

УДК 004.043

С.С. Заверуха

Тернопільський національний технічний університет імені Івана Пулюя, Україна

**ВИКОРИСТАННЯ БІНАРНИХ N-ВИМІРНИХ ВЕКТОРІВ ДЛЯ
ВСТАНОВЛЕННЯ МІРИ ПОДІБНОСТІ КОРИСТУВАЧІВ ІНФОРМАЦІЙНИХ
СИСТЕМ**

S.S. Zaverukha

**USING BINARY N-DIMENSIONAL VECTORS TO ESTABLISH THE SIMILARITY
OF INFORMATION SYSTEM USERS**

У еру розвитку сучасних комп'ютерних технологій постає проблема звуження соціальних кіл, що створює труднощі в комунікативному аспекті користувачам інформаційних систем. Ця, проблема обумовлена не тільки внутрішніми чинниками (характер, емоційна зрілість), а й зовнішніми (зміна місця проживання тощо.). Враховуючи, що людина являє собою соціальну істоту, яка має власний спектр інтересів, пошук важливих соціальних контактів на основі тем, що цікавлять є необхідною соціальною потребою.

Через складність пошуку важливих соціальних контактів відбувається процес асиміляції діаспор, малочисельних культур та субкультур. У певних випадках результатом уподібнення є доволі болісний процес інтеграції особистості у соціум, що призводить до стику двох культур. Як результат, в певних випадках процес інтеграції особистості в нове оточення є досить складним, що в подальшому може призвести до стику двох культур. Одним з варіантів вирішення даної проблеми є формування кластеру подібних користувачів на основі множин вподобань конкретних особистостей. Оскільки, кластерний аналіз являє собою статистичну процедуру збору та пошуку груп наближено схожих об'єктів, він дозволяє легше зрозуміти набір даних, та обрати відповідні методи обробки для певної групи об'єктів[1]. Для побудови кластеру можна скористатись графовим методом кластеризації, а саме методом ієрархічної кластеризації (кластерного дерева). Даний метод будує ієрархію кластерів з окремих елементів шляхом поступового злиття кластерів, орієнтуючись на матрицю відстаней між вузлами графа [2]. Порівняти отримані множини можна використавши бінарну міру подібності, що була запропонована Полем Жаккаром. Використовуючи різницю об'єднання та перетину множин ми отримуємо коефіцієнт подібності n-вимірних векторів, що в свою чергу відображає відповідність спектру інтересів користувачів один до одного[3].

Розробками алгоритмів для кластеризації користувачів на основі отриманих даних займається ряд компаній, такі як Rakuten, Facebook, Google та інші. Окрім цього створено програми, які за допомогою алгоритмів кластеризації, здійснюють розподіл користувачів на різні підгрупи. Проте, зважаючи на точність та недосконалість сучасних методів кластеризації n-вимірних векторів, дослідження даної технології все ще, залишається складною задачею.

Література

1. Chapter 8: Hierarchical Clustering. Introduction to HPC with MPI for Data Science. Springer [електронний ресурс] – режим доступу: <https://www.springer.com/gp/book/9783319219028>
2. Java Competitive Learning Application [Електронний ресурс] - Режим доступу: <http://homepages.feis.herts.ac.uk/~nngroup/software.html>
3. SimMetrics a sourceforge implementation of Jaccard index and many other similarity metrics [Електронний ресурс] - Режим доступу: <http://sourceforge.net/projects/>