УДК 004.93

© 2018 г. **А.В. Лапко** [1,2], д-р техн. наук,
В.А. Лапко [1,2], д-р техн. наук
[1](Институт вычислительного моделирования СО РАН, Красноярск)
[2](Сибирский государственный университет науки и технологий
им. академика М.Ф. Решетнева, Красноярск)

СИНТЕЗ И АНАЛИЗ КОЛЛЕКТИВА НЕПАРАМЕТРИЧЕСКИХ РЕШАЮЩИХ ФУНКЦИЙ В ДВУАЛЬТЕРНАТИВНОЙ ЗАДАЧЕ РАСПОЗНАВАНИЯ ОБРАЗОВ

Предлагаемая методика синтеза коллектива непараметрических уравнений разделяющих поверхностей в двуальтернативной задаче распознавания образов, основана на декомпозиции обучающей выборки по ее объему. Исследуются асимптотические свойства коллектива решающих функций. Проводится сравнение ее свойств с традиционной непараметрической оценкой уравнения разделяющей поверхности между классами.

Ключевые слова: распознавание образов, непараметрическая статистика, выборки большого объема, коллективы решающих функций, асимптотические свойства.

DOI: 10.22250/isu.2018.55.88-98

Введение

Непараметрические алгоритмы распознавания образов, основанные на оценках плотности вероятности типа Розенблатта – Парзена [1 – 5], получили широкое распространение в задачах исследования систем классификации при априорной неопределенности [6 – 12]. Вычислительная эффективность традиционных непараметрических алгоритмов распознавания образов значительно снижается с увеличением объема обучающих выборок. Поэтому возникает необходимость разработки непараметрических систем принятия решений, основанных на декомпозиции обучающих выборок большого объема и использовании технологии параллельных вычислений.

В данной работе исследуются асимптотические свойства коллектива непараметрических уравнений разделяющих поверхностей в двуальтернативной задаче распознавания образов. На этой основе определяются количественные зависимости его аппроксимационных свойств от числа составляющих коллектива решающих функций и степени неравномерности распределения элементов обучающей выборки между классами.

Построение коллектива непараметрических уравнений разделяющих поверхностей

Пусть $V=\left(x^i,\sigma\left(x^i\right),\ i=\overline{1,n}\right)$ – обучающая выборка объема n, составленная из признаков $x^i=\left(x^i_j,\ j=\overline{1,k}\right)$ классифицируемых объектов и соответствующих «указаний учителя» $\sigma\left(x^i\right)$ об их принадлежности к одному из двух классов Ω_1, Ω_2.

Объем n обучающей выборки V является большим, что снижает вычислительную эффективность непараметрических алгоритмов распознавания образов.

Предлагаемый подход при решении задачи распознавания образов в данных условиях реализуется при выполнении следующих действий:

1. Осуществить декомпозицию исходной выборки V на части $V_j=\left(x^i,\sigma\left(x^i\right),\ i\in I_j\right)$, $j=\overline{1,N}$, где I_j – множество номеров ситуаций из V, составляющих j-ю группу ситуаций. Будем считать, что количество элементов n_j множества I_j, $j=\overline{1,N}$ одинаково и равно $\bar{n}$, причем $N={n}/{\bar{n}}$.

2. По полученным данным V_j построить непараметрические решающие правила

$$\bar{m}_j(x):\begin{cases} x\in\Omega_1\,,\ если\ \bar{f}^{\,j}_{12}(x)\le 0 \\ x\in\Omega_2\,,\ если\ \bar{f}^{\,j}_{12}(x)>0\,,\ j=\overline{1,N}\,, \end{cases} \tag{1}$$

где непараметрические оценки уравнений разделяющих поверхностей определяются выражением

$$\bar{f}^{\,j}_{12}(x)=\bar{p}^{\,j}_2(x)-\bar{p}^{\,j}_1(x),\ \ j=\overline{1,N}. \tag{2}$$

