УДК: 519.7

© 2020 г. **А.В. Лапко** [1,2], д-р. техн. наук,
В.А. Лапко [1,2], д-р. техн. наук,
А.В. Бахтина[2]
[1](Институт вычислительного моделирования СО РАН, Красноярск),
[2](Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, Красноярск)

СРАВНЕНИЕ ЛИНЕЙНЫХ КОЛЛЕКТИВОВ МНОГОМЕРНЫХ НЕПАРАМЕТРИЧЕСКИХ РЕГРЕССИЙ

Предлагается методика синтеза и анализа линейных коллективов многомерных непараметрических регрессий. Построение составляющих коллектива осуществляется на основе данных различного объема. Исследуются асимптотические свойства коллектива. Анализируются результаты их сравнения со свойствами коллектива непараметрических регрессий при равномерном распределении статистических данных между его составляющими.

Ключевые слова: непараметрическая регрессия, линейный коллектив решающих функций, большие выборки, асимптотические свойства, параллельные вычисления.

DOI: 10.22250/isu.2020.63.84-94

Введение

Непараметрические регрессии, основанные на оценках плотности вероятности типа Розенблатта – Парзена, широко используются при восстановлении однозначных стохастических зависимостей [1 – 3]. На их основе создаются информационные средства, адаптируемые к различным условиям функционирования объектов исследования [4]. Однако при увеличении объема обучающей выборки вычислительная эффективность непараметрических статистик снижается. Подобные ситуации часто встречаются при обработке больших массивов аэрокосмической информации, данных исследования медико-биологических и экологических систем. В этих условиях использование традиционной непараметрической регрессии приводит к значительным временным затратам при формировании решений.

Перспективное направление «обхода» проблем больших выборок связано с использованием принципов декомпозиции исходных статистических данных по их объему и формированием на этой основе коллектива решающих функций. Известен ряд моделей стохастических зависимостей коллективного типа, к которым можно отнести, например, полупараметрические и частично линейные модели [2, 5], гибридные модели и линейные коллективы решающих функций [4]. Коллективы непараметрических регрессий используются для решения проблемы большой размерности в задаче восстановления стохастических зависимостей [6]. В работе [7] впервые исследованы аппроксимационные свойства коллектива одномерных непараметрических регрессий, основанных на декомпозиции статистических данных по объему. Полученные результаты обобщены на многомерный случай при равномерном распределении исходных данных между составляющими коллектива [8]. Подобные коллективы обеспечивают использование технологий параллельных вычислений при восстановлении стохастических зависимостей в условиях исходных данных большого объема.

Цель данной статьи состоит в исследовании аппроксимационных свойств коллективов многомерных непараметрических регрессий в условиях равномерного и неравномерного распределения исходных статистических данных между их составляющими.

Линейный коллектив многомерных непараметрических регрессий

Обозначим через выборку $V = \left(x^i, y^i, i = \overline{1, n}\right)$, составленную из n независимых наблюдений случайных величин $x = (x_1, \ldots, x_k)$, y и распределенных с плотностями вероятности $p(x_1, \ldots, x_k, y)$ и $p(x_1, \ldots, x_k) > 0$.

Выборку V разобьем на T групп наблюдений $V_j = \left(x^i, y^i, i \in I_j\right)$, $j = \overline{1, T}$, следуя специфики решаемой задачи. Здесь I_j – множество номеров наблюдений переменных (x, y), составляющих j-ю группу. Количество $n_j = \left|I_j\right|$ элементов в выборках V_j, $j = \overline{1, T}$ считаются разными.

