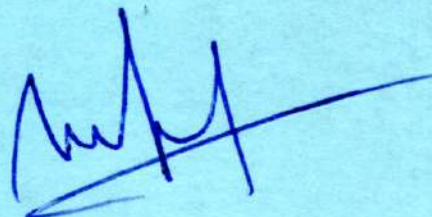


Міністерство освіти і науки України
Національний університет «Львівська політехніка»



ШВАРЦ МИХАЙЛО ЄВГЕНІЙОВИЧ

УДК 004.89

**ГІБРИДНІ МОДЕЛІ І МЕТОДИ ПРОГНОЗУВАННЯ РЕКОМЕНДАЦІЙ ДЛЯ
ІНТЕРНЕТ-МАГАЗИНУ**

01.05.03 - Математичне та програмне забезпечення обчислювальних
машин і систем

Автореферат

дисертації на здобуття наукового ступеня
кандидата технічних наук

Львів 2019

Дисертацією є рукопис.

Робота виконана у Національному університеті «Львівська політехніка»
Міністерства освіти і науки України.

Науковий керівник - доктор технічних наук, професор **Лобур Михайло Васильович**, Національний університет "Львівська політехніка" МОН України, завідувач кафедри систем автоматизованого проектування.

Офіційні опоненти - доктор технічних наук, професор **Глоба Лариса Сергіївна**, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» МОН України, м. Київ, завідувач кафедри інформаційно-телекомунікаційних мереж;

кандидат технічних наук, доцент **Боднарчук Ігор Орестович**, Тернопільський національний технічний університет імені Івана Пулюя МОН України, м. Тернопіль, завідувач кафедри комп'ютерних наук.

Захист відбудеться 24 жовтня 2019 р. о 16-00 годині на засіданні спеціалізованої вченої ради Д 35.052.05 у Національному університеті «Львівська політехніка» (79013, м. Львів-13, вул. С. Бандери, 12, ауд. 226 головного корпусу).

З дисертацією можна ознайомитися у науково-технічній бібліотеці Національного університету "Львівська політехніка" (79013, м. Львів, вул. Професорська, 1).

Автореферат розісланий: 24 вересня 2019 року

Вчений секретар

спеціалізованої вченої ради, д.т.н., проф.



Р.А. Бунь

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Людина в сучасному інформаційному суспільстві змушена опрацьовувати великі об'єми інформації. Користувач сучасного інформаційного суспільства потребує споживання конкретної персоналізованої інформації, але він не здатний цілеспрямовано опрацьовувати цю інформацію навіть в обмеженій предметній області. Рекомендаційна система спрямована на взаємодію з користувачем, якому надається величезна кількість об'єктів, серед яких він повинен зробити свій вибір. Користувачеві може не вистачати знань або досвіду, або того й іншого, щоб вибрати те, що відповідає його потребам. Користувач явно або побічно надає системі інформацію про свої уподобання. Таким чином, рекомендаційна система представляється у вигляді системи (програми), що використовує певний алгоритм фільтрації та наявну інформацію про потреби користувача, щоб рекомендувати йому набір об'єктів, які будуть найбільш корисними для нього. Рекомендаційні системи використовуються у середовищі електронної комерції, пошукових системах, системах електронної освіти. Широке використання рекомендаційні системи здобули в сучасних інтернет-магазинах як одному із основних структурних елементів електронної комерції. Рекомендаційні системи класифікують за способом відбору необхідного користувачеві матеріалу. В основному застосовується два базові підходи: колаборативна фільтрація і контентна фільтрація. Також існує гібридна фільтрація, яка поєднує в собі як колаборативну, так і контентну фільтрацію, або колаборативну фільтрацію на основі даних і моделей.

Розвиток електронної комерції в Україні відповідає Закону України «Про електронну комерцію». Відповідно до цього закону інтернет-магазин – це засіб представлення або реалізації товару, роботи чи послуги шляхом вчинення електронного правочину. Важливий внесок у розроблення структури і функцій інтернет-магазину здійснили такі вчені, як Плескач В.Л., Затонацька Е.Г. – розробили тактичні прийоми електронної комерції, дали характеристику поняття електронного магазину, розробили типову структуру інтернет-магазину, Шердані А. – розробив метод комплексного аналізу і порівняння економічної ефективності інтернет-магазинів, Шалева О.І. – дослідила основні категорії та інструментарій електронної комерції, технологію роботи інтернет-магазину, Тардаскіна Т.М., Стрельчук Є.М., Терешко Ю.В. – дослідили системи електронної комерції в споживчому секторі, виділили інтернет-магазин як один із основних елементів системи електронної комерції в споживчому секторі, Huang S.I. – дослідив застосування рекомендаційних систем в електронній комерції, Linden G., Smith J., York J. – розглянули застосування методу колаборативної фільтрації в інтернет-магазині Amazon.com, Resnick P., Iakovou N. – розробили одну із перших рекомендаційних систем GroupLens.

До класичних задач рекомендаційних систем, які потребують подальшого дослідження в області електронної комерції, у тому числі і при функціонуванні інтернет-магазину, належать підвищення точності і достовірності прогнозування рекомендацій, вирішення проблеми розрідженості матриці користувач-предмет, вирішення проблеми холодного старту, вирішення проблем масштабованості, синонімії, шахрайства, проблеми користувачів, профілі котрих постійно значно відрізняються від профілів інших користувачів.

Для інтернет-магазину додатковими задачами являються збільшення конверсії, збільшення лояльності користувачів, вирішення задачі супутніх продаж, вирішення задачі додаткових продаж, вирішення задачі після продажного супроводу користувачів, вирішення задачі «довгого хвоста». Окрім класичних задач, які перераховано вище, актуальною задачею підвищення ефективності роботи інтернет-магазину є задача пошуку груп користувачів з подібними уподобаннями і демографічними характеристиками. Розбиття користувачів на такі групи дозволяє підвищити точність прогнозування рекомендацій, вирішити задачу холодного старту, а також полегшує вирішення задач супутніх продаж, додаткових продаж і після продажного супроводу.