В выражении (2) непараметрическая оценка плотности вероятности распределения признаков x анализируемых объектов в s-м классе представляются статистикой типа [1, 2]

$$\bar{p}^{\,j}_s(x)=\frac{1}{\bar{n}_s\prod_{v=1}^{k}c^j_v}\sum_{i\in I^j_s}\prod_{v=1}^{k}\Phi\left(\frac{x_v-x^i_v}{c^j_v}\right),\ s=1,2,$$

где $\bar{n}_s$ – количество ситуаций s-го класса в обучающей выборке V_j, а $I_s^j \subset I_j$ – множество их номеров. Ядерные функции $\Phi(u)$ удовлетворяют условиям:

$$\Phi(u) = \Phi(-u), \quad 0 \le \Phi(u) < \infty,$$

$$\int \Phi(u)du = 1, \quad \int u^2\,\Phi(u)du = 1,$$

$$\int u^m\,\Phi(u)du < \infty, \quad 0 \le m < \infty .$$

Коэффициенты размытости c_ν^j, $\nu = \overline{1,k}$ ядерных функций убывают с ростом количества элементов $\bar{n}$ множеств I_j, $j = \overline{1, N}$. Здесь и далее бесконечные пределы интегрирования опускаются.

Оптимизация частных решающих правил (1) по коэффициентам размытости c_ν^j, $\nu = \overline{1, k}$ осуществляется в режиме «скользящего экзамена» из условия минимума статистической оценки вероятности ошибки распознавания образов

$$\bar{\rho}_j = \frac{1}{\bar{n}} \sum_{t \in I_j} 1\left(\sigma\left(x^t\right), \bar{\sigma}\left(x^t\right)\right), \; j = \overline{1, N},$$

$$1\left(\sigma\left(x^t\right), \bar{\sigma}\left(x^t\right)\right) = \begin{cases} 0, \textit{если } \sigma\left(x^t\right) = \bar{\sigma}\left(x^t\right) \\ 1, \textit{если } \sigma\left(x^t\right) \neq \bar{\sigma}\left(x^t\right), \end{cases}$$

где $\bar{\sigma}\left(x^t\right)$ – «решение» алгоритма (1) о принадлежности ситуации x^t к одному из двух классов. При формировании решения $\bar{\sigma}\left(x^t\right)$ ситуация x^t исключается из процесса обучения в непараметрической статистике (2).

3. Сформировать решающее правило второго уровня структуры системы классификации

$$\bar{\bar{m}}(x) : \begin{cases} x \in \Omega_1 \,, \textit{если } \bar{\bar{f}}_{12}(x) \le 0 \\ x \in \Omega_2 \,, \textit{если } \bar{\bar{f}}_{12}(x) > 0 \,, \end{cases} \tag{3}$$

где обобщенная решающая функция определяется выражением

$$\bar{\bar{f}}_{12}(x) = \frac{1}{N} \sum_{j=1}^{N} \bar{f}_{12}^{\,j}(x). \tag{4}$$

Нетрудно показать, что вычислительная эффективность предложенной двухуровневой системы классификации при использовании технологий параллельных вычислений повышается в N раз по сравнению с традиционным непараметрическим алгоритмом.

Без существенной потери общности рассмотрим асимптотические свойства $\bar{\bar{f}}_{12}(x)$ для одномерного случая. Будем полагать, что $p_1(x)$, $p_2(x)$ и первые две их производные ограничены и непрерывны. В дальнейших преобразованиях будем использовать технологию [2].

Определим выражение

$$M\left(\bar{\bar{f}}_{12}(x)\right)=\frac{1}{N}\sum_{j=1}^{N}M\left(\bar{p}_2^j(x)-\bar{p}_1^j(x)\right)=$$

$$=\frac{1}{N}\sum_{j=1}^{N}\left[\frac{1}{c^j}\int\Phi\left(\frac{x-t}{c^j}\right)p_2(t)\,dt-\frac{1}{c^j}\int\Phi\left(\frac{x-t}{c^j}\right)p_1(t)\,dt\right],$$

где M – знак математического ожидания, а $c^j=c(\bar{n})=c$.

