

АНАЛИЗ И СИНТЕЗ СИГНАЛОВ И ИЗОБРАЖЕНИЙ

УДК 519.7

НЕПАРАМЕТРИЧЕСКИЙ АЛГОРИТМ РАСПОЗНАВАНИЯ
ОБРАЗОВ В ЗАДАЧАХ АНАЛИЗА ДАННЫХ
ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ АНТРОПОГЕННЫХ
ТЕРРИТОРИЙ© А. В. Лапко^{1,2}, В. А. Лапко^{1,2}, А. В. Шаруева²¹Институт вычислительного моделирования СО РАН,
660036, г. Красноярск, Академгородок, 50, стр. 44²Сибирский государственный университет науки и технологий
им. академика М. Ф. Решетнева,
660037, г. Красноярск, просп. им. газеты «Красноярский рабочий», 31
E-mail: lapko@ict.krasn.ru

Приводятся результаты применения новой методики проверки гипотезы о независимости случайных величин при анализе данных дистанционного зондирования антропогенных территорий. Основу методики составляет непараметрический алгоритм распознавания образов, соответствующий критерию максимального правдоподобия. Определены линейные и нелинейные зависимости между спектральными признаками, которые характеризуют антропогенные территории. Рассматриваются результаты распознавания типов антропогенных территорий по спектральным данным дистанционного зондирования.

Ключевые слова: проверка гипотезы о независимости случайных величин, распознавание образов, непараметрическая оценка плотности вероятности, антропогенные территории, дистанционное зондирование.

DOI: 10.15372/AUT20250111

EDN: LAETBS

Введение. Задача проверки гипотезы о независимости случайных величин является одной из основных в теории математической статистики. Результаты проверки гипотез используются при синтезе эффективных моделей и алгоритмов принятия решений. Известно, что универсальным критерием проверки гипотезы о распределениях случайных величин, включая их независимость, является критерий Пирсона [1]. Однако его применение содержит трудно формализуемый этап разбиения области значений случайных величин на многомерные интервалы. Поэтому появилась необходимость разработки новых методов проверки гипотезы о независимости случайных величин, позволяющих обойти эту проблему. В работах [2–5] предложена методика проверки гипотезы о независимости случайных величин, основанная на использовании непараметрического алгоритма распознавания образов, соответствующего критерию максимального правдоподобия. Каждый класс в пространстве случайных величин определяется условиями их независимости и зависимости. Оценивание плотностей вероятностей в классах осуществляется по исходным данным с использованием статистик типа Розенблатта — Парзена. Предлагаемая методика обобщена на задачу проверки гипотезы о независимости многомерных случайных величин [2]. Рассмотрена её эффективность при усложнении однозначных функциональных зависимостей между сравниваемыми двумерными случайными величинами при различных объёмах статистических данных [3]. Предложенная методика развита при проверке гипотезы в условиях неоднозначных функций [4]. Её эффективность подтверждается результатами сравнения с традиционным критерием Пирсона [5].

Цель данной работы состоит в применении непараметрической методики проверки гипотезы о независимости случайных величин при анализе спектральных признаков данных дистанционного зондирования антропогенных территорий и использовании полученных результатов при распознавании их типов.

Непараметрическая методика проверки гипотезы о независимости случайных величин. Имеется выборка $V = (x_1^i, x_2^i, i = \overline{1, n})$ статистически независимых наблюдений случайных величин x_1, x_2 объёма n . Необходимо проверить гипотезу о независимости случайных величин x_1, x_2 :

$$H_0 : p(x_1, x_2) = p(x_1)p(x_2), \quad (1)$$

где $p(x_1, x_2)$ — совместная плотность вероятности, $p(x_1)p(x_2)$ — произведение плотностей вероятности одномерных случайных величин x_1, x_2 . С учётом гипотезы H_0 определим два класса Ω_1, Ω_2 . Класс Ω_1 характеризуется плотностью вероятности $p(x_1, x_2)$, а Ω_2 — $p(x_1)p(x_2)$.

На этой основе определим решающее правило распознавания классов Ω_1, Ω_2 , соответствующее критерию максимального правдоподобия:

$$m(x) : \begin{cases} x \in \Omega_1, & \text{если } p(x_1, x_2) > p(x_1)p(x_2); \\ x \in \Omega_2, & \text{если } p(x_1, x_2) < p(x_1)p(x_2), \end{cases} \quad (2)$$

где $x = (x_1, x_2)$.

