

УДК 681.325

В. В. ПУТРЕНКО¹, І. Г. КРАСОВСЬКА²¹ *Національний технічний університет України «Київський політехнічний інститут»*² *Інститут телекомунікацій і глобального інформаційного простору НАН України*

КЛАСТЕРИЗАЦІЯ ГЕОПРОСТОРОВИХ ДАНИХ ПРИ ІНТЕЛЕКТУАЛЬНОМУ АНАЛІЗІ

Розглянуто питання інтелектуального аналізу географічної інформації за допомогою процедури кластерного аналізу. Розроблено програмний модуль для проведення різних видів кластерного аналізу географічних об'єктів та автоматизованої побудови тематичних карт у програмному середовищі ArcGIS. На основі даних про накопичення відходів у регіонах України перевірено методику кластерного аналізу з використанням географічних координат центроїдів об'єктів для врахування їх просторового положення при тематичній класифікації. За результатами кластерного аналізу виділено групи регіонів України зі схожими показниками поводження з відходами та потенціальними небезпеками для навколишнього середовища.

Ключові слова: інтелектуальний аналіз, кластерний аналіз, геопросторові дані, ГІС, поводження з відходами.

Вступ

Сучасні підходи до управління територіальним розвитком ґрунтуються на складних інформаційних системах моделювання розвитку природних, техногенних та суспільних явищ та процесів, де визначну роль відіграють геоінформаційні системи (ГІС) як засіб накопичення, обробки та аналізу просторової інформації. Однією з інтегральних тем взаємодії суспільства та природи є прогнозування та попередження надзвичайних ситуацій, які загрожують збереженню природного середовища, біологічному різноманіттю, безпеці життя людей. В Україні існує декілька гострих екологічних проблем, які потенційно загрожують виникненням нових надзвичайних ситуацій різного територіального рівня та охопту, однією з яких є проблема поводження з відходами. Загальний обсяг накопичення відходів в Україні оцінюється в 35 млрд. тонн. Обсяг їх щорічного утворення досягає 780 – 800 млн. тонн. Переважну їх частину становлять промислові відходи – розкриті супутні породи, шлами – продукти збагачення корисних копалин, металургійні шлаки тощо. Значну небезпеку та проблему становлять місця накопичення відходів, які у більшості випадків не мають відповідних умов для їх збереження. В зв'язку з цим завдання прийняття науково обґрунтованих рішень у процесі управління відходами на різних територіальних та функціональних рівнях продовжує бути актуальним.

Аналіз останніх досліджень та публікацій

Кластеризація даних як інструмент оцінки поводження з відходами. Кластеризація територіальних одиниць на основі комплексних груп показників вже багато років використовується для типізації території та моделювання системоутворюючих зв'язків в економіці, суспільній географії та інших науках. Кластерний аналіз регіонів України за особливостями утворення та накопичення відходів дозволяє прогнозувати найбільші екологічні небезпеки, пов'язані з цим.

В галузі інтелектуального аналізу даних в Україні широко відомі роботи Інституту прикладного системного аналізу НТУУ «КПІ», Світового центру даних з геоінформатики та сталого розвитку М. З. Згуровського, Н. Д. Панкратова, А. І. Петренка, А. О. Болдака [1 – 4].

Дослідженнями моделей кластерного аналізу при розв'язанні геопросторових завдань у геоінформатиці займались В. С. Тикунов, О. А. Євтеєв, О. О. Іщук, О. О. Світличний, С. Н. Сербенюк та інші дослідники геоінформаційного напрямку [5-11].

В англомовній літературі основна увага була зосереджена на дослідженнях просторових методів неієрархічної або ієрархічної кластеризації, заснованих на основі розподілу щільності або на основі рівнорозподіленої сітки комірок відбору даних. Зокрема Д. Гуо досліджував процес регіоналізації просторових змінних з використанням динамічно обмеже-

ної агломераційної кластеризації [12]. Ж. Хан, М. Камбер, А. Тунг опрацювали загальні підходи до організації методів просторової кластеризації на основі просторової взаємодії об'єктів [13]. К. Коперські, Н. Стефанович запропонували метод двоступінчатої класифікації просторових даних [14]. С. Шюде присвятив роботу вивченню мультимасштабних кластерів у мережевому просторі [15].