Отже, тема дисертаційної роботи є актуальною.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана на кафедрі «Системи автоматизованого проектування» Національного університету «Львівська політехніка» та безпосередньо пов'язана з планами наукових досліджень у межах госпдоговірних тем та міжнародних наукових грантів, і відповідає науковому напрямку кафедри – автоматизація проектування та моделювання систем «розумного будинку» та виконана в межах госпдоговірних тем та міжнародних наукових грантів, а саме:

- міжнародний науково-дослідний проект TEMPUS-JPCR – «Розробка програми для нової спеціальності: “Магістр з інженерії проектування мікросистем”», термін виконання проекту 15.10.2012 - 14.10.2016, реєстраційний номер: № 530785-TEMPUS-1-2012-1-PL- TEMPUS-JPCR;

- грант молодих учених "Моделювання і створення нового класу кристалічних нанокомполімерів із контрольованою кристалізацією та їх дослідження в оптичному та субтерагерцовому діапазонах хвиль", термін виконання проекту 01.01.2016 - 31.12.2018, номер державної реєстрації 0116U004412.

Мета і завдання дослідження. Метою дисертаційної роботи є вироблення релевантних прогнозів рекомендацій для кінцевих користувачів інтернет-магазину за рахунок розроблення і удосконалення моделей та методів прогнозування рекомендацій спрямованих на підвищення точності, достовірності (релевантності) і швидкості прогнозування рекомендацій.

Мета дисертаційної роботи визначає необхідність розв'язання таких завдань:

- аналіз сучасного стану моделей і методів прогнозування рекомендацій для задач електронної комерції та інтернет-магазинів як суб'єктів електронної комерції;
- удосконалення методів розрахунку коефіцієнтів подібності векторів профілів користувачів і предметів для методу зваженої суми шляхом використання демографічних характеристик користувачів, якісних характеристик предметів і оберненої евклідової відстані між векторами профілів;
- розроблення методу пошуку груп користувачів, який адаптується до коефіцієнту розрідженості матриці користувач-предмет і використовує методи категоріальної, мішаної і числової кластеризації;
- розроблення методу мішаної кластеризації, який використовується для кластеризації категоріально-числових векторів профілів користувачів і автоматично вибирає центри кластерів;
- розроблення методу збільшення різноманітності рекомендованих предметів;

- розроблення методу прогнозування рекомендацій для користувачів інтернет-магазину на основі концепції асоціативних правил, який використовує алгоритм пошуку асоціативних правил за допомогою адаптивної зміни підтримки асоціативних правил;
- дослідження розроблених моделей і методів на тестовому наборі даних Movilens.

Об'єкт дослідження – процес вироблення рекомендацій інтернет-магазином для користувача.

Предмет дослідження – моделі, методи та алгоритми прогнозування рекомендацій для користувачів інтернет-магазину.

Методи дослідження – методи лінійної алгебри, методи кластерного аналізу, теорії реляційних баз даних, методи пошуку асоціативних правил та методи математичної статистики.

Наукова новизна отриманих результатів полягає у тому, що:

1. На основі концепції застосування в одному методі категоріальної, мішаної і числової кластеризації вперше розроблено метод пошуку груп користувачів, який адаптується до розрідженості матриці користувач-предмет.
2. Отримав подальший розвиток метод розрахунку коефіцієнтів подібності векторів профілів користувачів і векторів профілів предметів, який на відміну від існуючих, використовує демографічні характеристики користувачів, що дозволяє підвищити точність прогнозування рекомендацій і визначати коефіцієнти подібності для нового користувача і нового предмета.
3. Отримав подальший розвиток метод мішаної кластеризації, який використовується для кластеризації категоріально-числових векторів профілів користувачів і, на відміну від існуючих, автоматично вибирає центри кластерів і дозволяє зменшити час пошуку груп користувачів при високій точності виділення груп.
4. Отримав подальший розвиток метод збільшення різноманітності рекомендованих предметів, який дозволяє врахувати оцінки подібних предметів в околі предметів активного користувача і вирішує проблему "довгого хвоста".
5. Удосконалено метод прогнозування рекомендацій для користувачів інтернет-магазину, який на відміну від існуючих методів, використовує алгоритм пошуку асоціативних правил за допомогою адаптивної зміни підтримки асоціативних правил.

Практичне значення отриманих результатів дисертаційної роботи полягає у тому, що розроблені моделі і методи можуть бути застосовані в електронній комерції при створенні інтернет-магазинів, інтернет-аукціонів, веб-порталів, електронних дошок оголошень та інформаційно-пошукових системах.

Результати дисертаційних досліджень використано у навчальному процесі Національного університету «Львівська політехніка», зокрема на кафедрі систем автоматизованого проектування при викладанні дисциплін: «Інноваційні інформаційні технології» для підготовки магістрів за спеціальністю «Інформаційні технології проектування» і «Методи і системи штучного інтелекту» для підготовки бакалаврів за спеціальністю «Комп'ютерні науки».

Особистий внесок здобувача. Всі наукові результати теоретичних і практичних досліджень, які викладені в дисертації, отримані автором особисто. У друкованих

працях, опублікованих у співавторстві, дисертанту належать: метод пошуку груп користувачів, який адаптується до розрідженості матриці користувач-предмет і використовує категоріальну, мішану і числову кластеризацію [1]; вдосконалений метод розрахунку коефіцієнтів подібності векторів профілів користувачів і предметів шляхом введення демографічних і контентних характеристик користувачів і предметів [2]; метод надання рекомендацій для груп користувачів на основі кластеризації векторів профілів користувачів [3]; розроблено метод, який надає рекомендації новому користувачу, пропонує йому супутні предмети і використовує асоціативні правила [4]; метод прогнозування рекомендацій при проектуванні складних мікросистемних пристроїв [5]; результати дослідження методів розрахунку коефіцієнтів подібності векторів профілів користувачів і предметів за допомогою косинусної міри подібності, коефіцієнта кореляції Пірсона і оберненої евклідової відстані [6]; аналіз існуючого стану розвитку методів, засобів і алгоритмів побудови рекомендаційних систем [7]; розроблено метод прогнозування рекомендацій при проектуванні мікроелектромеханічних систем [8]; дослідження основних концепцій побудови і застосування рекомендаційних систем [9]; дослідження моделі прогнозування рекомендацій для груп користувачів на основі кластеризації векторів профілів користувачів, які входять у матрицю користувач-предмет [10]; удосконалений метод розрахунку коефіцієнтів подібності користувачів з врахуванням демографічних характеристик [11]; метод двоетапної категоріальної і числової кластеризації для виділення груп користувачів [12]; метод збільшення різноманітності рекомендованих предметів [13]; аналіз геометричного розрахунку числа Пі за допомогою методу Монте-Карло [14].

