

## АЛГОРИТМ ЗАБЕЗПЕЧЕННЯ ПІДТРИМКИ ПРОЦЕСУ СЕГМЕНТАЦІЇ СПОЖИВАЧІВ І ПОЗИЦІЮВАННЯ ТОВАРІВ ТА ПОСЛУГ

Кочубей І. Ю., Дорошенко А. В., Жураковська О. С.

## АЛГОРИТМ ОБЕСПЕЧЕНИЯ ПОДДЕРЖКИ ПРОЦЕССА СЕГМЕНТАЦИИ ПОТРЕБИТЕЛЕЙ И ПОЗИЦИОНИРОВАНИЯ ТОВАРОВ И УСЛУГ

Кочубей И. Ю., Дорошенко А. В., Жураковская О. С.

## ALGORITHM OF SUPPORTING THE PROCESS OF CONSUMER SEGMENTATION AND THE POSITIONING OF GOODS AND SERVICES

Kochubey I., Doroshenko A., Zhurakovska O.

Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
м. Київ, Україна  
[illa.kochubey1@gmail.com](mailto:illa.kochubey1@gmail.com)  
[o.zhurakovska@kpi.ua](mailto:o.zhurakovska@kpi.ua)

*Стаття присвячена актуальному питанню сегментації споживачів і позиціювання товарів та послуг. Розглянуті базові алгоритми класифікації та кластеризації. Розроблено алгоритм, який являється комбінацією кластеризації методом  $k$ -середніх та колаборативної фільтрації методом коефіцієнта кореляції Пірсона.*

**Ключові слова:** кластерний аналіз, колаборативна фільтрація

*Статья посвящена актуальному вопросу сегментации потребителей и позиционирования товаров и услуг. Рассмотрены базовые алгоритмы классификации и кластеризации. Разработан алгоритм, который является комбинацией кластеризации методом  $k$ -средних и колаборативной фильтрации методом коэффициента корреляции Пирсона.*

**Ключевые слова:** кластерный анализ, колаборативная фильтрация

*The article is attributed to the actual problem of the process of consumer segmentation and the positioning of goods and services. The basic classification and clustering algorithms are reviewed. An algorithm has been developed that is a combination of clustering using the  $k$ -means method and collaborative filtering using the Pearson correlation coefficient method.*

**Keywords:** cluster analysis, collaborative filtering

### ВСТУП

Сьогодні системи підтримки діяльності в сфері послуг та торгівлі користуються великим попитом. Такі системи дають можливість вирішити ряд проблем, що постають перед будь-якою організацією, що функціонує в сфері послуг та торгівлі.

В процесі аналізу ринку продажу товарів та послуг є актуальною проблема створення каталогів продукції, які забезпечують швидкий і наглядний доступ до товарів та послуг. Це вимагає здійснення сегментації ринку, виявлення цільової категорії користувачів та розробки маркетингових заходів для просування товарів та

послуг. У зв'язку із цим особливо актуальною є проблема вирішення задач класифікації та кластеризації, що дозволить знайти рішення поставлених задач.

### ПОСТАНОВКА ЗАДАЧІ

Для створеного сервісу з продажу товарів або послуг сформовано базу користувачів, які характеризуються профілем поведінки в системі. Цей профіль представлений множиною атрибутів користувачів та різних послідовностей їхніх дій при роботі із сервісом. Необхідно за цією інформацією здійснювати кластеризацію користувачів та формувати рекомендації користувачам.

В рамках даної статті розглядаються можливі методи вирішення проблеми сегментації ринку та виявлення цільової категорії клієнтів.

### АНАЛІЗ ДОСЛІДЖЕНЬ

В ході проведення досліджень даної теми було розглянуто ряд алгоритмів класифікації та кластеризації.

Метод колаборативної фільтрації відносно простий в реалізації та одночасно достатньо ефективний. Недоліком цього методу являється те, що даний метод не вирішує проблему «холодного» старту.

Метод кластеризації *k-means*, як і метод колаборативної фільтрації, нескладний у реалізації. Точність даного методу досить висока. Проте цей метод має ряд недоліків. Основним недоліком являється те що на вхід алгоритму *k-means* необхідно подавати кількість кластерів, тобто треба заздалегідь знати на скільки категорій треба розділити дані, що в реальних задачах не завжди є можливим. Дану проблему можна вирішити запускаючи алгоритм *k-means* декілька разів і потім порівняти отримані результати, наприклад, використовуючи індекс Девіса-Болдуїна. Також треба зазначити, що час роботи даного алгоритму досить високий.

