

УДК 681.58

Путренко В. В.

Національний технічний університет України «Київський політехнічний інститут»

КЛАСТЕРИЗАЦІЯ ГЕОПРОСТОРОВИХ ДАНИХ ЗЕМНОГО ПОКРИВУ УКРАЇНИ

У роботі визначено методика обробки даних земного покриття на прикладі території України з використанням карт самоорганізації Кахонена. Здійснено аналіз придатності для використання в якості джерела даних продукту GlobeLand30, розроблено методика вторинної обробки даних, проведення зонування і отримання показників зональної статистики, використання карт самоорганізації Кахонена для кластеризації адміністративних одиниць за співвідношенням видів землекористування.

Ключові слова: земний покрив, кластеризація, ДЗЗ, нейронні мережі.

Вступ. Отримання актуальної та достовірної геопросторової інформації про особливості території є запорукою проведення успішного планування територіального розвитку на переходу до сталих форм господарювання. В цьому контексті забезпечення

органів державного управління та планування якісними геопросторовими даними є одним із найважливіших завдань. Джерелами таких даних у більшості випадків виступають дані дистанційного зондування Землі (ДЗЗ), які після певної технічної підготовки можуть використовуватися для вирішення прикладних завдань моніторингу та аналізу території. В останні роки все більшого значення набувають дані космічної зйомки, а також широко використовуються результати лідарної зйомки та БПЛА різного класу для отримання знімків у режимі близькому до реального часу.

Постановка проблеми. Одним із базових продуктів, який може бути отримано на основі обробки даних дистанційного зондування Землі (ДЗЗ) є карта земного покриття або землекористування. Подібні дані становлять основу для організації та управління територією. Особливо важливо мати динамічний ряд для відслідковування тенденцій в зміні видів землекористування. Враховуючи важливість цих даних останні 30 років у всьому світі реалізуються міжнародні та регіональні програми та ініціативи, що направлені на дослідження земного покриття та створення карт землекористування. Найбільш відомими та в певному змісті піонерськими стало отримання світового покриття GLC2000 та успішна реалізація європейської програми CORINE, яка заклала єдині стандарти відображення типів землекористування в Європейському Союзі [3,4,5,10].

Метою дослідження є аналіз земного покриття території України для цілей державного управління та адміністрування з використанням методу карт самоорганізації Кахонена. **Завданнями** роботи є вивчення особливостей продукту Globeland 30, методів отримання даних земного покриття, розробка схеми та методики обробки даних для задач дослідження території України, отримання та інтелектуальний аналіз зональних даних для адміністративних одиниць України з використанням карт самоорганізації Кахонена.

Викладення матеріалу дослідження. З метою підтримки досліджень глобальних змін і розвитку системних моделей функціонування Землі, Міністерство науки і технологій Китаю запустило Національну високотехнологічну програму досліджень та розробок під назвою «Картографування за допомогою даних ДЗЗ і дослідження ключові технології побудови глобального

рослинного покриву» у 2010 році.

1. Програма GlobeLand30-2010. Одним з результатів проекту є картографічний продукт GlobeLand30-2010 глобального покриву з 30-метровий просторовим розрізненням, отриманий із зображень дистанційного зондування в 2010 році [9].

Набір даних охоплює територію площею від 80° Пн. ш. до 80° Пд. ш. та складається з 10 типів земного покриву. Для генерації даних земного покриття використовувалися класифіковані зображення із супутникових даних мультиспектральних зображень GlobeLand30-2010 з роздільною здатністю 30 м, в тому числі Landsat TM і ETM+ і мультиспектральні зображення китайської системи Environmental Disaster Alleviation Satellite (HJ-1). Були обрані безхмарні зображення, отримані протягом вегетаційного сезону в межах ± 1 рік відносно 2010 року.

HJ-1 китайський супутник дистанційного зондування, спрямований на моніторинг навколишнього середовища та запобігання і зменшення стихійних лих. HJ-1A супутник несе мультиспектральну камеру (CCD) і гіперспектральний сенсор (HSI).

