

СИСТЕМНІ ОСНОВИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ГЕОПРОСТОРОВИХ ДАНИХ

В.В. ПУТРЕНКО

Здійснено оглядове дослідження наукового напрямку інтелектуального аналізу геопросторових даних (ІАГД). Виявлено основні передумови формування цього напрямку і його зв'язок із геоінформатикою, системним аналізом та інтелектуальним аналізом даних. Проведено бібліографічне дослідження зарубіжних і вітчизняних публікацій в галузі ІАГД. У ході дослідження подано визначення ІАГД, виявлено основні завдання, функції та етапи його проведення, визначено коло перспективних напрямків розвитку та його зв'язок з підтримкою ухвалення рішень у регіональному управлінні. Із використанням ІАГД методів кластеризації гарячих точок проведено дослідження перевищення гранично допустимих концентрацій урану в підземних водах на території України на основі даних геологічних зйомок і виявлено зони обмеженнями використання підземних вод.

ВСТУП

Геоінформаційні технології стали важливою складовою інновацій, які формують сучасний вигляд інформаційного суспільства. Використання геопросторової інформації у вирішенні чисельних управлінських, бізнесових та інженерних завдань дозволяє підвищити їх ефективність та наочність процесу виконання. Поряд зі специфічним інструментарієм сучасних геоінформаційних систем (ГІС) базовим елементом стали геопросторові дані, що накопичуються у спеціалізованих сховищах.

Стрімкий розвиток геопросторової індустрії став можливим завдяки випереджаючому накопиченню геопросторових даних, які надходять із різних джерел. На сьогодні традиційне виробництво геопросторових даних, що ґрунтується на виконанні картографічних та кадастрових робіт, зазнає великих змін. Це пов'язано з подоланням високої вартості цих робіт, зміною суспільного замовлення та еволюцією сучасних інформаційних технологій. Постійно розширюється спектр галузей дистанційного зондування Землі, у якому постійно з'являються нові технічні засоби отримання даних та алгоритми їх подальшої обробки. Технології глобальних навігаційних супутникових систем (найбільш відома американська система GPS) стали джерелом постійного потоку геопросторових даних щодо місцеположення об'єктів та їх поведінки у просторі. Активно розвивається напрям волонтерського збору геопросторових даних з відкритою ліцензією завдяки таким проектам як OpenStreetMap [1].

Такий лавиноподібний потік накопичення структурованих та неструктурованих даних з геопозиційною складовою призвів до необхідності використання методів обробки Big Data та ІАГД, що дозволяє виявляти нові знання та закономірності у великих обсягах геоданих.

Тому дослідження особливостей методів ІАГД геопросторових даних є актуальним та перспективним напрямом наукових досліджень.

Особливе значення розробка методології та методики ІАГД має для управління територіальним розвитком, як елементом просторової організації суспільства. У цьому контексті ІАГД може надавати обґрунтовану інформацію і бути складовою систем підтримки прийняття рішень (СППР) на різному адміністративно-територіальному рівні, що допоможе в управлінні територією. На цих засадах сформовано мету та основні завдання дослідження.

Мета роботи — дослідження основ ІАГД, його структури та методів системного використання в процесі підтримки прийняття рішень.

Завданнями статті є формування з системних позицій визначення інтелектуального аналізу геопросторових даних, дослідження структури ІАГД та його місця в ГІС та СППР, аналіз складових процесу ІАГД, вивчення та прикладне застосування основних методів ІАГД.

ОГЛЯД ЛІТЕРАТУРИ

Методологічні підходи до ІАГД почали свій розвиток з середини ХХ сторіччя, коли разом із індустріальним розвитком економіки з'явилась необхідність у дослідженні оптимальних підходів до просторової організації суспільних процесів та охорони і раціонального використання природних ресурсів. Одночасно з цим з кінця 60-х років починається розвиток геоінформаційних систем як інформаційної основи для впровадження методів аналізу даних. Поява високопродуктивних обчислювальних машин разом з розвитком геопросторової індустрії та загостренням проблем економічного розвитку та екологічної ситуації призвели до активізації досліджень ІАГД та виділення його в окремий напрям.

В Україні таким поштовхом стала катастрофа на Чорнобильській АЕС, яка спонукала широке коло фахівців до використання геоінформаційних технологій щодо аналізу наслідків радіоактивного забруднення [2]. У галузі інтелектуального аналізу даних в Україні широко відомі роботи Інституту прикладного системного аналізу НТУУ «КПІ», Світового центру даних з геоінформатики та сталого розвитку вчених М.З. Згуровського, Н.Д. Панкратової, А.І. Петренко, А.О. Болдака [3–5]. Академічні та університетські дослідження інтелектуального аналізу відображено в роботах В.І. Зацерковного [4], Л.В. Саричевої [6], В.Ф. Ситника, М.Т. Краснюка, О.І. Черняка, П.В. Захарченка та ін. [7, 8].

У ХХІ столітті інтелектуальний аналіз геопросторових даних сформувався в окремий перспективний напрям наукових досліджень, який має свою структуру та особливості використання. Міжнародна дослідницька школа ІАГД сформувалась під впливом робіт таких вчених, як Г. Андрієнко, К. Клеве, М. Гахеган, Д. Гуо, М. Пьоке, Ж. Хан, К. Коперські, Н. Стефанович, М. Кулдорф, Ж. Менніс, Ж. Ліу, Х. Міллер, С. Опеншав, М. Чарлтон, К. Вімер, А. Крафт, С. Шекшар, С. Шюде [9-20] та ін.

