Методика последовательного формирования набора компонент многомерной случайной величины с использованием непараметрического алгоритма распознавания образов

И.В. Зеньков [1,3,4], *А.В. Лапко* [2,3], *В.А. Лапко* [2,3], *Е.В. Кирюшина* [1], *В.Н. Вокин* [1], *А.В. Бахтина* [3]
[1] *Сибирский федеральный университет, 660041, г. Красноярск, Россия, пр. Свободный, д. 79, стр. 3;*
[2] *Институт вычислительного моделирования СО РАН, 660036, Россия, г. Красноярск, Академгородок, д. 50, стр. 44;*
[3] *Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, 660037, г. Красноярск, пр. «Красноярский рабочий», д. 31;*
[4] *Федеральный исследовательский центр информационных и вычислительных технологий, 660049, г. Красноярск, просп. Мира, д. 53*

Аннотация

Определяется влияние априорных сведений о независимости случайных величин на аппроксимационные свойства непараметрической оценки плотности вероятности Розенблатта–Парзена. Предлагается новая методика формирования наборов независимых компонент многомерной случайной величины. Методика основывается на проверке гипотез о независимости сочетаний компонент многомерной случайной величины с использованием двухальтернативного непараметрического алгоритма распознавания образов ядерного типа, соответствующего критерию максимального правдоподобия. Классы соответствуют областям определения плотностей вероятностей наборов независимых и зависимых компонент многомерной случайной величины. Для оценивания плотностей вероятностей используются непараметрические статистики ядерного типа. Выбор коэффициентов размытости ядерных оценок плотностей вероятностей осуществляется из условия минимума среднего квадратического критерия. Последовательная процедура формирования набора независимых компонент начинается с анализа парных сочетаний компонент многомерной случайной величины. Для каждой пары компонент вычисляется оценка вероятности ошибки распознавания классов, соответствующих предположениям независимости и зависимости рассматриваемых компонент. Определяется пара компонент с максимальным отличием этих ошибок. Если полученные ошибки достоверно не отличаются, то в рассматриваемой многомерной случайной величине независимые компоненты отсутствуют. При достоверном отличии оценок вероятностей ошибок распознавания классов устанавливается пара независимых компонент. Эти компоненты входят в набор из трёх компонент многомерной случайной величины. Анализ их сочетаний осуществляется аналогично по представленным выше рекомендациям. Правилом остановки процесса формирования набора независимых компонент является отсутствие достоверного отличия между вероятностями ошибок распознавания ситуаций, принадлежащих принятым классам. В этом случае предыдущий набор независимых компонент является искомым результатом. В отличие от традиционной методики, основанной на применении критерия Пирсона, предлагаемый подход позволяет обойти проблему декомпозиции области значений случайных величин на многомерные интервалы.

Методика формирования набора независимых компонент многомерной случайной величины иллюстрируется результатами анализа спектральных признаков данных дистанционного зондирования лесных массивов с использованием космической съёмки со спутника Landsat-8.

Ключевые слова: распознавание образов, обработка информации, обработка оптических данных, проверка гипотезы, формирование набора независимых признаков, непараметрический алгоритм, ядерная оценка плотности вероятности, выбор коэффициентов размытости ядерных функций, данные дистанционного зондирования.

Цитирование: **Зеньков, И.В.** Методика последовательного формирования набора компонент многомерной случайной величины с использованием непараметрического алгоритма распознавания образов / И.В. Зеньков, А.В. Лапко, В.А. Лапко, Е.В. Кирюшина, В.Н. Вокин, А.В. Бахтина // Компьютерная оптика. – 2021. – Т. 45, № 6. – С. 926-933. – DOI: 10.18287/2412-6179-CO-902.

Citation: Zenkov IV, Lapko AV, Lapko VA, Kiryushina EV, Vokin VN, Bakhtina AV. A method of sequentially generating a set of components of a multidimensional random variable using a nonparametric pattern recognition algorithm. Computer Optics 2021; 45(6): 926-933. DOI: 10.18287/2412-6179-CO-902.

Введение

Проблема формирования набора независимых компонент многомерной случайной величины является актуальной при синтезе эффективных алгоритмов обработки статистических данных. Априорная информация о независимости случайных величин позволяет повысить аппроксимационные свойства непараметрической оценки плотности вероятности типа Розенблатта–Парзена по сравнению с ядерной статистикой для зависимых переменных [1 – 3]. Данный вывод подтверждается при анализе асимптотических свойств непараметрических решающих функций ядерного типа в задаче распознавания образов [4].

В настоящее время для проверки гипотезы о независимости случайных величин наиболее часто используется χ^2 критерия Пирсона в силу его универсальности и простоты. Его применение содержит трудно формализуемый этап декомпозиции области значений случайной величины на многомерные интервалы [5]. Поэтому возникает задача обхода существующей проблемы. Примером её успешного решения является методика проверки гипотезы о распределениях случайных величин на основе использования непараметрического алгоритма распознавания образов [6 – 8]. Непараметрический алгоритм распознавания образов, соответствующий критерию максимального правдоподобия, формируется по статистическим данным, которые характеризуют законы распределения сравниваемых случайных величин. Показана возможность замены поставленной гипотезы о тождественности законов распределения случайных величин на задачу проверки гипотезы о равенстве ошибки распознавания образов пороговому значению.

Цель данного исследования состоит в развитии предложенного подхода на задачу последовательного формирования набора независимых компонент многомерной случайной величины с использованием непараметрического алгоритма распознавания образов ядерного типа, соответствующего критерию максимального правдоподобия.

1. Свойства непараметрической оценки плотности вероятности независимых случайных величин

Имеется выборка $V=\left(x^i,\, i=\overline{1,n}\right)$ из n статистически независимых наблюдений k-мерной случайной величины $x=\left(x_v,\, v=\overline{1,k}\right)$ с априори неизвестной плотностью вероятности $p(x)$.

Для оценивания плотности вероятности в данных условиях используется непараметрическая статистика Розенблатта–Парзена [9, 10]

$$\bar{p}(x)=\frac{1}{n\prod_{v=1}^{k}c_v}\sum_{i=1}^{n}\prod_{v=1}^{k}\Phi\left(\frac{x_v-x_v^i}{c_v}\right), \tag{1}$$

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

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

Коэффициенты размытости ядерных функций $c_v=c_v(n)$ убывают с ростом n. Здесь и далее бесконечные пределы интегрирования опускаются.

Методы оптимизации статистики (1) по коэффициентам размытости c_v, $v=\overline{1,k}$, рассмотрены в работах [11 – 19].