2. Технічні характеристики продукту земного покриття. Продукт GlobeLand30-2010 було отримано з використанням ієрархічного методу вилучення. Кожен тип земного покриву був класифікований по одному, з обмеженнями маскою інших типів покриву. Робочий процес полягає в тому щоб отримати тільки один тип земельного покриву за один раз, а потім маскувати клас після визначення.

Відповідно до карти тайлів та карт кожного континенту було обрано 8 тайлів для оцінки точності. Загалом, 9 типів земного покриву та 150000 тестових прикладів було оцінено з точки зору точності. Загальна точність GlobeLand30-2010 досягає 83.51%. Індикатор Каппа дорівнює 0,78.

Тайлові дані GlobeLand30-2010 складаються з п'яти частин, а саме файлу результату класифікації, файлу інформацію про координати, індексного картографічного файлу класифікації зображень, файлу метаданих та ілюстративного файлу. Серед них:

- Файл класифікації результату: відповідає файлу зберігання інформації про класифікацію земного покриву.

- Файл інформації про координати відповідає файлу запису координатної інформації результатів класифікації даних.

- Індексний картографічний файл класифікації відповідає файлу векторного шару запису діапазону і часу кожної сцени класифікаційних зображень.

- Файл метаданих відповідає файлу запису інформації про метадані результатів класифікації;

- Ілюстративний файл відповідає файлу, що описує результат класифікації даних і пояснення авторських прав на дані.

Система класифікації включає в себе десять видів земного покриву, в тому числі 7 з них виявлено в Україні, а саме сільськогосподарські землі, ліси, луки, чагарник, водно-болотних угіддя, водойми, урбанізовані території.

3. Аналіз зональних даних. Підготовка даних для території України складалась з наступних етапів:

- Завантаження тайлів, що покривають територію України
- Обрізка країв тайлів за допомогою векторних масок
- Злиття мозаїки
- Обрізка мозаїки по контуру державного кордону України
- Експорт даних у формат GeoTIFF.

На наступному етапі необхідно сформувати геостатистичні дані для адміністративних одиниць України на різних територіальних рівнях. Перевагою такого підходу є можливість формування статистичних показників для різних територіальних одиниць або їх поєднань: річкові басейни, природні зони та ін.

Формування геостатистичних даних відбувається з використанням операцій зональної статистики, які реалізовано в модулі Spatial Analyst ПЗ ArcGis. Для обробки даних якісного розподілу класифікаційних ознак підходить інструмент «Зональна статистика в таблицю», який дозволяє узагальнити значення растру у межах зон другого набору даних.

Зональні інструменти дозволяють виконувати аналіз, вихідні дані якого є результатом обчислень, виконаних на всіх комірках, що належать кожній вхідній зоні. Зона може бути визначена як область однакових значення, але вона також може складатися з декількох відокремлених елементів, або регіонів, всі з яких мають одне значення. Зони можуть задаватися растром або наборами класів векторних об'єктів. Растр повинен мати тип ціле число, а просторові об'єкти повинні мати цілочисельне або строкове поле атрибутів.

Деякі зональні інструменти оцінюють кількість певних

Класифікація GlobeLand30

| Код | Тип | Зміст |
|-----------|---------------------------------|---|
| 10 | Сільськогосподарські землі | Землі, які використовуються для сільського господарства, багаторічних насаджень, в тому числі рисових полів, зрошуваних сільгоспугідь, овочівництва, фруктових садів. |
| 20 | Ліс | Землі, які вкриті деревнистою рослинністю з покривом більше 30%, в тому числі листяні і хвойні ліси, і рідколіся з покриттям 10 - 30% і т.д. |
| 30 | Луки | Землі, які покриті натуральною трав'янистою рослинністю з покриттям більше 10%, і т.д. |
| 40 | Деревно-чагарникова рослинність | Землі, які вкриті чагарниками з покриттям більше ніж 30%, в тому числі листяні і вічнозелені чагарники і пустельний степ з покриттям більше 10%, і т.д. |
| 50 | водно-болотні угіддя | Землі, які вкриті водно-болотною рослинністю і водоймами, в тому числі внутрішні болота, заболочені озера, заплави річок з водно-болотними угіддями, ліси / чагарники водно-болотних угідь, торф'яних болота і т.д. |
| 60 | Водні об'єкти | Водні об'єкти на суші, в тому числі річки, озера, водосховища, ставки для розведення риби і т.д. |
| 80 | Урбанізовані території | Землі модифіковані діяльністю людини, в тому числі всі види житла, промислові і гірничодобувні галузі, транспортні комунікації, внутрішні міські зелені зони і водойми і т.д. |

