$) байесова оценка $\hat{W}^*$ параметра W определяется из соотношения [I3]:

$$\hat{W}^* = \frac{M_W(\lambda_W/u)}{M_W(\lambda/u)} , \qquad (I8)$$

где математическое ожидание M_W определяется по апостериорному распределению (I0).

Если, например, принять C (W, $\hat{W}$) $= \frac{(W-\hat{W})^2}{W(1-W)}$, то $\lambda(W) = \frac{1}{W(1-W)}$ и из (I8) следует, что (индексы опускаем)

$$\hat{W}^* = \frac{\int_0^1 \frac{W^u(1-W)^{N'-u-1}}{W(1-W)}dW}{\int_0^1 \frac{W^{u-1}(1-W)^{N'-u-1}}{W(1-W)}dW} =$$

$$= \frac{\Gamma(u)\Gamma(N'-u-1)\Gamma(N'-2)}{\Gamma(N'-1)\Gamma(u-1)\Gamma(N'-u-1)} = \frac{u-1}{N'-2} = \frac{\ell-\xi+1}{N+N_a-2} \qquad (I9)$$

Сравнивая байесову оценку (I9) при принятой функции потерь с оценкой максимального правдоподобия $\hat{W} = \frac{\ell}{N}$, можно заметить, что $\hat{W} = \hat{W}^*$ лишь в случае

$$\frac{\xi-1}{N_a-2} = \frac{\ell}{N} \quad \text{или} \quad \xi = 1, N_a = 2 .$$

Поскольку ℓ - случайная величина, то равенство $\hat{W} = \hat{W}^*$ будет всегда справедливо только при $\xi = 1$, $N_a = 2$, что соответствует прямоугольному априорному распределению п.в. ошибок с плотностью I для двух областей решений.

Интервальные оценки параметра W

Доверительный интервал $[a, b]$ для параметра W можно определить из соотношения

$$P\{a \le W \le b\} = \int_a^b f(W)dW = \gamma_g , \qquad (20)$$

где $f(W)$ - апостериорная п.в. W; γ_g - двусторонний доверительный уровень. Принимая во внимание, что из (I0) следует равенство (индексы опускаем)

$$\int_0^y f(W)dW = I_y(u, N'-u) , \qquad (2I)$$

где $I_y(u, N'-u)$ - неполная бета-функция,

$$
I_y(u, N'-u) = \begin{cases} 0 & \text{при} \quad y < 0, \\ \dfrac{\Gamma(N')}{\Gamma(u)\Gamma(N'-u)} \displaystyle\int_0^y t^{u-1}(1-t)^{N'-u-1}\, dt & \text{при} \quad 0 \leq y \leq 1, \\ 1 & \text{при} \quad y > 1 \end{cases}
$$

можно переписать (20) в виде

$$
I_b(u, N'-u) - I_a(u, N'-u) = \gamma_g . \tag{22}
$$

Для практических целей наибольшее значение имеет верхняя доверительная граница b для параметра w, которая, как нетрудно увидеть, определяется из выражения

$$
I_b(u, N'-u) = \gamma , \tag{23}
$$

где γ – односторонний доверительный уровень.

На рис. I представлены графики $b = \varphi(N)$, построенные для доверительного уровня γ = 0,9 при различных значениях параметра $\hat{w}^*$ из (I9). Значения неполной бета-функции при различных параметрах были взяты из графиков Мёрфи [I4], приведенных в виде $I_{1-\beta}(m, n-m+1) = \alpha$. Пересчет в графики $b = \varphi(N)$ проводился с помощью соотношений (в правых частях соотношений приводятся параметры графиков Мёрфи):

$$
N = n-1, \quad \hat{w}^* = \frac{m-1}{n-1}, \quad \gamma = \alpha, \quad b = 1-\beta
$$

Для параметров априорного распределения $\xi = 1$ и $N_Q = 2$ параметры кривых на рисунке представляют собой частоты ошибок на выборке. Так, например, если на выборке объемом N =200 объектов дихотомический классификатор сделал ℓ = 30 ошибок, то при отсутствии априорной информации с вероятностью 90% можно утверждать, что вероятность ошибки на генеральной совокупности не превысит значения 0,185.

