

УДК 681.518.03

О. М. Роїк, д. т. н., проф.,

Ю. Я. Табачник, асп.

АНАЛІЗ МЕТОДІВ АВТОМАТИЧНОЇ КЛАСИФІКАЦІЇ ПРОСТОРОВИХ ДАНИХ У ГЕОІНФОРМАЦІЙНИХ СИСТЕМАХ

Постановка задачі

З розвитком комп'ютерних технологій з'явилася можливість перенести величезну кількість картографічної інформації з паперових носіїв (топографічні плани, карти тощо) у електронний вигляд. Програмні засоби, що використовуються для зберігання, відображення та аналізу географічних даних, носять назву географічних інформаційних систем (ГІС).

Однією із задач, що виникають під час проведення географічних, екологічних, економічних та інших досліджень за допомогою ГІС, є задача класифікації певної інформації. Вихідною інформацією для методів класифікації є багатовимірні дані, що описують досліджувані складні об'єкти, які знаходяться у атрибутивному багатовимірному просторі ознак. Інформація про ознаки об'єктів та зв'язки між ними знаходиться у таблицях вигляду «об'єкт–ознака» та «об'єкт–об'єкт» відповідно. Метою класифікації є розбиття множини об'єктів на деяку (як заздалегідь задану, так і не задану) кількість класів таким чином, щоб об'єкти одного класу були максимально подібними за обраними ознаками, а об'єкти різних класів – максимально відрізнялись.

Особливістю досліджень, у яких використовується інформація про географічні об'єкти, є оперування просторовими даними, тобто даними, що мають атрибут місцезнаходження. Для позна-

чення інформації про об'єкти, що одночасно знаходяться у атрибутивному та географічному просторі, використовується термін «просторові дані».

Зазвичай просторові дані є багатомірними. Багатомірність при цьому виникає під час аналізу наборів атрибутів просторових об'єктів (наприклад, зайнятості населення у різних галузях промисловості для карти адміністративно-територіального поділу). Сучасні ГІС практично не мають у своєму складі методів оброблення багатомірних даних. Тому в даній роботі основна увага приділена саме методам автоматичної класифікації багатомірних даних.

Виходячи з вищеведеного, впливає задача аналізу наявних методів автоматичної класифікації багатомірних даних з точки зору можливості їх застосування у геоінформаційних системах, та задача аналізу реалізації методів автоматичної класифікації та районування у наявних ГІС.

Формалізація задачі автоматичної класифікації

Позначимо всю множину об'єктів, що класифікуються, символом $O = \{o_1, \dots, o_n\}$, де o_i — i -й об'єкт, N — кількість об'єктів.

Метою класифікації є отримання певної кількості груп (класів) об'єктів. У межах кожного класу об'єкти мають бути максимально однорідними у деякому розумінні, а об'єкти різних класів — максимально відрізнятися.

Будемо називати j -м класом підмножину об'єктів

$$S_j = \{o_{ij} \in O \mid j \in 1, \dots, N_i\},$$

де $N_i = |S_j| \leq N$ — кількість об'єктів i -го класу (потужність i -го класу).

Під системою класів, отримання якої і є метою класифікації, будемо розуміти множину S із такими властивостями [1]:

1. $S = \{S_i \mid i = 1, \dots, K\}$, тобто S складається з K класів;
2. $\forall i, j \in \{1, \dots, K\} S_i \cap S_j = \emptyset$;
3. $\forall i \in \{1, \dots, K\} S_i \neq \emptyset$;
4. $\bigcup_{i=1}^K S_i = O$, $\sum_{i=1}^K N_i = N$.

Задати систему класів можливо N -мірним вектором належності об'єктів до одного з класів $v = (v_1, \dots, v_n)$, $v_i \in \{1, \dots, K\}$: $v_i = j \Leftrightarrow o_i \in S_j$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, K\}$.

