

2. Flowwright. (2023). The human element in workflow process automation: avoiding neglect. Available at: <https://www.flowwright.com/the-human-element-in-workflow-process-automation-avoiding-neglect>, p. 4–6.
3. Ciklum. (2025). AI agents explained: the future of task automation and productivity. Available at: <https://www.ciklum.com/blog/ai-agents-explained-the-future-of-task-automation-and-productivity/>, p. 6–9.
4. Alumio. (2025). How AI is transforming integration platforms in 2025. Available at: <https://www.alumio.com/blog/how-ai-is-transforming-integration-platforms-in-2025>, p. 3–8.

DOI 10.70286/ISU-11.02.2026.013

АНАЛІЗ МЕТОДІВ ФАКТОРИЗАЦІЇ ЕКОЛОГІЧНИХ ДАНИХ У ГЕОІНФОРМАЦІЙНИХ СИСТЕМАХ

Власов Ростислав Ігорович
аспірант

Волівач Антоніна Петрівна
кандидат технічних наук, доцент

Київський національний університет технологій та дизайну, Україна

Стрімке зростання обсягів багатовимірних даних, що надходять із супутникових платформ, наземних сенсорних мереж та автоматизованих станцій спостереження, зумовлене розвитком систем екологічного моніторингу. Такі дані характеризуються високою гетерогенністю, просторово-часовою залежністю та значною надлишковістю. Їх ефективна обробка у геоінформаційних системах (ГІС) потребує сучасних методів зменшення розмірності, виявлення прихованих закономірностей та підвищення точності класифікації та прогнозування.

В джерелі [1] автори зазначають, що одним із найбільш поширених підходів є застосування алгебраїчної факторизації. Такий підхід дозволяє представляти складні екологічні дані у вигляді компактних факторних моделей. Водночас класичні методи не завжди забезпечують достатню точність і стійкість в умовах шумів та пропусків даних, що обумовлює актуальність подальших досліджень у цьому напрямі.

При цьому методи матричної факторизації, зокрема сингулярне розкладання (SVD) та аналіз головних компонентів (PCA), належать до базових інструментів обробки екологічних табличних даних. Вони ґрунтуються на ортогональному перетворенні вхідної матриці та мінімізації середньоквадратичної похибки апроксимації [1]. Основними перевагами цих методів є обчислювальна ефективність, наявність аналітичного розв'язку та стійкість під час обробки великих обсягів даних.

Разом із тим, PCA та SVD мають низку суттєвих недоліків при застосуванні до екологічних даних. По-перше, отримані компоненти є лінійними комбінаціями початкових ознак, що ускладнює їх фізичну інтерпретацію. По-друге, ці методи чутливі до викидів і шумів, які характерні для даних екологічного моніторингу. По-третє, класичні матричні моделі не враховують багатовимірну просторово-часову структуру даних, що обмежує їх застосування в геоінформаційних системах. Середня точність класифікації на основі PCA становить 70–82%, проте при збільшенні рівня шуму вона знижується до 60–65% [1].

Альтернативою є ненегативна матрична факторизація (NMF), яка забезпечує кращу інтерпретованість факторів та підвищує точність класифікації на 5–10% порівняно з PCA. Негативна матрична факторизація представляє матрицю даних у вигляді добутку двох ненегативних матриць, що забезпечує адитивну інтерпретацію факторів [2, 3]. Це є важливою перевагою для екологічних задач, оскільки значення показників (концентрації, індекси забруднення тощо) фізично не можуть приймати від'ємні значення.

До переваг NMF також належать можливість виявлення прихованих тематичних або просторових патернів та підвищення точності класифікації у порівнянні з PCA. Водночас метод характеризується відсутністю єдиного глобального розв'язку, залежністю від початкових умов та зниженою стійкістю до шуму й пропусків даних. Крім того, NMF залишається чутливою до аномальних значень та неповних даних [5].