Будем считать, что плотность вероятности $p(x_1, \ldots, x_k)$ известна. Причем $p(x_1, \ldots, x_k, y)$ и $p(x_1, \ldots, x_k)$ допускают разложение в ряды Тейлора. Осуществим синтез частной непараметрической регрессии [8].

$$\bar{\varphi}_j(x) = \frac{1}{n_j \; p(x) \prod_{v=1}^{k} c_v^j} \sum_{i \in I_j} y^i \prod_{v=1}^{k} \Phi\left(\frac{x_v - x_v^i}{c_v^j} \right), \; j = \overline{1, T}, \qquad (1)$$

где ядерные функции $\Phi(u_v)$ удовлетворяют условиям H:

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

$$\int \Phi(u_v) du_v = 1, \quad \int u_v^2 \; \Phi(u_v) du_v = 1,$$

$$\int u_v^m \; \Phi(u_v) du_v < \infty, \quad 0 \le m < \infty, \; v = \overline{1, k},$$

а их коэффициенты размытости $c_v^j = c_v(n_j) \to 0$ с ростом n_j, $j = \overline{1, T}$. Здесь и далее бесконечные пределы интегрирования опускаются.

Синтез каждой статистики $\bar{\varphi}_j(x)$ осуществляется с использованием непараметрической оценки плотности вероятности $p(x, y)$ ядерного типа [9 – 11].

При аппроксимации $y = \varphi(x)$ воспользуемся статистикой вида

$$\bar{\bar{y}} = \bar{\bar{\varphi}}(x) = \frac{1}{n} \sum_{j=1}^{T} n_j \; \bar{\varphi}_j(x). \qquad (2)$$

Оптимизация частных непараметрических регрессий (1) по коэффициентам размытости c_v^j, $v = \overline{1, k}$, $j = \overline{1, T}$ ядерных функций осуществляется в режиме «скользящего экзамена» из условия минимума статистической оценки ошибки аппроксимации искомой зависимости.

Статистика (2) допускает использование технологии параллельных вычислений при оценивании кривой регрессии в условиях больших выборок.

Асимптотические свойства коллектива непараметрических регрессий

Положим, что интервалы изменения значений переменных x_v, $v = \overline{1, k}$ равные, тогда имеется возможность считать одинаковыми значения коэффициентов размытости $c_v^j = c_j$, $v = \overline{1, k}$, $j = \overline{1, T}$ ядерных функций в статистике (2).

В этом случае частные непараметрические регрессии (1) запишутся в виде

$$\bar{\varphi}_j(x) = \frac{1}{n_j \; p(x) \; c_j^k} \sum_{i \in I_j} y^i \prod_{v=1}^{k} \Phi\left(\frac{x_v - x_v^i}{c_j} \right), \; j = \overline{1, T}.$$

Асимптотические свойства коллектива непараметрических регрессий (2) определяются следующим утверждением.

Теорема. Пусть функция $\varphi(x)$ и плотности вероятности $p(x)$ разлагаются в ряд Тейлора по всем своим аргументам в каждой точке $x=(x_1,\dots,x_k)$; ядерные функции $\Phi(u_\nu)=\Phi(u)$, $\nu=\overline{1,k}$ удовлетворяют условиям H; последовательности коэффициентов размытости $c_j=c_j(n_j)$ таковы, что при $n_j\to\infty$, $n\to\infty$ значения $c_j\to 0$, а $nc_j^k\to\infty$, $j=\overline{1,T}$.

Тогда коллектив непараметрических регрессий $\overline{\overline{\varphi}}(x)$ является асимптотически несмещенной и состоятельной оценкой условного математического ожидания

$$\varphi(x)=\int_{-\infty}^{\infty} y\, p\left(y \big/ (x_1,\dots,x_k)\right) dy .$$

Доказательство теоремы. При анализе асимптотических свойств статистики (2) будем использовать технологию преобразований, предложенную в работе [12 – 14].

Определим условие асимптотической несмещенности коллектива (2) многомерных непараметрических регрессий. Для этого вычислим его математическое ожидание.