Вихідні об'єкти можуть бути задані у просторі ознак за допомогою матриці «об'єкт–ознака» (чи «об'єкт–властивість»), яка відображає значення M ознак для N об'єктів та містить N рядків і M стовпчиків.

$$X = \begin{pmatrix} o_1 \\ \dots \\ o_n \end{pmatrix} = (o^{(1)}, \dots, o^{(M)}) = \begin{pmatrix} o_1^{(1)} & \dots & o_1^{(M)} \\ \dots & o_k^{(j)} & \dots \\ o_N^{(1)} & \dots & o_N^{(M)} \end{pmatrix},$$

де $o_i = (o_i^{(1)}, \dots, o_i^{(M)})$ — i -й об'єкт у M -мірному просторі ознак, $o^{(j)}$ — j -та ознака,

$o^{(j)} = (o_1^{(j)}, \dots, o_N^{(j)})^T$, $o_k^{(j)}$ — значення j -ї ознаки на k -му об'єкті, $i \in \{1, \dots, N\}$, $j \in \{1, \dots, M\}$.

Іншим способом завдання вихідних об'єктів є використання матриці вигляду «об'єкт–об'єкт», що відображає результат співставлення об'єктів між собою у просторі ознак. Така матриця є квадратною та містить N рядків та N стовпчиків

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{pmatrix},$$

де a_{ij} — результат співставлення i -го та j -го об'єктів, $i, j \in \{1, \dots, N\}$.

Звичайно a_{ij} є мірою однорідності (чи неоднорідності) об'єктів. Значення елементів матриці a_{ij} знаходиться на проміжку $[0,1]$, де 0 означає абсолютну відмінність об'єктів, а 1 – абсолютну схожість.

У випадку, якщо a_{ij} є мірою неоднорідності, матриця A є симетричною з нулями на головній діагоналі.

Особливості методів автоматичної класифікації географічних даних

Об'єктом класифікації у геоінформаційних системах виступає операційно-територіальна одиниця (ОТО) [2]. ОТО вибирається залежно від задачі дослідження — у якості ОТО можуть розглядатися адміністративно-територіальні одиниці, населенні пункти, земельні ділянки тощо.

За остаточною метою виділяють оціночні та типологічні класифікації [3]. Оціночні класифікації використовуються для отримання множини класів $S = \{S_1, \dots, S_K\}$. Для будь-яких двох класів S_i та S_j цієї множини можна визначити, який з цих елементів краще чи гірше іншого. В результаті проведення оціночної класифікації класи ОТО у множині S впорядковуються таким чином: $S_1 < S_2 < \dots < S_K$.

Під позначенням $S_i < S_j$ розуміється, що будь-яка ОТО i -го класу краще ОТО j -го класу за комплексом показників. У межах кожного класу операційно-територіальні одиниці мають приблизно однакову оцінку.

Завершальним етапом проведення оціночної класифікації є перехід до єдиної ознаки. Підсумковий показник отримують таким чином, щоб класи ОТО з максимальним значенням цієї ознаки інтерпретувалися як «найпридатніші», а з мінімальним відповідно — як «найнепридатніші». Результати оціночних класифікацій зазвичай використовують для прийняття ділових рішень (про доцільність використання ділянки під забудову, можливість вирощування певних видів культур тощо).

Типологічні класифікації використовуються для отримання стійких груп ОТО у M -вимірному просторі ознак, тобто можливих типів ОТО. На відміну від оціночних класифікацій «згортка» показників у єдину ознаку не використовується, а підсумкові класи можуть містити ОТО з високими значеннями параметрів за одним з показників та низькими – за іншим. Результати типологічної класифікації використовуються для виявлення закономірності у сполученнях ознак.

