АНАЛІЗ АЛГОРИТМІВ ПОШУКУ В СТРУКТУРІ ДАНИХ ДЛЯ СТВОРЕННЯ ЗРУЧНОГО МЕХАНІЗМУ ОПИТУВАННЯ КОРИСТУВАЧІВ МЕРЕЖ

Абраменко Р.О.
Науковий керівник – доц. Ревенчук І.А.
Харківський національний університет радіоелектроніки
(61166, Харків, просп. Науки,14, каф. Радіотехніки, тел. (057) 702-00-00)

Undoubtedly, classification is the basic process in human intellectual activity. When we encounter a new phenomenon, we try to find an analogue in the field known to us. Considering a group of any objects, we involuntarily divide them into subgroups of closely related elements. Classification is present in the ordering of facts, phenomena and objects known to us. Based on the above, we can conclude that classification is a fundamental concept of science and practice.

Безсумнівно, класифікація – це основний процес в інтелектуальній діяльності людини. Зустрічаючись з новим явищем, ми намагаємося знайти йому аналог у відомій нам області. Розглядаючи групу яких-небудь об'єктів, ми мимоволі розділяємо їх на підгрупи близьких один одному елементів. Класифікація присутня при упорядкуванні відомих нам фактів, явищ, предметів. На підставі сказаного можна зробити висновок, що класифікація - це фундаментальне поняття науки і практики.

Кластерний аналіз – це сукупність методів, які дозволяють класифікувати багатомірні спостереження, кожне з яких описується набором вихідних перемінних $X_1, X_2 \dots X_m$. Метою кластерного аналізу є утворення груп, схожих між собою об'єктів, що прийнято називати кластерами.

Методи кластерного аналізу.

Існує ряд різних методів, які можна використовувати для проведення кластерного аналізу. Ці методи можна класифікувати наступним чином:

1. ієрархічні (агломеративні, дивізійні);
2. неієрархічні (часто відомі як k-методи кластеризації).

Ієрархічні методи.

В рамках цього підходу до кластерного аналізу існує ряд різних методів, що використовуються для визначення, які кластери повинні бути об'єднані на кожному етапі. Основні методи узагальнені нижче:

- метод найближчого сусіда (метод одного зв’язку). У цьому способі відстань між двома кластерами визначається як відстань між двома найближчими членами або сусідами;
- найдальший сусідський метод (метод повного зв'язку). У цьому випадку відстань між двома кластерами визначається як максимальна відстань між членами - тобто відстань між двома предметами, які віддалені один від одного;

- середній (між групами) метод зв’язку. Відстань між двома кластерами обчислюється як середня відстань між усіма парами об'єктів у двох кластерах;
- центроїдний метод. Тут обчислюється центроїд (середнє значення для кожної змінної) кожного кластеру та використовується відстань між центроїдами. Кластери, центроїди, які найближче один до одного, об'єднуються;
- метод Уорда. У цьому методі поєднуються всі можливі пари кластерів і обчислюється сума квадратних відстаней у кожному кластері. Потім це підсумовується для всіх кластерів. Вибирається комбінація, яка дає найменшу суму квадратів.

Неієрархічні методи кластеризації.

У цих методах заздалегідь задається бажана кількість кластерів і вибирається найкраще рішення. Неієрархічний аналіз кластерів, як правило, використовується, коли задіяні великі набори даних. Іноді це є переважним, оскільки дозволяє суб'єктам переходити з одного кластеру в інший (це неможливо в ієрархічному аналізі кластерів).

Види даних та міри відстані.

Дані, що використовуються при кластерному аналізі, можуть бути інтервальними, порядковими або категоричними. При кластерному аналізі потрібно мати певний спосіб, котрий характеризується трьома особливостями: мірою близькості будь-яких двох об'єктів, мірою близькості двох груп об'єктів, правилом вибору остаточного варіанту класифікації.

Список літератури

1. Пістунов І. М. Кластерний аналіз в економіці / І. М. Пістунов, О. П. Антонюк, І. Ю. Турчанінова. – Дніпропетровськ: НГУ, 2008. – 84 с.

2. Кластерный анализ [Електронний ресурс] – Режим доступу до ресурсу: http://statsoft.ru/home/textbook/modules/stcluan.html.