

## ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ КЛАСТЕРІЗАЦІЇ

Бурдега О.С.

Науковий керівник – к.т.н., доц. Колесницький О.К.

Задача кластеризації є однією з задач інтелектуального аналізу даних.

Кластеризація належить до проблеми сегментації, за цим підходом розподіляють записи до різних груп або сегментів. Багато методів кластеризації відрізняються між собою тим, що їх алгоритми на кожному кроці обчислюють різноманітні функціонали якості розбиття. Такі екстремальні завдання дозволяють визначити той кількісний критерій, завдяки якому можна було б віддати перевагу одне розбиття іншому.

У кожного з методів є як переваги, так і недоліки. Проводячи аналіз звертали увагу на такі характеристики як: робастність, інтерпретація, надійність, швидкість.

Під робастністю в статистиці розуміють нечутливість до різних відхилень і неоднородностям у вибірці, зв'язаним з тими чи іншими, у загальному випадку невідомими, причинами.

Інтерпретація забезпечує можливість розуміння моделі аналітиком.

Надійність методів кластеризації дає можливість роботи цих методів за наявності в наборі шумів і викидів.

Швидкість характеризує час, що потрібно на створення моделі і її використання.

Аналіз методів показав, що найперспективнішим є метод мережі Кохонена побудованої на основі нейронних мереж.

Вона може розпізнавати кластери в даних, а також встановлювати близькість кластерів. Таким чином, користувач може поліпшити своє розуміння структури даних. Мережі Кохонена можуть використовуватися і в тих задачах кластеризації, де кластери вже задані, - тоді перевага буде в тому, що мережа зможе виявити подібність між різними кластерами.

В алгоритмі функціонування мережі Кохонена використовується коефіцієнт швидкості навчання, яке поступово зменшується, для тонкої корекції на новій епосі. В результаті центр встановлюється у певній позиції, яка задовільним чином кластеризують приклади, для яких даний нейрон є переможцем. Мережа Кохонена навчається методом послідовних наближень. Вона здатна функціонувати в умовах перешкод, так як число кластерів фіксовано, ваги модифікуються повільно, настроювання ваг закінчується після навчання.

Після того, як мережа навчена розпізнаванню структури даних, її можна використовувати як засіб візуалізації при аналізі даних.