Класифікація методів автоматичного розбиття об'єктів на однорідні групи в залежності від наявності апіорної та попередньої вибіркової інформації має такий вигляд (табл.):

Систематизація методів автоматичної класифікації

Апіорні відомості про класи	Попередня інформація про вибірки відсутня	Попередня інформація про вибірки присутня
Певна інформація щодо закону розподілу досліджуваного вектора	Класифікація без навчання, розпізнавання образів «без навчання», ієрархічні класифікації	Непараметричні методи дискримінантного аналізу
Генеральні сукупності задані у вигляді параметричного сімейства законів розподілу ймовірностей з невідомими параметрами	Інтерпретація генеральної сукупності як суміші деяких генеральних сукупностей. Розділення даної суміші за допомогою методики оцінювання параметрів.	Параметричні методи дискримінантного аналізу
Досліджувані генеральні сукупності визначені однозначним описом відповідних законів	Класифікація при повністю визначених класах; розрізнення статистичних гіпотез	Навчальні вибірки у даному випадку не потрібні взагалі

Під час проведення автоматичної класифікації просторових даних наявність навчальних вибірок чи повністю визначених законів розподілу зустрічається дуже рідко. Тому найбільше поши-

рення у геоінформаційних системах набули методи розділення суміші розподілів та методи класифікації без навчання.

Методи розділення суміші розподілів

Дану групу методів можна використовувати для класифікації географічних даних, якщо зробити припущення, що розподіл об'єктів всередині кожного класу відбувається за нормальним законом. Вихідною інформацією для параметричної класифікації є подання географічних даних за допомогою таблиці «об'єкт-ознака». Модель суміші розподілів у випадку скінченної кількості класів M та однотипності компонентів має вигляд [4]

$$f_0(x) = \sum_{i=1}^M p_i F(x, \vartheta_i),$$

де $f_0(x)$ — густина генеральної сукупності, M — кількість компонентів суміші, p_i — ймовірність появи i -го компонента, $f(x, \vartheta_i)$ — щільність i -го компонента суміші, ϑ_i — вектор параметрів для i -го компонента суміші.

У застосуванні до задач автоматичної класифікації модель суміші розподілів передбачає, що i -й клас повністю описується i -ю компонентою суміші та ймовірністю її появи. Задача класифікації об'єктів полягає у визначенні класу, для якого поява даного об'єкту є найімовірнішою. Це можна записати таким чином:

$$o_i \in S_i \Leftrightarrow p_i f(o_i, \vartheta_i) = \max \{ p_t \cdot f(o_t, \vartheta_t) \mid t \in 1, \dots, M \}.$$

Найскладнішим етапом автоматичної класифікації на основі моделі суміші розподілів є процедура ідентифікації суміші, тобто алгоритм отримання кількості класів M та оцінок для p_i та ϑ_i , які необхідні для побудови правила прийняття рішень. Не усі суміші підлягають ідентифікації, оскільки не для усіх типів розподілу можна визначити єдині оцінки M , p_i та ϑ_i . Наприклад, неможливо ідентифікувати суміш рівномірних розподілів.

Найрозповсюдженішим алгоритмом оцінювання параметрів суміші є алгоритм оцінювання — максимізації (ЕМ-алгоритм) [5]. Даний метод є ітераційним та дозволяє визначити оцінки параметрів суміші p_i та ϑ_i , $i \in \{1, \dots, K\}$ для фіксованого числа класів (елементів) суміші за допомогою знаходження наближень до оцінок та максимізації з урахуванням отриманих наближень логарифмічної функції правдоподібності

$$L(p, \vartheta, k) = \ln \left(\prod_{i=1}^N f_0(o_i) \right) = \sum_{i=1}^N \ln f_0(o_i) = \sum_{i=1}^N \ln \left(\sum_{j=1}^K p_j F(o_i, \vartheta_j) \right). \quad (1)$$

ЕМ-алгоритм не дозволяє отримати оцінку кількості класів K_{res} . Для знаходження кількості класів можна використати такі методи:

1. Задати K_{min} та K_{max} та застосувати ЕМ-алгоритм для кожного елементу з діапазону $[K_{min}, K_{max}]$. Обчислити значення функції правдоподібності $L(p, \vartheta, k)$ для отриманих оцінок та проаналізувати графік $L(K)$. Формалізація останнього етапу пов'язана зі значними труднощами, тому під час розв'язання задач автоматичної класифікації цей метод не використовується.

