

ГЛИБОКІ НЕЙРОННІ МЕРЕЖІ ДЛЯ КЛАСТЕРИЗАЦІЇ МАСИВІВ ДАНИХ

Танянський О.С.

e-mail: oleksii.tanianskyi@nure.ua

Науковий керівник – д.т.н., доц. Шафроненко А.Ю.

Харківський національний університет радіоелектроніки, каф. ІНФ
м. Харків, Україна

The rapid advancement of computing technologies has led to an exponential increase in data volumes, creating an urgent need for innovative approaches to data analysis and clustering. Deep neural networks (DNNs) have emerged as powerful tools capable of automatically extracting salient features from raw data, thereby offering promising new directions in clustering applications.

This paper reviews the fundamental principles of applying deep neural networks to data clustering, discusses their advantages and limitations compared to traditional methods, and examines practical considerations in their implementation across various domains.

У сучасному світі, де дані є ключовим ресурсом, щоденне створення величезних і надзвичайно складних обсягів інформації створює серйозні виклики для класичних методів кластеризації. Традиційні алгоритми, такі як К-середні, ієрархічна кластеризація та методи на основі щільності, часто мають проблеми з багатовимірними даними, складними геометричними структурами та нестаціонарними середовищами. Ці методи зазвичай спираються на заздалегідь визначені параметри, такі як кількість кластерів або припущення щодо форм кластерів, які можуть не відповідати багатьом реальним програмам.

Глибокі нейронні мережі (DNN) з їхньою здатністю до нелінійної апроксимації та ієрархічного вивчення ознак пропонують переконливу альтернативу. Автоматично вивчаючи багатовимірні дані, DNN можуть розкривати внутрішні структури, які важко охопити за допомогою класичних методів кластеризації, що не тільки покращує продуктивність кластеризації, але й зменшує потребу в розширеному проектуванні функцій, роблячи глибоке навчання привабливим підходом для задач Data Mining.

Основою глибинної кластеризації є одночасне навчання представлення даних і їх розподілу на кластери в межах однієї моделі. Один із поширених підходів полягає у використанні автокодерів - нейронних мереж, що стискають дані у компактний латентний простір, а потім відновлюють вихідну інформацію. Приховані представлення, отримані через автокодування, зазвичай компактніші та допомагають виявити основні шаблони, що спрощують кластеризацію. Методи, як Deep Embedded Clustering (DEC) і Deep Clustering Network (DCN), поєднують навчання представлення та кластери-

зацію в одному процесі, що робить отриманий простір ознак більш відповідним для задач кластеризації. За допомогою цих методів мережа навчається не тільки для мінімізації помилок реконструкції, але й для посилення розділення між кластерами. Функції втрат зазвичай доповнюють спеціальними умовами, які накладають штрафи на великі відстані між об'єктами в одному кластері, водночас стимулюючи чіткий поділ між різними кластерами. Результатом є прихований простір, який краще відображає природне групування даних, забезпечуючи більш точні та надійні результати кластеризації.

Застосування глибоких нейронних мереж дозволяє автоматично здійснювати витяг ознак із сирих даних, що значно знижує необхідність попередньої обробки вхідних даних. Крім того, такі мережі можуть ефективно працювати з даними високої розмірності та невизначеною структурою, що є типовим для сучасних великих даних. Завдяки нелінійній природі трансформацій, що здійснюються в глибоких шарах, можливе виявлення складних, прихованих зв'язків, які важко виявити за допомогою класичних методів.

Глибокі нейронні мережі пропонують кілька ключових переваг перед традиційними алгоритмами кластеризації. По-перше, вони за своєю суттю здатні обробляти багатовимірні та неструктуровані дані, що часто зустрічаються в таких областях, як обробка зображень, обробка природної мови, тощо. Ієрархічна природа DNN дозволяє поступово виділяти абстрактні ознаки, ефективно зменшуючи розмірність і висвітлюючи найбільш дискримінаційні атрибути. По-друге, здатність DNN вивчати складні нелінійні зв'язки робить їх найбільш придатними для аналізу даних зі складною геометрією кластерів, що перетинаються.

Таким чином, глибокі нейронні мережі пропонують потужний інструмент для кластеризації даних завдяки своїй здатності автоматично отримувати релевантні функції та вивчати складні нелінійні зв'язки, що робить їх багатообіцяючим і цінним інструментом для вирішення задач інтелектуального аналізу даних.

Список використаних джерел:

1. Lei, T., Liu, P., Jia, X., Zhang, X., Meng, H., & Nandi, A. K. Automatic fuzzy clustering framework for image segmentation // IEEE Transactions on Fuzzy Systems. 2019. 28(9). P. 2078-2092.
2. Feng, Q., Chen, L., Chen, C. P., & Guo, L. Deep fuzzy clustering—a representation learning approach // IEEE Transactions on Fuzzy Systems. 2020. 28(7). P. 1420-1433.
3. Zhao, X., & Jia, M. A novel deep fuzzy clustering neural network model and its application in rolling bearing fault recognition / Measurement Science and Technology. 2018. 29(12). 125005p.