Проведем в интегралах последнего выражения замену переменных $(x-t)c^{-1}=u$ и, разлагая функции $p_s(x-cu)$, $s=1,2$ в ряд Тейлора в точке x, с учетом свойств ядерной функции при достаточно больших значениях $\bar{n}_1$, $\bar{n}_2$, получим

$$M\left(\bar{\bar{f}}_{12}(x)-f_{12}(x)\right)\sim\frac{c^2}{2}\left(p_2^{(2)}(x)-p_1^{(2)}(x)\right), \qquad (5)$$

где $p_s^{(2)}(x)$ – вторая производная плотности вероятности $p_s(x)$ по x, $s=1,2$.

Отсюда, из условия $c\to 0$ при $\bar{n}_1\to\infty$, $\bar{n}_2\to\infty$, следует свойство асимптотической несмещенности непараметрической статистики $\bar{\bar{f}}_{12}(x)$ вида (4).

Рассмотрим выражение

$$M\int\left(\bar{\bar{f}}_{12}(x)-f_{12}(x)\right)^2dx=M\int\left(\frac{1}{N}\sum_{j=1}^{N}\left(\bar{f}_{12}^j(x)-f_{12}(x)\right)\right)^2dx=$$

$$=\frac{1}{N^2}M\left(\sum_{j=1}^{N}\int\left(\bar{f}_{12}^j(x)-f_{12}(x)\right)^2dx+\sum_{j=1}^{N}\sum_{\substack{t=1\\t\neq j}}^{N}\int\left(\bar{f}_{12}^j(x)-f_{12}(x)\right)\left(\bar{f}_{12}^t(x)-f_{12}(x)\right)dx\right). \qquad (6)$$

Найдем асимптотическое выражение для среднеквадратического отклонения $\bar{f}_{12}^j(x)$ от $f_{12}(x)$:

$$M\int\left(f_{12}(x)-\bar{f}_{12}^j(x)\right)^2dx=M\int\left(p_1(x)-\bar{p}_1^j(x)\right)^2dx-$$

$$-2M\int\left(p_1(x)-\bar{p}_1^j(x)\right)\left(p_2(x)-\bar{p}_2^j(x)\right)dx+M\int\left(p_2(x)-\bar{p}_2^j(x)\right)^2dx. \qquad (7)$$

Асимптотическое выражение для среднеквадратического отклонения $\bar{p}_s^j(x)$ от $p_s(x)$ получено в работе [2, 13]:

$$M\int\left(p_s(x)-\bar{p}_s^j(x)\right)^2dx\sim\frac{\|\Phi(u)\|^2}{\bar{n}_s\,c}+\frac{c^4\left\|p_s^{(2)}(x)\right\|^2}{4},\ s=1,2,$$

где $\|\Phi(u)\|^2=\int\Phi^2(u)du$; $\left\|p_s^{(2)}(x)\right\|^2=\int\left(p_s^{(2)}(x)\right)^2dx$; $\bar{n}_s$ – количество элементов s-го класса в обучающей выборке V_j объема $\bar{n}$.

Следуя преобразованиям, аналогичным при доказательстве первой части теоремы, найдем асимптотическое выражение для функционала

$$M\int\left(p_1(x)-\bar{p}_1^{\,j}(x)\right)\left(p_2(x)-\bar{p}_2^{\,j}(x)\right)dx\sim\frac{c^4}{4}\int p_1^{(2)}(x)\,p_2^{(2)}(x)\,dx\,.$$

Тогда при достаточно больших значения $\bar{n}_1$, $\bar{n}_2$ асимптотическое выражение для среднеквадратического отклонения (7) представляется в виде

$$M\int\left(f_{12}(x)-\bar{f}_{12}^{\,j}(x)\right)^2dx\sim\frac{\left\|\Phi(u)\right\|^2\bar{n}}{\bar{n}_1\,\bar{n}_2\,c}+\frac{c^4}{4}\left\|p_1^{(2)}(x)-p_2^{(2)}(x)\right\|^2, \qquad (8)$$

где $\bar{n}=\bar{n}_1+\bar{n}_2$.