Апробація результатів дисертації. Основні результати дисертації доповідалися і обговорювалися на Міжнар. наук.-техн. конф. «Перспективні технології і методи проектування МЕМС» (Львів-Поляна, 2016, 2017, 2018), Міжнар. наук.-техн. конф. «Досвід розробки та застосування САПР в мікроелектроніці» (Львів-Поляна, 2017), Міжнар. наук.-техн. конф. «САПР у проектуванні машин. Питання впровадження і навчання» (Львів-Поляна, 2017, 2018), Наук.-практ. конф. «Проблеми та перспективи розвитку економіки і підприємництва та комп'ютерних технологій в Україні» (Львів, 2017), Всеукр. наук.-практ. конф. «Комп'ютерне моделювання та програмне забезпечення інформаційних систем і технологій» (Рівне, 2017).

Публікації. За результатами досліджень, які викладено в дисертації, опубліковано 14 наукових праць, з-поміж яких 4 статті опубліковано у фахових наукових виданнях України, 1 публікація в зарубіжному науковому періодичному виданні, яке входить в міжнародні наукометричні бази, 9 публікацій в матеріалах науково-технічних конференцій, 6 з яких міжнародні.

Структура та обсяг дисертації. Дисертаційна робота складається із вступу, чотирьох розділів, висновків, додатків та списку літературних джерел. Загальний обсяг дисертації складає 152 сторінок, з них 135 сторінок основного тексту, містить 41 рисунок та 10 таблиць. Список літератури містить 113 найменування.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету і завдання роботи, наукову новизну та практичну цінність отриманих результатів, показано зв'язок роботи з науковими програмами кафедри, планами і

темами. Наведено дані про впровадження результатів роботи, їх апробацію, публікації та особистий внесок здобувача.

У **першому розділі** виконано аналіз сучасного стану моделей і методів електронної комерції, виділено основні класи електронної комерції, показано, що інтернет-магазин є однією із основних систем електронної комерції, виконано класифікацію і виділено основні види інтернет-магазинів, наведено основні кроки роботи інтернет-магазину, розроблено структуру роботи інтернет-магазину, показано область застосування рекомендаційних систем у структурі роботи інтернет-магазину, наведено структури Веб-сайтів для інтернет-магазину, показано зв'язок структури Веб-сайту із процесом надання рекомендацій, наведено основні метрики ефективності роботи інтернет-магазину, здійснено класифікацію існуючих підходів до побудови рекомендаційних систем, виділено особливості застосування рекомендаційних систем у роботі інтернет-магазину.

Електронна комерція як самостійний напрямок інформаційно-комунікаційних технологій виникла і почала розвиватися з середини 90-х років XX століття. У першу чергу це пов'язано з винаходом Всесвітньої павутини (Word Wide Web) і широким впровадженням всесвітньої мережі Internet. Аналіз стану розвитку електронної комерції на сучасному етапі показав, що можна виділити такі моделі електронної комерції: бізнес-бізнес, бізнес-уряд, бізнес-споживач, споживач-уряд, споживач-споживач. Найбільшого поширення у світовій електронній комерції набула модель бізнес-споживач. Показано, що основним суб'єктом цієї моделі є інтернет-магазин.

Проведено класифікацію інтернет-магазинів. Показано, що інтернет-магазини можна класифікувати: за методом роздрібного продажу товарів у мережі; за бізнес-моделлю; за взаєминами з постачальниками. Можна виділити різні види інтернет-магазинів: сайт-магазин; сайт-посередник; сайт-вітрина. Виділено Веб-сайт як основний структурний елемент інтернет-магазину. Показано, що основними структурами Веб-сайтів інтернет-магазину є деревоподібна структура і тегова структура. В деревоподібній структурі має місце вкладеність однієї категорії в іншу, для кожної послуги або товару формується окрема гілка. Це стосується брендів, категорій, видів продукції. У структурі, яка складається із тегів створюються окремі сторінки тегів на основі певних параметрів: ціна, акції, особливості використання, властивості, специфічні характеристики і т.д. Показано, що перевага структури, яка складається з тегів, полягає у збільшенні кількості сторінок, які можуть залучити додатковий трафік. Такий результат досягається за рахунок розширення семантичного ядра: використання більшої кількості низькочастотних запитів.

Розроблено структуру роботи інтернет-магазину, яка включає такі основні кроки: 1) покупець заходить на сайт інтернет-магазину; 2) вибирає товар або товари, які його цікавлять; 3) кладе їх у кошик; 4) у кошику інтернет-магазин пропонує купити до цих товарів аксесуари або подібні товари; 5) покупець щось додатково вибирає і поповнює кошик; 6) покупець оформляє замовлення; 7) співробітник інтернет-магазину зв'язується з покупцем і підтверджує замовлення, потім передає його в службу доставки; 8) наступний крок – супровід проданого товару самим інтернет-магазином до клієнта.

Показано, що рекомендаційна система залучається до роботи на другому, четвертому і восьмому кроках. На другому кроці рекомендаційна система пропонує

користувачу товари на основі прогнозування рекомендацій вибраним методом, на четвертому кроці рекомендаційна система пропонує користувачу супутні товари і додаткові товари, на восьмому кроці рекомендаційна система працює з користувачем у режимі e-mail маркетингу. Дано характеристики кожного із цих режимів роботи і наведено особливості прогнозування рекомендацій у цих режимах.

У результаті дослідження застосування рекомендаційних систем у роботі інтернет-магазину виділено наступні метрики ефективності: конверсія, забезпечення лояльності користувачів, збільшення супутніх продажів і збільшення додаткових продажів. У дисертаційній роботі наведено вирази для розрахунку кожної із метрик і показано застосування рекомендаційних систем для збільшення цих показників.

Показано, що сучасні рекомендаційні системи можна класифікувати таким чином:

- контентно орієнтовані;
- системи колаборативної фільтрації;
- демографічні рекомендаційні системи;
- системи, які ґрунтуються на знаннях;
- гібридні рекомендаційні системи.

Показано, що найбільш широко в теперішній час використовуються системи колаборативної фільтрації і гібридні рекомендаційні системи.

Виділено основні проблеми систем колаборативної фільтрації: проблема холодного старту, проблема точності розрахунку коефіцієнтів подібності профілів користувачів і профілів предметів, проблема розрідженості матриці користувач-предмет, проблема масштабованості та проблема «довгого хвоста».