МЕТОДОЛОГІЧНІ ОСНОВИ НАПРЯМУ ІАГД

Інтелектуальний аналіз геопросторових даних має глибоке коріння в традиційних галузях геопросторового аналізу (просторова статистика, аналітична картографія, пошуковий аналіз даних) та різних методах інтелектуального аналізу даних у статистиці та інформатиці. Його мета полягає в інтеграції і подальшому вдосконаленні методів у різних галузях для аналізу великих і складних наборів геопросторових даних.

У якості визначення пропонуємо наступне:

Інтелектуальний аналіз геопросторових даних (spatial data mining) — напрям геоінформатики, направлений на системний аналіз геопросторових даних з метою пошуку закономірностей та відносин у даних, що засновані на місцеположенні та топологічних відношеннях.

ІАГД є зростаючим дослідницьким напрямом, який все ще перебуває на ранніх стадіях розвитку та оперує цілим рядом різних методів, що комбінують обчислювальні, статистичні та візуальні методики. Інтелектуальний аналіз даних та виявлення знань є ітераційним процесом, який містить кілька етапів, включаючи відбір даних, фільтрацію, первинну підготовку та перетворення; аналіз за допомогою обчислювальних алгоритмів і / або підходів із візуалізації, інтерпретації та оцінки результатів; формулювання або модифікацію гіпотез і теорій; корегування даних і методів аналізу; повторну оцінку результатів.

Інтелектуальний аналіз даних та виявлення знань має пошукову природу і є більш індуктивним процесом, ніж традиційні статистичні методи. Це природно вписується в початкову стадію дедуктивного процесу дослідження, де дослідники розробляють та модифікують теорії, що базуються на відкритій інформації за даними спостережень.

ІАГД знаходиться на перетині ГІС, Data Mining, статистики та прикладних галузей досліджень. З цієї точки зору використання інструментів інтелектуального аналізу даних разом із геоінформаційним блоком є значною перевагою обробки вхідної інформації для управління територією та використання ГІС в СППР. У процесі застосування геоінформаційних технологій у СППР територіального управління можна виділити три базові рівні. До першого інформаційного рівня відноситься вхідна інформація, яка використовується для обробки в ГІС. На сьогодні інтелектуальні системи можуть обробляти як структуровані, так і неструктуровані дані. Специфічними для використання в ГІС є бази геопросторових даних, дані дистанційного зондування Землі, дані з датчиків та сенсорів з відомими координатами, геотегована інформація з Інтернет та соціальних мереж, історичні картографічні дані, волонтерська географічна інформація тощо (рис. 1).

Поряд із ними широко використовуються механізми інтеграції просторової та непросторової інформації на основі процесів географічної прив'язки та геокодування. Це надає змогу повноцінно використовувати у ГІС інформацію із зовнішніх непросторових баз даних, які супроводжуються та наповнюються іншими органами управління. Необхідною складовою є бази знань, які структуровано за предметними галузями, що використовуються у СППР.

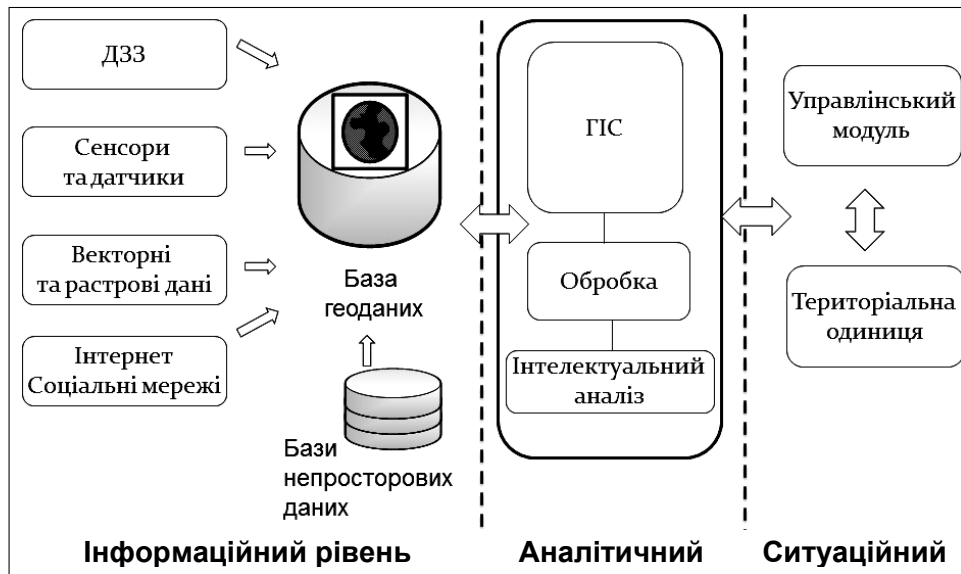


Рис. 1. ІАГД у структурі СППР територіального управління

Другий — аналітичний рівень спирається на використання геоінформаційної системи як платформи для обробки та аналізу вхідної інформації. Тут може бути виділено функції стандартизації та нормалізації, сумісної обробки, інтелектуального аналізу даних. На цьому етапі ГІС може взаємодіяти з іншими системами обробки даних та бути складовою СППР.