Если выполняется соотношение $p(x_1, x_2) < p(x_1)p(x_2)$, то в условиях x наиболее вероятным является предположение о независимости случайных величин x_1, x_2 . При справедливости соотношения $p(x_1, x_2) > p(x_1)p(x_2)$ алгоритм (2) относит ситуацию x к классу Ω_1 , в котором выполняется предположение о зависимости случайных величин x_1, x_2 . Пусть ρ_1 и ρ_2 — ошибки распознавания классов Ω_1 и Ω_2 с использованием решающего правила (2). Тогда появляется возможность заменить проверку гипотезы H_0 (1) на проверку гипотезы $\bar{H}_0$: $\rho_1 = \rho_2$. Нетрудно заметить, что значение ρ_1 определяется интегралом от плотности вероятности $p(x_1, x_2)$, значение которого зависит от уравнения разделяющей поверхности между классами Ω_1, Ω_2 . Вероятность ошибки распознавания ρ_2 в предположении независимости x_1 и x_2 вычисляется в виде произведения интегралов от $p(x_1)$ и $p(x_2)$. Поэтому выполнение гипотезы $\bar{H}_0$ возможно только при условии $p(x_1, x_2) = p(x_1)p(x_2)$.

Восстановление плотностей вероятностей $p(x_1, x_2), p(x_1)p(x_2)$ в решающем правиле (2) осуществляется по одной и той же выборке V соответственно в предположении о зависимости и независимости случайных величин x_1 и x_2 . Для оценивания этих плотностей вероятностей будем использовать их непараметрические оценки ядерного типа [6, 7]:

$$\bar{p}(x_1, x_2) = \frac{1}{nc_1c_2} \sum_{i=1}^n \Phi\left(\frac{x_1 - x_1^i}{c_1}\right) \Phi\left(\frac{x_2 - x_2^i}{c_2}\right), \quad (3)$$

$$\bar{p}(x_1)\bar{p}(x_2) = \frac{1}{n^2c_1c_2} \sum_{i=1}^n \sum_{j=1}^n \Phi\left(\frac{x_1 - x_1^i}{c_1}\right) \Phi\left(\frac{x_2 - x_2^j}{c_2}\right), \quad (4)$$

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

$$\Phi(u_v) = \Phi(-u_v), \quad 0 \leq \Phi(u_v) < \infty, \quad \int \Phi(u_v) du_v = 1,$$

$$\int u_v^m \Phi(u_v) du_v < \infty, \quad 0 \leq m < \infty, \quad v = 1, 2.$$

Здесь и далее бесконечные пределы интегрирования опускаются.

Значения коэффициентов размытости c_v , $v = 1, 2$, ядерных функций убывают с ростом объёма n выборки статистических данных V . На основании статистик (3), (4) определим непараметрический алгоритм распознавания образов, соответствующий критерию максимального правдоподобия (2), в виде

$$\bar{m}(x) : \begin{cases} x \in \Omega_1, & \text{если } \bar{p}(x_1, x_2) > \bar{p}(x_1)\bar{p}(x_2); \\ x \in \Omega_2, & \text{если } \bar{p}(x_1, x_2) < \bar{p}(x_1)\bar{p}(x_2). \end{cases} \quad (5)$$

Выбор коэффициентов размытости ядерных функций c_v , $v = 1, 2$, для произведения непараметрических оценок плотностей вероятностей (4) осуществляется из условия минимума критерия [8–12]:

$$\bar{W}(c_v) = \frac{1}{n^2 c_v^2} \sum_{j=1}^n \sum_{i=1}^n \int \Phi\left(\frac{x_v - x_v^j}{c_v}\right) \Phi\left(\frac{x_v - x_v^i}{c_v}\right) dx_v - \frac{2}{n^2 c_v} \sum_{j=1}^n \sum_{\substack{i=1 \\ i \neq j}}^n \Phi\left(\frac{x_v^j - x_v^i}{c_v}\right), \quad (6)$$

$$v = 1, 2.$$

Например, при ступенчатой ядерной функции

$$\frac{1}{c_v} \Phi\left(\frac{x_v - x_v^i}{c_v}\right) = \begin{cases} 0,5 c_v, & \text{если } |x_v - x_v^i| < c_v; \\ 0, & \text{если } |x_v - x_v^i| \geq c_v, \end{cases}$$