Однак більшість цих моделей носить нетопологічний характер та слабкий зв'язок з предметною областю дослідження, на чому акцентується увага в даній роботі.

Формулювання цілей статті, постановка задачі

Розробка автоматизованих алгоритмів багатовимірної кластеризації векторних геоінформаційних моделей даних на прикладі діагностування та управління безпеками, пов'язаними з накопиченням відходів у регіонах України.

Завданнями дослідження є опрацювання методичних основ проведення багатовимірних класифікацій, вибір форми реалізації програмних алгоритмів у геоінформаційному середовищі, апробація кластерних моделей на прикладі даних про утворення відходів у регіонах України, виділення основних ознак досліджених кластерних груп та небезпек, пов'язаних з накопиченням відходів у регіонах.

Виклад основних результатів

Процедура кластерного аналізу відноситься до базових методів математичної обробки даних, коли кожен об'єкт потрапляє до певної групи на основі аналізу його ознак у багатовимірному просторі. Кластерний аналіз входить до числа значної кількості статистичних та моделюючих програмних систем, таких як STATISTICA, MatLab та інших. У більшості сучасних геоінформаційних пакетів процедури спрощеної кластеризації (класифікації) об'єктів також присутні. Проте, кожен з програмних продуктів має свої обмеження.

Кластеризація (або кластерний аналіз) – це процес розділення набору даних (чи об'єктів) у набір значущих підкласів, що називаються **кластерами**. Кластери допомагають користувачам зрозуміти природне групування чи структуру в наборі даних. Кластеризація як спосіб групування багатовимірних об'єктів ґрунтується на поданні результатів окремих спостережень точками геометричного простору з наступним виділенням груп цих точок [16].

Сьогодні є дві загальні методології кластеризації: неієрархічні (прості, базові, опорні) алгоритми

розділення та ієрархічні алгоритми.

Неієрархічні (подільні) методи: розділення набору з N елементів в K кластери (низхідний (top-down) алгоритм).

Ієрархічні методи: *агломеративний (agglomerative) алгоритм* – пари схожих елементів визначаються як кластери з подальшим створенням більших кластерів на основі схожості попередніх варіантів; *дивізивний (divisive) алгоритм* – починають з усього набору як кластера і далі послідовно ділять набір на вузкі частини.

Для того щоб групувати подібні елементи, потрібний спосіб обчислення ступеня близькості між об'єктами, тобто міри відстані (дистанції) між двома довільними об'єктами A і B . Цю міру відстані позначимо як $\text{dist}(A,B)$. Для обчислення цієї величини набір атрибутів конкретного об'єкта кластеризації подають часто у вигляді вектора відповідного простору.

Властивості міри відстані:

$$\begin{aligned} \text{для всіх об'єктів } A \text{ і } B: \text{dist}(A,B) > 0; \\ \text{dist}(A,B) = \text{dist}(B,A); \end{aligned}$$

$$\begin{aligned} \text{для кожного з об'єктів } A: \text{dist}(A,A) = 0; \\ \text{dist}(A,C) < \text{dist}(A,B) + \text{dist}(B,C). \end{aligned}$$

Виходячи з обчисленої величини $\text{dist}(A,B)$, можна визначити наскільки елементи A і B «близькі» один до одного: чим менша відстань, тим «близькі» елементи і навпаки.

Часто в кластеризації паралельно з терміном *близькість* використовується термін «схожість» (*similarity*). Міра схожості $\text{sim}(A,B)$ – протилежність міри відстані.

У кластерному аналізі використовуються різні міри близькості (схожості). В інтелектуальному аналізі найбільш відомі такі міри відстані:

- відстань Матхатмана

$$\text{dist}(X,Y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|,$$

де $X = (x_1, x_2, \dots, x_n)$; $Y = (y_1, y_2, \dots, y_n)$; два об'єкти з набором значень атрибутів, які мають бути нормалізовані;

- евклідова відстань

$$\text{dist}(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} ;$$

- схожість за косинусом

$$\text{dist}(X,Y) = 1 - \text{sim}(X,Y),$$

$$\text{де } \text{sim}(X, Y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}}$$

Ця формула визначає косинус кута між векторами X і Y . В інформаційно-пошукових системах ця величина визначає ступінь близькості (коефіцієнт кореляції) пошукових образів X і Y .