Третій — ситуаційний рівень, пов'язаний із необхідністю прийняття рішень щодо аналізу конкретних ситуацій за допомогою інструментів СППР та ГІС. Важливим компонентом у цьому випадку стають механізми колективної участі в роботі за участі керівних осіб різних рівнів, експертів технічного персоналу. До таких механізмів відносяться ситуаційні центри, онлайн конференції, робочі наради тощо.

Процес ІАГД містить основні кроки перетворення даних. Кожний аналіз здійснюється виконавцем з певною метою на основі вхідних баз даних та баз знань, які відповідають проблемному напрямку. На першому кроці обробка даних здійснюється за допомогою інтерфейсу СКБД. На цьому етапі здійснюється вибірка даних з усієї сукупності збереження даних, проведення простих операцій аналізу даних, які можливо здійснити на основі вбудованих засобів СКБД. На другому кроці відбувається обробка даних за допомогою алгоритмів ІАГД або їх послідовне використання. Отримані результати аналізуються з метою виявлення нової інформації та знань. За результатами цього аналізу відбувається оцінювання результатів обробки та даних і в залежності від того, чи відповідають отримані результати критеріям аналізу, процедура аналізу повторюється з іншими параметрами, або отримані результати передаються для прийняття на їх основі рішень (рис. 2).

До ІАГД відносяться декілька напрямів, які дозволяють провести різностороннє дослідження даних з урахуванням геопросторової складової. До їх числа відносяться методи класифікації геопросторових даних, кластеризація та регіоналізація, правила асоціацій, геостатистика, геовізуалізація. Цей перелік продовжує розширюватися разом із розвитком ІАГД. Кожний із цих напрямів містить сукупність методів, за допомогою яких відбувається аналіз даних (рис. 3).

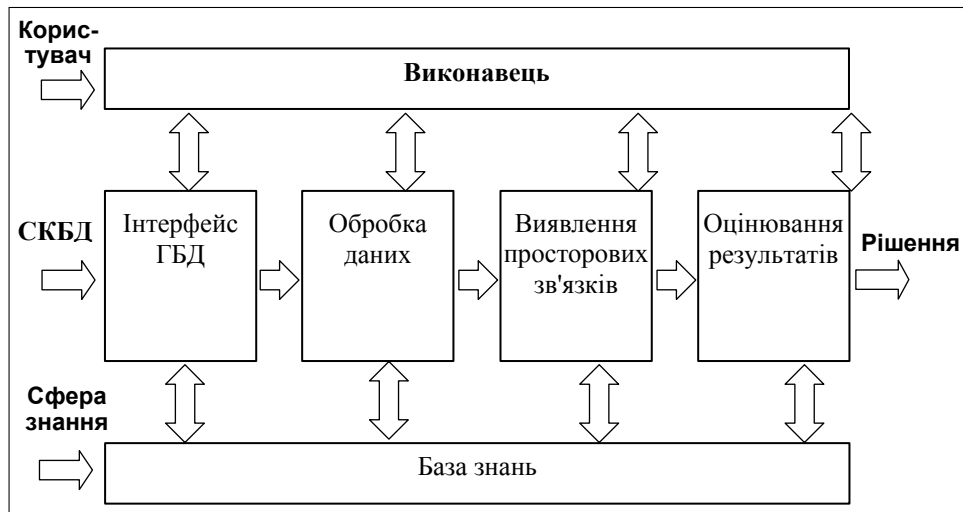


Рис. 2. Процес обробки даних в ІАГД



Рис. 3. Класифікаційна структура методів ІАГД

ПРОСТОРОВА КЛАСИФІКАЦІЯ ТА ПРОГНОЗУВАННЯ

Класифікація є процесом групування елементів даних за класами (категоріями) відповідно до їх властивостей (значення атрибутів). Класифікацію також називають керованою класифікацією, на відміну від некерованої класифікації (кластеризація). Керована класифікація потребує початкового набору даних для навчання (або налаштування) моделі класифікації, набору даних

перевірки для підтвердження (або оптимізації) конфігурації, і тестового набору даних для оцінки якості моделі.

До методів класифікації належать, наприклад, дерева рішень, штучні нейронні мережі (ШНМ), метод максимальної правдоподібності (ММП), лінійний дискримінантний аналіз (ЛДА), метод SVM (support vector machines — метод опорних векторів), методи найближчого сусідства і методи CBR (case-based reasoning — методи на основі міркування).

Просторові методи класифікації розширюють загальні методи класифікації враховуючи не тільки атрибути об'єкту, що класифікуються, але й атрибути сусідніх об'єктів та їх просторові відносини. Серед таких розробок варто звернути увагу на візуальний підхід до просторової класифікації, який було впроваджено в дерево рішень засноване на традиційному алгоритмі C4.5, що поєднується з картою візуалізації. Це дозволяє виявити просторові закономірності правил класифікації. Індукція дерева рішень також використовується для аналізу і прогнозування просторового вибору поведінки [9]. Дистанційне зондування є одним з основних напрямків, у якому зазвичай використовують методи класифікації для розподілу пікселів за категоріями. Широке використання при аналізі пікселів здобули штучні нейронні мережі.

Правила асоціації у видобутку даних були спочатку призначені для виявлення закономірностей між записами у великих транзакційних базах даних. Подібно до пошуку правил асоціації в транзакційних або реляційних базах даних, правила просторових асоціацій можуть бути знайдені в базі геопросторових даних, розглядаючи просторові властивості та предикати [16].