В соответствии с методикой декомпозиции исходной обучающей выборки V ее части V_j, $j=\overline{1,N}$ являются статистически независимыми. Поэтому, принимая во внимание (5), получим

$$M\int\left(\bar{f}_{12}^{\,j}(x)-f_{12}(x)\right)\left(\bar{f}_{12}^{\,t}(x)-f_{12}(x)\right)dx=$$
$$=\int M\left(\bar{f}_{12}^{\,j}(x)-f_{12}(x)\right)M\left(\bar{f}_{12}^{\,t}(x)-f_{12}(x)\right)dx\sim\frac{c^4}{4}\left\|p_1^{(2)}(x)-p_2^{(2)}(x)\right\|^2. \qquad (9)$$

С учетом (8), (9) асимптотическое выражение для (6) представляется в виде

$$\frac{\left\|\Phi(u)\right\|^2\bar{n}}{N\,\bar{n}_1\,\bar{n}_2\,c}+\frac{c^4}{4}\left\|p_1^{(2)}(x)-p_2^{(2)}(x)\right\|^2. \qquad (10)$$

Оценивая уравнения разделяющей поверхности $f_{12}(x)$ по выборке V без предварительной ее декомпозиции ($N=1$), полученный результат (10) совпадет с асимптотическим выражением среднеквадратического отклонения для традиционной непараметрической оценки решающей функции парзеновского типа [13].

Следуя предложенной методике преобразований, выражение (10) для k-мерного уравнения разделяющей поверхности запишется в виде

$$\frac{\left(\bar{n}_1+\bar{n}_2\right)\prod_{v=1}^{k}\int\Phi^2\left(u_v\right)du_v}{N\,\bar{n}_1\,\bar{n}_2\,c^k}+\frac{c^4}{4}B, \qquad (11)$$

где $B=\int\ldots\int\left(\sum_{v=1}^{k}\left(p_{2v}^{(2)}(x)-p_{1v}^{(2)}(x)\right)\right)^2dx_1\ldots dx_k$; $p_{jv}^{(2)}(x)$ – вторая производная плотности $p_j(x)$ по признаку x_v, $v=\overline{1,k}$, $j=1,2$.

Анализ свойств непараметрических оценок уравнений разделяющих поверхностей

Для анализа эффективности непараметрической оценки уравнения разделяющей поверхности, основанной на декомпозиции обучающей выборки, и традиционной непараметрической решающей функции [2] рассмотрим отношение соответствующих им асимптотических выражений среднеквадратических отклонений при оптимальных значениях коэффициентов размытости ядерных функций.

Определим минимальное значение W_2 выражения (10) при оптимальных значениях $\bar{c}$ коэффициентов размытости непараметрических оценок $\bar{f}_{12}^{j}(x)$ составляющих обобщенную решающую функцию (4). В принятых допущениях значение

$$\bar{c} = \left(\frac{\|\Phi(u)\|^2 \bar{n}}{\bar{n}_1 \bar{n}_2 \left\| p_1^{(2)}(x) - p_2^{(2)}(x) \right\|^2} \right)^{1/5}.$$

Тогда

$$W_2 = \left[\left(\frac{\|\Phi(u)\|^2 \bar{n}}{\bar{n}_1 \bar{n}_2} \right)^4 \left\| p_1^{(2)}(x) - p_2^{(2)}(x) \right\|^2 \right]^{\frac{1}{5}} \left(\frac{4+N}{4N} \right). \tag{12}$$

Нетрудно показать, что для традиционной непараметрической решающей функции

$$\tilde{f}_{12}(x) = \bar{p}_2(x) - \bar{p}_1(x), \tag{13}$$

восстанавливаемой по исходной выборке V объема n, оптимальное выражение коэффициента размытости $c^*(n)$ статистики (12) совпадает со значением $\bar{c}(\bar{n})$ при $\bar{n}_1 = n_1$, $\bar{n}_2 = n_2$. При этом минимальное асимптотическое выражение W_2' среднеквадратического отклонения $\tilde{f}_{12}(x)$ от $f_{12}(x)$ аналогично W_2 при $N = 1$.