Метод неієрархічної кластеризації K -means являє собою спробу визначити оптимальну кількість місць K , де сума відстаней від кожної точки до кожного з центрів K зведено до мінімуму, що називається глобальною оптимізацією. На практиці реалізуються початкові припущення про місця K і здійснюється локальна оптимізація для місць кластерів щодо сусідніх точок. Таким чином, дві K -means процедури можуть не виробляти ті ж результати, навіть якщо K є ідентичними із-за кількох основних методів локальної оптимізації.

Алгоритм K -середніх побудований на чотирьох основних операціях:

- вибір початкових точок для K кластерів;
- розрахунок несхожості між об'єктом і середнім значенням кластера;
- включення об'єкта в кластер, у якого середнє ближче до об'єкта;
- перерахунок середнього значення кластера із об'єктів, які належать йому, щоб всередині кластера несхожість зводилась до мінімуму. Крім першої операції інші три операції виконуються повторно в алгоритмі до виконання умови (поки немає більше точок, що змінюють кластери). Суть алгоритму полягає в мінімізації функції вартості, яка залежить від несхожості між мірою кожного спостереження з середнім значенням кластеру. Несхожість, як правило, моделюється як евклідова відстань в k -means. Функція витрат полягає в наступному:

$$\text{Minimize } \sum_{j=1}^n \sum_{k=1}^k a_j d_{jk} z_{jk},$$

де j, k позначає загальну кількість спостережень і кластерів;

a_j позначає вагу спостереження j ;

d_{jk} позначає відстань між спостереженням j і центром кластера k ;

$$z_{jk} = \begin{cases} 1, & \text{якщо спостереження } j \text{ є в кластері } k, \\ 0, & \text{в іншому випадку} \end{cases},$$

z_{jk} є індикатором приналежності спостереження до кластеру, за умови, що z_{jk} приймає значення 1 тільки один раз для призначеного кластера, і значення 0 для інших кластерів.

Неієрархічні методи кластеризації краще підходять до аналізу геопросторових даних, тому що дозволяють об'єднувати об'єкти одразу у географічні групи. В той же час ієрархічні методи мають більш розвинутий механізм для визначення кількості кластерів, що застосовується у дослідженні.

Для геоінформаційних продуктів притаманний одно- або двовимірний аналіз векторних або растрових моделей даних, коли дані групуються на основі одного чи двох параметрів та відображаються за допомогою картограм. З іншого боку, статистичні пакети мають досконалі модулі кластерного аналізу, але при цьому відсутня можливість автоматизованого картографування даних та використання географічної інформації для цілей поділу об'єктів на групи. В останній час ця проблема вирішується за рахунок інтеграції статистичних пакетів та геоінформаційних продуктів через протоколи взаємодії даних (експорт / імпорт даних, підтримка форматів), але така процедура вимагає наявності декількох програмних продуктів на робочому місці та досвідченого користувача. Тому існує потреба у створенні додаткових модулів кластерного аналізу, які працюють безпосередньо у геоінформаційному середовищі та використовують його переваги.

Такий програмний модуль було розроблено у вигляді інструменту кластеризації на основі розширень геоінформаційної системи ArcGis 10.1

Алгоритм використання програмного модулю зображено на рис. 1. До основних його переваг відноситься можливість проведення операцій кластеризації у багатовимірному просторі, залучення географічної інформації як додаткових координатних вимірів об'єкту, побудова автоматизованих картографічних моделей та легенд за результатами аналізу. Використання географічних координат дозволяє врахувати взаємне розташування об'єктів у географічному просторі, топологію їх зв'язків та відповідно здійснювати районування території на основі математично обґрунтованих даних.

Програмний модуль розроблено за допомогою мови $C\#$, набору засобів розробки ArcGis SDK для .NET архітектури. Модуль містить набір методів агрегації даних на основі методу ближнього сусіда, Уорда, середньої відстані, центральної відстані, дальнього сусіда, а також можливість задавати кількість вихідних кластерів. Результатом роботи модуля є тематична карта та легенда до неї з виділеними кластерами, таблиця з розбивкою за класами географічних об'єктів, дендрограма кластерного аналізу (рис. 2).