Правило просторової асоціації виражається у вигляді $A \Rightarrow B[s\%, c\%]$, де A та B є наборами просторових або непросторових предикатів, $s\%$ — підтримка правила, $c\%$ — впевненість у правилі.

Багато можливих просторових предикатів (наприклад, «близько до», «далеко від», «перетинання», «перекриття» тощо), можуть бути використані в правилах просторової асоціації. Але це потребує розгляду дуже великого обсягу обчислень різних просторових предикатів при виводі асоціативних правил з великих обсягів просторових даних. Ще одна потенційна проблема полягає у пошуку правил просторової асоціації, що може бути сформована великою кількістю правил, багато з яких є очевидними або загальновідомими. Тому необхідна наявність предметних знань, щоб відфільтрувати тривіальні правила і зосередитися тільки на нових і цікавих відкриттях.

Інтелектуальний аналіз просторової спільності розташування порівняно подібний, але технічно дуже відрізняється від правил асоціації. Враховуючи набір геопросторових даних і їх місце розташування, патерн спільності розташування є підмножинами об'єктів з певною частотою розташованих разом (наприклад, деякі види птахів, як правило, проживають поряд з певним типом дерев). Виміри та алгоритми для інтелектуального аналізу просторових спільностей розташування структур запропоновані в низці зарубіжних робіт [17–19].

Методологічно з правилами асоціації тісно пов'язаний аналіз топології, який дозволяє встановлювати правила поведінки об'єктів один відносно іншого, а також за допомогою функцій околу та оверлейних операцій виявляти подібні правила у геопросторових даних.

КЛАСТЕРИЗАЦІЯ ТА РЕГІОНАЛІЗАЦІЯ

Кластерний аналіз широко використовується як метод аналізу даних, який організовує набори елементів даних у групах (кластерах) таким чином, що елементи в одній групі більше схожі між собою та істотно відрізняються від елементів в інших угрупованнях. Методи *кластеризації* можна поділити на дві групи — неієрархічна та ієрархічна.

Для розгляду кластеризації геопросторових даних було проаналізовано три типи кластеризації, у тому числі геопросторову кластеризацію (кластеризація просторових точок), регіоналізацію (кластеризація з географічними обмеженнями примикання) і аналіз точкового розміщення (статистично вагомі скупчення точок «hot spot» за статистикою просторового сканування). Для першого типу просторової кластеризації схожість між точками даних або кластерами визначається за просторовими властивостями (місце і відстань). Просторові методи кластеризації можуть бути неієрархічними або ієрархічними, заснованими на основі розподілу щільності, або на основі рівнорозподіленої сітки комірок.

Регіоналізація є особливою формою кластеризації, яка прагне поєднати групи просторових об'єктів у просторово суміжні кластери (тобто регіони) у ході оптимізації цільової функції. Багато географічних додатків, таких як: кліматичне зонування, ландшафтний аналіз, сегментація даних дистанційного зондування, часто вимагають, щоб кластери були географічно суміжними. Існуючі методи районування, засновані на концепції кластеризації, можна розділити на три групи:

- багатовимірна (непросторова) кластеризація, яка потім потребує додаткової просторової обробки з метою упорядкування кластерів у регіони;
- кластеризація з виміром просторово зваженого відхилення, яка вважає просторові властивості фактором формування кластерів;
- кластеризація обмежена примиканням кластеризації, що слугує для формування просторової безперервності в процесі кластеризації [12].

Аналіз точкового розміщення, який також відомий як аналіз «гарячі місця», фокусується на виявленні незвичайних концентрацій подій у просторі, (наприклад, географічних кластерів хвороб, злочинності, або дорожньо-транспортних пригод). Загальною дослідницькою проблемою є визначення того, чи існує надлишок спостережуваних точкових подій (наприклад, випадків захворювання) для області (наприклад, у межах певної відстані до місця). Для пошуку подібних просторових кластерів використовується декілька статистичних методів, таких як метод Getis-Ord G_i^* [21], географічна машина аналізу (ГМА) [15] та сімейство просторово-часових методів статистики локального сканування. Все частіше статистика для виявлення просторових кластерів доступна для неевклідових просторів, зокрема мережевих просторів.

З метою радіоекологічної оцінки якості підземних вод для цілей водоспоживання було проведено дослідження розподілу вмісту урану в підземних водах на території України. Метою дослідження було виявлення територій, де спостерігається стійке перевищення гранично допустимої концентрації (ГДК) забруднювача, тобто перевищення концентрації має спостерігатися не тільки в одній пробі, а мати підтвердження на сусідніх ділянках.

За основу для проведення дослідження було обрано результати геологічної зйомки території України масштабу 1: 1 000 000 державного підприємства «Кіровгеологія», які містять відомості щодо вмісту природного урану в ґрунтових водах. База містить результати аналізу проб води у 6550 точках України, а також за її межами (на території Росії, Білорусі, Молдови).