значения составляющих критерия (6) задаются выражениями

$$\int \Phi\left(\frac{x_v - x_v^j}{c_v}\right) \Phi\left(\frac{x_v - x_v^i}{c_v}\right) dx_v = \begin{cases} (2c_v - |x_v^j - x_v^i|)/4, & \text{если } |x_v^j - x_v^i| < 2c_v; \\ 0, & \text{если } |x_v^j - x_v^i| \geq 2c_v, \end{cases}$$

$$\Phi\left(\frac{x_v - x_v^i}{c_v}\right) = \begin{cases} 0,5, & \text{если } |x_v - x_v^i| < c_v; \\ 0, & \text{если } |x_v - x_v^i| \geq c_v, \end{cases} \quad v = 1, 2.$$

По аналогии с выражением (6) нетрудно найти критерий выбора оптимальных коэффициентов размытости непараметрической статистики $\bar{p}(x_1, x_2)$ (3):

$$\begin{aligned} \bar{W}(c_1, c_2) = & \frac{1}{n^2 c_1^2 c_2^2} \sum_{j=1}^n \sum_{i=1}^n \prod_{v=1}^2 \int \Phi\left(\frac{x_v - x_v^j}{c_v}\right) \Phi\left(\frac{x_v - x_v^i}{c_v}\right) dx_v - \\ & - \frac{2}{n^2 c_1 c_2} \sum_{j=1}^n \sum_{\substack{i=1 \\ i \neq j}}^n \prod_{v=1}^2 \Phi\left(\frac{x_v^j - x_v^i}{c_v}\right), \end{aligned}$$

минимум которого определяет оптимальные коэффициенты размытости ядерной оценки плотности вероятности $\bar{p}(x_1, x_2)$.

Оптимизацию непараметрического решающего правила (5) по коэффициентам размытости ядерных функций можно упростить [13, 14], если положить в статистике (3) значения $c_v = s\bar{\sigma}_v$, $v = 1, 2$. Здесь $\bar{\sigma}_v$ — оценка среднего квадратического отклонения случайной величины x_v , вычисляемая по выборке V . Данное утверждение является очевидным, так как большей длине интервала значений x_v соответствует больший коэффициент размытости c_v ядерных функций $\Phi(u_v)$, $v = 1, 2$.

Обозначим оценки вероятностей ошибок распознавания классов Ω_1, Ω_2 с использованием решающего правила (5) через $\bar{\rho}_1, \bar{\rho}_2$.

Значения $\bar{\rho}_t$ вычисляются в режиме «скользящего экзамена» по выборке V в предположении, что её элементы принадлежат классу Ω_t :

$$\bar{\rho}_t = \frac{1}{n} \sum_{j=1}^n 1(\sigma(j), \bar{\sigma}(j)), \quad t = 1, 2, \quad (7)$$

где $\sigma(j)$ — указания типа $x^j = (x_1^j, x_2^j) \in \Omega_t$, а $\bar{\sigma}(j)$ — «решение» алгоритма (5) о принадлежности ситуации x^j к одному из классов Ω_t , $t = 1, 2$.

При вычислении $\bar{\rho}_t$ ситуация $x^j = (x_1^j, x_2^j)$ из выборки V , которая подаётся на контроль в алгоритм (5), исключается из процесса формирования статистик (3), (4).

Индикаторная функция в формуле (6) определяется выражением

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

Методика проверки гипотезы H_0 о независимости случайных величин предполагает выполнение следующих действий:

1. По исходным данным V осуществить синтез непараметрического алгоритма распознавания образов типа (5).

2. Вычислить оценки вероятностей ошибок $\bar{\rho}_1, \bar{\rho}_2$ распознавания классов Ω_1, Ω_2 .

3. Проверить достоверность отличия вероятностей ошибок распознавания классов ρ_1, ρ_2 при заданной вероятности β . На данном этапе используется традиционная методика проверки достоверности отличия вероятностей событий случайных величин [1] либо применяется частный вариант критерия Смирнова [15]. В последнем случае пороговое значение

$$D_\beta = \sqrt{-\frac{1}{n} \ln \left(\frac{\beta}{2} \right)}.$$

Вероятности ошибок распознавания классов ρ_1, ρ_2 не отличаются, если

$$|\bar{\rho}_1 - \bar{\rho}_2| < D_\beta. \quad (8)$$

Если выполняются условие (8), то гипотеза H_0 справедлива.