По определению имеем

$$M\left(\overline{\overline{\varphi}}(x)\right)=\frac{1}{n}\sum_{j=1}^{T} n_j\, M\left(\overline{\varphi}_j(x)\right)=$$

$$=\frac{1}{n\, p(x_1,\dots,x_k)}\sum_{j=1}^{T}\frac{1}{c_j^k}\sum_{i\in I_j}\int\dots\int y^i\prod_{\nu=1}^{k}\Phi\left(\frac{x_\nu-x_\nu^i}{c_j}\right) p\left(y^i,x_1,\dots,x_k\right)dy^i dx_1^i\dots dx_k^i =$$

$$=\frac{1}{n\, p(x_1,\dots,x_k)}\sum_{j=1}^{T}\frac{n_j}{c_j^k}\int\dots\int \varphi(t)\prod_{\nu=1}^{k}\Phi\left(\frac{x_\nu-t_\nu}{c_j}\right) p(t_1,\dots,t_k)dt_1\dots dt_k , \quad (3)$$

где $\varphi(t_1,\dots,t_k)=M\left(y \big/ (t_1,\dots,t_k)\right)$; M – знак математического ожидания.

В данных преобразованиях учитывается, что элементы статистической выборки V являются значениями одних и тех же случайных величин $(t_1,\dots,t_k,y)$ с плотностью вероятности $p(y,t_1,\dots,t_k)$.

В выражении (3) проведем замену переменных $(x_\nu-t_\nu)c_j^{-1}=u_\nu$ и разложим функции $\varphi\left(x_\nu-c_j t_\nu,\ \nu=\overline{1,k}\right)$, $p\left(x_\nu-c_j t_\nu,\ \nu=\overline{1,k}\right)$ в ряд Тейлора в точке $x=(x_1,\dots,x_k)$. Тогда, с учетом свойств ядерных функций, при достаточно больших значениях n_j, $j=\overline{1,T}$, получим асимптотическое выражение математического ожидания коллектива непараметрических регрессий (2):

$$M\left(\overline{\overline{\varphi}}(x)\right) \sim \varphi(x) + \sum_{j=1}^{T} \frac{n_j}{n} \left[\frac{c_j^2}{2\,p(x)} \sum_{v=1}^{k} \left(\varphi(x)\, p(x)\right)_v^{(2)} + \right.$$

$$\left. + \frac{c_j^4}{4\,p(x)} \left(\sum_{v=1}^{k} \sum_{\substack{r=1 \\ r \neq v}}^{k} \varphi_v^{(2)}(x)\, p_r^{(2)}(x) + \sum_{v=1}^{k} \varphi_v^{(2)}(x)\, p_v^{(2)}(x) \int u_v^4\, \Phi(u_v)\, du_v \right) \right], \quad (4)$$

где $\varphi_v^{(2)}(x)$, $p_v^{(2)}(x)$, $\left(\varphi(x)\, p(x)\right)_v^{(2)}$ – вторые производные функций $\varphi(x_1, \ldots, x_k)$, $p(x_1, \ldots, x_k)$ и их произведения по компоненте x_v. Слагаемые в асимптотическом выражении (4) степени малости порядка c_j^6, $j = \overline{1, T}$ не учитываются. При конечных значениях n_j / n из условия $c_j \to 0$ при $n_j \to \infty$, $j = \overline{1, T}$ следует свойство асимптотической несмещенности коллектива непараметрических регрессий $\overline{\overline{\varphi}}(x_1, \ldots, x_k)$ (2).

Проведем анализ асимптотических свойств среднеквадратического отклонения

$$M\left(\varphi(x) - \frac{1}{n} \sum_{j=1}^{T} n_j\, \overline{\varphi}_j(x) \right)^2 = M\left(\frac{1}{n} \sum_{j=1}^{T} n_j \left(\varphi(x) - \overline{\varphi}_j(x) \right) \right)^2 =$$

$$= \frac{1}{n^2} M\left[\sum_{j=1}^{T} n_j^2 \left(\varphi(x) - \overline{\varphi}_j(x) \right)^2 + \sum_{j=1}^{T} \sum_{\substack{t=1 \\ t \neq j}}^{T} n_j\, n_t \left(\varphi(x) - \overline{\varphi}_j(x) \right) \left(\varphi(x) - \overline{\varphi}_t(x) \right) \right]. \quad (5)$$