У результаті проведеного дослідження встановлено, що найбільш перспективними є гібридні системи прогнозування рекомендацій.

У **другому розділі** дисертаційної роботи наведено формальну постановку задачі прогнозування рекомендацій методом колаборативної фільтрації для інтернет-магазину, наведено особливості формальної моделі для прогнозування рекомендацій з урахуванням груп користувачів, виконано порівняння існуючих методів розрахунку подібності векторів профілів користувачів і предметів, удосконалено метод розрахунку коефіцієнтів подібності, який, на відміну від існуючих, використовує обернену евклідову відстань між векторами профілів користувачів і демографічні характеристики користувачів, показано можливість застосування цього методу для рішення задачі холодного старту.

Основним завданням рекомендаційної системи є прогнозування рекомендацій для користувача таких товарів, предметів та послуг, котрі найкраще будуть відповідати його потребам. Формальна модель функціонування рекомендаційної системи може бути представлена таким чином: нехай корисність j -го предмету для i -го користувача є числове значення $R_{ij}(u_i, i_j) \in \{r_{\min}, r_{\max}\}$, де r_{\min} – мінімальне значення оцінки корисності, r_{\max} – максимальне значення оцінки корисності. Значення R_{ij} зберігаються у матриці $\hat{A}(M, N)$. У більшості випадків $R_{ij}(u_i, i_j) \in \{0,5\}$, або $R_{ij}(u_i, i_j) \in \{0,10\}$. Завданням рекомендаційної системи є прогнозування найбільш можливо корисних предметів для активного користувача $\hat{R}_{al}(u_a, i_l)$, $l \in \{1, \dots, K\}$, ($K \leq N$), де N – загальна кількість предметів, інформація про які зберігається у матриці користувач-предмет рекомендаційної системи.

При цьому завдання рекомендаційної системи полягає у тому, щоб досягнути мінімальної різниці між дійсною оцінкою предмета користувачем R_{ij} і прогнозованою оцінкою \hat{R}_{ij} , тобто

$$|R_{ij} - \hat{R}_{ij}| \Rightarrow \min . \quad (1)$$

Рекомендаційна система інтернет-магазину повинна забезпечити прогнозування таких предметів, які дійсно задовольняють користувачів. Це може бути представлено наступним рівнянням

$$F(\hat{R}_{ij}, R_{ij}) = \min \left(\sum_{i=1}^M \sum_{j=1}^N |\hat{R}_{ij} - R_{ij}| \right). \quad (2)$$

Узагальнена схема функціонування рекомендаційної системи при використанні методу колаборативної фільтрації представлена на рис. 1.

Основними відомими методами прогнозування рекомендацій для колаборативної фільтрації є метод зваженої суми для моделі користувач-користувач і моделі предмет-предмет. Модель користувач-користувач має наступний вигляд

$$\hat{R}_{ai} = \bar{R}_a + \frac{\sum_{u \in U_z} (R_{ui} - \bar{R}_u) \cdot w_{au}}{\sum_{u \in U_z} |w_{au}|}, \quad (3)$$

де \bar{R}_a і \bar{R}_u – середні значення для вектора оцінок вектора активного користувача a і вектора оцінок користувача u ;

\hat{R}_{ai} – прогнозоване значення оцінки предмета i для активного користувача a ;

w_{ai} – ваговий коефіцієнт подібності вектора оцінок активного користувача a і вектора оцінок предмета u ;

U_z – множина векторів профілів користувачів, які мають спільно вибрані предмети.

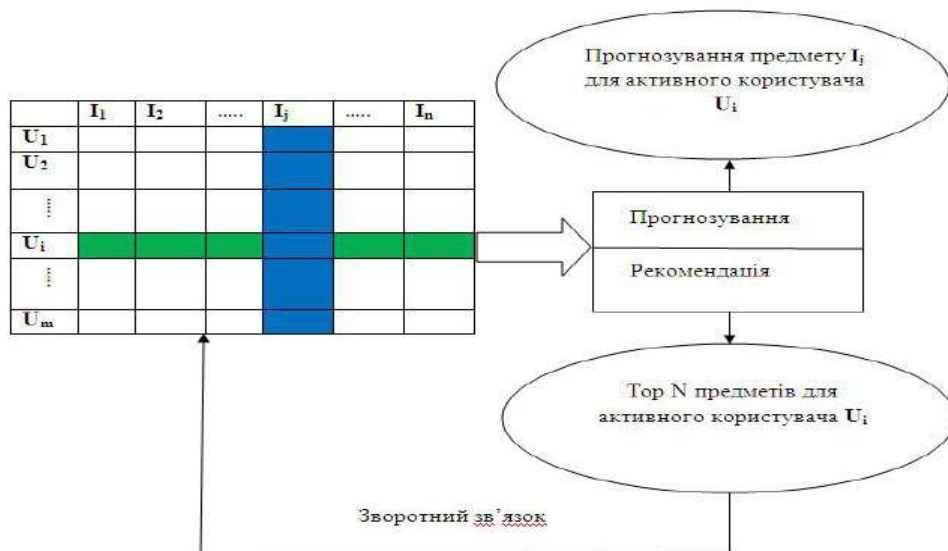


Рис. 1. Схема функціонування методу колаборативної фільтрації

Розрахунковий вираз для моделі предмет-предмет має такий вигляд:

$$\hat{R}_{ui} = \frac{\sum_{j \in I'} R_{uj} \cdot w_{uj}}{\sum_{j=1}^N |w_{uj}|}, \quad (4)$$

де \hat{R}_{ii} – прогнозоване значення оцінки предмету i для предмету u ; I' – множина предметів, які мають спільні оцінки користувачів.

Точність оцінок предметів залежить від точності розрахунку вагових коефіцієнтів подібності векторів. Показано, що найчастіше для розрахунку вагових коефіцієнтів подібності векторів використовують коефіцієнт кореляції Пірсона і косинусну міру близькості векторів і їх модифікації. Косинусна міра подібності широко використовується у методах інформаційного пошуку. В методах колаборативної фільтрації косинусна міра подібності використовується для розрахунку вагових коефіцієнтів подібності векторів профілів користувачів і векторів профілів предметів у багатовимірному просторі. Основним недоліком косинусної міри подібності є те, що вона не враховує відмінність між довжинами векторів профілів, а обчислює лише косинус кута між векторами. До основних недоліків коефіцієнта кореляції Пірсона належать: незадовільний результат при малій кількості спільно оцінених предметів; незадовільний результат при оцінці коефіцієнта близькості між векторами рейтингів, які містять лише один спільно оцінений предмет; ця міра не оцінює всі рейтингові оцінки, які містяться у парі векторів профілів. До спільного недоліку коефіцієнта кореляції Пірсона і косинусної міри близькості векторів належить те, що вони не працюють на стадії холодного старту.