2. Виконати перевірку серії гіпотез вигляду $H_0 : K_{res} = K$; $H_1 : K_{res} = K + 1$, починаючи з $K = K_{min}$.

Для врахування не лише атрибутивних, а й географічних ознак, доцільно використовувати модифікацію ЕМ-алгоритму — NEM-алгоритм [6]. Логарифмічна функція правдоподібності для NEM-алгоритму має вигляд

$$U = L(\vartheta) + \beta W(\tilde{\vartheta}),$$

де $L(\vartheta) = \sum_{i=1}^N \ln \left(\sum_{j=1}^K p_j F(o_i, \vartheta_j) \right)$ — функція правдоподібності (1), β — важливість географічного

простору, $W(\tilde{g}) = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N P_k^{(i)} P_k^{(j)} g_{ij}$, $G = (g_{ij})$ — матриця відстаней між ОТО у географічному просторі, $P_j^{(i)}$ — оцінки апостеріорних імовірностей належності i -ї ОТО до j -го класу, $i \in \{1, \dots, N\}$, $j \in \{1, \dots, N\}$.

Аналогічно до ЕМ-алгоритму, NEM-алгоритм ітераційно знаходить оцінки усіх параметрів, покращуючи їх на кожному кроці. Перевагою цього алгоритму є достатньо проста реалізація та адаптованість саме до задач автоматичної класифікації просторових даних. Одним з суттєвих недоліків алгоритму є необхідність підбирання параметра важливості географічного простору β , від значень якого дуже сильно залежить результат класифікації.

Ядерні методи автоматичної класифікації

Ядерні методи автоматичної класифікації відносяться до групи методів розпізнавання «без навчання». За допомогою цього класу методів виявляють скупчення об'єктів у просторі ознак. Базовими поняттями для ядерних методів класифікації є поняття ядра класу та відстані між класами.

Для класифікації просторових даних найчастіше використовуються кількісні ознаки. Розглянемо методи визначення відстані для даного типу ознак.

Найзагальнішим співвідношенням, що використовується для розрахунку відстані на M кількісних ознаках, є метрика Махаланобіса [5]

$$d_{mah}(o_i, o_j) = \sqrt{(o_i - o_j) \times \Lambda^T \times \Sigma^{-1} \times \Lambda \times (o_i - o_j)^T},$$

де Σ — коваріаційна матриця генеральної сукупності, Λ — матриця вагових коефіцієнтів (симетрична, невід'ємно-визначена).

$o_i = (o_i^{(1)}, \dots, o_i^{(M)})$, $o_i^{(x)}$ — значення показника $x \in \{1, \dots, M\}$ на i -му об'єкті;

$o_j = (o_j^{(1)}, \dots, o_j^{(M)})$, $o_j^{(x)}$ — значення показника $x \in \{1, \dots, M\}$ на j -му об'єкті.

Використання метрики махаланобіського типу є теоретично обґрунтованим, якщо ознаки розподілені за нормальним законом та мають однакову коваріаційну матрицю.

У геоінформаційних системах використовуються частинні випадки махаланобіського типу, а саме:

— зважена евклідова відстань $d_{we}(o_i, o_j) = \sqrt{\sum_{x=1}^M \omega_x (o_i^{(x)} - o_j^{(x)})^2}$,

де $\omega = (\omega_1, \dots, \omega_M)$ — вектор ваг;

— звичайна евклідова відстань $d_e(o_i, o_j) = \sqrt{\sum_{x=1}^M (o_i^{(x)} - o_j^{(x)})^2}$;

— манхеттенська відстань $d_e(o_i, o_j) = \sum_{x=1}^M |o_i^{(x)} - o_j^{(x)}|$.