Для проведення кластеризації було обрано адміністративно-територіальний поділ України на 24 області, Автономну Республіку Крим та міста республіканського підпорядкування Київ та Севасто-

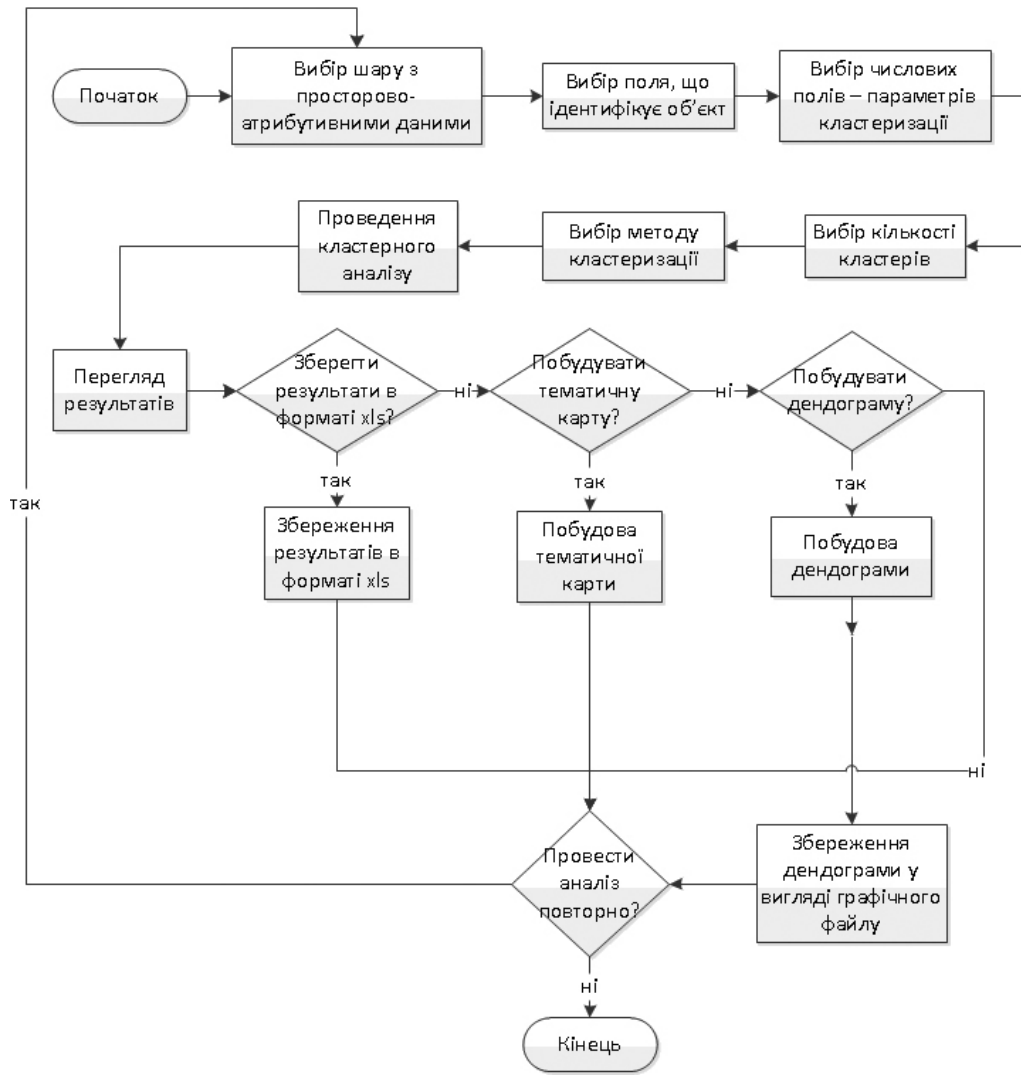


Рис. 1. Алгоритм використання модуля кластерного аналізу

Таблиця 1
Кластеризація регіонів із врахуванням геопросторових даних (метод – відстань Уорда)