На этой основе вычислим отношение $\dfrac{W_2}{W_2'} = \left(\dfrac{\bar{n}\, n_1 n_2}{n\, \bar{n}_1 \bar{n}_2} \right)^{\frac{4}{5}} \dfrac{4+N}{5N}$, где $n = n_1 + n_2$, $\bar{n} = n/N$. Определим $n_1 = \alpha n$, $n_2 = (1-\alpha)n$. Причем $\bar{n}_1 = \alpha \bar{n} = {}^{\alpha n}\!/_{N}$, $\bar{n}_2 = (1-\alpha)\bar{n} = {}^{(1-\alpha)n}\!/_{N}$.

Тогда отношение $R_2 = \dfrac{W_2}{W_2'} = \dfrac{4+N}{5N^{1/5}} > 1$, что подтверждает несколько большую эффективность в среднеквадратическом статистики (12) по сравнению с непараметрической решающей функцией (4).

Однако дисперсия оценки $\bar{\bar{f}}_{12}(x)$ меньше, чем для $\tilde{f}_{12}(x)$. В чем нетрудно убедиться, если сравнить соответствующие им главные дисперсионные составляющие $W_3 = \dfrac{\|\Phi(u)\|^2 \bar{n}}{N\,\bar{n}_1\,\bar{n}_2\,\bar{c}}$, $W_3' = \dfrac{\|\Phi(u)\|^2 n}{n_1\, n_2\, c^*}$ в среднеквадратических отклонениях для $\bar{\bar{f}}_{12}(x)$ и $\tilde{f}_{12}(x)$ от $f_{12}(x)$ при оптимальных значениях $\bar{c}$, c^* коэффициентов размытости ядерных функций.

Их отношение $R_3 = \dfrac{W_3}{W_3'} = N^{-1/5} < 1$.

Сравним аппроксимационные свойства $\tilde{f}_{12}(x)$ и $\bar{\bar{f}}_{12}(x)$ при условии неравномерности распределения элементов обучающей выборки между классами при синтезе статистики $\tilde{f}_{12}(x)$ (12). При этом будем считать, что при синтезе составляющих непараметрической оценки решающей функции (4) осуществляется равномерное распределение элементов обучающих выборок объема $\bar{n}$ между классами, т.е. $\bar{n}_1 = \bar{n}_2 = \dfrac{n}{2N}$.

В этом случае после несложных преобразований получим:

$$\bar{R}_2 = \frac{W_2(\alpha_1 = 0.5)}{W_2'(\alpha)} = (2(1-\alpha))^{4/5}\,\frac{4+N}{5N^{1/5}},\quad \bar{R}_3 = \frac{W_3(\alpha_1 = 0.5)}{W_3'(\alpha)} = (2(1-\alpha))^{4/5}\,N^{-1/5}.$$

При $\alpha = 0.5$ полученные результаты $\bar{R}_2$, $\bar{R}_3$ совпадают с R_2, R_3, что подтверждает корректность выполненных преобразований. Зависимости отношений $\bar{R}_j$, $j = \overline{1,3}$ от степени неравномерности распределения α элементов обучающей выборки V при синтезе статистики (12) приведены на рис. 1.