При статистически независимых компонентах x_v, $v=\overline{1,k}$, случайной величины x непараметрическая оценка плотности вероятности $p(x)$ запишется в виде

$$\bar{p}(x)=\prod_{v=1}^{k}\bar{p}_v(x_v), \tag{2}$$

где $$\bar{p}_v(x_v)=\frac{1}{n\,c_v}\sum_{i=1}^{n}\Phi\left(\frac{x_v-x_v^i}{c_v}\right). \tag{3}$$

Отметим, что статистика (2) допускает использование технологии параллельных вычислений.

Исследуем асимптотические свойства непараметрической оценки плотности вероятности типа (2). Для упрощения выкладок без существенной потери общности получаемых выводов будем оценивать плотность вероятности $p(x_1,x_2)=p_1(x_1)p_2(x_2)$ по выборке V при $k=2$.

Справедливо следующее утверждение [2].

Теорема. Пусть плотности вероятностей $p_v(x_v)$ случайных величин x_v, $v=1,2$, и первые их производные до второго порядка ограничены и непрерывны; ядерные функции $\Phi(u_v)$ удовлетворяют условиям H; последовательности $c_1=c_1(n)$, $c_2=c_2(n)$ коэффициентов размытости ядерных функций непараметрической оценки плотности вероятности $\bar{p}(x_1,x_2)$ таковы, что при $n\to\infty$ значения $c_1\to 0$, $c_2\to 0$, а $nc_1\to\infty$ и $nc_2\to\infty$.

Тогда асимптотические значения смещения и среднего квадратического отклонения $\bar{p}(x_1,x_2)=\bar{p}(x_1)\,\bar{p}(x_2)$ от $p(x_1,x_2)=p(x_1)\,p(x_2)$ определяются соответственно выражениями

$$M\left(\bar{p}_1(x_1)\,\bar{p}_2(x_2)\right)\sim p_1(x_1)\,p_2(x_2)+\frac{c_2^2}{2}p_2^{(2)}(x_2)\,p_1(x_1)+\frac{c_1^2}{2}p_1^{(2)}(x_1)\,p_2(x_2)+\frac{c_1^2\,c_2^2}{4}p_1^{(2)}(x_1)\,p_2^{(2)}(x_2)+0\left(c_1^4,c_2^4\right), \tag{4}$$

$$\begin{aligned}M\iint\left(p_1(x_1)\,p_2(x_2)-\bar{p}_1(x_1)\,\bar{p}_2(x_2)\right)^2dx_1\,dx_2 &\sim \frac{\left(\left\|\Phi(u)\right\|^2\right)^2}{n^2\,c_1\,c_2}+\frac{\left\|\Phi(u)\right\|^2\left\|p_2(x_2)\right\|^2}{n\,c_1}+\\ &+\frac{\left\|\Phi(u)\right\|^2\left\|p_1(x_1)\right\|^2}{n\,c_2}+\iint\left(\frac{p_2(x_2)\,p_1^{(2)}(x_1)\,c_1^2}{2}+\frac{p_1(x_1)\,p_2^{(2)}(x_2)\,c_1^2}{2}\right)^2dx_1\,dx_2,\end{aligned} \tag{5}$$

где M – знак математического ожидания. Здесь $p_2^{(2)}(x_v)$ – вторая производная $p(x_v)$ по компоненте x_v, $v=1,2$;

$$\left\|\Phi(u)\right\|^2=\int\Phi^2(u)\,du;$$

$$\left\|p^{(2)}(x_v)\right\|^2=\int\left(p^{(2)}(x_v)\right)^2 dx_v.$$

Из анализа выражения (5) следует, что при выполнении условий $c_1\to 0$, $c_2\to 0$, $nc_1\to\infty$ и $nc_2\to\infty$ статистика $\bar{p}_1(x_1)\,\bar{p}_2(x_2)$ обладает свойством сходимости в среднем квадратическом, а с учётом её асимптотической несмещённости (4) является состоятельной.

Исследуем значимость влияния сведений о независимости случайных величин на аппроксимационные свойства непараметрической оценки плотности вероятности $p_1(x_1)\,p_2(x_2)$.

Сравним минимальные значения асимптотических выражений для средних квадратических отклонений непараметрической оценки $\bar{p}(x_1,x_2)=\bar{p}_1(x_1)\,\bar{p}_2(x_2)$, формируемой на основе статистики (2), и оценки плотности вероятности $\bar{p}(x_1,x_2)$ (1) при $k=2$.

Для этого вычислим выражение (5) при оптимальных значениях коэффициентов размытости ядерных функций:

$$c_v^*=\left[\left\|\Phi(u)\right\|^2\Big/\left(n\left\|p_v^{(2)}(x_v)\right\|^2\right)\right]^{1/5},\ v=1,\ 2,$$

непараметрических оценок $\bar{p}_1(x_1)$ и $\bar{p}_2(x_2)$ плотностей вероятностей $p_1(x_1)$, $p_2(x_2)$ [10].

Данное преобразование является объёмным. Поэтому в качестве примера вычислим значение фрагмента выражения (5)

$$\frac{\left\|\Phi(u)\right\|^2\left\|p_2(x_2)\right\|^2}{n\,c_1}+\frac{c_1^4}{4}\left\|p_2(x_2)\right\|^2\left\|p_1^{(2)}(x_1)\right\|^2 \qquad (6)$$

при оптимальном коэффициенте размытости c_1^*. Второе слагаемое в (6) соответствует выражению

$$\frac{c_1^4}{4}\iint\left(p_2(x_2)\,p_1^{(2)}(x_1)\right)^2 dx_1\,dx_2.$$

При оптимальном c_1^* выражение (6) преобразуется к виду

$$\frac{5}{4}\left\|p_2(x_2)\right\|^2\left[\left(\frac{\left\|\Phi(u)\right\|^2}{n}\right)^4\left\|p_1^{(2)}(x_1)\right\|^2\right]^{1/5}.$$

$$W_2=\left(\frac{\left\|\Phi(u)\right\|^2}{n}\right)^{\frac{4}{5}}\left[\left(\frac{\left\|\Phi(u)\right\|^2}{n}\right)^{\frac{4}{5}}\times\left(\left\|p_1^{(2)}(x_1)\right\|^2\left\|p_2^{(2)}(x_2)\right\|^2\right)^{\frac{1}{5}}+\frac{5}{4}\left(\left(\left\|p_1^{(2)}(x_1)\right\|^2\right)^{\frac{1}{5}}\left\|p_2(x_2)\right\|^2+\right.\right.$$
$$\left.\left.+\left(\left\|p_2^{(2)}(x_2)\right\|^2\right)^{\frac{1}{5}}\left\|p_1(x_1)\right\|^2\right)+\frac{1}{2}\left(\left\|p_1^{(2)}(x_1)\right\|^2\left\|p_2^{(2)}(x_2)\right\|^2\right)^{-\frac{2}{5}}\times\prod_{v=1}^{2}\int p_v(x_v)\,p_v^{(2)}(x_v)\,dx_v\right].$$