Кластер	1	2	3	4	5	6	7
Об'єкти	Волинська	Київська	Донецька	Чернігівська	Полтавська	Дніпропетровська	Автономна Республіка Крим
	Рівненська	м. Київ	Луганська	Тернопільська	Харківська	Запорізька	Львівська
	Житомирська	Черкаська		Хмельницька	Миколаївська		Івано-Франківська
	Закарпатська	Кіровоградська		Сумська	Херсонська		
	Чернівецька			Одеська	м. Севастополь		
				Вінницька			

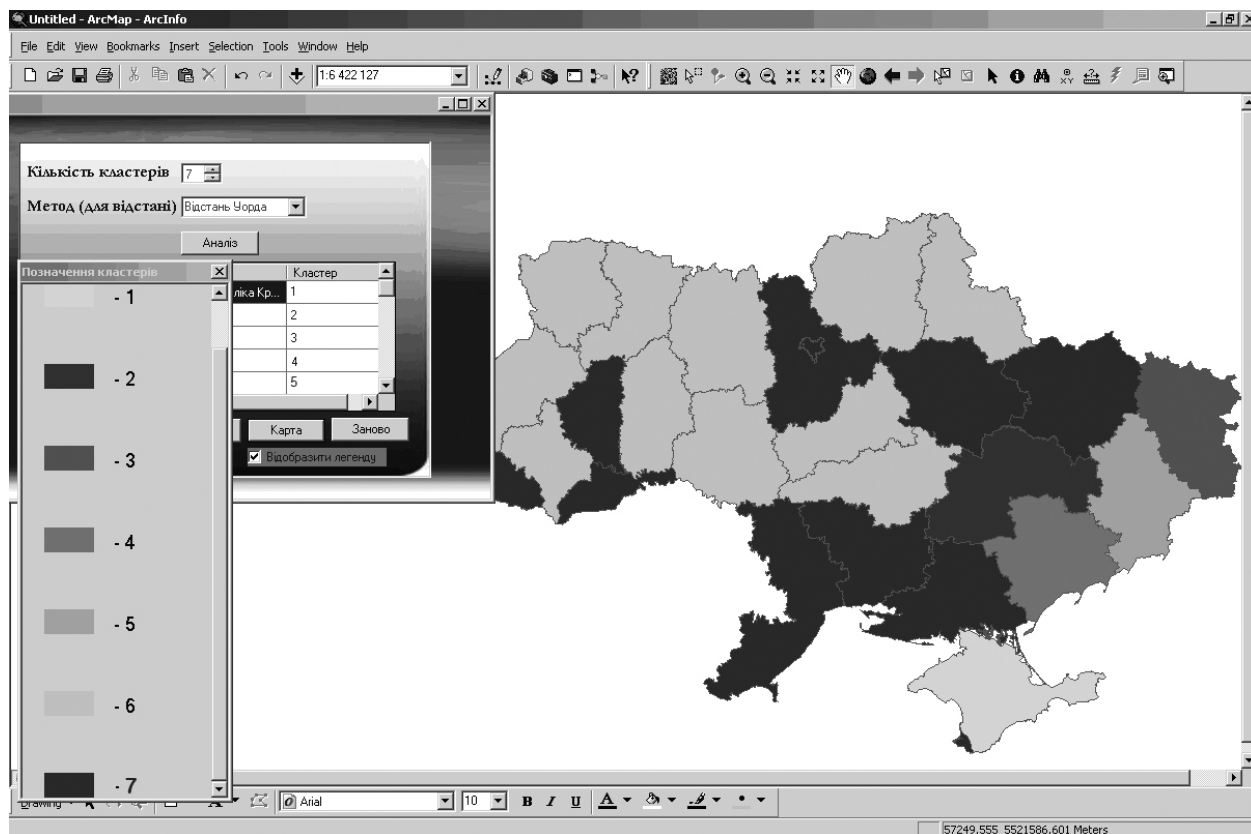


Рис. 2. Вікно ArcGis з результатами кластерного аналізу

поль. В якості вхідних даних для кластеризації використано статистичні дані за 2013 р. про утворення та накопичення на території регіонів відходів 1-4 класів небезпеки, що дало можливість проведення аналізу у 8-вимірному просторі [17]. Основним завданням кластеризації за цими показниками є виявлення основних типів регіонів за особливостями утворення відходів та небезпек з ними пов'язаних.

