

УДК 004.93

НЕПАРАМЕТРИЧЕСКИЕ АЛГОРИТМЫ ОЦЕНИВАНИЯ СОСТОЯНИЙ ПРИРОДНЫХ ОБЪЕКТОВ*

А. В. Лапко^{1,2}, В. А. Лапко^{1,2}

¹Институт вычислительного моделирования СО РАН,
660036, г. Красноярск, Академгородок, 50, стр. 44

²Сибирский государственный университет
науки и технологий им. академика М. Ф. Решетнева,
660037, г. Красноярск, просп. им. газеты «Красноярский рабочий», 31
E-mail: lapko@icm.krasn.ru

Рассматриваются модификации непараметрического алгоритма распознавания образов, соответствующего критерию максимального правдоподобия с дополнительными решающими функциями. Синтез предлагаемых алгоритмов основан на анализе отношений оценок плотностей вероятности распределения случайных величин в классах и их функционалов с вводимыми порогами. Выбор порогов определяется особенностями задачи классификации. Полученные результаты применяются при оценивании состояний древостоев лесных массивов по данным дистанционного зондирования.

Ключевые слова: распознавание образов, ядерная оценка плотности вероятности, выбор коэффициента размытости, решающее правило с градациями преимущества, дистанционное зондирование, состояния лесных массивов.

DOI: 10.15372/AUT20180504

Введение. Объёмы освоения природных ресурсов в России постоянно растут, что требует повышения эффективности их контроля. Наиболее совершенными средствами исследования ресурсов являются методы дистанционного зондирования с привлечением аэрокосмических аппаратов. Современные методы комплексного анализа данных дистанционного зондирования — основа создания региональных геоинформационных систем, обеспечивающих оперативность получения требуемой информации и поддержку принятия управленческих решений. Их применение позволяет значительно повысить точность и оперативность принятия решений по результатам дистанционного зондирования [1, 2]. Поэтому методы комплексного анализа данных дистанционного зондирования являются базовыми для решения важнейших задач эффективного освоения природных ресурсов северных территорий и Сибири [3, 4].

Цель предлагаемой работы — модификация непараметрического алгоритма распознавания образов ядерного типа для оценивания состояний природных объектов по данным дистанционного зондирования. Рассматривается применение информационных средств при оценивании состояний древостоев лесных массивов по степени их усыхания, что имеет важное значение при планировании противопожарных мероприятий.

Непараметрическая оценка плотности вероятности ядерного типа. Основу синтеза непараметрических алгоритмов распознавания образов составляют оценки плотности вероятности типа Розенблатта — Парзена [5, 6]. Их исследованию посвящены работы [7–15].

Пусть имеется выборка $V_j = (x_v^i, v = \overline{1, k}, i = \overline{1, n_j})$ объёма n_j , составленная из значений спектральных признаков $x_v^i, v = \overline{1, k}$, характеризующих элемент S^i земной по-

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 18-01-00251).

верхности по состояниям Ω_j природного объекта. В качестве оценки плотности вероятности x в области определения состояния (класса) Ω_j примем статистику ядерного типа

$$\bar{p}_j(x) = \left(n_j \prod_{v=1}^k c_v \right)^{-1} \sum_{i=1}^{n_j} \prod_{v=1}^k \Phi \left(\frac{x_v - x_v^i}{c_v} \right). \quad (1)$$

Здесь $\Phi(u_v)$ — ядерные функции, удовлетворяющие следующим условиям:

$$\begin{aligned} \Phi(u_v) = \Phi(-u_v), \quad 0 \leq \Phi(u_v) < \infty, \quad \int_{-\infty}^{\infty} \Phi(u_v) du_v = 1, \\ \int_{-\infty}^{\infty} u_v^2 \Phi(u_v) du_v = 1, \quad \int_{-\infty}^{\infty} u_v^m \Phi(u_v) du_v < \infty, \quad 0 \leq m < \infty, \quad v = \overline{1, k}. \end{aligned}$$