Выполняя подобные преобразования с другими фрагментами критерия (5) при значениях c_v^*, $v=1,2$ и обобщая их результаты, получим

В этих условиях минимальное среднее квадратическое отклонение непараметрической оценки плотности вероятности $\bar{p}(x_1,x_2)$ (1) при $k=2$ и $c_1=c_2$ определяется выражением

$$W_2'=\frac{5}{2^{7/3}}\left[\left(\frac{\left(\left\|\Phi(u)\right\|^2\right)^2}{n}\right)^4\times\right.$$
$$\left.\times\left(\iint\left(p_1^{(2)}(x_1,x_2)+p_2^{(2)}(x_1,x_2)\right)^2 dx_1\,dx_2\right)^2\right]^{\frac{1}{6}},$$

где $p_v^{(2)}(x_1,x_2)$ – вторая производная плотности вероятности $p(x_1,x_2)$ по переменной x_v, $v=1,2$.

При конечных значениях $p_v(x_v)$, $p_v^{(2)}(x_v)$, $v=1,2$ с ростом объёма n статистических данных значения W_2 минимального среднего квадратического отклонения $\bar{p}_1(x_1)\,\bar{p}_2(x_2)$ стремятся к нулю пропорционально величине $r=n^{-4/5}$. Порядок подобной сходимости выше, чем для W_2', значения которого убывают пропорционально величине $n^{-4/6}$. Данная закономерность свойственна и для дисперсий непараметрических оценок плотности вероятности $\bar{p}(x_1,x_2)$, $\bar{p}_1(x_1)\,\bar{p}_2(x_2)$.

Полученные результаты обосновывают значимость информации о независимости случайных величин при синтезе непараметрической оценки их плотности вероятности.

2. Формирование набора независимых компонент многомерной случайной величины

Разработанная методика предполагает выполнение следующей последовательности действий.

1. Сформировать из исходной выборки наблюдений многомерной случайной величины $V=\left(x_v^i,\ v=\overline{1,k},\ i=\overline{1,n}\right)$ массив данных $V(j,r)=\left(x_j^i,\ x_r^i,\ i=\overline{1,n}\right)$, где x_j, x_r – компоненты $x=\left(x_v,\ v=\overline{1,k}\right)$. Проверить гипотезу

$$H_0(j,r):\ p(x_j,x_r)\equiv p(x_j)\,p(x_r) \qquad (7)$$

о независимости случайных величин x_j, x_r.

Для проверки гипотезы $H_0(j,r)$ (7) решить двухальтернативную задачу распознавания образов. Классы Ω_1, Ω_2 соответствуют областям определения плотностей вероятностей $p(x_j)\,p(x_r)$, $p(x_j,x_r)$. В этих условиях байесовское решающее правило, соответствующее критерию максимального правдоподобия, имеет вид

$$m(x_j,x_r):\begin{cases}(x_j,x_r)\in\Omega_1,\\ \text{если } p(x_j,x_r)<p(x_j)\,p(x_r),\\ (x_j,x_r)\in\Omega_2,\\ \text{если } p(x_j,x_r)>p(x_j)\,p(x_r).\end{cases}\tag{8}$$

В отличие от традиционной постановки задачи распознавания образов при синтезе решающего правила (8) в условиях исходной неопределённости явно отсутствует обучающая выборка. Оценивание плотностей вероятностей $p(x_j)\,p(x_r)$, $p(x_j,x_r)$ осуществляется по выборке $V(j,r)$. Для этого используются непараметрические статистики типа Розенблатта–Парзена [9, 10]

$$\bar{p}(x_j,x_r)=\frac{1}{n\,c_j\,c_r}\sum_{i=1}^{n}\Phi\left(\frac{x_j-x_j^i}{c_j}\right)\Phi\left(\frac{x_r-x_r^i}{c_r}\right),\tag{9}$$

$$\bar{p}(x_j)\,\bar{p}(x_r)=\frac{1}{n^2\,c_j\,c_r}\sum_{i=1}^{n}\sum_{l=1}^{n}\Phi\left(\frac{x_j-x_j^i}{c_j}\right)\Phi\left(\frac{x_r-x_r^l}{c_r}\right).\tag{10}$$

Значения коэффициентов размытости c_j, c_r ядерных функций убывают с ростом объёма n выборки статистических данных $V(j,r)$.

Тогда с учётом выражений (8 – 10) непараметрическое решающее правило классификации случайных величин x_j, x_r запишется как

$$\bar{m}(x_j,x_r):\begin{cases}(x_j,x_r)\in\Omega_1,\\ \text{если } \bar{p}(x_j,x_r)<\bar{p}(x_j)\,\bar{p}(x_r),\\ (x_j,x_r)\in\Omega_2,\\ \text{если } \bar{p}(x_j,x_r)>\bar{p}(x_j)\,\bar{p}(x_r).\end{cases}\tag{11}$$

В этой модификации непараметрического алгоритма распознавания образов (11) оптимальные коэффициенты размытости c_j, c_r ядерных оценок плотностей вероятностей $\bar{p}(x_j,x_r)$, $\bar{p}(x_j)\,\bar{p}(x_r)$ выбираются по результатам анализа их аппроксимационных свойств.

Выбор коэффициентов размытости ядерных функций в непараметрической оценке плотности вероятности, например, $\bar{p}(x_r)$ осуществить из условия минимума оценки среднего квадратического критерия

$$W(c_r)=\int\left(\bar{p}(x_r)-p(x_r)\right)^2 dx_r.$$

Статистическая оценка $W(c_r)$ имеет вид [11, 20]

$$\bar{W}(c_r)=\frac{1}{n^2c_r^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\int\Phi\left(\frac{x_r-x_r^j}{c_r}\right)\Phi\left(\frac{x_r-x_r^i}{c_r}\right)dx_r-$$
$$-\frac{2}{n^2c_r}\sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n}\Phi\left(\frac{x_r^j-x_r^i}{c_r}\right).\tag{12}$$

По аналогии с выражением (12) определить критерии выбора оптимальных коэффициентов размытости статистик $\bar{p}(x_j)$, $\bar{p}(x_j,x_r)$.