Основними характерними особливостями розподілу відходів в Україні є їх накопичення у регіонах видобутку корисних копалин та промислового виробництва у східних, центральних та деяких західних областях України. Пов'язані з цим небезпеки залежать від обсягів утворення та накопичення відходів. Відходи 1 групи небезпеки накопичені переважно у Донбасі, а також Кіровоградській, Сумській та Харківській областях, а за їх утворенням перші місця займають Луганська, Харківська, Полтавська та Донецька області. Відходи 2 групи накопичені у АР Крим, Донецькій та Луганській області, а за їх утворенням очолюють список АР Крим, Дніпропетровська, Сумська та Донецька області. Найбільша кількість відходів 3 групи накопичена у Запорізькій та Донецькій області, а за їх щорічним утворенням перші місця займає Полтавська, Миколаївська та Донецька області. Абсолютні обсяги накопичення відходів 4 класу небезпеки зафіксовані у Дніпропетровській, Донецькій та Луганській областях, які ра-

зом з Кіровоградщиною також мають найбільші обсяги щорічного утворення відходів цієї групи.

Після порівняльного аналізу методів кластеризації було вирішено обрати метод Уорда, який ґрунтується на принципі мінімізації дисперсії двох кластерів, що об'єднуються на кожному кроці [16].

$$V_l = \sum_i \sum_j (x_{ij} - x_{jl})^2,$$

де l – номер кластера;

i – номер об'єкта ($i=1,2, \dots, n_l$)

n_l – кількість об'єктів в l -м кластері,

J – номер ознаки ($j = 1,2, \dots, k$),

k – кількість ознак, що характеризують кожний об'єкт.

В дослідженні аналізовано випадок кластеризації у 8- та 10-вимірному просторі. В другому випадку використано в якості додаткових показників географічні координати центрів регіонів, які дозволяють врахувати близькість розміщення адміністративних одиниць та провести їх зонування.

У 8-вимірному просторі виділяється 5 кластерів регіонів. Перший кластер містить АР Крим, 2 – Дніпропетровська та Луганська область, 3 – Запорізька та Донецька область, 4 – Сумська, Кіровоградська, Івано-Франківська, Чернігівська, Рівнен-

ська, Черкаська, Житомирська, Львівська, Хмельницька, Волинська, Вінницька області. 5 – Харківська, Херсонська, Київська, Одеська, м. Київ, Тернопільська, Закарпатська, Чернівецька, м. Севастополь, Полтавська, Миколаївська.

У 10-вимірному просторі виділяється 7 кластерів, які демонструють географічне та тематичне поєднання (див. табл. 1). За цими даними було розроблено карту території України, на якій виділено сім груп областей (рис. 3). Перший кластер містить Донецьку та Луганську області з високим рівнем накопичення відходів, зокрема відходів 1 та 4 класів, що свідчить про високий ступінь небезпеки, пов'язаної з високотоксичними речовинами у місцях їх складування та небезпечними процесами, що пов'язані з відходами 4 класу. До 2 кластеру відноситься Дніпропетровська та Запорізька області з високим ступенем небезпеки хімічних відходів 1 класу та відходів металургії. 3 кластер містить АР Крим, Львівську та Івано-Франківську області, які вирізняються наявністю відходів 2 – 3 класів небезпеки, що пов'язано з

видобутком та промисловим виробництвом хімічних сполук. Склад 4 кластеру містить Київ, Київську, Черкаську та Кіровоградську області, де переважають токсичні відходи гірничого видобутку та відходи 1-2 класів небезпеки при хімічному виробництві. 5 кластер містить Харківську, Полтавську, Миколаївську, Херсонську області та м. Севастополь, де переважають відходи 3-4 ступенів небезпеки. 6 кластер складається з Чернігівської, Тернопільської, Хмельницької, Сумської, Одеської, Вінницької областей з відносно низьким рівнем небезпеки відходів, які переважно відносяться до 3 та 4 групи небезпеки. Останній, 7 кластер складається з Волинської, Рівненської, Житомирської на півночі та Закарпатської, Чернівецької областей на заході України з низькими обсягами утворення та складування відходів. Отримані кластери у своїй більшості мають компактну географічну структуру, яка дозволяє проводити не лише типізацію території, але здійснювати районування за певним набором ознак.