Универсальность предложенного подхода состоит в возможности его использования при проверке гипотезы о независимости наборов случайных величин $x = (x_v, v = \overline{1, k})$:

$$H_0: p(x_v, v = \overline{1, k}) = \prod_{v=1}^k p(x_v).$$

В этих условиях решается двухальтернативная задача распознавания образов типа (5) в многомерном пространстве признаков.



Рис. 1. Фрагменты спутниковой съёмки Sentinel-2. Антропогенные территории: карьер (а), пригородная застройка (b)

Зависимости между спектральными признаками данных дистанционного зондирования антропогенных территорий. Объектами исследования являются антропогенные территории в окрестностях г. Красноярска: карьер с координатами ($55^{\circ}57'–55^{\circ}58'$) с. ш., $92^{\circ}54'$ в. д. на высоте 375 м над уровнем моря (рис. 1, а); пригородная застройка с координатами $56^{\circ}03'$ с. ш., $92^{\circ}54'$ в. д. на высоте 195 м над уровнем моря (рис. 1, b).

Исходная информация формировалась по фрагментам съёмки спутника Sentinel-2 от 26.08.2021 г., выделенным белым контуром, по девяти спектральным каналам. Их характеристика приведена в табл. 1.

Спектральные признаки x_v , $v = \overline{1,9}$, соответствуют спектральным каналам B_v , $v = \overline{2,8}$, и $B11$, $B12$ (см. табл. 1), а их значения — количеству уровней яркости спектральных каналов, которые определяются радиометрическим разрешением спутника Sentinel-2. Для этого спутника потенциальный диапазон уровней яркости изменяется от 0 до 4095.

Карьер. Рассмотрим результаты применения разработанной методики проверки гипотезы о независимости случайных величин при анализе данных дистанционного зон-

Таблица 1

Характеристики используемых спектральных каналов спутника Sentinel-2

Признаки x_v , $v = \overline{1,9}$	Канал	Спектральные диапазоны съёмки	Длина волны, нм
x_1	$B2$	Синий (blue)	458–523
x_2	$B3$	Зелёный (green peak)	543–578
x_3	$B4$	Красный (red)	650–680
x_4	$B5$	Крайний красный (red edge)	698–713
x_5	$B6$	Крайний красный (red edge)	733–748
x_6	$B7$	Крайний красный (red edge)	773–793
x_7	$B8$	Ближний инфракрасный (NIR)	785–899
x_8	$B11$	Коротковолновый инфракрасный (SWIR)	1565–1655
x_9	$B12$	Коротковолновый инфракрасный (SWIR)	2100–2280

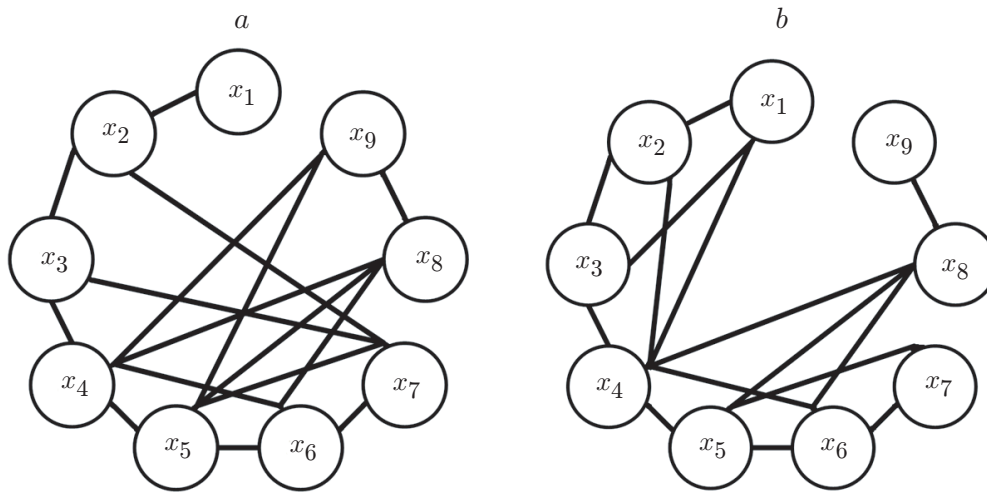


Рис. 2. Иллюстрация сильной линейной зависимости между парами спектральных признаков (x_i, x_j) , характеризующихся оценками коэффициентов корреляции $\bar{r}(i, j) \geq 0,9$ для карьера (a) и пригородной застройки (b)

дирования территории карьера. По данным спутниковой съёмки сформирована выборка $V_1 = (x_v^i, v = \overline{1, 9}, i = \overline{1, n_1})$ объёмом $n_1 = 3377$ (см. рис. 1, a).