Сначала определим асимптотическое выражение

$$M\left(\varphi(x) - \overline{\varphi}_{\mathrm{j}}(x) \right)^2 = M\left(\overline{\varphi}_j^2(x) \right) - 2\, \varphi(x)\, M\left(\overline{\varphi}_j(x) \right) + \varphi^2(x). \quad (6)$$

С учетом используемой выше технологии исследований проведем преобразования

$$M\left(\overline{\varphi}_j^2(x) \right) = \frac{1}{n_j^2\, c_j^{2k}\, p^2(x)} \left[\sum_{i \in I_j} M\left(\left(y^i\right)^2 \prod_{v=1}^{k} \Phi^2\left(\frac{x_v - x_v^i}{c_j} \right) \right) + \right.$$

$$\left. + \sum_{i \in I_j} \sum_{\substack{r \in I_j \\ \mathrm{r} \neq \mathrm{i}}} M\left(y^i \prod_{v=1}^{k} \Phi\left(\frac{x_v - x_v^i}{c_j} \right) y^r \prod_{v=1}^{k} \Phi\left(\frac{x_v - x_v^r}{c_j} \right) \right) \right] =$$

$$= \frac{1}{n_j^2\, c_j^{2k}\, p^2(x)} \left[n_j \int \ldots \int \varphi^2(t_1, \ldots, t_k) \prod_{v=1}^{k} \Phi^2\left(\frac{x_v - t_v}{c_j} \right) p(t_1, \ldots, t_k)\, dt_1 \ldots dt_k + \right.$$

$$\left. + n_j \left(n_j - 1\right) \left(\int \ldots \int \varphi(t_1, \ldots, t_k) \prod_{v=1}^{k} \Phi\left(\frac{x_v - t_v}{c_j} \right) p(t_1, \ldots, t_k)\, dt_1 \ldots dt_k \right)^2 \right].$$

Пренебрегая величинами малости $0(1/n_j)$, $0\left(1/\left(n_j c_j^{k-2}\right)\right)$, $0\left(c_j^6\right)$, найдем асимптотическое выражение

$$M\left(\overline{\varphi}_j^2(x)\right) \sim \varphi^2(x) + \frac{1}{n_j c_j^k p(x)} \varphi^2(x) \prod_{v=1}^{k} \int \Phi^2(u_v) du_v +$$

$$+ \frac{c_j^4}{4 p^2(x)} \left(\sum_{v=1}^{k} (\varphi(x) p(x))_v^{(2)} \right)^2 + c_j^2 \frac{\varphi(x)}{p(x)} \sum_{v=1}^{k} (\varphi(x) p(x))_v^{(2)} +$$

$$+ c_j^4 \frac{\varphi(x)}{2 p(x)} \left(\sum_{v=1}^{k} \sum_{\substack{r=1 \\ r \neq v}}^{k} \left(\varphi_v^{(2)}(x) p_r^{(2)}(x) \right) + \sum_{v=1}^{k} \left(\varphi_v^{(2)}(x) p_v^{(2)}(x) \int u_v^4 \Phi(u_v) du_v \right) \right). \quad (7)$$

Нетрудно показать, что асимптотическое выражение для математического ожидания $\overline{\varphi}_j(x)$ соответствует

$$M\left(\overline{\varphi}_j(x)\right) \sim \varphi(x) + \frac{c_j^2}{2 p(x)} \sum_{v=1}^{k} (\varphi(x)\, p(x))_v^{(2)} +$$

$$+ \frac{c_j^4}{4 p(x)} \left(\sum_{v=1}^{k} \sum_{\substack{r=1 \\ r \neq v}}^{k} \varphi_v^{(2)}(x)\, p_r^{(2)}(x) + \sum_{v=1}^{k} \varphi_v^{(2)}(x)\, p_v^{(2)}(x) \int u_v^4\, \Phi(u_v) du_v \right). \quad (8)$$