Для здійснення цього дослідження було обрано аналіз гарячих точок за методом Getis-Ord G_i^* , який працює шляхом аналізу кожного об'єкта в контексті сусідніх об'єктів. Щоб бути статистично суттєвою гарячою точкою, об'єкт має бути високого значення і бути оточеним іншими об'єктами з такими ж високими значеннями. Це дозволяє виділити ділянки зі стабільним перевищенням допустимих концентрацій. Локальна сума для об'єкта та його сусідів порівнюється пропорційно з сумою всіх об'єктів. Коли локальна сума дуже відрізняється від очікуваної локальної суми, і ця відмінність є занадто великою, щоб бути результатом випадкового процесу, то виходить статистично значуща z -оцінка, тобто стандартне відхилення показника

$$G_i^* = \frac{\sum_{j=1}^n \omega_{i,j} x_j - \bar{X} \sum_{j=1}^n \omega_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n \omega_{i,j}^2 - \left(\sum_{j=1}^n \omega_{i,j} \right)^2}{n-1}}}$$

де x_j — атрибутивне значення для точки j , w_{ij} — просторова вага між точками i та j , n — загальна кількість точок та:

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n}, \quad S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}.$$

Статистична величина G_i^* , повернута для кожного об'єкта в наборі даних, є z -оцінкою. Для статистично значущих позитивних z -оцінок, чим більше z -оцінка, тим інтенсивніша кластеризація високих значень (гаряча точка). Для статистично значущих негативних z -оцінок, чим менше z -оцінка, тим інтенсивніша кластеризація низьких значень (холодна точка).

На рис. 4. показано карту точок проб ґрунтових вод на вміст урану, які були оцінені за допомогою аналізу гарячих точок. На карті червоними кольорами позначені гарячі точки та синіми кольорами холодні точки кластеризації. Побудова цієї моделі дозволила виявити статистично значущі зони високих значень природного урану та побудувати карту перевищення гранично допустимих концентрацій урану у ґрунтових водах, що може негативно впливати на здоров'я населення (рис. 5). У цьому випадку територію України поділено за статистичним показником ступеня небезпеки перевищення ГДК урану в ґрунтових водах. На карті можна побачити, що достатньо висока радіоекологічна небезпека використання ґрунтових вод спостерігається в Центральній та Південно-Східній частині України.