У роботі запропоновано міру подібності, яка складається з двох складових: складової, яка враховує подібність векторів демографічних характеристик користувачів і складової, яка враховує подібність векторів числових рейтингових профілів користувачів. Цей підхід використовується для розрахунку вагових коефіцієнтів подібності для методу користувач-користувач і методу предмет-предмет. Подібність векторів числових рейтингових профілів користувачів обчислюється за такою формулою

$$w_{u_i, u_j} = \frac{1}{1 + D(u_i, u_j)}, \quad (5)$$

де u_i – вектор профілю користувача u_i ; u_j – вектор профілю користувача u_j ; $D(u_i, u_j)$ – евклідова відстань між векторами u_i і u_j .

Демографічну інформацію про користувача можна отримати при реєстрації користувача в системі, із соціальних мереж, із аналізу контенту інших сайтів. На основі аналізу демографічної інформації формується демографічний вектор профілю користувача

$$U_{\text{dem}} = \{\text{вік, стать, рід занять}\}. \quad (6)$$

Після цього компоненти вектора демографічних характеристик категоризуються за схемою, наведеною у табл.1.

Якісну характеристику предметів можна отримати із аналізу опису предметів. При цьому елементами вектору якісних характеристик предмета є атрибути предмета.

Всього категоризований вектор демографічного профілю має 27 компонент. Компонента категоризованого вектора профілю користувача містить 1 у тій позиції, яка відповідає категорії демографічної характеристики і 0 у решті позицій. Розмірність вектора демографічного профілю користувача і вектора якісних характеристик предмета може бути збільшена в залежності від потреб поставленої задачі.

Після категоризації компоненти вектора (6) мають двійкове бітове значення. Незважаючи на те, що розмірність вектора зростає до 27, двійковий бітовий вміст вектора дозволяє ефективно визначати подібність між категоризованими

демографічними векторами профілів користувачів. Для цього використовується коефіцієнт подібності Жакарда:

$$J = \frac{|A \cap B|}{|A \cup B|}, \quad (7)$$

де A і B – вектори, які можуть містити довільні дійсні значення, двійкові значення або алфавітно-цифрову інформацію.

Таблиця 1. Категоризація вектора демографічних характеристик користувача

Вік				Стать		Рід занять		
вік < 18	18 < вік < 30	30 < вік < 50	50 < вік	чол.	жін.	Лікар	інженер

Для двійкових векторів формула розрахунку коефіцієнта подібності (7) має вигляд:

$$J = \frac{M_{11}}{M_{10} + M_{01} + M_{11}}, \quad (8)$$

де M_{11} – загальна кількість елементів, у яких компоненти векторів A і B мають значення 1;

M_{10} – загальна кількість елементів, де компоненти вектора A мають значення 1 і компоненти вектора B мають значення 0;

M_{01} – загальна кількість елементів, де компоненти вектора A мають значення 0 і компоненти вектора B мають значення 1;

M_{00} – загальна кількість елементів, де компоненти векторів A і B мають значення 0.

Кожен компонент векторів A і B повинен знаходитись в одній із чотирьох категорій:

$$M_{11} + M_{10} + M_{01} + M_{00} = n, \quad (9)$$

де n – розмірність векторів A і B .

Обчислення коефіцієнтів у виразі (9) вимагає виконання операцій над двійковим вмістом демографічних векторів профілів, що вимагає значно менших часових затрат ніж обчислення коефіцієнтів подібності для рейтингових векторів профілів користувачів.

Для прогнозування рекомендацій у формулі (3) використовують два варіанти обчислення модифікованого коефіцієнта подібності

$$w'_{ij} = sim(i, j) + J_{ij} \times sim(i, j), \quad (10)$$

$$w'_{ij} = J_{ij} + J_{ij} \times sim(i, j), \quad (11)$$

де $sim(i, j) = w_{ij}$ – коефіцієнт подібності між числовими рейтинговими векторами профілів користувачів (5);

J_{ij} – коефіцієнт подібності між демографічними векторами профілів користувачів.

Вираз (10) враховує домінуюче значення подібності користувачів за рейтинговими векторами профілів. Вираз (11) враховує домінуюче значення подібності користувачів за демографічними векторами профілів.

Запропонований метод дозволяє вирішити проблему нового користувача (холодного старту). Для нового користувача вираз (11) має такий вигляд:

$$w'_{ij} = J_{ij}. \quad (12)$$

В подальшому отримані коефіцієнти подібності використовуються при прогнозуванні рекомендацій у формулах (3) та (4).

У **третьому розділі** дисертаційної роботи розроблено гібридний метод пошуку груп користувачів, який адаптується до розрідженості матриці користувач-предмет; розроблено новий метод мішаної кластеризації, який враховує категоріальні і числові складові вектора профілю користувача і адаптується до розрідженості матриці користувач-предмет; показано застосування методів прогнозування рекомендацій для груп користувачів; розроблено метод прогнозування рекомендацій на основі пошуку асоціативних правил; розроблено теоретико-множинні моделі прогнозування рекомендацій для супутніх продаж і додаткових продаж; розроблено метод прогнозування рекомендацій для супутніх продаж, додаткових продаж і режиму після продажної роботи, розроблено метод збільшення різноманітності товарів, які пропонує інтернет-магазин.

Актуальною задачею для будь-якого інтернет-магазину є пошук груп користувачів з подібними споживчими інтересами. Ця задача вирішується в дисертаційній роботі шляхом розроблення адаптивного гібридного методу пошуку груп користувачів за допомогою пошуку кластерів. При цьому кожен кластер є аналогом групи користувачів. У гібридному методі пошуку груп користувачів використовуються метод числової кластеризації, метод мішаної кластеризації та метод категоріальної кластеризації. Вибір методу кластеризації здійснюється за допомогою оцінки розрідженості матриці користувач-предмет.