Подальший розвиток отримали робастні та регуляризовані методи факторизації, зокрема Robust PCA та Sparse NMF, які демонструють підвищену стійкість до шуму та пропусків даних. Robust PCA базується на представленні матриці даних, як суми низькорангової та розрідженої складових [4]. Такий підхід дозволяє ефективно відокремлювати структурну інформацію від шумових та аномальних компонент.

Перевагою робастних моделей є підвищена стійкість до зашумлених вимірювань і пропусків даних. Серед недоліків варто виділити підвищену обчислювальну складність та необхідність підбору параметрів регуляризації, що ускладнює практичну реалізацію у великомасштабних ГІС. У ряді досліджень показано, що використання таких моделей дозволяє зменшити втрати точності більш ніж удвічі порівняно з класичними підходами [4].

Одним із методів, що дозволяє без втрати структурної інформації аналізувати просторово-часові екологічні дані, є тензорна факторизація (CP, Tucker). Тензорні моделі факторизації є природним узагальненням матричних методів для багатовимірних даних і дозволяють зберігати просторову, часову та тематичну структуру екологічних даних, що є критично важливим для геоінформаційних систем [6].

Методи CP- та Tucker-факторизації демонструють високу точність класифікації та стійкість до шуму. Основними недоліками є ускладнена інтерпретація факторів при зростанні порядку тензора та значні обчислювальні витрати для великих масивів даних. Водночас тензорні моделі забезпечують

найвищу точність класифікації – 90–95% і дозволяють оптимізувати обчислювальні ресурси при обробці багатовимірних даних [6].

За результатами проведеного аналізу було сформовано узагальнену порівняльну характеристику методів факторизації [1-6], що застосовуються в екологічних геоінформаційних системах (таблиця 1).

Таблиця 1. Порівняльна характеристика методів факторизації екологічних даних

Метод	Тип даних	Точність, %	Стійкість до шуму	Застосування в ГІС
PCA	Табличні	70–82	Низька	Обмежене
NMF	Табличні	78–88	Середня	Середнє
Robust PCA	Зашумлені	83–90	Висока	Високе
Tensor (CP/Tucker)	Просторово-часові	85–95	Дуже висока	Дуже високе

Отже, у ході дослідження встановлено, що найбільш перспективним напрямом розвитку екологічних геоінформаційних систем є застосування удосконалених алгебраїчних методів факторизації, зокрема регуляризованих матричних і тензорних моделей. Їх використання дозволяє підвищити точність класифікації, зменшити вплив шуму та забезпечити ефективну обробку багатовимірних екологічних даних. При цьому, вибір методу алгебраїчної факторизації для екологічних геоінформаційних систем повинен враховувати компроміс між інтерпретованістю, стійкістю до шуму та обчислювальною складністю. Класичні матричні методи доцільно застосовувати для попереднього аналізу, тоді як регуляризовані та тензорні моделі є більш перспективними для побудови високоточних класифікаційних систем.

Список використаних джерел

1. Golub G.H., Van Loan C.F. Matrix Computations. – Baltimore: Johns Hopkins University Press, 2013, 784 p.
2. Gan J., Liu T., Li L., Zhang J. Non-negative Matrix Factorization: A Survey // The Computer Journal. — 2021. — Vol. 64, Issue 7. — P. 1080–1092.
3. Feng J., Duan T., Zhou Y., Chang X., Li Y. An improved nonnegative matrix factorization with the imputation method model for pollution source apportionment during rainstorm events // Journal of Environmental Management. — 2023. — Vol. 328:116888.
4. Candès E.J., Li X., Ma Y., Wright J. Robust principal component analysis? Journal of the ACM, 2011, Vol. 58(3).
5. Cichocki A., Zdunek R., Phan A.H., Amari S. Nonnegative Matrix and Tensor Factorizations. – Wiley, 2009, 553 p.
6. Kolda T.G., Bader B.W. Tensor decompositions and applications. SIAM Review, 2009, Vol. 51(3), pp. 455-500.