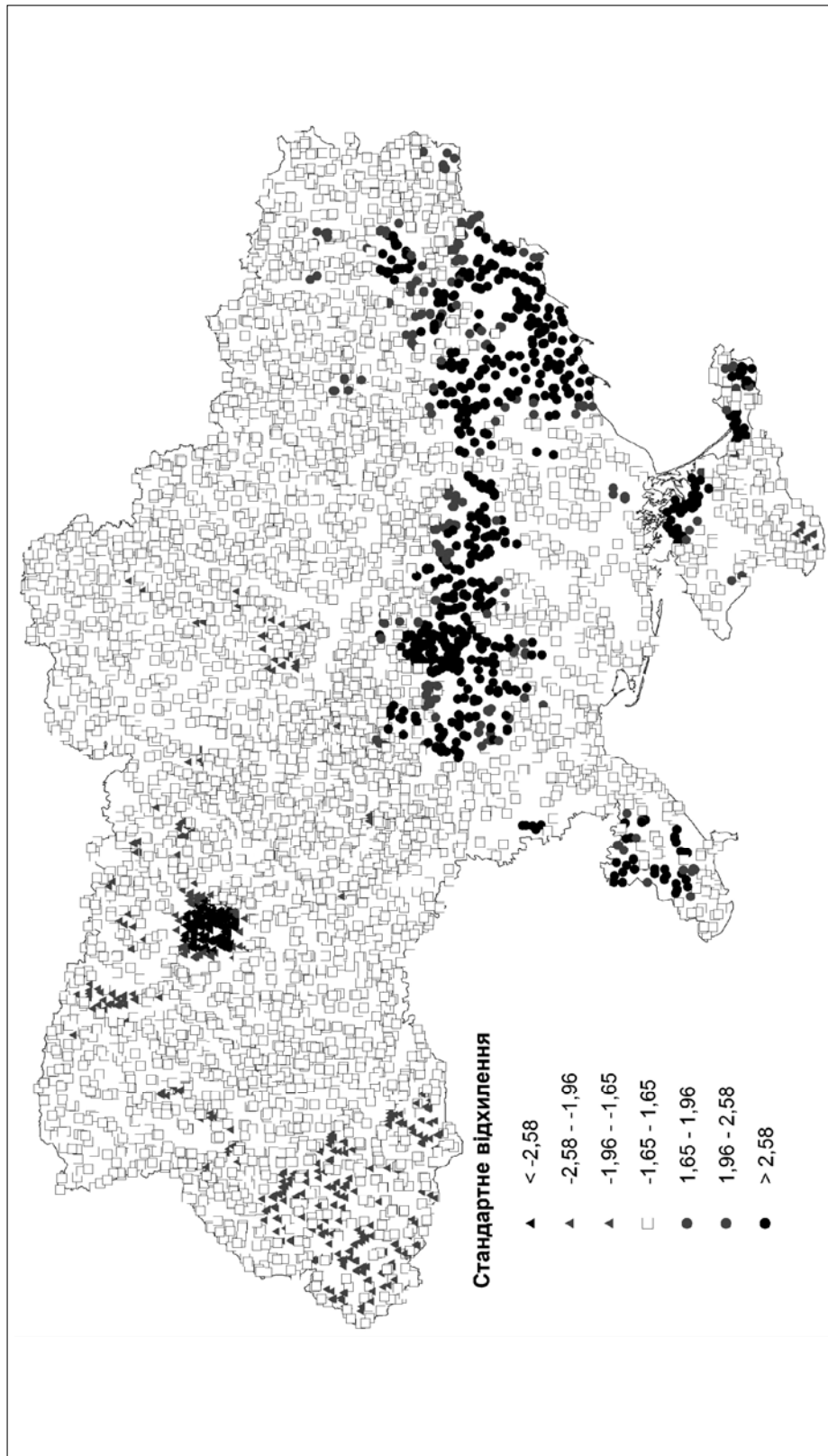


Рис. 4. Аналіз даних розподілу урану в ґрунтових водах методом горячих точок

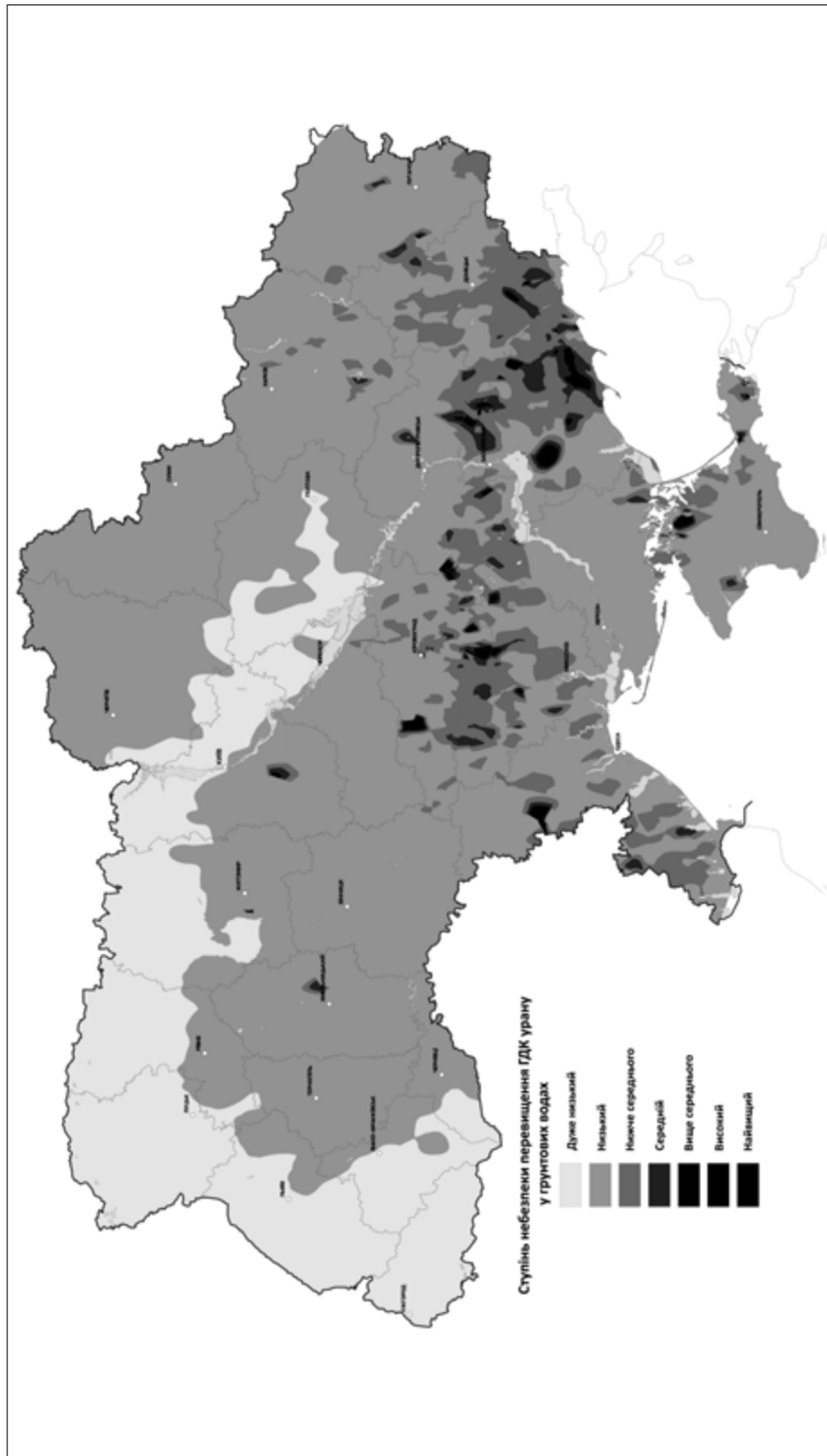


Рис. 5. Степінь небезпеки перевищення ГДК урану в ґрунтових водах

Таким чином, подібні методи кластеризації можуть бути корисними у широкому колі природничих та соціально-економічних досліджень. Іншим різновидом групи статистичних методів є використання ГМА.

У якості тестової статистики в ГМА використовується кількість точок (наприклад, випадки епідемічних хвороб) в межах певної області (кругової області навколо центру комірки). Щоб визначити, чи має значущість кількість точок в області, використовується процедура Монте-Карло для генерації великої кількості (наприклад, 500) випадкових наборів даних, кожний з яких представляє реалізацію нульової гіпотези в тій же самій області. Значення статистики випробувань розраховують для кожного випадкового набору даних i , таким чином, розподіл значень тестової статистики за наявної нульової гіпотези є похідним. Порівнюючи фактичне значення тестової статистики (тобто кількість точок) і отримані результати розподілу, отримуємо рівень значущості для тестової статистики в області. Потенційною проблемою з ГМА є те, що важко встановити рівень для завдання мультитестування. Обчислювальне навантаження також є недоліком, але в більшій чи меншій мірі всім статистичним додаткам сканування буде потрібна значна обчислювальна потужність для пошуку та тестування локальних кластерів.