Установлено, что сильной линейной зависимостью при оценке значений коэффициентов корреляции $\bar{r}(i, j) \geq 0,9$ обладают пары спектральных признаков (рис. 2, a):

$$(x_8, x_9), (x_5, x_9), (x_4, x_9), (x_4, x_8), (x_5, x_8), (x_6, x_8),$$

$$(x_2, x_7), (x_3, x_7), (x_5, x_7), (x_6, x_7), (x_4, x_6), (x_5, x_6),$$

$$(x_4, x_5), (x_3, x_4), (x_2, x_3), (x_1, x_2).$$

На рис. 2, a представлен граф $G(X, A)$, вершинами которого являются спектральные признаки $X = (x_v, v = \overline{1, 9})$. Между двумя вершинами графа x_i и x_j существует ребро $a_{ij} \in A$, если спектральные признаки x_i, x_j имеют сильную линейную зависимость при $\bar{r}(i, j) \geq 0,9$.

Предлагаемая методика проверки гипотезы о независимости случайных величин подтверждает наличие линейных зависимостей между 31 парой спектральных признаков при $\bar{r}(i, j) \geq 0,8$. Дополнительно определяют нелинейные зависимости между парами:

$$(x_1, x_9), (x_1, x_8), (x_1, x_7), (x_1, x_5), (x_1, x_4).$$

При этом данный вывод является достоверным и для всех пар спектральных признаков, так как $|\bar{\rho}_1 - \bar{\rho}_2| > 0,036$ при $\beta = 0,025$ и значении порога $D_\beta = 0,036$.

Полученные результаты анализа зависимостей между спектральными признаками территории карьера можно объяснить её относительной однородностью по сравнению с другими природными объектами. Имеются основания использовать показатели зависимости между спектральными признаками в качестве критерия обнаружения антропогенных территорий типа карьер по данным дистанционного зондирования.

Пригородная застройка. Объект исследования (см. рис. 1, b) характеризовался выборкой спектральных данных дистанционного зондирования

$$V_2 = (x_v^i, v = \overline{1, 9}, i = \overline{1, n_2})$$

объёмом $n_2 = 5049$.

Сильную линейную зависимость $\bar{r}(i, j) \geq 0,9$ имеют пары спектральных признаков (рис. 2, b):

$$\begin{aligned} &(x_8, x_9), (x_4, x_8), (x_5, x_8), (x_6, x_8), (x_5, x_7), \\ &(x_6, x_7), (x_4, x_6), (x_5, x_6), (x_4, x_5), (x_1, x_4), \\ &(x_2, x_4), (x_3, x_4), (x_1, x_3), (x_2, x_3), (x_1, x_2). \end{aligned}$$

Количество сильных линейных зависимостей между парами спектральных признаков для карьера и пригородной застройки составляет 16 и 15 соответственно. Для этих антропогенных территорий совпадают пары спектральных признаков с сильной линейной зависимостью:

$$\begin{aligned} &(x_8, x_9), (x_4, x_8), (x_5, x_8), (x_6, x_8), (x_5, x_7), \\ &(x_6, x_7), (x_4, x_6), (x_5, x_6), (x_4, x_5), (x_3, x_4), \\ &(x_2, x_3), (x_1, x_2). \end{aligned}$$

Особенностью для карьера является наличие сильных зависимостей между парами спектральных признаков:

$$(x_5, x_9), (x_4, x_9), (x_2, x_7), (x_3, x_7).$$

Эти пары спектральных признаков не свойственны для данных дистанционного зондирования пригородных застроек. Они отличаются наличием пар спектральных признаков:

$$(x_1, x_4), (x_2, x_4), (x_1, x_3).$$

Предлагаемая методика проверки гипотезы о независимости случайных величин подтверждает наличие линейной зависимости между 29 парами спектральных признаков и дополнительно определяет нелинейные зависимости между парами:

$$\begin{aligned} &(x_7, x_9), (x_4, x_9), (x_3, x_9), (x_2, x_9), \\ &(x_1, x_9), (x_1, x_8), (x_1, x_7). \end{aligned}$$

Полученные результаты являются достоверными для всех пар спектральных признаков, так как соблюдается условие $|\bar{\rho}_1 - \bar{\rho}_2| > D_\beta$ при $D_\beta = 0,029$ и риске $\beta = 0,025$ отвергнуть гипотезу H_0 равенства значений ρ_1, ρ_2 .