Оптимизацию непараметрического решающего правила (11) по коэффициентам размытости ядерных функций c_j, c_r можно упростить, если положить в статистиках (9), (10) значения $c_j=c\bar{\sigma}_j$, $c_r=c\bar{\sigma}_r$. Здесь $\bar{\sigma}_j$, $\bar{\sigma}_r$ – оценки средних квадратических отклонений случайных величин x_j, x_r, вычисляемых по выборке $V(j,r)$. Данное утверждение является очевидным, так как большей длине интервала значений случайной величины соответствует больший коэффициент размытости ядерной функции. Подобный подход использовался при построении «быстрых» процедур оптимизации непараметрических оценок плотности вероятности ядерного типа [13 – 19]. Поэтому появляется возможность оптимизацию непараметрического алгоритма распознавания образов (11) проводить лишь по одному параметру c коэффициентов размытости ядерных функций.

Определить оценки вероятностей ошибок $\bar{\rho}_1(j,r)$, $\bar{\rho}_2(j,r)$ распознавания элементов выборки $V(j,r)$ решающим правилом (11) при оптимальных коэффициентах размытости ядерных функций статистик $\bar{p}(x_j)$, $\bar{p}(x_r)$, $\bar{p}(x_j,x_r)$.

Значение $\bar{\rho}_t(j,r)$ вычисляется в режиме «скользящего экзамена» по выборке $V(j,r)$ в предположении, что её элементы принадлежат классу Ω_t,

$$\bar{\rho}_t(j,r)=\frac{1}{n}\sum_{l=1}^{n}1\left(\delta(l),\bar{\delta}(l)\right),\ t=1,2,\tag{13}$$

где $\delta(l)$ – указания типа $(x_j^l,x_r^l)\in\Omega_t$, а $\bar{\delta}(l)$ – «решение» алгоритма (11) о принадлежности ситуации x^l к одному из классов Ω_t, $t=1,2$.

Индикаторная функция в формуле (13) определяется выражением

$$1\left(\delta(l),\bar{\delta}(l)\right)=\begin{cases}0,\ \text{если } \delta(l)=\bar{\delta}(l)\\ 1,\ \text{если } \delta(l)\neq\bar{\delta}(l).\end{cases}$$

При вычислении $\bar{\rho}_t(j,r)$ в соответствии с методикой «скользящего экзамена» ситуация (x_j^l,x_r^l) из выборки $V(j,r)$, которая подаётся на контроль в алгоритм (11), исключается из процесса формирования статистик (2), (3).

2. Повторить этап 1 для парных сочетаний случайных величин (x_j,x_r), $j,r=\overline{1,k}$, $r>j$. Количество таких парных сочетаний равно значению $k(k-1)/2$. Из них выбрать пару (x_q,x_s), которой соответствуют условия

$$\left|\bar{\rho}_1(q,s)-\bar{\rho}_2(q,s)\right|=\max_{(j,r)\in J_2}\left|\bar{\rho}_1(j,r)-\bar{\rho}_2(j,r)\right|,$$

где J_2 – множество номеров неповторяющихся парных сочетаний компонент (x_j, x_r). Причём значения $\bar{\rho}_1(q,s)$, $\bar{\rho}_2(q,s)$ достоверно отличаются.

3. Сформировать из исходных статистических данных V выборку $V(q,s,j)=\left(x_q^i, x_s^i, x_j^i, i=\overline{1,n}\right)$, где x_j является одной из компонент x_v, $v=\overline{1,k}$, $v\neq q$, $v\neq s$.

Основываясь на последовательности действий этапа 1, проверить гипотезу

$$H_0(q,s,j):\ p(x_q, x_s, x_j)\equiv p(x_q)\,p(x_s)\,p(x_j)$$

о независимости случайных величин x_q, x_s, x_j.

Для этих условий построить непараметрическое решающее правило классификации случайных величин x_q, x_s, x_j, которое запишется в виде

$$\bar{m}(x_q, x_s, x_j):\begin{cases}(x_q, x_s, x_j)\in\Omega_1,\\ \text{если}\ \bar{p}(x_q,x_s,x_j)<\bar{p}(x_q)\,\bar{p}(x_s)\,\bar{p}(x_j),\\ (x_q,x_s,x_j)\in\Omega_2,\\ \text{если}\ \bar{p}(x_q,x_s,x_j)>\bar{p}(x_q)\,\bar{p}(x_s)\,\bar{p}(x_j).\end{cases}$$

Класс Ω_1 характеризуется областью определения плотности вероятности $p(x_q)p(x_s)p(x_j)$, а Ω_2 – $p(x_q, x_s, x_j)$. Ядерные оценки плотности вероятности $\bar{p}(x_q, x_s, x_j)$ и $\bar{p}(x_q)\,\bar{p}(x_s)\,\bar{p}(x_j)$ формируются по выборке $V(q,s,j)$ по аналогии со статистиками (2), (3) соответственно.

Следуя рекомендациям этапа 1, выбрать оптимальные коэффициенты размытости непараметрических оценок плотностей вероятностей $\bar{p}(x_q, x_s, x_j)$, $\bar{p}(x_q)\,\bar{p}(x_s)\,\bar{p}(x_j)$ из условия минимума статистических критериев типа (12) в трёхмерном пространстве компонент x_q, x_s, x_j многомерной случайной величины $x=\left(x_v, v=\overline{1,k}\right)$.

На основе полученной информации в режиме «скользящего экзамена» вычислить оценки вероятностей ошибок $\bar{\rho}_1(q,s,j)$, $\bar{\rho}_2(q,s,j)$ принадлежности ситуаций выборки $V(q,s,j)$ первому и второму классам.

4. Повторить этап 3 для сочетаний компонент (x_q, x_s, x_j), $j=\overline{1,k}$, $j\neq q$, $j\neq s$ многомерной случайной величины $x=\left(x_v, v=\overline{1,k}\right)$. Количество таких сочетаний равно значению $k-2$. Из них выбрать триаду (x_q, x_s, x_λ), которой соответствует условие

$$\left|\bar{\rho}_1(q,s,\lambda)-\bar{\rho}_2(q,s,\lambda)\right|=\max_{(q,s,j)\in J_3}\left|\bar{\rho}_1(q,s,j)-\bar{\rho}_2(q,s,j)\right|,$$

где J_3 – множество номеров триад (x_q, x_s, x_j), $j=\overline{1,k}$, $j\neq q, j\neq s$.

Проверить достоверность отличия оценок вероятностей ошибок распознавания образов $\bar{\rho}_1(q,s,\lambda)$ и $\bar{\rho}_2(q,s,\lambda)$. Если гипотеза об их равенстве отвергается, то количество независимых компонент случайной величины x может быть больше 3. В этом случае необходимо перейти к следующему этапу анализа четырёхмерных случайных величин $(x_q, x_s, x_\lambda, x_j)$, $j=\overline{1,k}$, $j\neq q$, $j\neq s$, $j\neq\lambda$, по аналогии с предыдущими этапами предлагаемой методики.