Оценка качества решения при выборе классификатора

Сформулируем эту задачу в точных терминах. Имеются два классификатора V_1 и V_2 , для которых определены значения

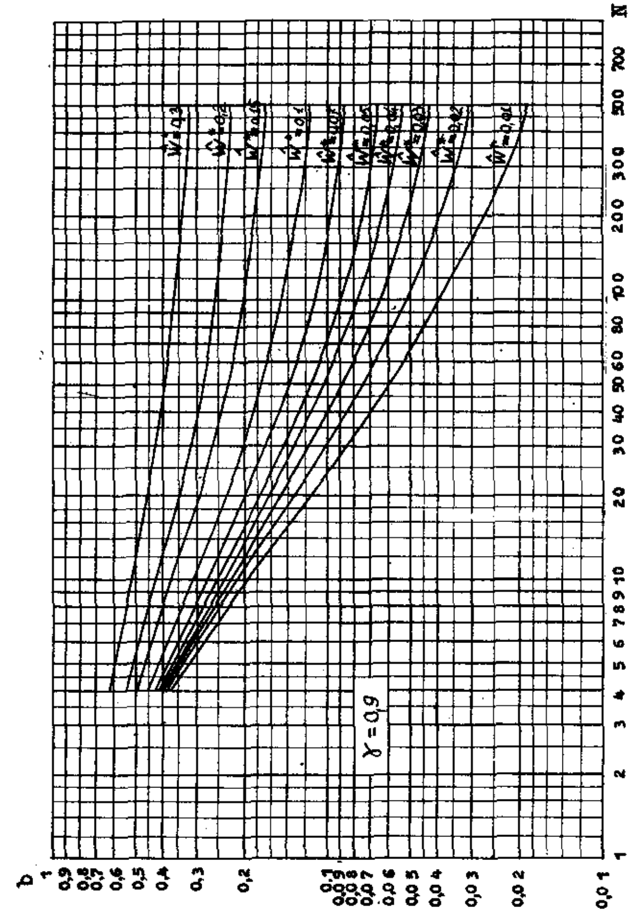

Рис. I. Зависимость верхнего доверительного предела вероятности ошибки от объема выборки.

критериев качества R_1 и R_2, причем $R_1 < R_2$. Заменяя классификатор V_2 классификатором V_1 при доверительном уровне γ, требуется определить нижнюю границу выигрыша по качеству α, то есть решить относительно α уравнение:

$$P\{R_2 - R_1 \geqslant \alpha\} = \gamma. \tag{24}$$

Для решения этой задачи в общем виде необходимо знание апостериорных законов распределения критериев качества, которыми, за исключением частных случаев, мы не располагаем. Таким частным случаем, например, является критерий Неймана-Пирсона, для которого известно апостериорное распределение, совпадающее с соответствующим распределением вероятности ошибки. В этом случае выражение (24) примет вид

$$P\{W_2 - W_1 \geqslant \alpha\} = \gamma;\ \alpha, \gamma \in [0,1], \tag{25}$$

где W_1, W_2 – у.в. ошибок второго рода при одинаковых у.в. ошибок первого рода для классификаторов V_1 и V_2, соответственно.

Пусть $t = W_2 - W_1$. В предположении независимости случайных переменных W_1 и W_2 п.в. $\rho(t)$ случайной переменной t можно представить сверткой

$$\rho(t) = \int^{1-t} f_1(z) f_2(z_2 + t)\, dz, \tag{26}$$

где $f_i(W)$ – апостериорная плотность вероятностей W_i; $i = 1,2$.