Статистика локального сканування, розроблена Кульдорфом обчислює відношення правдоподібності для кожної локальної області. Для подолання проблеми множинного тестування, статистика локального сканування використовує максимальне відношення правдоподібності (яка є максимальним коефіцієнтом ймовірності серед усіх місць), як тестову статистику. Таким чином, спосіб статистики локального сканування визначає найімовірніший кластер, хоча множина вторинних кластерів також додається. Для цього алгоритм спочатку обчислює відношення правдоподібності для кожної із сукупності зон і знаходить максимум. Для виводу рівня значущості, генеруються копії наборів даних за нульової гіпотези, умови якої розповсюджуються на загальну кількість точок. Для кожної реплікації, значення тестової статистики обчислюють знову (тобто максимальне відношення правдоподібності знаходиться над усіма перерахованими місцями). Тоді фактичне значення тестової статистики порівнюється з тестовими значеннями всіх реплікацій для отримання рівня значущості для найбільш ймовірного кластеру, а також вторинних кластерів.

Характеризувати розподіл точок можна з використанням статистичних індексів. Прикладом такого підходу є діаграма Морішіта. Індекс Морішіта обчислюється для області, що розбита на рівнокутні комірки рівного розміру, за формулою [22]:

$$I_{\delta} = Q \frac{\sum_{i=1}^Q n_i(n_i - 1)}{N(N-1)},$$

де N — повне число точок мережі моніторингу; Q — кількість осередків розбиття; n_i ($i=1,2,\dots,Q$) — число точок мережі моніторингу, що потрапили в i -ту комірку. Цей індекс характеризує ймовірність того, що під час вибору двох випадкових точок вони опиняться в одній комірці. Існують три типи характерних значень індексу Морішіта, комбінації яких дозволяють судити про характеристики мережі моніторингу:

- Величина індексу Морішіта із зростанням розміру комірки зростає і прагне до 1. Тоді розподіл точок можна вважати рівномірним.
- Величина індексу Морішіта не залежить від розміру комірки і приблизно дорівнює ≈ 1 (коливається близько 1). Це означає, що розподіл точок випадковий і не має кластерних структур.
- Величина індексу Морішіта із зростанням розміру комірки зменшується або зростає вище 1 — розподіл точок мережі кластерний.

ГЕОВІЗУАЛІЗАЦІЯ

Цей напрям стосується розвитку теорії та методології полегшення конструювання знань на основі візуального дослідження та аналізу геопросторових даних та реалізації візуальних інструментів для подальшого пошуку знань, синтезу, зв'язків та використання [11]. Як галузь, що активно розвивається, геовізуалізація залучає інтереси різних наукових областей і розвивається разом із різноманітним набором напрямів досліджень. Основна відмінність між традиційною картографією та геовізуалізацією полягає в тому, що картографія фокусується на розробці та використанні карт для передачі інформації та суспільного споживання, а геовізуалізація підкреслює розвиток інтерактивних карт і пов'язаних з ними інструментів для дослідження даних, побудови гіпотез і створення нових знань.

Геовізуалізація також має тісні зв'язки з дослідницьким аналізом даних (EDA) і пошуковим просторовим аналізом даних (ESDA), які пов'язують статистичні графіки і карти і спираються на вміння людини-експерта взаємодіяти з даними, візуально виявляти закономірності і формулювати гіпотези/моделі. Однак, щоб впоратися з сучасними великими і різноманітними наборами геопросторових даних і полегшити відкриття і розуміння складної інформації, геовізуалізації необхідно вирішити ряд завдань, у тому числі:

- ефективну обробку дуже великих наборів даних (big data);
- обробку декількох сценаріїв і роботу з багатьма змінними одночасно, що дозволяє виявити складні зв'язки між даними;
- розробку ефективних користувальницьких інтерфейсів і інтерактивної стратегії полегшення пошукового процесу.

Для обробки великих масивів даних і візуалізації загальних закономірностей, візуальні підходи часто поєднуються з обчислювальними методами (кластеризація, класифікація та правила асоціації) для узагальнення даних, виділення структур і допомоги користувачам під час вивчення і розуміння закономірностей [9]. Для візуалізації різних точок зору і багатьох змінних часто необхідно об'єднувати методи візуалізації з методами зменшення розмірності, такими як багатовимірне шкалювання, аналіз головних компонент (PCA), карти самоорганізації, або інші методи пошуку оптимального відображення даних. Багатовимірне картографування є актуальною проблемою дослідження, для якого було розроблено численні підходи, такі як: спеціально розроблені символи; мультизв'язані перегляди; підходи, що ґрунтуються на кластеризації [13]. Дослідницькі зусилля третьої хвилі з'явилися як активний субнапрямок — візуальна аналітика.