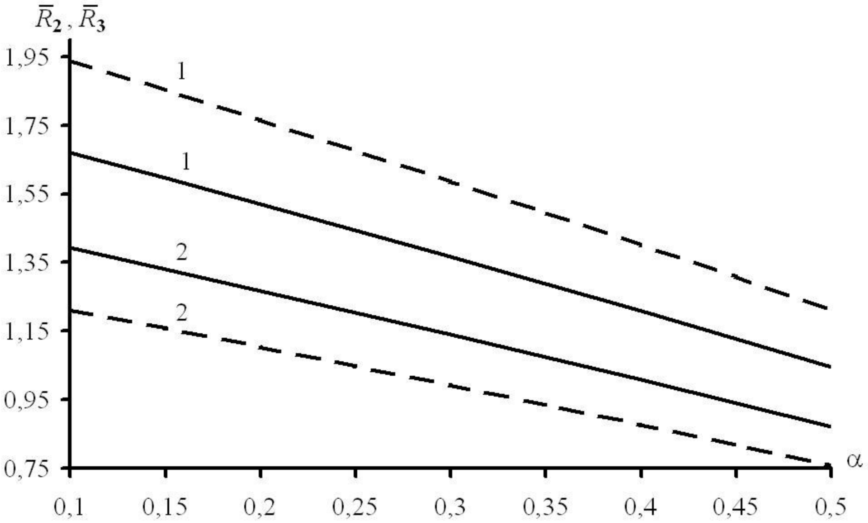

Рис. 1. Зависимости отношений $\bar{R}_2$, $\bar{R}_3$ (линии 1, 2 соответственно) от степени неравномерности распределения α. Сплошные линии определяют исследуемые закономерности при разбиении обучающей выборки на $N = 2$ частей, а штриховые – $N = 4$.

Существуют пороговые значения $\bar{\alpha}$ степени неравномерности распределения α, при которых сохраняются соотношения $\bar{R}_3 < 1$. С ростом N значения $\bar{\alpha}$

имеют тенденцию к снижению. Например, при $N=2$, $N=4$ эти значения соответствуют 0.4 и 0.3.

С увеличением количества N частей обучающей выборки, на которые она разбивается в процессе синтеза коллектива непараметрических решающих функций, наблюдается существенный рост значений $\overline{R}_2$ по сравнению с темпом снижения значений $\overline{R}_3$ (рис. 2). С уменьшением α значения исследуемых отношений возрастают при сохранении тенденции их изменения.

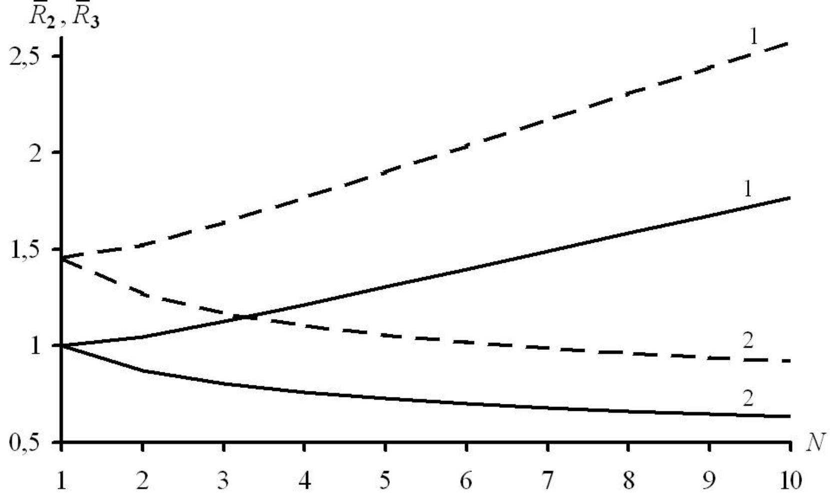

Рис. 2. Зависимости отношений $\overline{R}_2$, $\overline{R}_3$ (кривые 1, 2 соответственно) от количества N частных решающих функций коллектива (4) при $\alpha=0.5$ (сплошные кривые), $\alpha=0.2$ (штриховые).

Выбор количества составляющих коллектива непараметрических решающих функций

Так как статистика (4) имеет меньшую дисперсию по сравнению с традиционной непараметрической оценкой уравнения разделяющей поверхности, то отношение их минимальных асимптотических выражений будем рассматривать в качестве основного критерия при выборе значений N. В многомерном случае при $x=\left(x_v\,,\, v=\overline{1,k}\right)$ главная составляющая асимптотического выражения дисперсии $\bar{\bar{f}}_{12}(x)$ соответствует первому слагаемому в выражении (11). Ее минимальное значение достигается при оптимальных значениях коэффициентов размытости

$$\tilde{c}=\left(\frac{k\,\overline{n}\,\prod\limits_{v=1}^{k}\int\Phi^2(u_v)du_v}{\overline{n}_1\,\overline{n}_2\,B}\right)^{1/(k+4)} \qquad (14)$$

непараметрических оценок $\bar{f}_{12}(x)$ составляющих коллектива $\bar{\bar{f}}_{12}(x)$ [7].