Тогда, подставляя выражения (8) и (7) в (6), получим

$$M\left(\overline{\varphi}_j(x) - \varphi(x)\right)^2 \sim \frac{\varphi^2(x)}{n_j c_j^k p(x)} \prod_{v=1}^{k} \int \Phi^2(u_v) du_v + \frac{c_j^4}{4 p^2(x)} \left(\sum_{v=1}^{k} (\varphi(x) p(x))_v^{(2)} \right)^2. \quad (9)$$

С учетом свойства асимптотической несмещенности (4) и статистической независимости выборок V_j, V_t второе слагаемое выражения (5) при достаточно больших значениях n_j, $j = \overline{1, T}$ представляется в виде

$$\frac{1}{n^2} \left(\sum_{j=1}^{T} \sum_{\substack{t=1 \\ t \neq j}}^{T} n_j\, n_t\, M\left(\varphi(x) - \overline{\varphi}_j(x)\right) M\left(\varphi(x) - \overline{\varphi}_t(x)\right) \right) \sim$$

$$\sim \frac{1}{n^2} \sum_{j=1}^{T} \sum_{\substack{t=1 \\ t \neq j}}^{T} \frac{n_j\, n_t\, c_j^2\, c_t^2}{4 p^2(x)} \left(\sum_{v=1}^{k} (\varphi(x) p(x))_v^{(2)} \right)^2 + 0\left(c_j^2\, c_t^4\right) + 0\left(c_j^4\, c_t^2\right). \quad (10)$$

Подставляя результаты (9), (10) в (5), получим асимптотическое выражение среднеквадратического отклонения коллектива непараметрических регрессий (2):

$$M\left(\varphi(x)-\bar{\bar{\varphi}}(x)\right)^2 \sim \frac{\varphi^2(x)}{n^2 p(x)}\left(\|\Phi(u)\|^2\right)^k \sum_{j=1}^{T}\frac{n_j}{c_j^k}+\frac{\left(\sum_{v=1}^{k}\left(\varphi(x)p(x)\right)_v^{(2)}\right)^2}{4n^2p^2(x)}\left(\sum_{j=1}^{T} n_j c_j^2\right)^2, \quad (11)$$

где $\|\Phi(u)\|^2 = \int \Phi(u)\, du$.

Тогда, принимая во внимание соотношения (4) и (11), при конечных значениях T и n_j/n из условия $c_j \to 0$, а $n\, c_j^k \to \infty$ при $n \to \infty$, $n_j \to \infty$, $j=\overline{1,T}$, следует свойство состоятельности коллектива непараметрических регрессий $\bar{\bar{\varphi}}(x)$.

Полученный результат (11) для одномерного случая $(k=1)$ соответствует выводам работы [7], что подтверждает корректность выполненных преобразований.

Анализ аппроксимационных свойств линейных коллективов непараметрических регрессий

Проведем сравнение аппроксимационных свойств коллективов непараметрических регрессий, синтез которых основан на равномерном и неравномерном распределении исходных данных между группами наблюдений V_j, $j=\overline{1,T}$. Для этого исследуем отношение соответствующих им асимптотических выражений среднеквадратических отклонений при оптимальных значениях коэффициентов размытости ядерных функций.

В этом случае минимальное значение W_2 запишется в виде

$$\frac{\left(\|\Phi(u)\|^2\right)^k}{n^2}\sum_{j=1}^{T}\frac{n_j}{c_j^k}\int\ldots\int\frac{\varphi^2(x)}{p(x)}dx_1\ldots dx_k +,$$

$$+\frac{\left(\sum_{j=1}^{T} n_j c_j^2\right)^2}{4n^2}\int\ldots\int\frac{\left(\sum_{v=1}^{k}\left(\varphi(x)p(x)\right)_v^{(2)}\right)^2}{p^2(x)}dx_1\ldots dx_k, \quad (12)$$

которое получено путем интегрирования результата (11).