Если выразить левую часть неравенства (25) через $\rho(t)$ из (26) с учетом того, что $f_2(t+z) = 0$ при $z > 1-t$, а это позволяет изменить верхний предел интегрирования по переменной z на I, то получим

$$P\{t \geqslant \alpha\} = \int_\alpha^{1} \rho(t)\, dt = \int_\alpha^1 \Big[\int_0^1 f_1(z) f_2(t+z)\, dz\Big] dt. \tag{27}$$

Поменяв местами внешнюю и внутреннюю переменные интегрирования и введя подстановку $t + z = x$, имеем:

$$P\{t \geqslant \alpha\} = \int_0^1 f_1(z) \Big[\int_{\alpha+z}^1 f_2(x)\, dx\Big] dz = \int_0^{1-\alpha} \big[1 - I_{\alpha+z}(u_2, N_2' - u_2)\big] \times$$

$$\times f_1(z)\, dz = \int_0^{1-\alpha} I_{1-\alpha-z}(N_2' - u_2, u_2) f_1(z)\, dz = M_{W_1}\{I_{1-\alpha-W_1}(N_2' - u_2, u_2)\}. \tag{28}$$

Здесь и ниже числовые индексы соответствуют номерам классификаторов.

Подставив (28) в (25), получим окончательно уравнение для оценки нижней доверительной границы выигрыша по качеству при выборе лучшего из двух классификаторов по критерию Неймана-Пирсона с доверительным уровнем γ

$$M_{W_1}\{I_{1-\alpha-W_1}(N_2' - u_2, u_2)\} = \gamma. \tag{29}$$

Очевидно, что выражение (29) можно использовать также для определения нижней доверительной границы α уменьшения произвольной у.в. ошибки или увеличения у.в. правильной классификации по их байесовым оценкам $\hat{w}_1^*$ и $\hat{w}_2^*$ для классификаторов V_1 и V_2, соответственно.

Левая часть выражения (29) представляет собой определенный интеграл от неполной бета-функции, который не выражен через элементарные функции и не табулирован. Вычисление его значений возможно лишь с помощью ЭЦВМ. Соответствующая программа, разработанная Г. Кузьминой для ЭЦВМ М-220, была использована для построения графиков типа показанных на рис. 2-4.

На рис. 2 представлены графики $\gamma = \varphi_1(\hat{w}_1^*)$, построенные для $\hat{w}_2^* = 0,1\ (0,1)\ 0,9;\ 0,98$ при $\alpha = 0$, $N = 100$, где

$$\hat{w}_\nu^* = \frac{u_\nu - 1}{N - 2}, \quad \nu = 1,2 \text{ – номер классификатора.}$$

По графикам рис. 2 можно определять при заданном $\hat{w}_2^*$ значение $\hat{w}_1^*$, обеспечивающее с произвольным доверительным уровнем γ преимущество классификатора V_1 по сравнению с V_2. Так, например, если $\hat{w}^*$ – оценки у.в. ошибок, то при $\hat{w}_2^* = 0,2$ и $\gamma = 0,9$ V_1 лучше V_2 при $\hat{w}_1^* \leqslant 0,105$. Напомним, что в качестве $\hat{w}^*$ могут фигурировать байесовы оценки как у.в. ошибок, так и у.в. правильной классификации, а при отсутствии априорной информации в качестве оценок $\hat{w}^*$ могут использоваться относительные частоты событий на выборке.

На рис. 3 представлены графики $\alpha = \varphi_2(\hat{w}_1^*)$, построенные для $\hat{w}_2^* = 0,1\ (0,1)\ 0,9;\ 0,98$ при $\gamma = 0,9$ и $N = 100$ в представляющей наибольший интерес области изменений $\alpha = 0 \div 0,2$. Из рис. 3 видно, что эти графики являются отрезками прямых линий, хотя в выражении (29) трудно подметить такую закономерность. Отсюда следует, что для соответствующих значений параметров выигрыш по α линейно зависит от разности между байесовыми оценками параметра w для сравниваемых классификаторов.