Формальна модель прогнозування рекомендацій для груп користувачів полягає у наступному. Нехай $U = \{U_1, U_2, \dots, U_n\}$ – множина векторів профілів користувачів, $G = \{G_1, G_2, \dots, G_M\}$ – множина груп користувачів, $G_i = \{U_{1G_i}, U_{2G_i}, \dots, U_{kG_i}\}$ – множина профілів користувачів для групи G_i . Необхідно здійснити прогноз рекомендацій для груп користувачів $\hat{R}_{G_i} = \text{Predict}(G_i)$. При цьому кожен користувач із певної групи $U_j \in G_i$ отримує однакову рекомендацію. Узагальнену схему роботи методу прогнозування рекомендацій для спільнот користувачів наведено на рис. 2. Для пошуку груп подібних користувачів використовуються методи кластеризації.

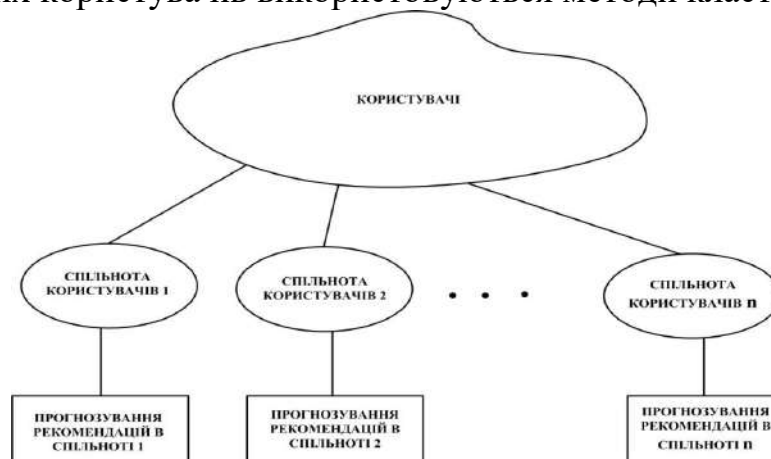


Рис. 2. Узагальнена схема роботи методу прогнозування рекомендацій для спільнот користувачів

Особливістю матриці користувач-предмет є те, що вона містить значну кількість нульових елементів. Кількість ненульових елементів не перевищує 10% від загальної кількості елементів матриці предмет-користувач. Тому для кластеризації користувачів у групи доцільно використовувати демографічні характеристики користувачів. Основними демографічними атрибутами користувачів є такі: вік, стать, освіта, рід занять. Вік – це числовий атрибут, стать, освіта, рід занять – категоріальні атрибути. Атрибут вік попередньо категоризується за схемою, яку наведено в табл. 1.

Нехай рейтинговий вектор профілю i -го користувача задається вектором

$$\mathbf{U}_i = (u_{1i}, u_{2i}, \dots, u_{mi}), \quad (13)$$

де u_{ji} – рейтингова оцінка j -го предмета i -м користувачем.

Розширимо цей вектор за допомогою демографічних атрибутів користувача

$$\mathbf{U}_i^{\text{ext}} = (u_{1i}, u_{2i}, \dots, u_{mi}, d_{1i}, d_{2i}, d_{3i}, d_{4i}), \quad (14)$$

де $d_{1i}, d_{2i}, d_{3i}, d_{4i}$ – категоріальні демографічні атрибути користувача.

У методі мішаної кластеризації відбувається кластеризація векторів профілів користувачів, які містять мішані демографічно-числові вектори профілів.

Структурну схему гібридного методу кластеризації для пошуку груп подібних користувачів наведено на рис. 3.

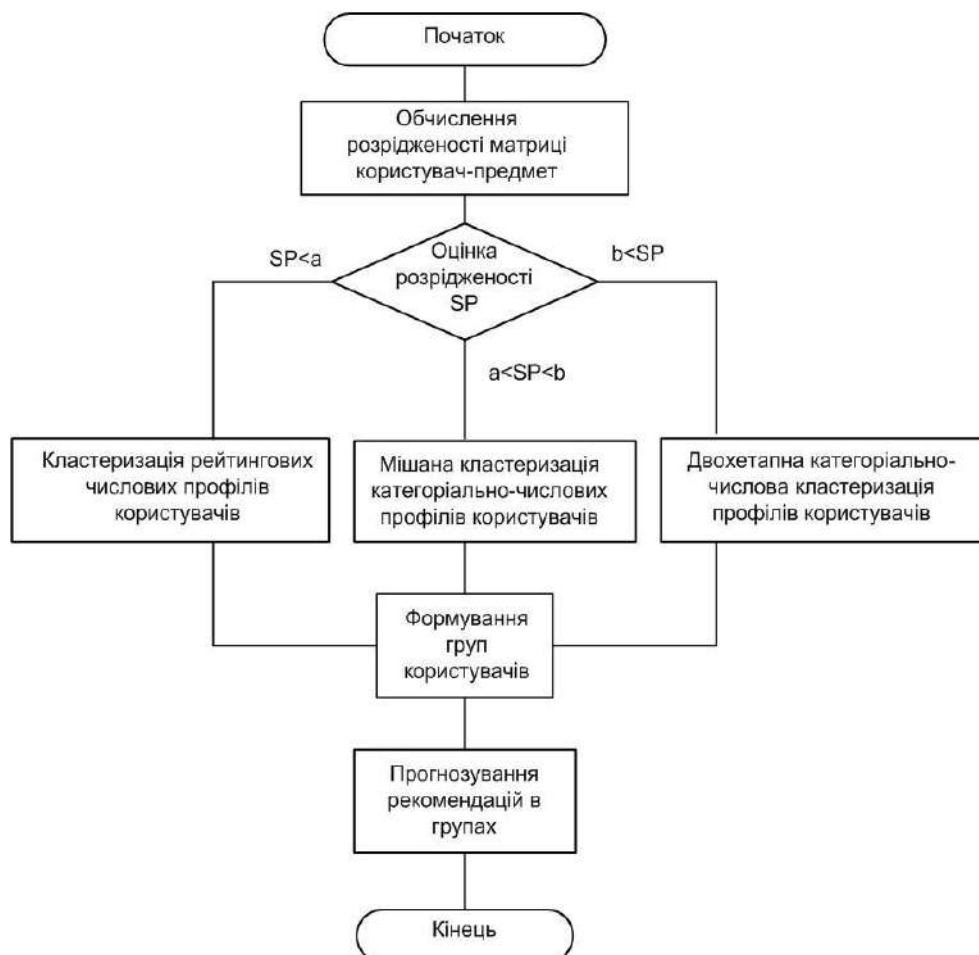


Рис. 3. Структурна схема гібридного методу кластеризації для пошуку груп подібних користувачів

Оцінка розрідженості матриці користувач-предмет відбувається за виразом:

$$SP = 1 - \frac{nR}{nUSER * nITEM}, \quad (15)$$

де nR – кількість відмінних від нуля елементів матриці користувач-предмет; $nUSER$ – кількість користувачів системи; $nITEM$ – кількість предметів в системі.