При подтверждении гипотезы о равенстве значений $\bar{\rho}_1(q,s,\lambda)$ и $\bar{\rho}_2(q,s,\lambda)$ с заданным риском β процесс выбора независимых компонент случайной величины x останавливается. В качестве независимых случайных величин определяются компоненты x_q, x_s.

3. Анализ результатов вычислительных экспериментов

Рассмотрим применение разработанной методики формирования набора независимых компонент случайной величины при анализе данных дистанционного зондирования. Исследуется территория лесных массивов Бурлинского ленточного бора Алтайского края. Исходная информация определяется в виде обучающей выборки для оценивания состояний лесных массивов по данным дистанционного зондирования. Под состоянием лесных массивов понимается хвойный, лиственный и повреждённые древостои. Данные дистанционного зондирования объёма $n=1700$ формировались по фрагментам спутниковой съёмки Landsat-8 с пространственным разрешением 30 метров: x_1 (синий) – (0,450 – 0,515) мкм; x_2 (зелёный) – (0,525 – 0,600) мкм; x_3 (красный) – (0,630 – 0,680) мкм; x_4 (ближний инфракрасный NIR) – (0,845 – 0,885) мкм; x_5 (ближний инфракрасный SWIR 2) – (1,560 – 1,660) мкм; x_6 (ближний инфракрасный SWIR 3) – (2,100 – 2,300) мкм.

Количественные характеристики законов распределения случайных величин x_v, $v=\overline{1,6}$, при объёме статистических данных $n=1700$ приведены в табл. 1.

Табл. 1. Количественные характеристики законов распределения спектральных признаков

Значения	Спектральные признаки					
	x_1	x_2	x_3	x_4	x_5	x_6
$\bar{x}_v$	8574	7782	6814	15331	9343	6859
$\bar{\sigma}_v$	98	146	193	2629	988	461

В табл. 1 символами $\bar{x}_v$, $\bar{\sigma}_v$ обозначены оценки математического ожидания и среднего квадратического отклонения случайной величины x_v соответственно.

Следуя предложенной методике формирования набора независимых компонент многомерной случайной величины, для парных сочетаний спектральных признаков (x_j, x_r) по выборке $V(j,r)$ осуществим синтез непараметрического алгоритма распознавания образов (11) и определим оценки вероятностей ошибок классификации $\bar{\rho}_1(j,r)$, $\bar{\rho}_2(j,r)$, $j,r=\overline{1,6}$, $r>j$. Полученные результаты представлены в табл. 2.

Символом $\bar{R}(j,r)$ в табл. 2 обозначена оценка коэффициента корреляции между спектральными признаками x_j и x_r. Из анализа результатов вычислительных экспериментов для парных сочетаний анализируемых случайных величин следует справедливость соотношения $\bar{\rho}_1(j,r)>\bar{\rho}_2(j,r)$. В соответствии с предложенной методикой данное соотношение показывает наличие зависимости исследуемых спектральных признаков. Для проверки достоверности полученного утверждения проверим гипотезу

$$H_0':\ \bar{\rho}_1(j,r)=\bar{\rho}_2(j,r)$$

с риском её отклонения, равным $\beta=0{,}05$, используя критерий Колмогорова–Смирнова. Определим значение

$$D_{34}=\min_{(j,r)\in J_2}\left|\bar{\rho}_1(j,r)-\bar{\rho}_2(j,r)\right|=0{,}514,$$

которое соответствует признакам x_3, x_4. Здесь J_2 – множество номеров сочетаний спектральных признаков, приведённых в табл. 2. Для проверки гипотезы H_0' при $n=1700$ пороговое значение [21]

$$D_\beta=\sqrt{-\frac{1}{n}\ln\frac{\beta}{2}}=0{,}047.$$

Так как выполняется соотношение $D_{34}>D_\beta$, то гипотеза H_0' отвергается. Дополнительным подтверждением полученного вывода является обнаружение линейной зависимости между спектральными признаками

(x_1, x_2), (x_1, x_3), (x_1, x_6), (x_2, x_3), (x_2, x_5), (x_2, x_6), (x_3, x_6), (x_4, x_5), (x_5, x_6),

для которых оценка коэффициента корреляции $\bar{R}\in[0{,}74;\ 0{,}89]$. С учётом результатов использования предложенной методики следует ожидать, что между спектральными признаками

(x_1, x_4), (x_1, x_5), (x_3, x_4), (x_3, x_5), (x_4, x_6)

существует нелинейная зависимость. Этим сочетаниям спектральных признаков соответствуют оценки коэффициентов корреляции $\bar{R}\in[-0{,}233;\ 0{,}549]$.

Табл. 2. Результаты вычислительных экспериментов

Значения	Спектральные признаки				
	x_1, x_2	x_1, x_3	x_1, x_4	x_1, x_5	x_1, x_6
$\bar{\rho}_1(j,r)$	0,918	0,873	0,787	0,793	0,815
$\bar{\rho}_2(j,r)$	0,077	0,121	0,207	0,203	0,179
$\bar{R}(j,r)$	0,890	0,842	0,085	0,549	0,753
Значения	Спектральные признаки				
	x_2, x_3	x_2, x_4	x_2, x_5	x_2, x_6	x_3, x_4
$\bar{\rho}_1(j,r)$	0,897	0,787	0,840	0,892	0,754
$\bar{\rho}_2(j,r)$	0,096	0,208	0,156	0,102	0,240
$\bar{R}(j,r)$	0,721	0,370	0,739	0,800	–0,233
Значения	Спектральные признаки				
	x_3, x_5	x_3, x_6	x_4, x_5	x_4, x_6	x_5, x_6
$\bar{\rho}_1(j,r)$	0,800	0,855	0,891	0,874	0,925
$\bar{\rho}_2(j,r)$	0,195	0,139	0,105	0,121	0,071
$\bar{R}(j,r)$	0,412	0,804	0,749	0,295	0,846

Заключение

Априорные сведения о независимости случайных величин позволяют значительно повысить аппроксимационные свойства непараметрической оценки плотности вероятности типа Розенблатта–Парзена. Предлагаемая методика формирования набора независимых признаков на основе непараметрического алгоритма распознавания образов позволяет обойти трудно формализуемую проблему декомпозиции области значений многомерной случайной величины, которая свойственна критерию Пирсона. Основу методики составляет процедура последовательного анализа оценок вероятностей ошибок распознавания классов, соответствующих областям определения плотностей вероятностей случайных величин в предположении их независимости и зависимости. На каждом этапе этой процедуры последовательно рассматриваются наборы компонент многомерной случайной величины по мере увеличения их количества. Критерием завершения формирования набора независимых компонент является заданное их количество либо допустимая ошибка распознавания классов.