Рис. 3. Результати кластеризації за просторово-атрибутивними ознаками поводження з відходами

Висновки

Застосування методів кластерного аналізу у поєднанні з аналітичними можливостями геоінформаційних систем значно розширює методичний інструментарій сучасного інтелектуального аналізу даних. Створені засоби проведення кластерного

аналізу у середовищі ArcGis дозволяють класифікувати просторові об'єкти за просторово-атрибутивними параметрами. До основних переваг багатовимірного аналізу в середовищі ГІС слід віднести автоматизоване картографування кластерів, використання додаткової географічної інформації, варіабельність у прийнятті рішень щодо аналізу. В

результаті апробації програмного модуля кластеризації при дослідженні даних про поведінку з відходами було зроблено декілька висновків:

У порівнянні з кластеризацією за атрибутивними параметрами застосування методу із просторово-атрибутивними параметрами надало можливість виділення більш рівномірних за кількістю адміністративних одиниць та більш територіально зв'язаних кластерів.

Отримані результати дослідження доповнюють результати зарубіжних досліджень просторової кластеризації, тому що дозволяють використовувати разом із ієрархічною моделлю кластеризації топологічні моделі, засновані на координатному просторі, що дозволяє виділяти кластери із врахуванням умов сусідства під час однієї процедури обробки та оцінювати взаємний вплив існуючих небезпек.

Серед 7 виділених кластерів перші 4 відносяться до регіонів з відносно високим рівнем небезпеки виникнення НС, 3 наступних кластери мають помірну небезпеку, пов'язану з поведінкою з відходами.

Подальші дослідження пов'язані з розвитком підходів до кластеризації просторових об'єктів із врахуванням їх різної ваги на основі геометричних та семантичних атрибутів.

Література

1. *Аналіз сталого розвитку: глобальний і регіональний контексти [Текст] : монографія / Міжнародна рада з науки (ICSU) та ін. ; наук. кер. проекту М. З. Згуровський / М. З. Згуровський, О. А. Акімова, А. О. Болдак, П. Л. Вавулін, І. М. Джигирей, К. В. Єфремов, В. В. Путренко. – К. : НТУУ «КПІ», 2014. – Ч. 2. Україна в індикаторах сталого розвитку (2013). – 172 с.*

2. *Интеллектуальный анализ и системное согласование научных данных в междисциплинарных исследованиях [Текст] / М. З. Згуровський, А. О. Болдак, К. В. Єфремов // Кибернетика и системный анализ. – 2013. – № 4. – С. 62-75.*

3. *Згуровський, М. Системний аналіз: Проблеми, методологія, прикладні [Текст] / М. Згуровський, Н. Панкратова. – К. : Наук. думка, 2005. – 743 с.*

4. *Петренко, А. І. Grid та інтелектуальна обробка даних Data Mining [Текст] / А. І. Петренко // Систем. дослідж. та інформ. технології. – 2008. – № 4. – С. 97-110.*

5. *Геоінформатика: учебник для студ. высш. учебн. заведений [Текст]. В 2 кн. Кн 1. / Е. Г. Кап-*

ралов, А. В. Кошкарев, В. С. Тикунов и др. ; под ред. В. С. Тикунова. – 2-е изд. перераб. и доп. – М. : Издательский центр "Академия", 2008. – 384 с.