Оптимизацию непараметрической оценки плотности вероятности $\bar{p}_j(x)$ по коэффициентам размытости ядерных функций c_v , $v = \overline{1, k}$, можно упростить, если положить $c_v = c\bar{\sigma}_v$, $v = \overline{1, k}$, где $\bar{\sigma}_v$ — оценка среднеквадратического отклонения случайной величины x_v в классе Ω_j . Данное утверждение является очевидным, так как большей длине интервала значений x_v соответствует больший коэффициент размытости c_v ядерной функции $\Phi(u_v)$, $v = \overline{1, k}$. Подобный подход использовался при построении «быстрых» процедур оптимизации непараметрических оценок плотности вероятности ядерного типа [12–15].

В таком случае непараметрическая оценка плотности вероятности (1) запишется в виде

$$\bar{p}_j(x) = \left(n_j \prod_{v=1}^k (c\bar{\sigma}_v) \right)^{-1} \sum_{i=1}^{n_j} \prod_{v=1}^k \Phi \left(\frac{x_v - x_v^i}{c\bar{\sigma}_v} \right). \quad (2)$$

Поэтому появляется возможность осуществить оптимизацию непараметрического алгоритма оценивания состояний природного объекта лишь по одному параметру c коэффициентов размытости ядерных функций.

Непараметрический алгоритм распознавания образов. Пусть имеется обучающая выборка $V = (x^i, \sigma(i), i = \overline{1, n})$ для решения многоальтернативной задачи распознавания образов, составленная из значений спектральных признаков $x^i = (x_v^i, v = \overline{1, k})$ элементов земной поверхности S^i и соответствующих им указаний $\sigma(i)$ о состоянии исследуемого природного объекта. Обозначим через Ω_j , $j = \overline{1, M}$, например, состояния природного объекта Ω_j , $j = \overline{1, 3}$, которые соответствуют тёмнохвойному лесу, усыхающему и сухому древостоям лесных массивов на исследуемой территории.

Для оценивания состояний природного объекта будем использовать непараметрические алгоритмы распознавания образов ядерного типа, соответствующие критерию максимального правдоподобия [16, 17]

$$\bar{m}(x): \quad x \in \Omega_j, \quad \text{если } \bar{p}_j(x) = \max_{t=\overline{1, M}} \bar{p}_t(x), \quad (3)$$

где $\bar{p}_j(x)$ — непараметрическая оценка плотности вероятности типа (2).

Оптимизация непараметрического алгоритма распознавания образов по параметру c коэффициентов размытости ядерных функций осуществляется в режиме «скользящего экзамена» по обучающей выборке V из условия минимума оценки вероятности ошибки классификации

$$\bar{\rho}(c) = \frac{1}{n} \sum_{j=1}^n 1(\sigma(j), \bar{\sigma}(j)),$$

где

$$1(\sigma(j), \bar{\sigma}(j)) = \begin{cases} 0, & \text{если } \sigma(j) = \bar{\sigma}(j), \\ 1, & \text{если } \sigma(j) \neq \bar{\sigma}(j); \end{cases}$$

$\bar{\sigma}(j)$ — «решения» алгоритма $\bar{m}(x^j)$ (3) при $x = x^j$. Если ситуация $x^j \in \Omega_t$, то она исключается в процессе оптимизации из непараметрической оценки плотности вероятности $\bar{p}_t(x)$ в решающем правиле (3).

При использовании непараметрического алгоритма (3) возникают условия «надёжных» и «менее надёжных» решений. В условиях $x = x'$ для надёжных решений $x' \in \Omega_j$ значение непараметрической оценки плотности вероятности $\bar{p}_j(x')$ должно значительно превышать $\bar{p}_t(x') \neq 0, t \in I(x')$, где $I(x')$ — множество номеров классов, в которых $\bar{p}_t(x')$ отлично от нуля при $t \neq j$. Понятие «значительно» порождает неопределённость в принятии решений о принадлежности ситуации x' к тому или иному классу. Для учёта объективно возникающей проблемы введём порог q , который будем сопоставлять с отношением $\bar{q} = \bar{p}_j(x')/F(\bar{p}_t(x'), t \in I(x'))$, где $\bar{p}_j(x')$ характеризуется статистикой (2) при $x = x'$. Например, $F(\bar{p}_t(x'), t \in I(x'))$ может определяться значениями