Тогда, подставляя (14) в первое слагаемое выражения (11), запишем минимальное асимптотическое выражение дисперсии коллектива $\bar{\bar{f}}_{12}(x)$ в виде

$$\overline{W_3} = \frac{1}{N\, k^{k/_{k+4}}} \left[\left(\frac{\overline{n} \prod_{\nu=1}^{k} \int \Phi^2(u_\nu) du_\nu}{\overline{n}_1 \, \overline{n}_2} \right)^4 B^k \right]^{1/(k+4)} .$$

Нетрудно заметить, что для традиционной непараметрической решающей функции $\tilde{f}_{12}(x)$ в принятых условиях $c_\nu(n) = c$, $\nu = \overline{1, k}$ оптимальный коэффициент размытости $\tilde{c}(n)$ совпадает со значением $\overline{c}(\overline{n})$ при $\overline{n}_1 = n_1$, $\overline{n}_2 = n_2$. Причем минимальное асимптотическое выражение $\tilde{W}_3$ дисперсии статистики $\tilde{f}_{12}(x)$ типа (13) аналогично $\overline{W_3}$ при $N = 1$ и $\overline{n}_1 = n_1$, $\overline{n}_2 = n_2$. Здесь n_1 и n_2 количество элементов обучающей выборки соответственно первого и второго классов $(n_1 + n_2 = n)$.

Пусть априорные вероятности классов равны, т.е. $\overline{n}_1 = \dfrac{n_1}{N}$, $\overline{n}_2 = \dfrac{n_2}{N}$ и $n_1 = n_2$, тогда отношение $R'_3 = \dfrac{\overline{W_3}}{\tilde{W}_3} = N^{-k/_{k+4}} < 1$.

При $k = 1$ отношение R'_3 совпадает с полученным ранее результатом R_3, что подтверждает корректность выполненных преобразований.

Введем пороговое значение $\alpha < 1$ для отношения R'_3 и из условия $N^{-k/(k+4)} \le \alpha$ определим количество составляющих коллектива (4) $\overline{N} = \left| \alpha^{-(k+4)/k} \right| + 1$, а также соответствующие ему значения отношения $R'_3 = \overline{N}^{-k/(k+4)}$. Здесь символом $|\beta|$ обозначена целая часть числа β.

При этом выполняется требуемое условие $R'_3 < \alpha$ задачи синтеза структуры коллектива непараметрических решающих функций, так как с ростом количества его составляющих значение R'_3 уменьшается.

Для данного значения $\overline{N}$ вычислим отношение $R'_2 = \dfrac{\overline{W_2}}{\tilde{W}_2}$ асимптотических выражений среднеквадратических отклонений $\overline{W_2}$, $\tilde{W}_2$ анализируемых оценок решающих функций $\overline{\overline{f}}_{12}(x)$, $\tilde{f}_{12}(x)$ соответственно.

Подставляя оптимальные значения $\tilde{c}$ (14) в выражение (11), получим:

$$\overline{W_2} = \frac{4 + N\, k}{4\, N^{k/(k+4)}} \left[\left(\frac{\overline{n} \prod_{\nu=1}^{k} \int \Phi^2(u_\nu) du_\nu}{\overline{n}_1 \, \overline{n}_2} \right)^4 B^k \right]^{1/(k+4)} .$$

Для традиционной непараметрической решающей функции $\tilde{f}_{12}(x)$ (13) минимальное асимптотическое выражение $\tilde{W}_2$ среднеквадратического отклонения $\tilde{f}_{12}(x)$ от байесовской решающей функции $f_{12}(x)$ может быть получено из $\overline{W_2}$ при $N=1$ и $\bar{n}_1 = n_1$, $\bar{n}_2 = n_2$.

С учетом принятых выше допущений, отношение R_2' при $N=\overline{N}$ запишется в виде (14)

$$R_2' = \frac{\overline{W_2}}{\tilde{W}_2} = \frac{4+\overline{N}\,k}{(4+k)\overline{N}^{k/k+4}},$$

которое при $k=1$ совпадает с приведенным выше результатом R_2.