При великій розрідженості матриці користувач-предмет ($SP > 0,7$) звичайні методи числової кластеризації непридатні. Тому використовується алгоритм двоетапної категоріально-числової кластеризації. На першому етапі здійснюється категоріальна кластеризація методом ROCK. У результаті першого етапу кластеризації отримуємо кластери значно меншого розміру, які відповідають демографічним характеристикам користувачів. На другому етапі виконується пошук груп користувачів за допомогою методу числової кластеризації (рис. 4).



Рис. 4. Структурна схема двоетапного методу категоріально-числової кластеризації

Якщо $0,1 < SP < 0,7$, тоді кластеризацію виконуємо методом мішаної кластеризації; якщо $SP < 0,1$, тоді кластеризацію виконуємо методом числової кластеризації; якщо $SP > 0,7$, тоді кластеризацію здійснюємо двоетапним методом категоріально-числової кластеризації. У подальшому обчислюємо ступінь розрідженості кожного отриманого кластера. Якщо для отриманого кластера $0,1 < SP < 0,7$, або $SP > 0,7$ тоді виконуємо подальшу кластеризацію до отримання розрідженості $SP < 0,1$. В отриманих кінцевих кластерах виконуємо прогнозування рекомендацій методом зваженої суми, або розрахунок узагальненої рекомендації для всієї групи користувачів.

У дисертаційній роботі розроблено метод мішаної кластеризації. Для спрощення опису методу будемо позначати вектор U_i^{ext} за допомогою вектора $X_i = (x_{1i}, x_{2i}, \dots, x_{ni})$. Таким чином отримуємо мішаний вектор профілю користувача, який містить числові і категоріальні значення.

Кластеризація мішаних векторів профілів користувачів здійснюється за допомогою методу мішаної кластеризації. Метод мішаної кластеризації базується на розрахунку щільності розміщення мішаних векторів профілів користувачів і визначає кількість і положення центрів кластерів. Щільність визначається як кількість векторів профілів користувачів, які знаходяться в околі радіусом d_c біля кожного користувача :

$$\rho_i = \sum_{j=1}^N f(d_{ij} - d_c);$$

$$f(x) = \begin{cases} 1, & x = d_{ij} - d_c \leq 0, \\ 0, & x = d_{ij} - d_c > 0, \end{cases} \quad (16)$$

де d_{ij} – відстань між i -м та j -м векторами профілів користувачів; d_c – порогове значення; N – кількість користувачів.

Відстань між векторами профілів користувачів обчислюємо як середню зважену суму відстаней між атрибутами векторів профілів користувачів:

$$d_{ij} = D(X_i, X_j) = \frac{\sum_{k=1}^N w_{ij}^k l_{ij}^k}{\sum_{k=1}^N w_{ij}^k}, \quad (17)$$

де $w_{ij}^k = 0$, якщо атрибут d_{ij}^k відсутній і $w_{ij}^k = 1$ у протилежному випадку, l_{ij}^k – відстань між i -м та j -м атрибутами об'єктів X_i і X_j . Відстань l_{ij}^k обчислюється окремо для числових і категоріальних атрибутів векторів X_i і X_j .

Для числових атрибутів:

$$l_{ij}^k = \frac{|x_i^k - x_j^k|}{\max z_{ij}^k - \min z_{ij}^k}, \quad (18)$$

де $\max z_{ij}^k$ – максимальне значення у множині атрибутів векторів X_i і X_j , $\min z_{ij}^k$ – мінімальне значення у множині атрибутів векторів X_i і X_j .

Для категоріальних атрибутів:

$$l_{ij}^k = \begin{cases} 0, & x_i^k = x_j^k; \\ 1, & x_i^k \neq x_j^k. \end{cases} \quad (19)$$

Для знаходження положення і кількості центрів кластерів визначаємо мінімальні відстані між множинами об'єктів з різними щільностями

$\delta_i = \min_j (d_{ij}), \rho_j > \rho_i$. Будуються два вектори в порядку спадання значень $\rho_a > \rho_b$, $\delta_a > \delta_b$. Будемо вважати, що $\rho_1 \div \rho_a$ – це "великі" значення щільностей, $\rho_b \div \rho_n$ – це "малі" значення щільностей. Подібні умови накладемо на $\delta_1 \div \delta_a$ і $\delta_b \div \delta_n$. Якщо виконується умова $\rho_i \in (\rho_1, \rho_a)$ і $\delta_i \in (\delta_1, \delta_a)$, тоді X_i центр наступного кластера. Якщо виконується умова $\rho_i \in (\rho_b, \rho_n)$ і $\delta_i \in (\delta_1, \delta_a)$, тоді X_i об'єкт "шуму" і в подальших розрахунках не враховується. У подальшому об'єкти розподіляються по кластерах за критерієм мінімуму відстані до центрів кластерів.

Для числової кластеризації розроблено адаптивну модифікацію методу k -середніх. Метод включає два етапи. На першому етапі визначається орієнтовна кількість кластерів. На другому етапі метод визначає остаточну кількість і розміщення центрів кластерів і належність кожного об'єкта до певного кластера.

Контроль закінчення процесу кластеризації здійснюється за допомогою оцінки відношення середнього узагальненого об'єму усіх отриманих кластерів до середньої величини міжкластерної відстані.

Для прогнозування узагальнених рекомендацій для груп користувачів використовують алгоритм адитивної корисності, алгоритм мультиплікативної корисності предметів у групі, стратегія узгодженого голосування, стратегія найменшого задоволення, стратегія найбільшого задоволення користувача та стратегія перевищення заданого порогового значення.

Для рішення задач прогнозування рекомендацій для активного споживача, продажу супутніх товарів, додаткових товарів і після продажного e-mail маркетингу розроблено удосконалений метод пошуку асоціативних правил Apriori.