$$\bar{p}_\lambda(x) = \max_{t \in I(x')} \bar{p}_t(x) \quad (4)$$

либо

$$F(\bar{p}_t(x'), t \in I(x')) = \sum_{t \in I(x')} \bar{p}_t(x').$$

Порог q назначается исследователем в соответствии со спецификой решаемой задачи, тем самым задаётся надёжность принимаемых решений. Большие значения \bar{q} по сравнению с q определяют более надёжное решение $x' \in \Omega_j$. Данный подход увеличивает количество априори вводимых классов с M до \bar{M} , где \bar{M} зависит от значения M и количества порогов r . При $r = 1$ значение $\bar{M} = 2M$.

Будем обозначать новые классы, которые различаются соотношениями между порогом q и значением \bar{q} в условиях $x = x'$, через $\Omega_j(q)$ и $\bar{\Omega}_j(q)$. Причём класс $\Omega_j(q)$ соответствует более надёжным решениям $x' \in \Omega_j$, а $\bar{\Omega}_j(q)$ — менее надёжным в условиях принятого порога q . В этом случае непараметрический алгоритм распознавания образов (3) с градациями преимущества при $r = 1$ запишется в виде

$$\bar{m}_1(x'): \begin{cases} x' \in \Omega_j(q), & \text{если } \bar{q} \geq q \text{ либо } \bar{p}_j(x') > 0, \bar{p}_t(x') = 0, t = \overline{1, M}, t \neq j, \\ x' \in \bar{\Omega}_j(q), & \text{если } q > \bar{q} \geq 1, j = \overline{1, M}. \end{cases} \quad (5)$$

Если количество вводимых порогов увеличивается до значений $q_v, v = \overline{1, r}$, то алгоритм (5) представляется как

$$\bar{m}_r(x'): \begin{cases} x' \in \Omega_j(q_v), & \text{если } q_v > \bar{q} \geq q_{v-1} \text{ либо } \bar{p}_j(x') > 0, \bar{p}_t(x') = 0, t = \overline{1, M}, t \neq j, \\ x' \in \bar{\Omega}_j(q_r), & \text{если } q_{r-1} > \bar{q} \geq 1, j = \overline{1, M}. \end{cases}$$

Для визуального восприятия пространственного распределения состояний природных объектов количество порогов r не должно быть большим.

Предлагаемая методика реализована в программном комплексе NSARSD v. 1.0 в среде Delphi. Его функциональные возможности позволяют решать следующие задачи: оценивание достаточно полного набора количественных показателей законов распределений случайных величин $x_v, v = \overline{1, k}$, в классах $\Omega_j, j = \overline{1, M}$; проверка гипотез о распределениях x_v ,

$v = \overline{1, k}$, в двух классах $\Omega_j, \Omega_t, t = \overline{1, M}, j \neq t$, на основе непараметрических алгоритмов распознавания образов [17, 18]; оценивание принадлежности ситуации $x = x'$ к тому или иному классу и пространственное распределение состояний изучаемых природных объектов с последующей графической интерпретацией результатов классификации. Исходные данные представляются в виде таблиц Excel.

Пространственное распределение состояний древостоев лесных массивов по степени их усыхания. Рассмотрим практическое применение разработанной информационной системы при оценивании состояний древостоев лесных массивов по степени их усыхания по данным дистанционного зондирования.

Предлагаемая методика оценивания состояний природных объектов и их пространственного распределения была использована при исследовании лесных массивов юга Красноярского края по информации, полученной с аппарата MODIS/Terra. Каждый элемент земной поверхности характеризовался шестью спектральными признаками $x = (x_1, \dots, x_6)$, которым свойственны следующие длины волн (нанометры): 620–670 (канал 1), 841–876 (2), 459–479 (3), 545–565 (4), 1230–1250 (5), 1628–1652 (6).