властивостей геометрії, або форми, вхідних даних зони, і їм не потрібні інші вхідні дані. Інші зональні інструменти використовують вхідні дані зони для визначення місць розташування, для яких будуть обчислюватися інші параметри, наприклад, статистика, площі або частота значень. Існує також зональний інструмент, який використовується для заповнення зазначених зон мінімальними значеннями, що знаходяться на межі зони.

4. Використання карт самоорганізації Кохонена для кластеризації даних. Математичною основою моделі кластеризації адміністративно-територіальних одиниць України є штучна нейронна мережа на основі карт самоорганізації Кохонена. Карта самоорганізації, зазвичай, представляється як двовимірна матриця масивів вузлів нейронів. У нашому випадку даний масив формує гексагональну решітку, що є ефективним рішенням для візуального аналізу даних.

Опишемо процедуру побудови даної нейронної мережі [6]. Вхідний простір даних $X \subset \mathfrak{R}^n$ проектується, найчастіше, на двовимірну решітку вузлів нейронів. Кожному вузлу

$i, i = \overline{1, k}$, ставиться у відповідність параметричний вектор моделі

$m_i = \{m_{1i}, m_{2i}, \dots, m_{ni}\} \in \mathfrak{R}^n$, де k – кількість вузлів нейронів.

У той же час вхідний вектор $x = \{x_1, x_2, \dots, x_n\} \in X \subset \mathfrak{R}^n$

пов'язаний одночасно з усіма нейронами через скалярні ваги m_{ji} .

Перед початком проведення навчання нейронної мережі всі ваги m_{ji} ініціалізуються за допомогою генератора випадкових величин.

Після ініціалізації мережі запускається так званий процес конкуренції, який полягає у визначенні найбільш подібних нейронів до векторів вхідних даних. Найбільш поширеним підходом до даного процесу є обчислення Евклідової відстані:

$$\|x - m_i\| = \sqrt{\sum_{j=1}^n (x_j - m_{ji})^2}, \quad i = \overline{1, k}, \quad j = \overline{1, n}. \quad (1)$$

Після подання вхідних векторів на входи нейронів,

відбувається змагання нейронів шару Кохонена за правилом «переможець отримує все», згідно якого обирається нейрон-переможець, що є найближчим до вхідного вектора:

$$c = \arg \min_{i=1,k} \{ \|x - m_i\| \}, \quad (2)$$

Після визначення нейрона-переможця здійснюється визначення околу його впливу на сусідні нейрони. Нейрон-переможець знаходиться в центрі топологічного околу. При збудженні, він впливає на просторово близькі до нього нейрони, проте даний вплив зменшується із збільшенням відстані між нейронами. Даний процес призводить до визначення топологічного сусідства нейронів.

На заключному етапі побудови карти самоорганізації відбувається процес синаптичної адаптації. Даний процес полягає у процедурі корегування вектора синаптичних ваг нейрона-переможця та його сусідів відповідно до вектора вхідного показника:

$$m_i(t+1) = m_i(t) + v(t)h_{ci}(t)[x(t) - m_i(t)], \quad (3)$$

де $v(t)$ – параметр швидкості навчання; $h_{ci}(t)$ – функція

топологічного сусідства з центром у нейроні-переможці c . Результат налаштування такого типу нейронних мереж можна представити у вигляді карти Кохонена (рис. 1).