Оптимальные значения коэффициентов размытости ядерных функций $\bar{c}_j$ для непараметрических регрессий $\bar{\varphi}_j(x)$ определяются из условия минимума выражения, получаемого в результате интегрирования (9), по c_j.

В результате получим: $\bar{c}_j = \left(\dfrac{k\, A\left(\|\Phi(u)\|^2\right)^k}{n_j\, B}\right)^{1/(k+4)}$, где $A = \int \ldots \int \varphi^2(x) p^{-1}(x) dx_1 \ldots dx_k$, $B = \int \ldots \int \left(p^{-1}(x) \sum_{v=1}^{k} (\varphi(x) p(x))_v^{(2)} \right)^2 dx_1 \ldots dx_k$.

Подставляя значения $\bar{c}_j$, $j = \overline{1, T}$ в выражение (12), получим

$$W_2 = \frac{1}{n^2} \left(\left(A \left(\|\Phi(u)\|^2 \right)^k \right)^4 B^k \right)^{1/(k+4)} \times$$

$$\times \left[\frac{1}{k^{k/(k+4)}} \sum_{j=1}^{T} n_j^{(2k+4)/(k+4)} + \frac{1}{4} k^{4/(k+4)} \left(\sum_{j=1}^{T} n_j^{(k+2)/(k+4)} \right)^2 \right]. \qquad (13)$$

Тогда минимальное асимптотическое выражение среднеквадратического отклонения коллектива непараметрических регрессий в условиях равномерного распределения исходных данных между ее составляющими при $n_j = n/T$ и $c_j = c$, $j = \overline{1, T}$ следует из выражения (13):

$$\overline{W_2} = \left(\left(A \left(\|\Phi(u)\|^2 \right)^k \right)^4 B^k \right)^{1/(k+4)} \left(\frac{T}{n} \right)^{4/(k+4)} \left(\frac{4 + k\,T}{4T\,k^{k/(k+4)}} \right).$$

На этой основе определим отношение минимальных значений асимптотических выражений среднеквадратических отклонений

$$R_2 = \frac{W_2}{\overline{W_2}} = \frac{4T^{k/(k+4)}}{n^{(2k+4)/(k+4)}(4 + kT)} \left(\sum_{j=1}^{T} n_j^{(2k+4)/(k+4)} + \frac{k}{4} \left(\sum_{j=1}^{T} n_j^{(k+2)/(k+4)} \right)^2 \right).$$

Как и следовало ожидать, при $n_j = n/T$, $j = \overline{1, T}$ значение $R_2 = 1$.

Определим аппроксимационные свойства $\tilde{\varphi}(x)$ и коллектива непараметрических регрессий при равномерном распределении исходных данных между его составляющими. Можно показать, что асимптотическое выражение среднеквадратического отклонения непараметрической регрессии $\tilde{\varphi}(x)$ следует из $\overline{W_2}$ при $T = 1$:

$$\widetilde{W}_2 = \left(\left(\frac{A \left(\|\Phi(u)\|^2 \right)^k}{n} \right)^4 B^k \right)^{1/(k+4)} \left(\frac{4 + k}{4k^{k/(k+4)}} \right).$$

Вычислим отношение $\overline{R}_2 = \widetilde{W}_2 / \overline{W_2}$. После несложных преобразований получим выражение

$$\overline{R}_2 = \frac{(4 + k) T^{k/(k+4)}}{4 + kT},$$

значение которого меньше единицы (рис. 1). Причем справедливо соотношение $\overline{R}_2 < R_2 < 1$.

Для оценивания влияния неравномерности распределения исходных статистических данных на аппроксимационные свойства коллектива непараметрических регрессий (2) проведем вычислительный эксперимент. Его результаты при n=1000, T=10 и $k \in [2, 10]$ приведены на рис. 1.

