

АНАЛІЗ ВПЛИВУ ПАРАМЕТРІВ КЛАСТЕРИЗАЦІЇ НА РЕЗУЛЬТАТИ ОБЧИСЛЕНЬ

Шевченко С.В., Рудковський К.Д.

*Національний технічний університет
«Харківський політехнічний інститут», Харків*

В роботі розглянуті питання впливу параметрів процесу кластеризації на її результати. Алгоритми кластеризації зазвичай мають низку параметрів, що суттєво впливають на їх продуктивність та результати. Вибір цих параметрів може суттєво вплинути на точність та інтерпретацію результатів кластеризації. В роботі розглядається вплив параметрів кластеризації на результати обчислень та аналізується їх налаштування для отримання значущих результатів.

Основною метою кластеризації є розбиття набору даних на групи або кластери, де дані всередині кожного кластера схожі між собою і відрізняються від даних в інших кластерах. Якість результатів кластеризації залежить від різних факторів, у тому числі від вибору параметрів кластеризації. Ці параметри впливають на кількість і форму кластерів, швидкість збіжності результатів застосування алгоритму та їх стабільність.

В роботі наведені дані досліджень ключових параметрів, що суттєво впливають на результати кластеризації.

1. Вибір кількості кластерів k має вирішальне значення в алгоритмах кластеризації. Вище значення k може призвести до отримання великої кількості малих кластерів, тоді як нижче значення k дає більш широкі кластери. Придатне значення k залежить від конкретного набору даних і задачі, що розв'язується.

2. Міра відстані або подібності, що використовується в алгоритмах кластеризації, визначає схожість або відмінність між екземплярами даних. Різні міри відстані або подібності можуть призвести до різних результатів кластеризації. Вибір відстані або міри подібності повинен відповідати характеристикам даних.

3. Метод ініціалізації в алгоритмах кластеризації визначає початкове розміщення центрів кластерів. Різні методи ініціалізації можуть призвести до різних результатів кластеризації. Вибір методу ініціалізації впливає на швидкість збіжності та стабільність алгоритму.

4. Критерії збіжності в алгоритмах кластеризації суттєво впливають на кінцеві результати кластеризації. Встановлення занадто низьких або занадто високих критеріїв збіжності може призвести до недостатньої або надмірної відповідності даних. Вибір критеріїв збіжності та умов зупинки має вирішальне значення для отримання значущих результатів кластеризації.

Для забезпечення надійних та інтерпретованих результатів кластеризації повинні проводитися ретельні дослідження впливу параметрів кластеризації таких, як кількість кластерів, міра відстані або подібності, метод ініціалізації та критерії збіжності.