Объём обучающей выборки задавался значением $n = 2432$, а контрольной — $n' = 10201$. Снимок, из которого извлекалась контрольная выборка, представлен на рис. 1. Состояния природного объекта $\Omega_j, j = \overline{1, 3}$, соответствовали тёмнохвойному лесу, усыхающему и сухому древостою. Оценка вероятности ошибки определения состояний древостоев лесных массивов составила $\bar{\rho}(\bar{c}) = 0,024$ при оптимальном параметре $\bar{c} = 0,67$ коэффициента размытости ядерных функций В. А. Епанечникова [5].

Имеется контрольная выборка $V' = (x_1^i, \dots, x_k^i, y^i, i = \overline{1, n'})$, составленная из значений спектральных признаков $x^i = (x_1^i, \dots, x_k^i)$ элементов S^i земной поверхности, относительно которых неизвестны состояния древостоев лесных массивов, и координаты $y^i = (y_1^i, y_2^i)$ этих элементов. Необходимо, используя непараметрическое решающее правило (5) при $M = 3$, осуществить пространственное отображение состояний древостоев на исследуемой территории.

Установим каждому состоянию древостоев $\Omega_j, j = \overline{1, 3}$, определённый цвет. Например, тёмнохвойному — зелёный, усыхающему — жёлтый и сухому — синий. Если ситуация x принадлежит к состоянию тёмнохвойных деревьев, то при уменьшении уровня градаций преимущества q зелёный цвет осветляется. По аналогии с уменьшением значений q в условиях принадлежности x сухому древостою синий цвет осветляется. Полученную цветовую

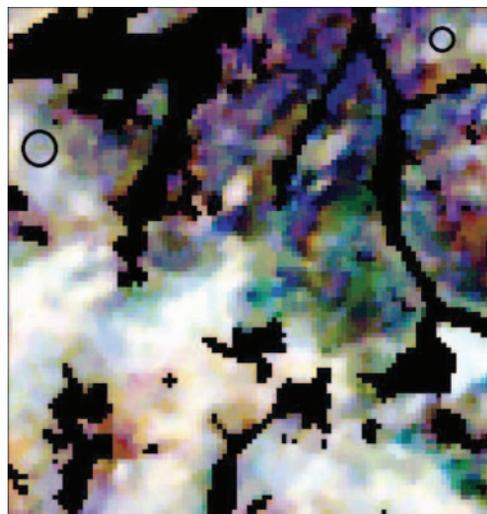


Рис. 1. Снимок территории, полученный с аппарата MODIS/Terra

гамму в пространстве $y = (y_1, y_2)$ определим значениями $\bar{\Omega}_j(y)$, $j = \overline{1, M}$. Тогда алгоритм пространственного распределения состояний изучаемого природного объекта представляется следующим соотношением:

$$y^i \in \bar{\Omega}_j(y), \quad \text{если } x^i \in \bar{\bar{\Omega}}_j, \quad i = \overline{1, n'}, \quad j = \overline{1, 3}.$$

Здесь $\bar{\bar{\Omega}}_j$ включает области $\Omega_j(q)$, $\bar{\Omega}_j(q)$ принадлежности ситуации x^i к j -му классу в соответствии со значениями q и \bar{q} .

По данному алгоритму формируется пространственное распределение состояний древостоев анализируемого лесного массива, которое реализовано с использованием предложенной методики и на основе программного продукта Erdas Imagine (рис. 2).

Результаты, представленные на рис. 2, отличаются значениями порога q , который определяет уровень надёжности принимаемых решений. При больших значениях q наиболее контрастно выделяются области Ω_j , $j = \overline{1, 3}$, и соответствующие им площади уменьшаются. При этом увеличивается площадь территорий для областей $\bar{\Omega}_j(q)$, $j = \overline{1, 3}$.