Метод кореляції Пірсона дає більш якісні результати порівняно з алгоритмом колаборативної фільтрації. Проте він також не вирішує проблему «холодного старту»[1].

Метод асоціативних правил, хоча і простий в реалізації проте дає не якісні результати та не вирішує проблему «холодного» старту[1].

Метод контентної фільтрації являється менш ефективним порівняно з вище згаданими методами та складнішим у реалізації[1].

Метод *TF-IDF* вирішує проблему «холодного» старту та являється досить точним. Проте він достатньо складний у реалізації[1].

Щоб обрати який алгоритм обрати для вирішення проблеми сегментації ринку та виявлення цільової категорії клієнтів можна використати метод *ELECTRE* та метод аналізу ієрархій (MAI).

Автори [2] використовують метод *ELECTRE* для прийняття рішення при виборі алгоритму класифікації та кластеризації, а автори [3] використовують метод аналізу ієрархії для вирішення цієї проблеми.

Згідно з [2], необхідно визначити множину альтернатив та отримати експертні оцінки за деякими критеріями.

Включимо до множини альтернатив деякі базові алгоритми кластеризації:

- кластеризація *k-середніх*;
- колаборативна фільтрація методом коефіцієнта кореляції Пірсона;
- *TF-IDF*;
- асоціативні правила;

- метод гібридної фільтрації;
- метод *SVD*;
- метод найближчого сусіда.

Множина критеріїв за якими здійснюється оцінювання альтернатив:

- принципи роботи;
- складність реалізації;
- ефективність;
- точність;
- ціна реалізації;
- взаємодія з іншими алгоритмами;
- можливість розв'язання проблеми «холодного» старту.

Після оцінювання альтернатив за вказаними критеріями для вибору найкращих альтернатив застосовується метод *ELECTRE*. В результаті отримане ядро: алгоритм *TF-IDF* та коефіцієнт кореляції Пірсона.

Згідно з [3], для вирішення цієї проблеми пропонується застосувати метод аналізу ієрархії. Множина альтернатив та множина критеріїв формуються аналогічно. В результаті визначено найкращу альтернативу – колаборативна фільтрація методом коефіцієнта кореляції Пірсона.

Підсумовуючи вище сказане для вирішення проблеми сегментації ринку та виявлення цільової категорії клієнтів було рекомендовано та обрано наступні алгоритми: коефіцієнт кореляції Пірсона та кластеризація методом *k*-середніх.

Розбиття множини векторів характеристик об'єктів на групи з урахуванням подібності цих об'єктів називають кластеризацією. Метод *k*-середніх являється одним із найпростіших серед всіх алгоритмів кластеризації. На вхід даному алгоритму необхідно подавати кількість кластерів на яку треба розділити вхідну вибірку.

На першій ітерації алгоритм випадковим чином розподіляє об'єкти по кластерах. Потім для кожного кластеру обраховується центр мас. Після чого обраховується відстань від кожного об'єкта до центрів мас. Об'єкти перерозподіляються по кластерах таким чином, щоб відстань між об'єктом і центром була найменшою. Наступний крок – знову обраховуються центри мас кластерів. Алгоритм триває доти поки об'єкти будуть змінювати належність до кластерів.

В даному алгоритмі можна використати наступні метрики для обрахування відстані між центрами мас кластерів та об'єктами:

- евклідова відстань;
- степенева відстань;
- квадрат евклідової відстані;
- відстань Чебишева;
- «манхеттенська» відстань;

Було прийняте рішення використовувати евклідову відстань (1) при реалізації алгоритму кластеризації.

$$\rho(x, x') = \sqrt{\sum_i^n (x_i - x'_i)^2} \quad (1)$$

Приклад кластеризації двовимірних векторів зображено на рис. 1.