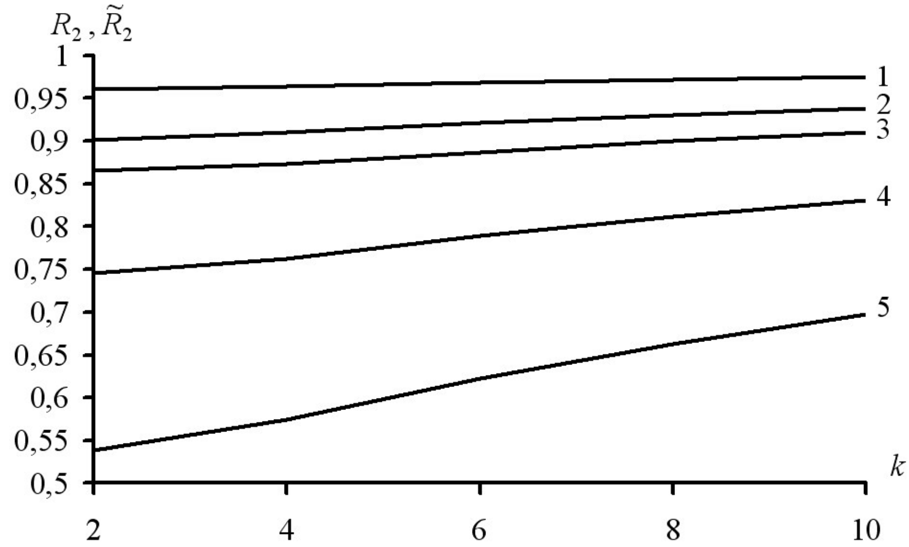

Рис. 1. Зависимость отношения R_2 от размерности k переменных x в условиях неравномерного распределения исходных статистических данных между составляющими коллектива непараметрических регрессий (2): кривые 1, 2 соответствуют условиям: n_j=50, $j=\overline{1,5}$ и n_j=150, $j=\overline{6,10}$; n_j=25, $j=\overline{1,5}$ и n_j=175, $j=\overline{6,10}$; кривые 3, 4 характеризуются условиями: n_j=50, $j=\overline{1,9}$ и n_{10}=550; n_j=25, $j=\overline{1,9}$ и n_{10}=775; кривая 5 иллюстрирует зависимость отношения $\overline{R}_2$ от размерности k.

Исходные данные разбиваются на две части разного объема ($n(1)$, $n(2)$). На основе данных каждой j-й части формируются группы наблюдений в количестве m_j равного объема $r_j = n(j)/m_j$, $j=1, 2$. Отношение r_2/r_1 принимается в качестве показателя неравномерности распределения данных между составляющими коллектива непараметрических регрессий в вычислительном эксперименте. В принятых обозначениях кривые 1, 2, 3, 4 на рис. 1 соответствуют значениям r_2/r_1 равными 3, 7, 11, 31. С увеличением показателя неравномерности распределения исходных данных между составляющими коллектива непараметрических регрессий (2) его среднеквадратическое отклонение по сравнению с равномерным распределением снижается, что проявляется в уменьшении значений R_2 (рис. 1). Данный факт объясняется особенностью структуры коллектива непараметрических регрессий (2).

С уменьшением объема данных n_j при синтезе непараметрической регрессии $\overline{\varphi}_j(x)$ ухудшаются ее аппроксимационные свойства. Однако отмеченная негативная тенденция компенсируется уменьшением «веса» (n_j/n) модели $\overline{\varphi}_j(x)$ в коллективе (2). При неравномерном распределении исходных данных между составляющими коллектива (2) повышается значимость непараметрических регрессий $\overline{\varphi}_j(x)$, построение которых осуществляется по данным большего объема.

Проведем анализ условий, когда при распределении исходных данных n значения n_j, $j=\overline{1, T-1}$ снижаются, а объем данных n_T при синтезе непараметрической регрессии $\overline{\varphi}_T(x)$ увеличивается. В подобной ситуации показатель неравномерности распределения исходных данных возрастает, и в конечном итоге коллектив (2) вырождается в традиционную непараметрическую регрессию $\widetilde{\varphi}(x)$, восстанавливаемую по статистическим данным объема $n_T = n$. Этот вывод подтверждается результатами анализа кривых 3, 4 и 5 на рис. 1. Кривая 5 соответствует значениям $\overline{R}_2$<1, что возможно при условии $\widetilde{W}_2 < \overline{W}_2$.