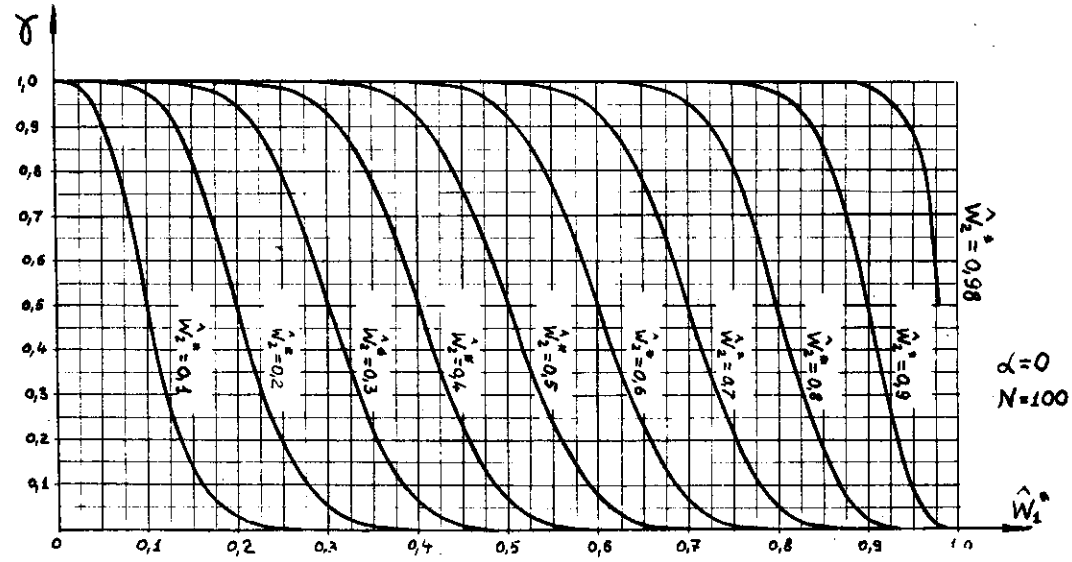

Рис. 2. Зависимость доверительной вероятности правильной классификации от величины ошибки классификации одного из двух классификаторов.

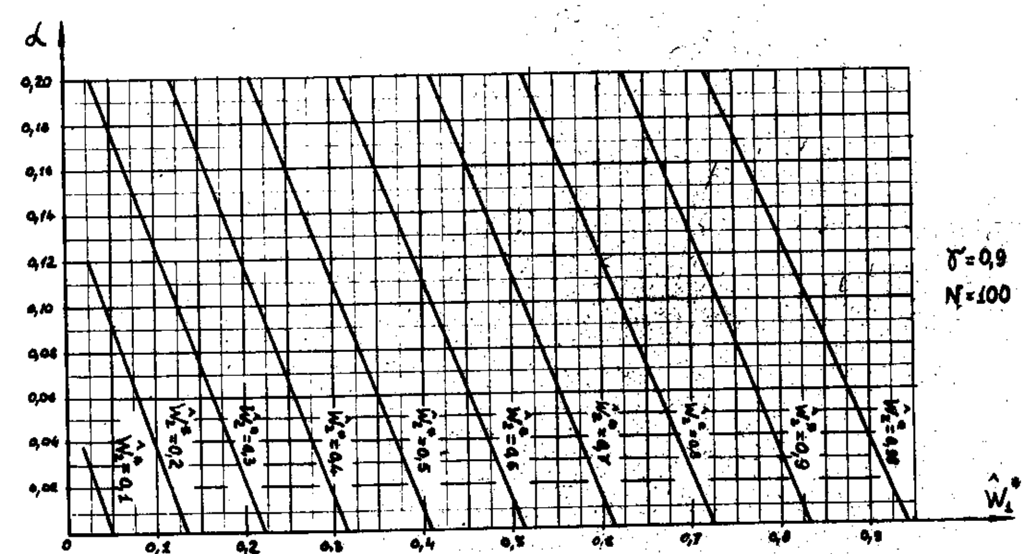

Рис. 3. Зависимость выигрыша качества при замене классификатора V_2 классификатором $V_{1,2}$