Распознавание типов антропогенных территорий по спектральным данным дистанционного зондирования. По результатам анализа спектральных признаков элементов антропогенных территорий (карьер, пригородная застройка) определены их наборы (x_4, x_5, x_6, x_8) , (x_5, x_6, x_8) , в которых пары являются взаимно линейно зависимыми с оценками коэффициентов корреляции больше 0,9. Признаки в этих наборах в значительной степени несут повторяющуюся информацию об объекте исследования. Поэтому в процессе организации вычислительных экспериментов при оценивании антропогенных территорий последовательно исключались спектральные признаки из обнаруженных наборов. Будем решать задачу распознавания рассматриваемых типов антропогенных территорий при следующих условиях:

1. Пространство спектральных признаков определяется их полным набором $x = (x_v, v = \overline{1, 9})$. Имеется обучающая выборка $V = (x^i, \sigma(i), i = \overline{1, n})$, где $n = n_1 + n_2$, $n_1 = 3377$ (карьер), а $n_2 = 5049$ (пригородная застройка).

2. Обучающая выборка

$$V(\bar{4}, \bar{5}) = (x^i(\bar{4}, \bar{5}), \sigma(i), i = \overline{1, n})$$

формируется в пространстве признаков x за исключением (x_4, x_5) .

3. Обучающая выборка

$$V(\bar{5}, \bar{6}) = (x^i(\bar{5}, \bar{6}), \sigma(i), i = \overline{1, n})$$

формируется в пространстве признаков x за исключением (x_5, x_6) .

4. Обучающая выборка

$$V(\bar{6}, \bar{8}) = (x^i(\bar{6}, \bar{8}), \sigma(i), i = \overline{1, n})$$

формируется в пространстве признаков x за исключением (x_6, x_8) .

5. Обучающая выборка

$$V(\bar{7}, \bar{8}) = (x^i(\bar{7}, \bar{8}), \sigma(i), i = \overline{1, n})$$

формируется в пространстве признаков x за исключением (x_7, x_8) .

6. Обучающая выборка

$$V(\bar{4}, \bar{5}, \bar{6}) = (x^i(\bar{4}, \bar{5}, \bar{6}), \sigma(i), i = \overline{1, n})$$

формируется в пространстве признаков x за исключением (x_4, x_5, x_6) .

7. Обучающая выборка

$$V(\bar{6}, \bar{7}, \bar{8}) = (x^i(\bar{6}, \bar{7}, \bar{8}), \sigma(i), i = \overline{1, n})$$

формируется в пространстве признаков x за исключением (x_6, x_7, x_8) .

Для распознавания типов антропогенных территорий используется непараметрическое решающее правило (4), коэффициенты размытости которого определяются из условия минимума оценки вероятности ошибки распознавания образов в режиме «скользящего экзамена». Полученные результаты приведены в табл. 2, где приняты следующие обозначения наборов признаков:

$$x = (x_v, v = \overline{1, 9}); \quad x(\bar{t}, \bar{j}) = (x_v, v = \overline{1, 9}, v \neq t, v \neq j);$$

$$x(\bar{t}, \bar{j}, \bar{q}) = (x_v, v = \overline{1, 9}, v \neq t, v \neq j, v \neq q).$$

Определим значения $D_{12} = |\bar{\rho} - \bar{\rho}(\bar{t}, \bar{j})|$ либо $D_{12} = |\bar{\rho} - \bar{\rho}(\bar{t}, \bar{j}, \bar{q})|$. Здесь $\bar{\rho}$ — оценка вероятности ошибки распознавания типов антропогенных территорий в пространстве признаков $x = (x_v, v = \bar{1}, \bar{9})$, а $\bar{\rho}(\bar{t}, \bar{j})$, $\bar{\rho}(\bar{t}, \bar{j}, \bar{q})$ — оценки вероятностей ошибок распознавания карьера и пригородной застройки в пространстве x за исключением признаков (x_t, x_j) либо (x_t, x_j, x_q) .