Между спектральными признаками, полученными при дистанционном зондировании лесных массивов с использованием космической съёмки со спутника Landsat-8, существуют линейные и нелинейные зависимости. Эти сведения являются основой для выбора информативных спектральных признаков в задаче синтеза эффективных алгоритмов оценивания состояния лесных массивов по данным дистанционного зондирования.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ, Правительства Красноярского края и Красноярского краевого фонда науки в рамках научного проекта № 20-41-240001.

Литература

1. **Лапко, А.В.** Свойства непараметрической оценки многомерной плотности вероятности независимых случайных величин / А.В. Лапко, В.А. Лапко // Информатика и системы управления. – 2012. – Т. 31, № 1. – С. 166-174.
2. **Лапко, А.В.** Непараметрическая оценка плотности вероятности независимых случайных величин / А.В. Лапко, В.А. Лапко // Информатика и системы управления. – 2011. – Т. 29, № 3. – С. 118-124.
3. **Лапко, А.В.** Влияние априорной информации о независимости многомерных случайных величин на свойства их непараметрической оценки плотности вероятности / А.В. Лапко, В.А. Лапко // Системы управления и информационные технологии. – 2012. – Т. 48, № 2.1. – С. 164-167.
4. **Лапко, А.В.** Свойства непараметрической решающей функции при наличии априорных сведений о независимости признаков классифицируемых объектов / А.В. Лапко, В.А. Лапко // Автометрия. – 2012. – Т. 48, № 4. – С. 112-119.
5. **Пугачёв, В.С.** Теория вероятностей и математическая статистика: учебное пособие / В.С. Пугачёв. – М: Физматлит, 2002. – 496 с.
6. **Лапко, А.В.** Непараметрические алгоритмы распознавания образов в задаче проверки статистической гипотезы о тождественности двух законов распределения

случайных величин / А.В. Лапко, В.А. Лапко // Автометрия. – 2010. – Т. 46, № 6. – С. 47-53.
7. **Лапко, А.В.** Сравнение эмпирической и теоретической функций распределения случайной величины на основе непараметрического классификатора / А.В. Лапко, В.А. Лапко // Автометрия. – 2012. – Т. 48, № 1. – С. 45-49.
8. **Лапко, А.В.** Методика проверки гипотез о распределениях многомерных спектральных данных с использованием непараметрического алгоритма распознавания образов / А.В. Лапко, В.А. Лапко // Компьютерная оптика. – 2019. – Т. 43, № 2. – С. 238-244. – DOI: 10.18287/2412-6179-2019-43-2-238-244.
9. **Parzen, E.** On estimation of a probability density function and mode / E. Parzen // The Annals of Mathematical Statistics. – 1962. – Vol. 33, Issue 3. – P. 1065-1076. – DOI: 10.1214/aoms/1177704472.
10. **Епанечников, В.А.** Непараметрическая оценка многомерной плотности вероятности / В.А. Епанечников // Теория вероятности и её применения. – 1969. – Т. 14, № 1. – С. 156-161.
11. **Rudemo, M.** Empirical choice of histogram and kernel density estimators / M. Rudemo // Scandinavian Journal of Statistics. – 1982. – Vol. 9, No. 2 – P. 65-78.
12. **Hall, P.** Large-sample optimality of least squares cross-validation in density estimation / P. Hall // Annals of Statistics. – 1983. – Vol. 11, Issue 4. – P. 1156-1174. – DOI: 10.1214/aos/1176346329.
13. **Silverman, B.W.** Density estimation for statistics and data analysis / B.W. Silverman. – London: Chapman and Hall, 1986. – 175 p.
14. **Sheather, S.** A reliable data-based bandwidth selection method for kernel density estimation / S. Sheather, M. Jones // Journal of Royal Statistical Society Series B. – 1991. – Vol. 53, Issue 3. – P. 683-690. – DOI: 10.1111/j.2517-6161.1991.tb01857.x.
15. **Sheather, S.J.** Density estimation / S.J. Sheather // Statistical Science. – 2004. – Vol. 19, Issue 4. – P. 588-597. – DOI: 10.1214/088342304000000297.
16. **Terrell, G.R.** Oversmoothed nonparametric density estimates / G.R. Terrell, D.W. Scott // Journal of the American Statistical Association. – 1985. – Vol. 80, Issue 389. – P. 209-214. – DOI: 10.1080/01621459.1985.10477163.
17. **Jones, M.C.** A brief survey of bandwidth selection for density estimation / M.C. Jones, J.S. Marron, S.J. Sheather // Journal of the American Statistical Association. – 1996. – Vol. 91, Issue 433. – P. 401-407. – DOI: 10.2307/2291420.
18. **Scott, D.W.** Multivariate density estimation: Theory, practice, and visualization / D.W. Scott. – New Jersey: John Wiley & Sons, 2015. – 384 p.
19. **Лапко, А.В.** Модифицированный алгоритм быстрого выбора коэффициентов размытости ядерных оценок многомерных плотностей вероятностей / А.В. Лапко, В.А. Лапко // Измерительная техника. – 2020. – № 11. – С. 9-13. – DOI: 10.32446/0368-1025it.2020-11-9-13.
20. **Лапко, А.В.** Анализ методов оптимизации непараметрической оценки плотности вероятности по коэффициенту размытости ядерных функций / А.В. Лапко, В.А. Лапко // Измерительная техника. – 2017. – № 6. – С. 3-8. – DOI: 10.32446/0368-1025it.2017-6-3-8.
21. **Шаракшанэ, А.С.** Сложные системы / А.С. Шаракшанэ, И.Г. Железнов, В.А. Ивницкий. – М.: Высшая школа, 1977. – 248 с.

Сведения об авторах

Зеньков Игорь Владимирович, 1963 года рождения, в 1985 г. окончил Красноярский институт цветных металлов по специальности «Технология и комплексная механизация открытой разработки месторождений полезных ископаемых», доктор технических наук, профессор, профессор кафедры систем автоматики, автоматизированного управления и проектирования Сибирского федерального университета, ведущий научный сотрудник Красноярского филиала Федерального исследовательского центра информационных и вычислительных технологий. Область научных интересов: решение задач горнодобывающей промышленности с использованием ресурсов дистанционного зондирования; информационное обеспечение мониторинга технологических, логистических параметров предприятий горной промышленности; дистанционное зондирование.
E-mail: *zenkoviv@mail.ru* .