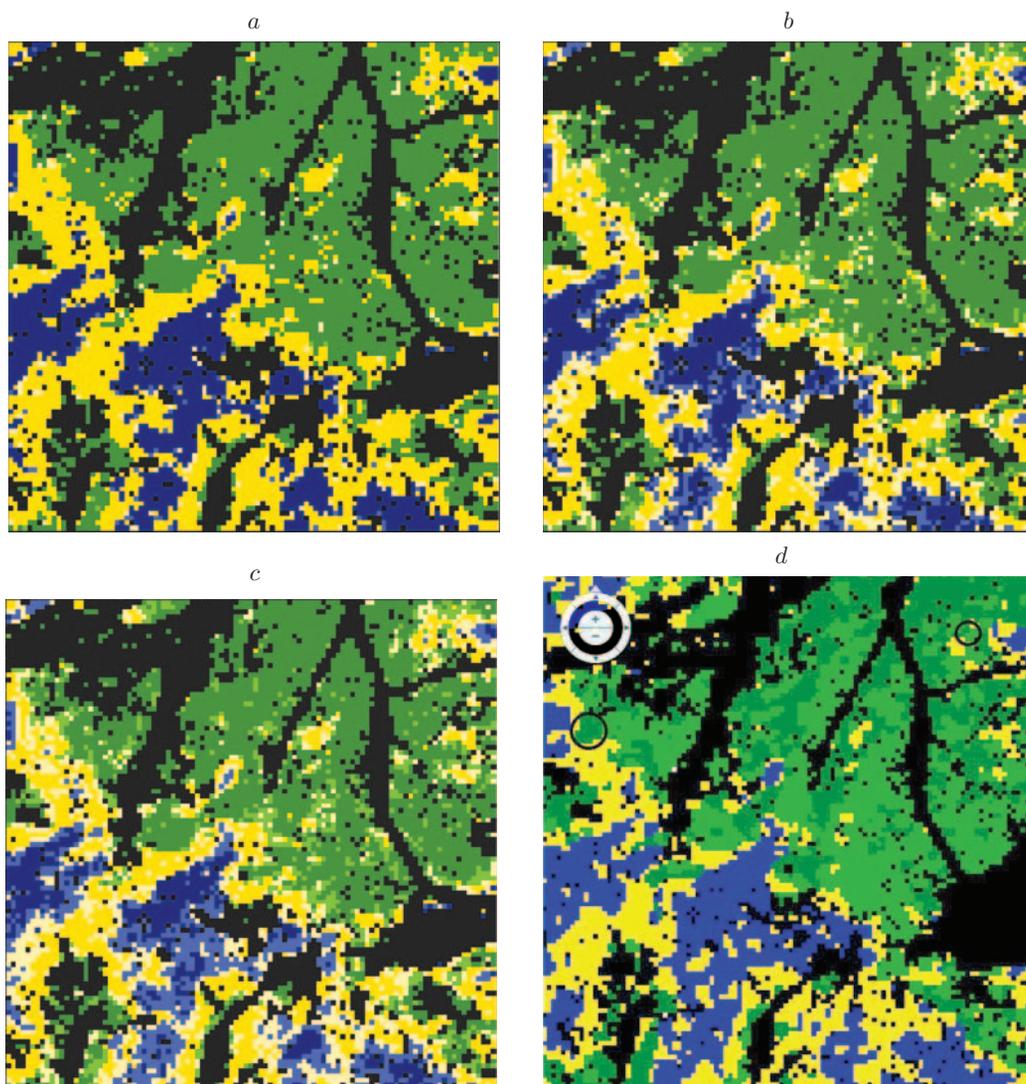


Рис. 2. Пространственное отображение состояний древостоев лесных массивов на изучаемой территории: при значениях $q = 1, 2$ (a), $q = 3$ (b), $q = 5$ (c), с применением программного продукта Erdas Imagine (d)