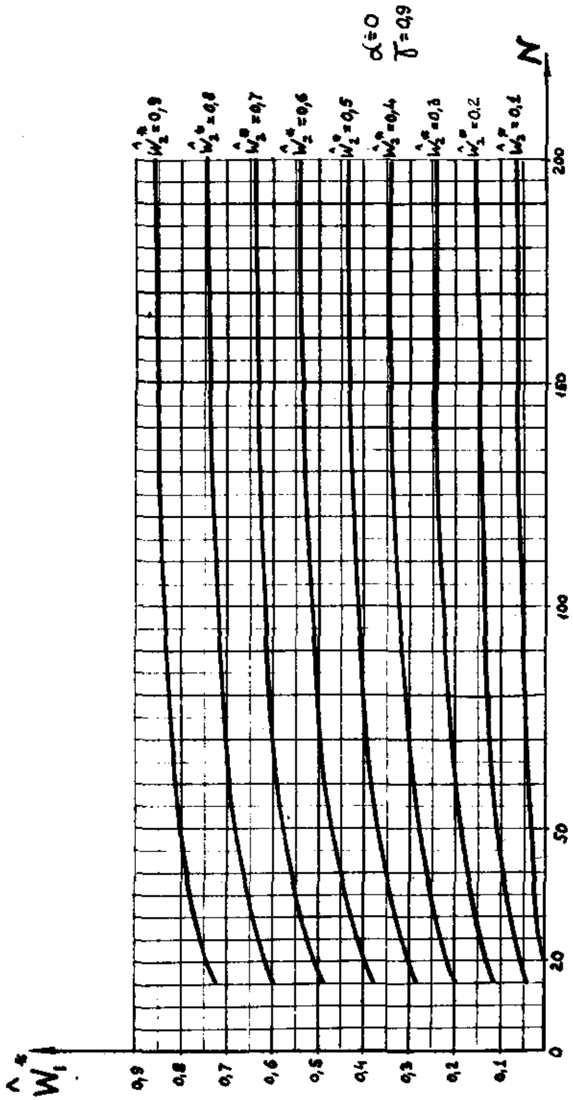

Рис. 4. Зависимость качества классификатора V_1 от объема выборки.

Возможное использование графиков на рис. 3 поясним на примере. Пусть байесовы оценки у.в. правильных решений $\hat{w}_1^* = 0,6$ и $\hat{w}_2^* = 0,8$. Тогда в соответствии с рис. 3 можно утверждать, что классификатор V_2 с $\gamma = 0,9$ обеспечивает повышение у.в. правильной классификации на I2%.

На рис. 4 приведены графики $\hat{w}_1^* = \varphi_3(N)$, построенные для $\hat{w}_2^* = 0,I\ (0,I)\ 0,9$ при $\alpha = 0$, $\gamma = 0,9$.Область лежащая ниже кривой для некоторого значения параметра $\hat{w}_2^*$, соответствует значениям $\hat{w}_1^*$ для классификатора V_1, обеспечивающих его улучшение по параметру w по сравнению с V_2. Так, например, если w - у.в. ошибок и $\hat{w}_2^* = 0,3$, а $\hat{w}_1^* = 0,2$, то при выборке объема $N = 80$ можно с доверительным уровнем 0,9 утверждать, что V_1 лучше V_2 по параметру w, а при выборке объема $N = 50$ такого утверждения сделать нельзя.