Лапко Александр Васильевич, 1949 года рождения, в 1971 году окончил Фрунзенский политехнический институт по специальности «Автоматика и телемеханика», доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук, профессор кафедры космических средств и технологий Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева. Область научных интересов: непараметрическая статистика; распознавание образов и анализ изображений; моделирование и оптимизация неопределённых систем, дистанционное зондирование. E-mail: *lapko@icm.krasn.ru* .

Лапко Василий Александрович, 1974 года рождения, в 1996 году окончил Красноярский государственный технический университет по специальности «Управление и информатика в технических системах», доктор технических наук, профессор, ведущий научный сотрудник Института вычислительного моделирования Сибирского отделения Российской академии наук, заведующий кафедрой космических средств и технологий Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева. Область научных интересов: непараметрическая статистика; распознавания образов и анализ изображений; моделирование неопределённых систем, дистанционное зондирование. E-mail: *valapko@yandex.ru* .

Кирюшина Елена Васильевна, 1963 года рождения, в 1985 г. окончила Красноярский институт цветных металлов по специальности «Технология и комплексная механизация открытой разработки месторождений полезных ископаемых», кандидат технических наук, доцент, доцент кафедры открытых горных работ Сибирского федерального университета. Область научных интересов: решение задач горнодобывающей промышленности с использованием ресурсов дистанционного зондирования; информационное обеспечение мониторинга технологических, логистических параметров предприятий горной промышленности; дистанционное зондирование. E-mail: *kiryushinaev@mail.ru* .

Вокин Владимир Николаевич, 1954 года рождения, в 1976 г. окончил Красноярский институт цветных металлов по специальности «Технология и комплексная механизация открытой разработки месторождений полезных ископаемых», кандидат технических наук, доцент, профессор кафедры открытых горных работ Сибирского федерального университета. Область научных интересов: решение задач горнодобывающей промышленности с использованием ресурсов дистанционного зондирования; информационное обеспечение мониторинга технологических, логистических параметров предприятий горной промышленности; дистанционное зондирование. E-mail: *vokin@krasmail.ru* .

Бахтина Анна Владимировна, 1978 года рождения, в 2020 году окончила Сибирский университет науки и технологий имени академика М.Ф. Решетнева по направлению 09.04.02 «Информационные системы технологии», профиль «Информационные системы обработки данных дистанционного зондирования», аспирант Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева. Область научных интересов: непараметрическая статистика; распознавание образов и анализ изображений; моделирование неопределённых систем, дистанционное зондирование. E-mail: *anna-denisyuk@yandex.ru* .

ГРНТИ: 28.23.15
Поступила в редакцию 5 апреля 2021 г. Окончательный вариант – 24 мая 2021 г.

A method of sequentially generating a set of components of a multidimensional random variable using a nonparametric pattern recognition algorithm

I.V. Zenkov [1,3,4], A.V. Lapko [2,3], V.A. Lapko [2,3], E.V. Kiryushina [1], V.N. Vokin [2], A.V. Bakhtina [3]
[1]Siberian Federal University,
660041, Krasnoyarsk, Russia, Svobodny Av. 79,
[2]Institute of Computational Modelling SB RAS,
660036, Krasnoyarsk, Russia, Akademgorodok 50,
[3]Reshetnev Siberian State University of Science and Technology,
660037, Krasnoyarsk, Russia, Krasnoyarsky Rabochy Av. 31;
[4]Krasnoyarsk branch of the Federal Research Center for Information and Computational Technologies,
660049, Krasnoyarsk, Russia, Mira Av. 53

Abstract

We study in which way a priori information on the independence of random variables affects the approximation accuracy of a nonparametric estimate of the Rosenblatt–Parzen probability density. A new technique for generating sets of independent components of a multidimensional random variable is proposed. The methodology is based on testing the hypotheses of the independence of combinations of the multidimensional random variable components using a two-alternative nonparametric kernel algorithm for pattern recognition corresponding to the maximum likelihood criterion. Classes correspond to the domains of definition of the probability densities of sets of independent and dependent components of the multidimensional random variable. Nonparametric statistics of the kernel type are used to estimate the probability densities. The choice of the bandwidths of the kernel estimates of the probability densities is made from the condition of the minimum root-mean-square criterion. The sequential procedure for generating a set of independent components begins with the analysis of paired combinations of components of a multidimensional random variable. For each pair of components, the probability of an error in recognizing classes corresponding to the assumptions of independence and dependence of the considered components is estimated. A pair of components with the maximum difference between these errors is determined. If the errors obtained do not differ significantly, then there are no independent components in the considered multivariate random variable. If there is a significant difference in the probability estimates of class recognition errors, a pair of independent components is established. These components are included in a three-component set of a multidimensional random variable. The analysis of their combinations is carried out in the same way, following the above-described procedure. The process of generating the set of independent components is stopped when no reliable difference occurs any more between the probabilities of errors in recognizing situations belonging to the accepted classes. In this case, the previous set of independent components is the desired result. In contrast to the traditional methodology based on the Pearson criterion, the proposed approach allows us to bypass a problem of the decomposition of the range of values of random variables into multidimensional intervals.

The method of generating a set of independent components of a multidimensional random variable is illustrated by the results of the analysis of spectral features of remote sensing data of forest tracts using space imagery from the Landsat-8 satellite.

Keywords: pattern recognition, information processing, optical data processing, hypothesis testing, forming a set of independent features, nonparametric pattern recognition algorithm, kernel probability density estimate, bandwidths selection of the kernel functions, remote sensing data.

Citation: Zenkov IV, Lapko AV, Lapko VA., Kiryushina EV, Vokin VN, Bakhtina AV. A method of sequentially generating a set of components of a multidimensional random variable using a nonparametric pattern recognition algorithm. Computer Optics 2021; 45(6): 926-933. DOI: 10.18287/2412-6179-CO-902.

Acknowledgements: The research was funded by the Russian Foundation for Basic Research, Government of the Krasnoyarsk Territory, and Krasnoyarsk Regional Science Foundation, project No. 20-41-240001.

References

[1] Lapko AV, Lapko VA. Properties of nonparametric estimates of multidimensional probability density of independent random variables [In Russian]. Informatika i Sistemy Upravleniya 2012; 31(1): 166-174.

