

ПОБУДОВА КЛАСТЕРІВ З ВИКОРИСТАННЯМ ГЕНЕТИЧНОГО АЛГОРИТМУ

Манаєва О.О.

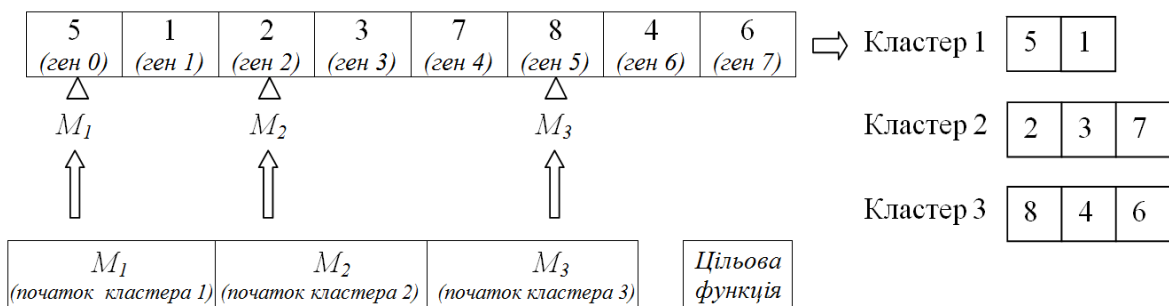
Науковий керівник – к. т. н., проф. Кондратенко Н.Р.

Задача кластеризації на сьогодні є перспективним напрямком досліджень. Відомо, що для її розв'язання не існує універсальних методів, що дозволяють швидко знайти абсолютно точні рішення. Процедура знаходження оптимального розв'язку методами повного перебору практично нездійсненна за винятком тих випадків, коли n (число об'єктів) та m (число кластерів) невелике. Тому для розв'язання цієї задачі доцільно використати наближені методи, зокрема генетичні алгоритми.

Задача кластеризації полягає у виділенні серед множини $I = \{I_1, I_2, \dots, I_n\}$ деяких об'єктів m однорідних кластерів, таких, щоб об'єкти всередині кластерів були в певному сенсі подібними, а об'єкти з різних кластерів – відмінними.

Запропонуємо генетичний алгоритм розв'язання задачі в даній постановці. Провівши аналогію між оптимізаційним процесом та процесом еволюції в природі, будемо розглядати кожен можливий розв'язок поставленої задачі як особину.

Для формалізованого опису особини будемо застосовувати структуру даних – хромосому, - яка має певні особливості, пов'язані зі специфікою розв'язуваної задачі. Запропонована хромосома є неоднорідною, тобто містить дві складові, які попри зовнішню подібність мають принципово різний зміст. Перша складова являє собою лінійний масив розмірності n , що містить гени - номери об'єктів - у певній послідовності:



Для виділення в межах хромосоми окремих кластерів введемо до складу нашої хромосоми другу складову – набір маркерів розбиття M_1, \dots, M_m . Значення маркеру $M_i = k$ сигналізує про те, що з k -того гену основної складової починається новий кластер. Окрім того, кожній хромосомі з простору можливих розв'язків ставиться у відповідність значення цільової функції.

Введемо початкову популяцію – набір особин-розв'язків, - над якою будемо послідовно проводити операції схрещування, мутації та відбору кращих за критерієм цільової функції особин.

Багатократним повторенням описаної процедури досягається мінімізація цільової функції та прийнятна точність отриманого розв'язку.