Заметим, что кривые на рис. 4 идут достаточно полого,особенно на участке N = I00 - 200. Это обстоятельство позволяет сделать важный для практики вывод, что при ограниченных объемах выборок, находящихся в распоряжении исследователя,можно основную часть выборки использовать для синтеза (обучения или самообучения) классификатора и оставлять для контроля разумно малый объем выборки.

Заключение

Применению классической теории статистических решений для оценки качества классификатора препятствует неполное или полное отсутствие априорной информации относительно многомерных распределений наблюдений. В целях обеспечения возможности практического применения критериев теории статистических решений для количественного анализа качества классификаторов в условиях априорной неопределенности необходимо располагать апостериорными распределениями правильной и ошибочной классификации.

Показано, что при принятой в работе форме представления априорной информации (в частности, и для случая, когда последняя отсутствует) апостериорные вероятности правильной и ошибочной классификации подчиняются закону бета-распределения независимо от структуры пространства наблюдений, вероятностных мер в этом пространстве и типа классификаторов.

Параметры апостериорного распределения в общем случае легко определяются по параметрам априорного распределения, общим параметрам классификаторов и числам ошибок классификации на случайной независимой выборке произвольного (известного) объема.

Построенные таким способом апостериорные распределения позволяют получить простые выражения для апостериорного риска и оптимальных точечных оценок вероятностей правильной и ошибочной классификации, построить доверительные интервалы для этих вероятностей, осуществить оптимальный выбор классификатора из группы классификаторов.

Поставлена задача оценки нижней доверительной границы выигрыша по качеству, обеспечиваемого при выборе одного из двух классификаторов, и получено решение этой задачи для выигрыша по условиям вероятностям ошибочных (или правильных) решений.

Приведенные результаты могут быть использованы для статистического анализа качества классификаторов и сравнения различных типов и вариантов классификаторов друг с другом.

Литература

1. ГЛУШКОВ В.М. Введение в кибернетику, Киев, 1964, изд. АН УССР.

2. СЕБАСТИАН Г.С. Процессы принятия решений при распознавании образов, Киев, "Техника", 1965.

3. ЯКУБОВИЧ В.А. Некоторые общие теоретические принципы построения обучаемых опознающих систем. - "Вычислительная техника и вопросы программирования", изд. ЛГУ, Л., 1965, 4.

4. ЦЫПКИН Я.З. Адаптация и обучение в автоматических системах, М., "Наука", 1968.

5. ИВАХНЕНКО А.Г. Самообучающиеся системы распознавания и автоматического управления. Киев, "Техника", 1969.

6. ВАСИЛЬЕВ В.И. Распознающие системы. Киев, "Наукова думка", 1969.

7. ЗАГОРУЙКО Н.Г. Структура проблемы распознавания слуховых образов и методы ее решения.-"Распознавание слуховых образов и методы ее решения". Новосибирск, "Наука", 1970.

8. SPRAGINS I.D. A note on the iternative application of Bayes Rule. IEEE Trans.Inform. Theory II,1965, Oct., N 4.

9. УИЛКС С. Математическая статистика. М., "Наука", 1967.

10. ANDERSON T.W. Some nonparametric multivariate prob - lems based on Statistic. equivalent blocks. Multivariate analaysis Academic Press, N.-Y. and L., 1966.

11. HENRICHON E.G., FU K.S. A nonparametric partit. procadure for parrern classification. IEEE Trans, 1969,vol.C-18.

12. КЕНДАЛЛ М.Д., СТЬЮАРТ А. Теория распределений.М.,"Нака", 1966.

13. БЛЕКУЭЛЛ Д., ГИРШИК М.А. Теория игр и статистических решений. М., ИЛ,1958.

14. MURPHY R.B. Nonparametric tolerance limits. The annals of math. Statistics, 1948, vol.XIX, N 4.

Поступила в ред.-изд.отд.
29. X. 1971 г.