[2] Lapko AV, Lapko VA. Nonparametric estimation of probability density of independent random variables [In Rus-

sian]. Informatika i Sistemy Upravleniya 2011; 29(3): 118-124.
[3] Lapko AV, Lapko VA. Effect of a priori information about independence multidimensional random variables on the properties of their nonparametric density probability estimates [In Russian]. Sistemy Upravleniya i Informatsionnyye Tekhnologii 2012; 48(2.1): 164-167.
[4] Lapko AV, Lapko VA. Properties of the nonparametric decision function with a priori information on independence of attributes of classified objects. Optoelectronics, Instrumentation and Data Processing 2012; 48(4): 416-422. DOI: 10.3103/S8756699012040139.
[5] Pugachev VS. Probability theory and mathematical statistics: textbook [In Russian]. Moscow: "Fizmatlit" Publisher; 2002.
[6] Lapko AV, Lapko VA. Nonparametric algorithms of pattern recognition in the problem of testing a statistical hypothesis on identity of two distribution laws of random variables. Optoelectronics, Instrumentation and Data Processing 2010; 46(6): 545-550. DOI: 10.3103/S8756699011060069.
[7] Lapko AV, Lapko VA. Comparison of empirical and theoretical distribution functions of a random variable on the basis of a nonparametric classifier. Optoelectronics, Instrumentation and Data Processing 2012; 48(1): 37-41. DOI: 10.3103/S8756699012010050.
[8] Lapko AV, Lapko VA. A technique for testing hypotheses for distributions of multidimensional spectral data using a nonparametric pattern recognition algorithm. Computer Optics 2019; 43(2): 238-244. DOI: 10.18287/2412-6179-2019-43-2-238-244.
[9] Parzen E. On estimation of a probability density function and mode. Ann Math Statistic 1962; 33(3): 1065-1076. DOI: 10.1214/aoms/1177704472.
[10] Epanechnikov VA. Non-parametric estimation of a multivariate probability density. Theory Probab its Appl 1969; 14(1): 153-158. DOI: 10.1137/1114019.
[11] Rudemo M. Empirical choice of histogram and kernel density estimators. Scand Stat Theory Appl 1982; 9(2): 65-78.
[12] Hall P. Large-sample optimality of least squares cross-validation in density estimation. Annals of Statistics 1983; 11(4): 1156-1174. DOI: 10.1214/aos/1176346329.
[13] Silverman BW. Density estimation for statistics and data analysis. London: Chapman and Hall; 1986.
[14] Sheather S, Jones M. A reliable data-based bandwidth selection method for kernel density estimation. J R Stat Soc Series B 1991; 53(3): 683-690. DOI: 10.1111/j.2517-6161.1991.tb01857.x.
[15] Sheather SJ. Density estimation. Stat Sci 2004; 19(4): 588-597. DOI: 10.1214/088342304000000297.
[16] Terrell GR, Scott DW. Oversmoothed nonparametric density estimates. J Am Stat Assoc 1985; 80(389): 209-214. DOI: 10.1080/01621459.1985.10477163.
[17] Jones MC, Marron JS, Sheather SJ. A brief survey of bandwidth selection for density estimation. J Am Stat Assoc 1996; 91(433): 401-407. DOI: 10.2307/2291420.
[18] Scott DW. Multivariate density estimation: Theory, practice, and visualization. New Jersey: John Wiley and Sons; 2015.
[19] Lapko AV, Lapko VA. Modified algorithm for rapid choice of spread coefficients for kernel estimates of multidimensional probability densities. Measurement Techniques 2021; 63(11): 856-861. DOI: 10.1007/s11018-021-01873-w.
[20] Lapko AV, Lapko VA. Analysis of optimization methods for nonparametric estimation of the probability density with respect to the blur factor of kernel functions. Measurement Techniques 2017; 60(6): 515-522. DOI: 10.1007/s11018-017-1228-x.
[21] Sharakshaneh AS, Zheleznov IG, Ivnitskij VA. Complex system [In Russian]. Moscow: "Vysshaya shkola" Publisher; 1977.

Authors' information

Igor Vladimirovich Zenkov (b. 1963), graduated from Krasnoyarsk Institute of Non-ferrous Metals on speciality "Technology and Complex Mechanization of Opencast Mining of Mineral Deposits" in 1985. Doctor of Science in Technology, professor, professor of Automation Systems, Automated Control and Design department at the Siberian Federal University; leading researcher at the Krasnoyarsk Branch of the Federal Research Center for Information and Computational Technologies. Research interests: solving problems in the mining industry using remote sensing resources; information support for monitoring technological, logistic parameters of mining enterprises; remote sensing.
E-mail: zenkoviv@mail.ru .

Alexander Vasilievich Lapko (b. 1949), graduated from Frunze Polytechnic Institute on speciality "Automation and telemechanics" in 1971. Doctor of Science in Technology, professor, honored worker of science of the Russian Federation, chief researcher of the Institute of Computational Modeling of the Siberian Branch of the Russian Academy of Sciences; Professor of Space Facilities and Technologies department of the Reshetnev Siberian State University of Science and Technology. Research interests: nonparametric statistics; pattern recognition and image analysis; modeling and optimization of uncertain systems; remote sensing. E-mail: lapko@icm.krasn.ru .

Vasiliy Aleksandrovich Lapko (b. 1974), graduated from Krasnoyarsk State Technical University on speciality "Management and Informatics in Technical Systems" in 1996. Doctor of Science in technology, professor, leading researcher at the Institute of Computational Modeling of the Siberian Branch of the Russian Academy of Sciences; Head of Space Facilities and Technologies department of the Reshetnev Siberian State University of Science and Technology. Research interests: nonparametric statistics; pattern recognition and image analysis; modeling of uncertain systems; remote sensing. E-mail: valapko@yandex.ru .

Elena Vasilievna Kiryushina (b. 1963), graduated from the Krasnoyarsk Institute of Non-ferrous Metals on Speciality "Technology and Complex Mechanization of Opencast Mining of Mineral Deposits" in 1985. Candidate of Science in Technology, associate professor, associate professor of the Open Mining department at the Siberian Federal University. Research interests: solving problems in the mining industry using remote sensing resources; information support for monitoring technological, logistic parameters of mining enterprises; remote sensing. E-mail: kiryushinaev@mail.ru .

Vladimir Nikolaevich Vokin (b. 1954), graduated from the Krasnoyarsk Institute of Non-ferrous Metals on Speciality "Technology and Complex Mechanization of Opencast Mining of Mineral Deposits" in 1976. Candidate of Science in Technology, associate professor, associate professor of Open Mining department at the Siberian Federal University. Research interests: solving problems in the mining industry using remote sensing resources; information support for monitoring technological, logistic parameters of mining enterprises; remote sensing. E-mail: vokin@krasmail.ru .

Anna Vladimirovna Bakhtina (b. 1978), graduated from the Reshetnev Siberian State University of Science and Technology in the direction of 09.04.02 "Information systems of technology", profile "Information systems for processing remote sensing data", postgraduate student of the Reshetnev Siberian State University of Science and Technology. Research interests: nonparametric statistics; pattern recognition and image analysis; modeling of uncertain systems; remote sensing. E-mail: anna-denisyuk@yandex.ru .

Received April 5, 2021. The final version – May 24, 2021.