

РОЗРОБКА ПРОГРАМНОГО МОДУЛЯ ДЛЯ АНАЛІЗУ ТА ПІДБОРУ ІНФОРМАЦІЇ ВІДПОВІДНО ДО ПРОФІЛЮ КОРИСТУВАЧА

Моргун І. С.

Науковий керівник – к.т.н., доц. Суприган О. І.

Система, що забезпечує підбір інформації кожному користувачу індивідуально, повинна виконувати такі функції:

- опрацювання індивідуальних даних про користувача та формування критеріїв підбору відповідної інформації.
- по закінченню аналізу виводити запропоновану інформацію відповідно критеріям підбору.
- можливість зберігання інформації, наданої користувачу.

По даній темі постала задача аналізу великої кількості даних. Дуже помічним виявився кластерний аналіз, який являє собою сукупність методів класифікації багатовимірних спостережень чи об'єктів, які базуються на визначенні поняття віддалі між досліджуваними об'єктами з наступним виділенням в них подібних груп. При цьому не вимагається апріорної інформації про розподіл генеральної сукупності. Однорідність сукупності задається правилом обчислення певної метрики, що характеризує ступінь подібності j -ї та k -ї одиниць сукупності. Такою метрикою може бути відстань між ними S або коефіцієнт подібності R_{jk} . Близькі, схожі за вибраними метриками одиниці вважаються належними до одного типу, однорідними. Вибір метрики є вузловим моментом кластерного аналізу, від якого залежить кінцевий варіант поділу сукупності на класи. Основною метою кластерного аналізу є розділення багатовимірної сукупності вхідних даних на однорідні групи так, щоб об'єкти всередині групи були подібними між собою згідно з деяким критерієм, а об'єкти із різних груп відрізнялися один від одного. Причому класифікація об'єктів проводиться одночасно за декількома ознаками на основі введення певної міри сумарної близькості за всіма ознаками класифікації.

Отже, оптимальним вирішенням поставленої задачі є дивізімний метод, так як відповідні класи заздалегідь не сформовані, а будуть створені за допомогою сформованих критеріїв шляхом поділу початкового кластера. На кожному кроці ділення кластера буде здійснюватись суто за сформованими критеріями підбору.

Дивізімний метод виявляє досить високу стійкість по відношенню до шумів і викидів, некоректного вибору метрики, включенню незначущих змінних в набір, який бере участь в кластеризації.