Отметим, что отношения R_j', $j=2,3$ получены в результате анализа асимптотических свойств статистик $\bar{\bar{f}}_{12}(x)$ и $\tilde{f}_{12}(x)$. Поэтому применение предлагаемой методики синтеза коллектива $\bar{\bar{f}}_{12}(x)$ (4) рекомендуется для исследуемых условий больших выборок, когда использование принципа их декомпозиции является обоснованным.

Заключение

Коллектив непараметрических оценок решающих функций в двуальтернативной задаче распознавания образов, основанный на декомпозиции обучающей выборки по ее объему, обеспечивает использование технологии параллельных вычислений. Установлены условия асимптотической сходимости коллектива и их зависимость от количества его составляющих. По сравнению с традиционной непараметрической решающей функцией предлагаемый коллектив имеет меньшую дисперсию, но несколько большее значение асимптотического выражения среднеквадратического отклонения. Это объясняется увеличением смещения исследуемого коллектива, так как снижаются объемы обучающих выборок при оценивании его составляющих. Неравномерность распределения элементов обучающей выборки между классами снижает аппроксимационные свойства коллектива непараметрических решающих функций.

ЛИТЕРАТУРА

1. *Parzen E.* On estimation of a probability density function and mode // Ann. Math. Statistic. – 1962. – Vol. 33, N 3. – P. 1065-1076.
2. *Епанечников В.А.* Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и ее применения. – 1969. – Т. 14, №1. – С. 156-161.
3. *Лапко А.В., Лапко В.А.* Синтез структуры смеси непараметрических оценок плотности вероятности многомерной случайной величины // Системы управления и информационные

технологии. – 2011. – Т.43, №1. – С. 12-15.

4. *Лапко А.В., Лапко В.А.* Свойства непараметрической оценки плотности вероятности многомерных случайных величин в условиях больших выборок // Информатика и системы управления. – 2012. – Т.32, №2. – С. 121-126.
5. *Лапко А.В., Лапко В.А.* Регрессионная оценка плотности вероятности и ее свойства // Системы управления и информационные технологии. – 2012. – Т. 49, №3.1. – С. 152-156.
6. *Лапко А.В., Ченцов С.В.* Непараметрические системы обработки информации. – М.: Наука, 2000.
7. *Лапко А.В., Лапко В.А.* Коллектив непараметрических решающих функций в двуальтернативной задаче распознавания образов // Системы управления и информационные технологии. – 2009. – Т. 37, № 3.1. – С. 156 – 160.
8. *Лапко А.В., Поликарпов Л.С., Манчук В.Т. и др.* Автоматизация научных исследований в медицине. – Новосибирск: Наука, 1996.
9. *Лапко А.В., Лапко В.А.* Анализ непараметрических алгоритмов распознавания образов в условиях пропуска данных // Автометрия. – 2008. – Т. 44, № 3. – С. 65–74.
10. *Лапко А.В., Лапко В.А.* Непараметрическая оценка уравнения разделяющей поверхности в условиях больших выборок и ее свойства // Системы управления и информационные технологии. – 2010. – Т. 39, № 1.2. – С. 300–304.
11. *Лапко А.В., Ченцов С.В.* Многоуровневые непараметрические системы принятия решений. – Новосибирск: Наука, 1997.
12. *Лапко А.В.* Непараметрические методы классификации и их применение. – Новосибирск: Наука, 1993.
13. *Лапко А.В., Лапко В.А., Соколов М.И., Ченцов С.В.* Непараметрические системы классификации. – Новосибирск: Наука, 2000.
14. *Лапко А.В., Лапко В.А.* Синтез структуры семейства непараметрических решающих функций в задаче распознавания образов // Автометрия. – 2011. – Т. 47, № 4. – С. 76-82.

E-mail:
Лапко Александр Васильевич – lapko@icm.krasn.ru;
Лапко Василий Александрович – lapko@icm.krasn.ru.