Отметим, что оценка вероятности ошибки распознавания рассматриваемых типов антропогенных территорий в полном пространстве $x_v, v = \bar{1}, \bar{9}$, спектральных признаков $\bar{\rho} = 0,012$. Это значение больше, чем для условий исключения наборов признаков (x_4, x_5) , (x_5, x_6) и (x_4, x_5, x_6) . Установлено увеличение оценки ошибки распознавания образов по сравнению с $\bar{\rho}$ при исключении наборов спектральных признаков (x_6, x_8) , (x_7, x_8) и (x_6, x_7, x_8) . Однако отличие вероятности ошибки распознавания образов ρ от ошибок, соответствующих сокращённым наборам, не является достоверным, так как при объёмах выборок n_1 и n_2 порог $D_\beta = 0,03$ больше значений D_{12} (табл. 2) при риске $\beta = 0,025$ отвергнуть гипотезу H_0 .

Наблюдается снижение ошибки распознавания образов по сравнению с полным набором $x_v, v = \bar{1}, \bar{9}$, если из обучающей выборки исключаются пары (x_4, x_5) , (x_5, x_6) . В этих условиях значения $\bar{\rho}(\bar{4}, \bar{5}) = 0,011$, $\bar{\rho}(\bar{5}, \bar{6}) = 0,01$ и $\bar{\rho} = 0,012$. При этом оценки коэффициентов корреляции между спектральными признаками для карьера (x_4, x_5) и (x_5, x_6) соответствуют большим значениям 0,977 и 0,993. Снижение ошибки распознавания типов антропогенных территорий при исключении признаков (x_4, x_5, x_6) до значения 0,008 по сравнению с $\bar{\rho} = 0,012$ объясняется значениями его парных оценок коэффициентов корреляции в интервале $[0,973; 0,993]$.

Подобные закономерности сохраняются и для территории пригородной застройки. Для этой территории значение $\bar{r}(4, 5) = 0,957$, а $\bar{r}(5, 6) = 0,993$. Причём парные оценки коэффициента корреляции в наборе (x_4, x_5, x_6) принадлежат интервалу $[0,931; 0,993]$.

При наличии набора спектральных признаков, характеризующихся сильной линейной зависимостью между его парами, появляется возможность сокращения количества признаков при распознавании типов антропогенных территорий. В обнаруженном наборе, например (x_4, x_5, x_6, x_8) , можно исключить признаки (x_4, x_5, x_6) , а вместо них оставить x_8 . При этом ошибка распознавания рассматриваемых типов антропогенных территорий в пространстве шести признаков составляет значение 0,008.

При анализе табл. 2 отмечается увеличение ошибки оценивания антропогенных территорий по сравнению с $\bar{\rho} = 0,012$ при исключении спектральных признаков (x_6, x_8) , (x_7, x_8)

Таблица 2

**Результаты распознавания типов антропогенных территорий
в пространстве спектральных признаков**

Набор признаков	Оценки вероятностей ошибок распознавания	Оценки оптимальных коэффициентов размытости	D_{12}
x	0,012	0,23	—
$x(\bar{4}, \bar{5})$	0,011	0,23	0,001
$x(\bar{5}, \bar{6})$	0,01	0,2	0,002
$x(\bar{6}, \bar{8})$	0,017	0,2	0,005
$x(\bar{7}, \bar{8})$	0,017	0,19	0,005
$x(\bar{4}, \bar{5}, \bar{6})$	0,008	0,19	0,004
$x(\bar{6}, \bar{7}, \bar{8})$	0,016	0,19	0,004

и набора (x_6, x_7, x_8) , соответствующие им значения оценок вероятностей ошибок распознавания образов равны 0,017; 0,017; 0,016.

Проведём анализ соответствующих этим признакам оценок коэффициентов корреляции. Для территории карьера значения $\bar{r}(6, 8) = 0,904$, $\bar{r}(7, 8) = 0,814$, $\bar{r}(6, 7) = 0,926$. Для пригородных застроек эти оценки коэффициентов корреляции соответствуют значениям: 0,943; 0,87; 0,938.

Из полученных результатов следует, что увеличение оценок вероятностей ошибок распознавания типов антропогенных территорий связано с относительным снижением значений оценок коэффициентов корреляции между спектральными признаками. Данный вывод обосновывает возможность снижения размерности пространства, в котором распознаются типы антропогенных территорий путём исключения из набора признаков с сильной линейной зависимостью их части.