Оскільки карти самоорганізації здійснюють стиснення інформації із збереженням найбільш важливих топологічних та/або метричних зв'язків між вхідними елементами даних, можна також вважати, що з їх допомогою створюються абстракції (узагальнення) певного виду. Ці дві властивості карт самоорганізації, візуалізацію та узагальнення, можна використовувати у вирішенні таких складних задач, як аналіз, управління та передача інформації.

В якості вхідних даних для побудови карт використовували результати зональної статистики розподілу видів земного покриву для кожного адміністративного району за всіма видами покривів. Результатом класифікації стало розбиття всієї множини районів на 8 кластерів.

Кожен кластер характеризує особливості землекористування

на рівні адміністративних районів, що може бути використано при створенні програм місцевого розвитку громад, врахуванні політики сталого розвитку територій.

Висновки. В результаті можна зазначити, що наявність глобального продукту аналізу земного покриття Globeland 30 надає можливість для здійснення ефективної політики в галузі регіонального управління та територіального планування з урахуванням реально існуючого розподілу типів землекористування.

Використання методів зонування та зональної статистики, кластеризації в багатовимірному просторі ознак на основі карт самоорганізації Кахонена дозволяє використовувати дані про розподіл типів земного покриття при регіональних оцінках різних територіальних рівнів та формуванні політики раціонального землекористування.

Рецензент – кандидат географічних наук, доцент С. К. Дрич

Література:

1. Горбаченко В. И. Сети и карты Кохонена [Електронний ресурс] Режим доступу : <http://gorbachenko.self-organization.ru/index.html>.

2. Питенко А. А. Нейросетевой анализ в геоинформационных системах. Дис. ... канд. техн. наук: 05.13.16/ А.А. Питенко. – Красноярск, 2000. – 97 с.

3. Bossard M., Feranec J., Otahel J. The revised and supplemented Corine Land Cover nomenclature // Techn. Rept EEA. — 2000. — N38. — 110 p.

4. Brian O'Connor, Cristina Secades, Johannes Penner, Ruth Sonnenschein, Andrew Skidmore, Neil D. Burgess & Jon M. Hutton Earth observation as a tool for tracking progress towards the Aichi Biodiversity Targets // Remote Sensing in Ecology and Conservation/ – у John Wiley & Sons Ltd, 2015/ – 19 – 27

5. Buttner G., Feranec J., Jaffrain G., et al. The CORINE Land Cover 2000 Project // EARSeL Proceedings 3(3). — 2004. — P. 331–346.

6. National Geomatics Center of China, 2014 National Geomatics Center of China 30 M Global Land Cover Data Product (GlobeLand30) (2014).

В. В. Путренко

КЛАСТЕРИЗАЦИЯ ГЕОПРОСТРАНСТВЕННЫХ ДАНЫХ ЗЕМНОГО ПОКРОВА УКРАИНЫ

В работе определена методика обработки данных земного покрова на примере территории Украины с использованием карт самоорганизации Кахонена. Осуществлен анализ пригодности для использования в качестве источника данных продукта GlobeLand30, разработана методика вторичной обработки данных, проведено зонирования и получения показателей зональной статистики, использование карт самоорганизации Кахонена для кластеризации административных единиц по соотношению видов землепользования.

Ключевые слова: земной покров, кластеризация, ДЗЗ, нейронные сети.

V. Putrenko

SPATIAL DATA CLUSTERING OF UKRAINE LAND COVER

The method detecting of land cover data on the example of Ukraine with the use of self-organizing maps was defined in the paper. The analysis of the suitability for use as a source GlobeLand30 product data was done. The technique of secondary data, carried out, the zoning and obtaining indicators zonal statistics, the use of self-organizing maps for clustering administrative units on the base of the ratio of land uses.

Keywords: land cover, clustering, remote sensing, neural networks.

Надійшла до редакції 26 березня 2016 р.