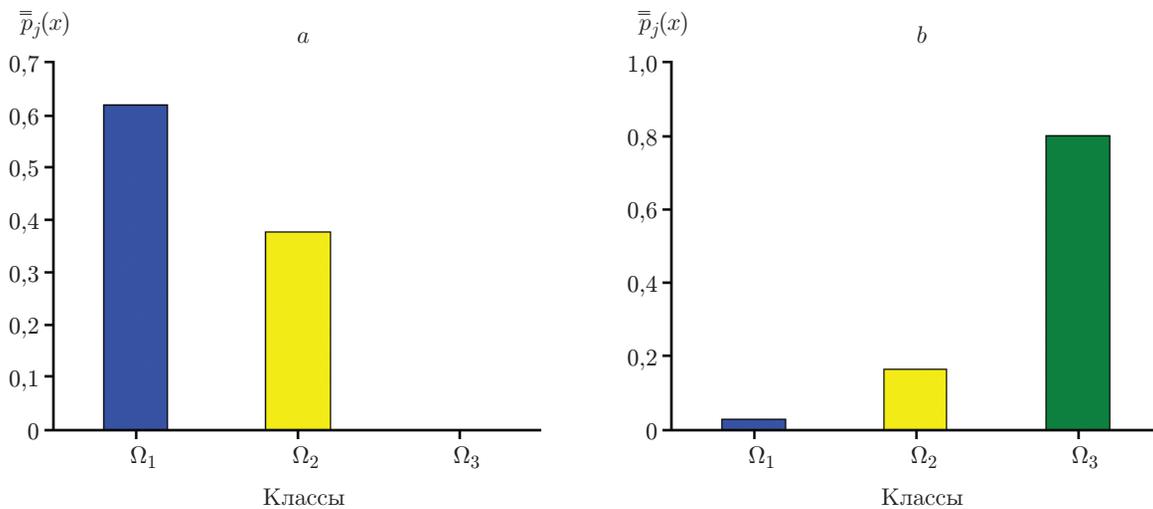


Рис. 3. Нормированные значения непараметрических оценок плотностей вероятностей в классах Ω_j , $j = \overline{1, 3}$, для ситуаций: a — $x' = (x_1 = 46, x_2 = 20, x_3 = 14, x_4 = 77, x_5 = 53, x_6 = 26)$; b — $x' = (x_1 = 46, x_2 = 18, x_3 = 13, x_4 = 45, x_5 = 30, x_6 = 13)$

На рис. 1 и 2 выделяются фрагменты, которые не принадлежат к классам $\Omega_j(q)$, $\bar{\Omega}_j(q)$, $j = \overline{1, 3}$. Они представлены чёрным цветом. Очевидно, что этим фрагментам соответствуют спектральные данные, не принадлежащие ни к одному из классов. В таких условиях значения $\bar{p}_j(x) = 0$, $j = \overline{1, 3}$.

Программный комплекс NSARSD v. 1.0 позволяет количественно оценить принадлежность ситуации $x = (x_1, \dots, x_6)$ к классам Ω_j , $j = \overline{1, 3}$, используя нормированные значения непараметрических оценок плотностей вероятностей $\bar{p}_j(x)$, $j = \overline{1, 3}$ (рис. 3). Высота столбцов на рисунке определяется значениями

$$\bar{p}_j(x) = \bar{p}_j(x) / \sum_{t=1}^3 \bar{p}_t(x), \quad j = \overline{1, 3}.$$

В каждой ситуации $x = x'$ дополнительно имеется возможность изучать зависимость состояний двевостоев лесных массивов от значений спектральных признаков, которые представляются на плоскости с координатами x_v , x_t при фиксированных значениях x_g , $g = \overline{1, k}$, $g \neq v$, $g \neq t$. Спектральные признаки x_v , x_t и значения остальных спектральных каналов определяются исследователем.

Результаты пространственного распределения состояний древостоев позволяют в режиме диалога специалиста и разработанных информационных средств определять значимость спектральных признаков, изменяя их наборы и учитывая значения оценок вероятностей ошибок распознавания состояний.