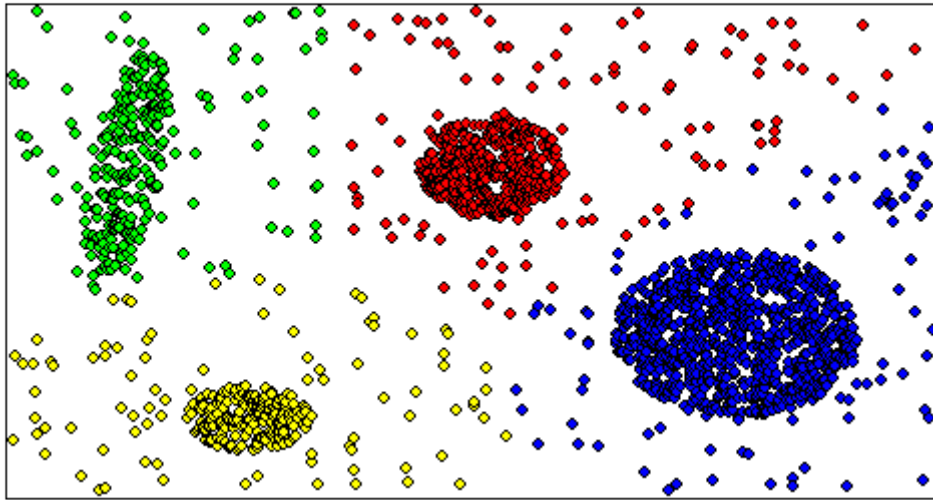


Рис. 1. Кластеризація у двовимірному просторі

На вхід реалізованому алгоритму подається множина векторів і кількість кластерів. Елементом кожного вектора є значення відповідної властивості об'єкту. В загальному випадку кількість властивостей, які описують об'єкт дорівнює  $n$ . Отже кластеризація проводиться в  $R^n$  просторі.

Наприклад, необхідно розділити на групи множину користувачів книгарні. Тоді властивостями будуть:

- вік;
- стать;
- місце проживання;
- кількість переглядів книг жанру «Фантастика»;
- кількість переглядів книг жанру «Дитяча література»;
- тощо.

На виході отримаємо набір кластерів користувачів, де користувачі з одного кластеру максимально подібні між собою за даними властивостями.

Після етапу кластеризації, настає етап фільтрації, що є кінцевою частиною комбінованого алгоритму.

Для того, щоб обрахувати подібність двох користувачів конкретного кластеру, будемо використовувати коефіцієнт кореляції Пірсона, за рахунок того, що можна виміряти лінійну залежність між ними. Особливістю реалізованого алгоритму є те, що ми не визначаємо значення функції між всіма користувачами системи, а лише між користувачами попередньо сформованого кластеру, що дає змогу значно пришвидшити алгоритм за рахунок зменшення вхідних даних для алгоритму. Ще одним важливим фактором є те, що фільтрація відбувається між подібними користувачами. Це значить, що будуть порівнюватись користувачі з однаковими вподобаннями. Такі дії дають можливість вдало поєднати його з алгоритмом  $k$ -середніх.

Покажемо формулу алгоритму КФ реалізованого методом ККП (2).

$$r_{u,i} = k \sum_{u' \in U} sim(u, u') r_{u',i} \quad (2)$$

$sim$  – міра, яка показує на стільки два користувача подібні між собою;

$k$  – коефіцієнт нормування (3).

$$k = \frac{1}{\sum_{u' \in U} |sim(u, u')|} \quad (3)$$

$U$  – множина всіх користувачів;

$r$  – оцінка, яку виставив користувач;

Отже розроблено алгоритм, який являється комбінацією кластеризації методом  $k$ -середніх та колаборативної фільтрації методом коефіцієнта кореляції Пірсона.

## ВИСНОВКИ

Проаналізовано проблему сегментації користувачів веб-каталогу. Зроблено висновок про доцільність застосування колаборативної фільтрації, попередньо кластеризувавши користувачів за спільними вподобаннями. Розроблене алгоритмічне забезпечення застосовано для вирішення задачі сегментації користувачів. Отриману в результаті множину кластерів доцільно використовувати при формуванні рекомендації користувачам та в процесі просування продукції на ринку.

## ЛІТЕРАТУРА

1. Шитиков В. К., Мاستицкий С. Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R.: электронна книга, адреса доступа: <https://github.com/ranalytics/data-mining> 2017. 351 с.
2. Литвак Р. Б., Дорошенко А. В. Метод ELECTRE як інструмент вибору алгоритмів системи формування пропозицій користувачам книгарні. *Підсумки розвитку наукової думки: 2018: зб. наук. праць «ΛΟΓΟΣ» з матеріалами міжнар. наук.-практ. конф., м. Івано-Франківськ, 5 грудня, 2018 р. Вінниця : ГО «Європейська наукова платформа». 2018. Т.4. с128. С. 88-95.*
3. Литвак Р. Б., Кочубей І. Ю. Застосування МАІ для вирішення задачі вибору алгоритму системи формування пропозицій користувачам книгарні. *Aktuelle Themen im Kontext der Entwicklung der modernen Wissenschaften: der Sammlung wissenschaftlicher Arbeiten «ΛΟΓΟΣ» zu den Materialien der internationalen wissenschaftlich-praktischen Konferenz, Dresden, 23 Januar, 2019. Dresden : NGO «Europäische Wissenschaftsplatform», 2019. B. 5. 122 s. С. 62-68.*