ВИСНОВКИ

ІАГД є перспективним напрямом наукових досліджень, який розвивається на перетині геоінформатики, методів Data Mining та статистичних методів обробки даних із широким залученням методології системного аналізу. Основною його метою використання є пошук закономірностей та відносин у великих масивах даних, які містять просторово координовану прив'язку. Тому використання ІАГД у складі СППР у територіальному управлінні та прогнозуванні є важливим та актуальним інструментом обґрунтування управлінських рішень. Процес реалізації ІАГД відрізняється стадійністю та ітераційністю з метою пошуку оптимальних результатів аналізу. До основних напрямів ІАГД відносяться класифікація, кластеризація, правила асоціацій, геостатистика та геовізуалізація, які разом формують методологію системного інтелектуального аналізу з метою підтримки прийняття рішень. ІАГД підвищує ефективність обробки даних разом із іншими методами аналізу даних на основі різних інформаційних платформ.

Практичне застосування методів кластеризації для аналізу даних щодо розподілу вмісту природного урану у пробах підземних вод дозволило виявити території зі стабільним перевищенням норм ГДК, що може негативно впливати на стан здоров'я населення.

Подальші дослідження пов'язані з дослідженням методичного апарату окремих напрямів ІАГД та їх інтеграцією на системній основі.

ЛІТЕРАТУРА

1. *Goodchild M.F.* Citizens as sensors: The world of volunteered geography // *Journal of Geography*. — 2007. — **69**(4). — P. 211–221.
2. *Іщук О.О., Коржнев М.М., Кошляков О.Е.* Просторовий аналіз і моделювання в ГІС: Навч. посібник /; за ред. акад. Д. М. Гродзинського. — К.: Вид.-поліграфічний центр «Київський університет», 2003. — 200 с.
3. *Згуровський М.З., Болдак А.О., Єфремов К.В.* Интеллектуальный анализ и системное согласование научных данных в междисциплинарных исследованиях // *Кибернетика и системный анализ*. — 2013. — № 4. — С. 62–75.
4. *Зацерковний В.І.* Геоінформаційні технології як засіб інтелектуалізації управління територіальними об'єктами // *Вісник Астрономічної школи*. — 2014. — **10**, № 1–2. — 2014. — С. 97–101.
5. *Петренко А.І.* Grid та інтелектуальна обробка даних Data Mining // *Системні дослідження та інформаційні технології*. — 2008. — № 4. — С. 97–110.
6. *Сарычева Л.В.* Пространственно-временной подход в задачах кластеризации // *Штучний інтелект*. — 2006. — № 3. — С. 646–653.
7. *Ситник В.Ф., Краснюк М.Т.* Интеллектуальный анализ даних (дейтамайнінг): навч. посібник. — К.: КНЕУ, 2007. — 376 с.
8. *Черняк О.І., Захарченко П.В.* Интеллектуальный анализ даних: Підручник. — К.: Знання, 2014. — 599 с.
9. *Andrienko G., Andrienko N.* Data mining with C4.5 and interactive cartographic visualization // In N. W. G. T. Paton (Ed.), *User interfaces to data intensive systems*. Los Alamitos, CA: IEEE Computer Society. — 1999. — P. 162–165.
10. *Cleve C., Kelly M., Kearns F.R., Morltz M.* Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-

- resolution aerial photography // *Computers Environment and Urban Systems*. — 2008. — **32**(4). — P. 317–326.
11. Guo D., Gahegan M., MacEachren A.M., Zhou B. Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach // *Cartography and Geographic Information Science*. — P. 113–132.
 12. Guo D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP) // *International Journal of Geographical Information Science*. — 2005. — **22**(7). — P. 801–823.
 13. Han J., Kamber M., Tung A.K.H. Spatial clustering methods in data mining: A survey / 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 / K. Koperski, J. Han and N. Stefanovic // In 1998 international symposium on spatial data handling SDH'98, Vancouver, BC, Canada. — 1998. — P. 45–54.
 15. Kulldorff M. A spatial scan statistic // *Communications in Statistics. — Theory and Methods*. — 1997. — **26**(6). — P. 1481–1496.
 16. Mennis J., Liu J. W. Mining association rules in spatio-temporal data: An analysis of urban socioeconomic and land cover change // *Transactions in GIS*. — 2005. — **9**(1). — P. 5–17.
 17. Miller H., Han J. Geographic data mining and knowledge discovery: An overview // *Geographic data mining and knowledge discovery*. CRC Press, Taylor and Francis Group. — 2009. — P. 1–26.
 18. Openshaw S., Charlton M., Wymer C., Craft A. A mark 1 geographical analysis machine for the automated analysis of point data sets // *International Journal of Geographical Information Science*. — 1987. — **1**(4). — P. 335–358.
 19. Shekhar S., Huang Y. Discovering spatial co-location patterns: A summary of results / In C. Jensen, M. Schneider, B. Seeger, & V. Tsotras (Eds.). // *Advances in spatial and temporal databases, proceedings, lecture notes in computer science*. Berlin: Springer-Verlag. — 2001. — P. 236–256.
 20. Shiode S., Shiode N. Detection of multi-scale clusters in network space // *International Journal of Geographical Information Science*. — 2009. — **23**. — P. 75–92.
 21. Ord J.K., Getis A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application // *Geographical Analysis*. — 1995. — № 27(4). — P. 13–22.
 22. Cressie N. *Statistics for spatial data*. — NY: John Wiley & Sons, 1991. — 900 p.

Надійшло 15.05.2015