Установлено, что значение отношения $\overline{R}_2$ при фиксированных параметрах T, k и условиях распределения между составляющими коллектива исходных данных не зависит от их объема.

Заключение

Линейный коллектив многомерных непараметрических регрессий в условиях неравномерного распределения статистических данных между его составляющими обладает свойствами асимптотической несмещенности, сходимости в среднеквадратическом и состоятельности. Получены минимальные значения асимптотических выражений среднеквадратических отклонений линейных коллективов непараметрических регрессий. Неравномерное распределение исходных данных между составляющими коллектива по сравнению с равномерным способствует снижению значений среднеквадратического отклонения предлагаемой модели.

Полученные результаты имеют важное значение при проектировании систем обработки больших массивов статистических данных с использованием технологии параллельных вычислений.

ЛИТЕРАТУРА

1. *Надарая Э.А.* Замечания о непараметрических оценках плотности вероятности и кривой регрессии // Теория вероятностей и ее применение. – 1970. – Т.15, №1. – С. 139-142.
2. *Härdle W., Müller M., Sperlich S., Werwatz A.* Nonparametric and semiparametric models. – New York: Springer – Verlag, 2004.
3. *Катковнив В.Я.* Сходимость линейных и нелинейных непараметрических оценок ядерного типа // Автоматика и телемеханика. – 1983. – №4. – С. 108-120.
4. *Лапко А.В.* Непараметрические системы обработки неоднородной информации. – Новосибирск: Наука, 2007.
5. *Kristensen D.* Semiparametric modeling and estimation // Quantile. – 2009. – №7. – P. 53 – 83.
6. *Lafferty J., Liu H., Wasserman L.* Sparse additive models // Journal of the Royal Statistical Society: Series B. – 2009. – Vol. 71, №5. – P. 1009-1030.
7. *Лапко А.В., Лапко В.А.* Анализ свойств коллектива одномерных непараметрических регрессий, основанных на различных условиях декомпозиции исходных данных // Информатика и системы управления. – 2014. – Т.41, №3. – С. 84 – 92.
8. *Лапко А.В., Лапко В.А.* Синтез структуры коллектива непараметрических регрессий // Информатика и системы управления. – 2015. – Т.44, №2. – С. 53 – 59.
9. *Епанечников В.А.* Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и ее применения. – 1969. – Т.14, №1. – С. 156-161.
10. *Parzen E.* On estimation of a probability density function and mode // Ann. Math. Statistic. – 1962. – Vol. 33, No3. – P. 1065-1076.
11. *Лапко А.В., Лапко В.А.* Регрессионная оценка плотности вероятности и ее свойства // Системы управления и информационные технологии. – 2012. – №3-1 (49). – С. 152-156.
12. *Лапко А.В., Лапко В.А*. Синтез структуры смеси непараметрических оценок плотности вероятности многомерной случайной величины // Системы управления и информационные технологии. – 2011. – №1 (43). – С. 12-15.
13. *Лапко А.В., Лапко В.А*. Непараметрическая оценка уравнения разделяющей поверхности в условиях больших выборок и ее свойства // Системы управления и информационные технологии. – 2010. – №1-2 (39). – С. 300-304.
14. *Лапко А.В., Лапко В.А.* Анализ непараметрических алгоритмов распознавания образов в условиях пропуска данных // Автометрия. – 2008. – Т. 44, №3. – С. 65-74.

E-mail:
Лапко Александр Васильевич – lapko@icm.krasn.ru;
Лапко Василий Александрович – lapko@icm.krasn.ru;
Бахтина Анна Владимировна – anna-denisyuk@yandex.ru.