Ядром класу називається певний об'єкт (який може бути як реальним, так і умовним), увесь комплекс характеристик якого є еталоном для даного класу. Зазвичай алгоритми, що засновані на описі класів ядрами, використовують узагальнену процедуру класифікації об'єктів за мінімальністю відстаней [4]:

1. Обрати метрику d .
2. Знайти ядра класів $c_1, \dots, c_K$, $c_i = (c_i^{(1)}, \dots, c_i^{(M)})$, $i \in \{1, \dots, K\}$.
3. Віднести об'єкт до відповідного ядра за мінімальністю відстані

$$o_t \in S_i \Leftrightarrow d(o_t, c_i) \leq d(o_t, c_j) \quad \forall i, j \in \{1, \dots, K\}, \quad t \in \{1, \dots, N\}.$$

На виході алгоритму буде отримана система класів $S = \{S_1, \dots, S_K\}$.

В залежності від способу подання об'єктів на вхід алгоритму ядерні процедури автоматичної класифікації об'єктів поділяються на послідовні (об'єкти подаються по одному) та паралельні (об'єкти подаються всі разом). Перевагою послідовного алгоритму класифікації є порівняно низька обчислювальна вартість, але при цьому результат класифікації може залежати від порядку подання об'єктів на вхід алгоритму. Паралельні методи вільні від такого недоліку, але, у зв'язку з більшою кількістю даних, що оброблюються одночасно, є складнішими та потребують вищих обчислювальних потужностей.

Ієрархічні методи автоматичної класифікації.

Ієрархічні методи класифікації також відносяться до групи методів розпізнавання образів «без навчання». Подібні методи орієнтовані на послідовне об'єднання вихідних об'єктів до деякої заздалегідь відомої чи невідомої кількості класів (агломеративні методи автоматичної класифікації), або на роз'єднання одного чи декількох класів до потрібного ступеня деталізації (дивізімні методи).

Результатом роботи ієрархічних алгоритмів класифікації є послідовність розділень множини O на множини класів $\{S^n \mid n = 1, \dots, I + 1\}$, де I — кількість проведених ітерацій, S^n — n -та система методів.

Серед агломеративних методів ієрархічної класифікації найбільше розповсюдження у геоінформації отримали колмогорівський та «швидкий» методи [7].

Колмогорівський агломеративний алгоритм починає свою роботу з формування $K_1 = N$ класів (тобто кожен об'єкт на нульовій ітерації виділяється у окремий клас). На кожному наступному кроці алгоритму відбувається об'єднання двох найближчих у певній метриці об'єктів в один. Останній $(N - 1)$ -й крок алгоритму об'єднує два сформовані на попередніх етапах класи в один клас, що включає всі наявні об'єкти

$$S_n = \{S_i^n \neq \emptyset \mid j = 1, \dots, N - n + 1\}, \quad O = \bigcup_{i=1}^{N-n+1} S_i^n, \quad S_i^n \cap S_j^n = \emptyset \\ \forall i, j = \{1, \dots, N - n + 1\}, \quad i \neq j, \quad n \in \{1, \dots, N\}.$$

«Швидкий» агломеративний алгоритм заснований на використанні деякої відомої чи такої, що змінюється у процесі класифікації, послідовності порогових значень $c_1, \dots, c_L$ (при цьому можлива ситуація, коли $c_n = c = \text{const} \quad \forall n = \{1, \dots, L\}$). Найпростішим підходом до автоматизації вибору порогових значень c_n є використання середньої відстані між об'єктами та номеру кроку (наприклад, $c_n = \frac{n}{N} \sum_{i=1}^N d(o_i, o_j)$).

Дивізімні методи автоматичної класифікації, що застосовуються у геоінформаційних системах, використовують апарат теорії графів (методи на основі розділення графу близькості). Аналогічно до «швидкого» агломеративного алгоритму, дивізімні методи потребують задавання послідовності порогових значень $c_1, \dots, c_L$. На кожній ітерації алгоритму $n = \{1, \dots, L\}$ відбувається виключення з графу ребер з більшими значеннями, після чого виконується перевірка, на скільки підграфів (компонент зв'язності) розпався вихідний граф. Після завершення роботи алгоритму кожний підграф є окремим класом.