Заключение. Непараметрические алгоритмы распознавания образов ядерного типа с градациями преимущества являются эффективным средством оценивания состояний природных объектов по данным дистанционного зондирования. Их применение позволяет принимать решение в неоднозначных условиях распределения спектральных данных. На этой основе появляется возможность более широкой пространственной интерпретации результатов оценивания состояний природных объектов. Результаты оценивания состояний древостоев лесных массивов по степени их усыхания, по данным дистанционного зондирования, являются основой принятия решений по обеспечению пожарной безопасности лесных массивов. Дальнейшее развитие этого подхода предполагает использование регрессионной оценки плотности вероятности при синтезе решающих правил распознавания образов в условиях больших объёмов обучающих и контрольных выборок.

СПИСОК ЛИТЕРАТУРЫ

1. **Борзов С. М., Потатуркин О. И.** Исследование эффективности спектрально-пространственной классификации данных гиперспектральных наблюдений // *Автометрия*. 2017. **53**, № 1. С. 32–42.
2. **Борзов С. М., Потатуркин О. И.** Классификация гиперспектральных изображений при различных способах формирования обучающих выборок // *Автометрия*. 2018. **54**, № 1. С. 89–97.
3. **Зеньков И. В., Им С. Т., Лапко А. В. и др.** Развитие и применение информационных технологий исследования природных ресурсов территорий Сибири на основе данных дистанционного зондирования. Красноярск: СибГАУ, 2017. 280 с.
4. **Kharuk V. I., Im S. T., Petrov I. A. et al.** Fir decline and mortality in the southern Siberian Mountains // *Reg. Environ. Change*. 2017. **17**, N 3. P. 803–812.
5. **Епанечников В. А.** Непараметрическая оценка многомерной плотности вероятности // *Теория вероятностей и ее применения*. 1969. **14**, № 1. С. 156–161.
6. **Parzen E.** On estimation of a probability density function and mode // *Ann. Math. Statist.* 1962. **33**, N 3. P. 1065–1076.
7. **Лапко А. В., Лапко В. А.** Непараметрическая оценка плотности вероятности парзеновского типа с неявно заданной формой ядерной функции // *Измерительная техника*. 2016. № 6. С. 14–17.
8. **Lapko A. V., Lapko V. A.** Nonparametric estimate of a Parzen-type probability density with an implicitly specified form of the kernel // *Measur. Techn.* 2016. **59**, N 6. P. 571–576.
9. **Лапко А. В., Лапко В. А.** Регрессионная оценка многомерной плотности вероятности и её свойства // *Автометрия*. 2014. **50**, № 2. С. 50–56.
10. **Duong T.** Ks: Kernel density estimation and kernel discriminant analysis in R // *Journ. Statist. Software*. 2007. **21**, N 7. P. 1–16.
11. **Botev Z. I., Grotowski J. F., Kroese D. P.** Kernel density estimation via diffusion // *Ann. Statist.* 2010. **38**, N 5. P. 2916–2957.
12. **Sheather S. J.** Density estimation // *Statist. Sci.* 2004. **19**, N 4. P. 588–597.
13. **Scott D. W.** *Multivariate Density Estimation: Theory, Practice, and Visualization*. New Jersey: John Wiley & Sons, 2015. 384 p.
14. **Chen S.** Optimal bandwidth selection for kernel density functionals estimation // *Journ. Probability and Statist.* 2015. **2015**, N 1. P. 1–21.
15. **Borrajo M. I., González-Manteiga W., Martínez-Miranda M. D.** Bandwidth selection for kernel density estimation with length-biased data // *Journ. Nonparametr. Statist.* 2017. **29**, N 3. P. 636–668.
16. **Лапко А. В., Лапко В. А.** Анализ асимптотических свойств непараметрической оценки уравнения разделяющей поверхности в двувальтернативной задаче распознавания образов // *Автометрия*. 2010. **46**, № 3. С. 48–53.
17. **Лапко А. В., Лапко В. А., Соколов М. И., Ченцов С. В.** Непараметрические системы классификации. Новосибирск: Наука, 2000. 240 с.
18. **Лапко А. В., Лапко В. А.** Многоуровневые непараметрические системы обработки информации. Красноярск: СибГАУ, 2013. 270 с.

Поступила в редакцию 8 июня 2018 г.