6. *Евтеев, О. А. Проектирование и составление социально-экономических карт [Текст] : учебник / О. А. Евтеев. – М. : Изд-во Моск. ун-та, 1999. – 224 с.*

7. *Іщук, О. О. Просторовий аналіз і моделювання в ГІС [Текст] : навчальний посібник / О. О. Іщук, М. М. Коржнев, О. Е. Кошляков ; за ред. акад. Д. М. Гродзинського. – К. : Видавничо-поліграфічний центр "Київський університет", 2003. – 200 с.*

8. *Світличний, О. О. Основи геоінформатики [Текст] : навчальний посібник / О. О. Світличний, С. В. Злотницький ; за заг. ред. О. О. Світличного. – Суми : ВТД Університетська книга, 2006. – 295 с.*

9. *Сербенюк, С. Н. Картография и геоинформатика – их взаимодействие [Текст] / С. Н. Сербенюк. – М., 1990. – 159 с.*

10. *Тикунов, В. С. Классификации в географии: ренессанс или увядание? [Текст] / В. С. Тикунов. – М., Смоленск : Изд-во СГУ, 1997. – 367 с.*

11. *Ситник, В. Ф. Интеллектуальный анализ данных (дейтамайнінг) [Текст] : навч. посібник / В. Ф. Ситник, М. Т. Краснюк. – К. : КНЕУ, 2007. – 376 с.*

12. *Guo, D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP) [Text] / D. Guo // International Journal of Geographical Information Science. – 2005. – № 22(7). – P. 801–823.*

13. *Han J. Spatial clustering methods in data mining [Text] : A survey / J. Han, M. Kamber, A. K. H. Tung ; In H. J. Miller & J. Han (Eds.). – Geographic data mining and knowledge discovery. London and New York: Taylor and Francis. – 2001. – P. 33–50.*

14. *Koperski, K. An efficient two-step method for classification of spatial data [Text] / K. Koperski, J. Han, N. Stefanovic // In 1998 international symposium on spatial data handling SDH'98. – Vancouver, BC, Canada, 1998. – P. 45–54.*

15. *Shiode, S. Detection of multi-scale clusters in network space [Text] / S. Shiode, N. Shiode // International Journal of Geographical Information Science. – 2009. – № 23. – P. 75–92.*

16. *Murray, Alan T. Cluster discovery techniques for exploratory spatial data analysis [Text] / Alan T. Murray, Vladimir Estivill-Castro // International Journal of Geographical Information Science. – 1998. – Vol. 12, Issue 5. – P. 431–443.*

17. *Довкілля України - 2013 [Текст]. – Державна служба статистики України, 2014. – 205 с.*

КЛАСТЕРИЗАЦИЯ ГЕОПРОСТРАНСТВЕННЫХ ДАННЫХ ПРИ ИНТЕЛЛЕКТУАЛЬНОМ АНАЛИЗЕ

В. В. Путренко, И. Г. Красовская

Рассмотрены вопросы интеллектуального анализа географической информации с помощью процедуры кластерного анализа. Разработан программный модуль для проведения различных видов кластерного анализа географических объектов и автоматизированного построения тематических карт в программной среде ArcGIS. На основе данных о накоплении отходов в регионах Украины проверена методика кластерного анализа с использованием географических координат центроидов объектов для учета их пространственного положения при тематической классификации. По результатам кластерного анализа выделены группы регионов Украины с похожими показателями обращения с отходами и потенциальными опасностями для окружающей среды.

Ключевые слова: интеллектуальный анализ, кластерный анализ, геопространственные данные, ГИС, обращения с отходами.

GEOSPATIAL DATA CLUSTERING IN THE PROCESS OF INTELLECTUAL ANALYSIS

V. V. Putrenko, I. G. Krasovskaya

The problems of data mining of geographic information with using the procedure of cluster analysis were considered. A software module for various types of cluster analysis of geographical objects and the automated construction of thematic maps was programmed in ArcGIS. Based on the data about accumulation of waste in the regions of Ukraine method of cluster analysis using the geographical coordinates of the centroids of objects to take into account their position in content classification was tested. The results of cluster analysis identified groups of regions of Ukraine with similar indicators for waste management and the potential hazards to the environment.

Keywords: spatial data mining, cluster analysis, geospatial data, GIS, waste management.

Путренко Віктор Валентинович – канд. геогр. наук, докторант, Національний технічний університет України «Київський політехнічний інститут».

Красовська Інеса Григорівна – канд. техн. наук, ст. наук. співр., Інститут телекомунікацій і глобального інформаційного простору НАН України, Київ, Україна, e-mail: ines75ma@ukr.net.