Перевагами ієрархічних алгоритмів автоматичної класифікації є можливість їх застосування без наявності апріорної інформації про властивості класів (ядер класів чи навчальних вибірок). Суттєвою перевагою даного класу алгоритмів під час класифікації географічних даних є також легкість реалізації наочної візуалізації ходу та результатів класифікації на спеціальному графіку — деднрограмі.

До недоліків ієрархічних процедур слід віднести значну обчислювальну вартість реалізації. «Швидкі» алгоритми частково компенсують цей недолік, проте їх точність не в усіх випадках є достатньою.

Реалізація методів автоматичної класифікації у сучасних ГІС.

Будь-яку класифікацію, виходячи з розмірності простору ознак, можна поділити на одновимір-

ну ($M = 1$) та багатовимірну ($M > 1$). Оскільки класифікація просторових об'єктів є однією з основних функцій геоінформаційної системи, усі сучасні реалізації ГІС мають у своєму арсеналі декілька методів одномірної класифікації [8]. Найчастіше такими методами є «метод рівних площ» та «метод рівних інтервалів».

Як і у випадку багатовимірних класифікацій, результатом роботи методу одновимірної класифікації є розбиття множини O на класи, що не перетинаються. При цьому повинна виконуватись умова $S_1 \leq S_2 \leq \dots \leq S_k$.

Метод рівних площ дозволяє отримувати класи з приблизно однаковою кількістю об'єктів у кожному, тобто

$$|S_i| = N, \forall i \in \{1, \dots, K\}, \quad N_i \in \left\{ \left\lceil \frac{N}{K} \right\rceil, \left\lceil \frac{N}{K} \right\rceil + 1 \right\},$$

де $\lceil \rceil$ — операція виділення цілої частини.

За допомогою методу рівних інтервалів можна виконувати розбиття всього діапазону вимірювання заданої ознаки на відрізки рівної довжини, тобто

$$\forall o \in O \setminus o_{\min} \forall i \in \{1, \dots, K\} \quad o \in S_i \Leftrightarrow (i-1) \times \frac{o_{\max}^{(1)} - o_{\min}^{(1)}}{K} \leq o \leq i \times \frac{o_{\max}^{(1)} - o_{\min}^{(1)}}{K}, \quad o_{\min} \in S_i,$$

$$o_{\min}^{(1)} = \min \{ o_i^{(1)} \mid i \in 1, \dots, N \}, \quad o_{\max}^{(1)} = \max \{ o_i^{(1)} \mid i \in 1, \dots, N \}.$$

Багатовимірні алгоритми класифікації у сучасних геоінформаційних системах практично відсутні. Певні аналоги алгоритмів багатовимірної класифікації присутні у таких геоінформаційних системах, як IDRISI for Windows та ER Mapper, повноцінна реалізація методу багатовимірної класифікації – у ГІС ArcInfo 8 та вище.

Геоінформаційна система IDRISI for Windows включає в себе ряд аналітичних модулів, які можуть застосовуватись для проведення автоматичної класифікації. Це модуль багатокритеріальної оцінки території з використанням обмежень та неперервних факторів та модуль імовірнісної класифікації [9].

Геоінформаційна система ER Mapper 5.5 та вище має в своєму складі як методи кластеризації (безумовної класифікації), що використовують алгоритми ISOCCLASS та ISODATA, так і методи керованої класифікації. Функція керованої класифікації використовує відому статистику по заданій області чи класу. Система може бути визначена на основі еталонної області, заданої вручну, отриманої з файлу даних чи в результаті проведення попередніх класифікацій [9].