Заключение. Предлагаемая методика проверки гипотез о независимости случайных величин позволяет обнаруживать стохастические зависимости между спектральными признаками. Результаты её применения определяются типом антропогенной территории и вводимыми параметрами, характеризующими исследуемую область значений случайных величин.

При наличии набора спектральных признаков, характеризующихся сильной линейной зависимостью между его парами, появляется возможность сокращения количества спектральных признаков при распознавании типов антропогенных территорий с уменьшением оценки вероятности ошибки распознавания карьера и пригородной территории.

Увеличение оценок вероятностей ошибок распознавания типов антропогенных территорий связано с относительным снижением значений оценок коэффициентов корреляции между спектральными признаками.

СПИСОК ЛИТЕРАТУРЫ

1. **Пугачев В. С.** Теория вероятностей и математическая статистика. Изд. 2-е, испр. и доп. М: Физматлит, 2002. 496 с.
2. **Зеньков И. В., Лапко А. В., Лапко В. А. и др.** Непараметрический алгоритм распознавания образов в задаче проверки гипотезы о независимости случайных величин // Компьютерная оптика. 2021. **45**, № 5. С. 767–772. DOI: 10.18287/2412-6179-CO-871.
3. **Лапко А. В., Лапко В. А., Бахтина А. В.** Исследование методики проверки гипотезы о независимости двумерных случайных величин с использованием непараметрического классификатора // Автометрия. 2021. **57**, № 6. С. 90–100. DOI: 10.15372/AUT20210610.
4. **Лапко А. В., Лапко В. А., Бахтина А. В.** Применение непараметрического алгоритма распознавания образов в задаче проверки гипотезы о независимости переменных неоднозначных функций // Измерительная техника. 2022. № 1. С. 17–22. DOI: 10.32446/0368-1025it.2022-01-17-22.
5. **Лапко А. В., Лапко В. А., Бахтина А. В.** Сравнение методик проверки гипотезы о независимости случайных величин, основанных на непараметрическом классификаторе и критерии Пирсона // Автометрия. 2023. **59**, № 5. С. 36–46. DOI: 10.15372/AUT20230504.
6. **Parzen E.** On Estimation of a Probability Density Function and Mode // Ann. Math. Statistic. 1962. **33**, N 3. P. 1065–1076.
7. **Епанечников В. А.** Непараметрическая оценка многомерной плотности вероятности // Теория вероятностей и её применения. 1969. **14**, вып. 1. С. 156–161.
8. **Rudemo M.** Empirical Choice of Histograms and Kernel Density Estimators // Scand. Journ. Statist. 1982. **9**, N 2. P. 65–78.

9. **Hall P.** Large Sample Optimality of Least Squares Cross-Validation in Density Estimation // Ann. Statist. 1983. **11**, N 4. P. 1156–1174.
10. **Bowman A. W.** A comparative study of some kernel-based nonparametric density estimators // Journ. Statist. Comput. and Simul. 1985. **21**, Iss. 3–4. P. 313–327. DOI: 10.1080/00949658508810822.
11. **Jiang M., Provost S. B.** A hybrid bandwidth selection methodology for kernel density estimation // Journ. Statist. Comput. and Simul. 2014. **84**, Iss. 3. P. 614–627. DOI: 10.1080/00949655.2012.721366.
12. **Shimazaki H., Shinomoto S.** Kernel bandwidth optimization in spike rate estimation // Journ. Comput. Neurosci. 2009. **29**, Iss. 1–2. P. 171–182. DOI: 10.1007/s10827-009-0180-4.
13. **Silverman B. W.** Density Estimation for Statistics and Data Analysis / Monographs on Statistics and Applied Probability. Vol. 26. London—N. Y.: Chapman & Hall, 1986. 175 p.
14. **Scott D. W.** Multivariate Density Estimation: Theory, Practice, and Visualization. 2nd Ed. N. J.: John Wiley & Sons, 2015. 384 p.
15. **Шаракшанэ А. С., Железнов И. Г., Ивницкий В. А.** Сложные системы. М.: Высш. шк., 1977. 248 с.

Поступила в редакцию 11.01.2024

После доработки 10.02.2024

Принята к публикации 04.06.2024
