

УДК 004.4

К.Добруцький, О.Назаревич, канд. техн. наук, ст. викл.

Тернопільський національний технічний університет імені Івана Пулюя

**ОГЛЯД ТА ПОРІВНЯЛЬНИЙ АНАЛІЗ АЛГОРИТМІВ ТА МЕТОДІВ
КЛАСТЕРИЗАЦІЇ ТА РЕГРЕСІЇ**

К.Dobruts'kyu, O.Nazarevych

**REVIEW AND COMPARATIVE ANALYSIS ALGORITHMS AND METHODS OF
CLUSTERING AND REGRESSION**

Кластеризація – це автоматичне розбиття елементів деякої множини (об'єкти, дані, вектора характеристик) на групи (кластери) за принципом схожості [1]. Кластерний аналіз – це сукупність методів, що дозволяють класифікувати баговимірні спостереження, кожне з яких описується набором вихідних змінних X_1, X_2, \dots, X_n [2]. Метою кластерного аналізу є утворення груп схожих між собою об'єктів, які прийнято називати кластерами.

Методи кластерного аналізу дозволяють вирішувати наступні завдання:

- проведення класифікації об'єктів з урахуванням ознак, що віддзеркалюють сутність, природу об'єктів;
- побудова нових класифікацій для маловивчених явищ, коли необхідно встановити наявність зв'язків усередині сукупності та спробувати привнести до неї структуру.

В задачі кластерного аналізу формою подання вихідних даних служить прямокутна таблиця, кожний рядок якої представляє результат виміру k -ознак, що розглядаються, по одному з об'єктів.

Кластеризація включає в себе наступні етапи (рисунк 1):

- виділення характеристик;
- визначення метрики;
- розбиття об'єктів на групи;
- представлення результатів.

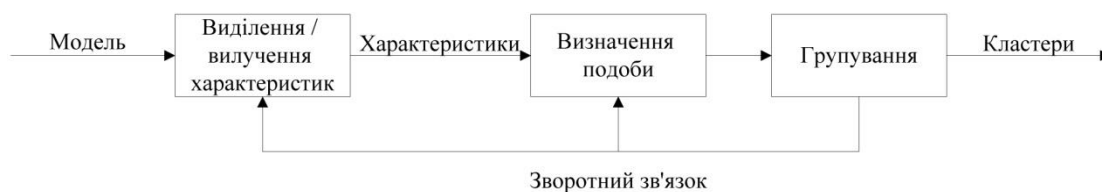


Рисунок 1 – Загальна схема кластеризації

Сутність ієрархічних агломеративних методів у тому, що на першому кроці кожний об'єкт вибірки розглядається як окремий кластер. Процес об'єднання кластерів відбувається послідовно: на підставі матриці відстаней або матриці подібності поєднуються найбільш близькі об'єкти. Послідовність об'єднання легко піддається геометричній інтерпретації й може бути представлена у вигляді графа-дерева (дендрограми) [2].

Література.

1. Meesad P. A hybrid intelligent system for medical diagnosis / P. Meesad, G. Yen // Proc. of IJCNN'01. Washington. 2001. P. 2558 – 2563.

2. Кластерный анализ [Электронный ресурс]. – Режим доступа к ресурсу: <http://www.statsoft.ru/home/textbook/modules/stcluan.html>.