У одній з найбільш функціональних ГІС ArcInfo, починаючи з версії ArcInfo 8, з'явився додатковий геостатистичний модуль розширення Geostatistical Analyst та єдина функція багатомірної класифікації. Проте можливості Geostatistical Analyst є достатньо специфічними, а функція класифікації орієнтована лише на якісні ознаки і виконує лише виділення груп об'єктів з повним співпадінням значень усіх ознак.

На ринку програмних продуктів існує значна кількість пакетів, що можуть виконувати багатовимірну класифікацію просторових даних (без врахування їхньої географії). Серед таких продуктів можна виділити SAS, Statgraphics, SpatialStats тощо. Проте відсутність можливості оброблення саме географічної компоненти просторових даних робить такі пакети малопридатними для проведення автоматичної класифікації даних у геоінформаційних системах.

Ще одним недоліком використання зовнішніх статистичних пакетів для аналізу географічних даних є відсутність стандарту на обмін інформацією між програмами: в кожному конкретному випадку необхідно або розробляти конвертор даних між ГІС та програмою аналізу, або вручну переносити вихідні дані та результати аналізу між програмами.

Найсерйознішим кроком у розв'язанні проблеми використання в ГІС інструментів розвинутого статистичного аналізу був зроблений спеціалістами компаній ERSI та MathSoft. Компанія MathSoft, розробник програмного пакету MathCAD для символьних обчислень та системи статистичного аналізу S-PLUS, випустила модуль для ERSI ArcView GIS – S-Plus for ArcView GIS. Фактично новий модуль розширення забезпечує доступ з ArcView GIS до функцій пакету S-PLUS, який продовжує залишатися самостійним продуктом. Які окремих продукт також реалізований модуль S+SpatialStats – доповнення до пакету S-PLUS, що призначене саме для розв'язання задач просторової статистики.

Висновки

1. Проаналізовано методи автоматичної класифікації багатовимірних даних, наведені їх переваги і недоліки. Показано, які саме методи автоматичної класифікації доцільно використовувати для класифікації просторових даних з точки зору обчислювальної складності та придатності до аналізу як атрибутивного, так і географічного простору ознак.

2. Розглянуто можливості сучасних реалізацій геоінформаційних систем стосовно аналізу просторових даних. Показано, що аналіз багатовимірних даних для практично усіх систем є або неможливим, або пов'язаним зі значними труднощами. Виходячи з цього впливає необхідність створення модуля автоматичної класифікації просторових даних для використання в одній з наявних ГІС.

СПИСОК ЛІТЕРАТУРИ

1. Горелик А. Л., Скрипник В. А. Методы распознавания. — М.: Высш. Шк., 1989. — 232 с.
2. Тикунов В. С. Классификация в географии — ренессанс или увядание? Смоленск: Изд. Смоленского гуманитарного университета, 1997. — 367с.
3. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989. — 607 с.
4. Айвазян С. А. и др. Прикладная статистика: Исследование зависимостей. — М.: Финансы и статистика, 1985. — 587 с.
5. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. — М.: Юнити, 1998. — 1024 с.
6. Ambroise C., Dang V., Govaert G. Clustering of spatial data by the EM algorithm // Quantitative Geology and Geostatistics. — 1997. — № 9.— P. 493—504.
7. Янковой А. Г. Многомерный анализ в системе STATISTICA. — Одесса: Оптимум, 2001. Вып. 1. — 216 с.
8. Э. Митчел. Руководство по ГИС-анализу. Пространственные модели и взаимосвязи. — К.: ЕКОММ, 2000. — 180 с.
9. Бусыгин Б. С., Гаркуша И. Н., Серединин Е. С., Гаевенко А. Ю. Инструментарий геоинформационных систем (справочное пособие). — К.: ИРГ «ВБ», 2000. — 172 с.

Рекомендована кафедрою інтелектуальних систем

Надійшла до редакції 27.04.04
Рекомендована до друку 14.09.04

Роїк Олександр Митрофанович — завідувач кафедри інформаційного менеджменту; **Табачник Юрій Якович** — аспірант кафедри інтелектуальних систем.

Вінницький національний технічний університет