Асоціативні правила – це імплікація

$$X \Rightarrow Y,$$

де $X \in G, Y \in G, X \cap Y = \emptyset, G$ – множина досліджуваних предметів (книги, фільми, наукові статті і т. д.).

Ця імплікація означає, що при виборі предметів, які належать множині X , вибираються предмети, які належать множині Y .

Попередньо матриця користувач-предмет перетворюється у матрицю користувач-інтерес за таким алгоритмом: 1) якщо $R_{ij} < 3$ для інтервалу оцінки $[0,5]$ або $R_{ij} < 6$ для

інтервалу оцінки $[0,10]$, тоді $R_{ij} = 0$; 2) якщо $R_{ij} \geq 3$ для інтервалу $[0,5]$ або $R_{ij} \geq 6$ для інтервалу $[0,10]$, тоді $R_{ij} = 1$. Для множини предметів з оцінкою інтересу 1, які належать вектору профілю активного користувача U_a будується множина асоціативних правил $\{X_{ai} \Rightarrow Y_{aj}\}$, де $X_{ai}, Y_{aj} \in U_a$. Множина Y_{aj} – це множина предметів, які рекомендаційна система рекомендує користувачу. Із цієї множини користувачу пропонується Top N предметів з найвищою рейтинговою оцінкою. Розроблено модифікацію алгоритму Apriori, яка дозволяє керувати підтримкою (support) і достовірністю (confidence) асоціативних правил з метою зменшення об'єму обчислень і часу пошуку асоціативних правил. Попередньо задається мінімальне значення support, бажане значення confidence і діапазон кількості асоціативних правил. На відміну від класичного алгоритму пошуку асоціативних правил у запропонованій модифікації методу для кожного предмету, для якого $R_{ij} = 1$, шукаються асоціативні правила $i_j \Rightarrow Y$, де Y – множина предметів, до яких виявили інтерес N користувачів з околу активного користувача U_a , $i_j \in I_N$. Множина I_N – це множина предметів, які увійшли у профілі N користувачів з околу активного користувача. Профілі в околі активного користувача шукаються за допомогою методу k -найближчих. Подібним чином працює алгоритм прогнозування рекомендацій для вирішення задачі супутніх продаж. При цьому за допомогою асоціативних правил прогнозуються предмети у множині Y , де множина X – це предмети, які мають високу популярність серед користувачів.

Розроблено модель і метод збільшення диверсифікації товарів і послуг, які надає інтернет-магазин. Стосовно інтернет-магазину диверсифікація означає розширення асортименту товарів і послуг, які інтернет-магазин пропонує потенційному покупцеві. Рекомендаційна система пропонує покупцеві top-N предметів, товарів або послуг. Однак алгоритм прогнозування рекомендацій передбачає прогноз предметів з подібними характеристиками або великою популярністю. За рахунок великої розрідженості матриці користувач-предмет рекомендовані предмети можуть мати майже однакові характеристики і відрізнятися тільки ціною. Окрім того при функціонуванні будь-якого суб'єкта комерції взагалі і електронної комерції у тому числі виникає проблема «довгого хвоста». Проблема «довгого хвоста» означає, що рекомендаційна система пропонує переважно предмети з високою популярністю. У середньому це складає 20% від усіх предметів, які може запропонувати інтернет-магазин. При цьому про 80% предметів у покупця відсутня інформація. Диверсифікація у значній мірі допомагає вирішити ці проблеми і збільшити прибуток інтернет-магазину.

Для рішення задачі підвищення диверсифікації в дисертаційній роботі запропоновано використовувати такий вираз:

$$FDIV(i, U, R) = \alpha F_1(i) + (1 - \alpha) F_2(i, R) * (\beta F_3(i, U) + (1 - \beta) F_4(i, U)). \quad (20)$$

де i – предмет, для якого оцінюється диверсифікація; U – профіль активного користувача; R – множина прогнозованих предметів; N – множина предметів, для яких оцінюється диверсифікація; $\alpha, \beta \in (0,1]$ – керуючі параметри; $F_1(i)$ – оцінка пріоритетності прогнозованого рейтингу для i -го предмета; $F_2(i, R)$ – оцінка диверсифікації i -го предмета відносно множини R ; $F_3(i, U)$ – оцінка переваги i -го предмета для активного користувача; $F_4(i, U)$ – оцінка новизни предмета i по відношенню до множини предметів, які належать профілю користувача U .

Нехай R множина прогнозованих предметів методом колаборативної фільтрації. Ця множина ітеративно збільшується за рахунок предметів із множини N , для котрих функція диверсифікації приймає максимальне значення.

У **четвертому розділі** дисертаційної роботи розроблено інформаційне забезпечення для тестування моделей і методів прогнозування рекомендацій для інтернет-магазину, розроблено структуру математичного та програмного забезпечення, яка дозволяє вибрати метод прогнозування рекомендацій, метод пошуку груп користувачів, метод прогнозування рекомендацій у групі користувачів та метод прогнозування рекомендацій для формування додаткових продаж, метод розрахунку точності прогнозування, величину поділу тестової матриці користувач-предмет на прогнозовану і тестову.

Контроль точності прогнозування рекомендацій здійснюється за допомогою методу середньої абсолютної похибки (MAE), середньоквадратичної похибки (RMSE) і нормованого значення середньої абсолютної похибки (NMAE).

Для тестування розроблених моделей і методів використовується тестова матриця користувач-предмет MovLens, яка містить 1000209 числових рейтингових оцінок, які виставлені 6040 користувачами для 3900 фільмів. Номінальні значення рейтингових оцінок $R_{ij} \in [0,5]$. Розмірність матриці користувач-предмет складає 23556000 елементів. Коефіцієнт розрідженості, який розрахований за формулою (15) складає 0,9567. Для проведення експериментів для гібридного методу пошук груп користувачів розрідженість матриці зменшувалась за допомогою зміни кількості рейтингових оцінок у матриці користувач-предмет для дослідження кожного методу (методу мішаної кластеризації, двоетапного методу категоріальної і числової кластеризації та методу кластеризації числових профілів користувачів) методом прогнозування рекомендацій користувач-користувач. Тестову матрицю користувач-предмет організовано у вигляді реляційної бази даних і включає таблиці: користувач, рейтинг, фільм, вік користувача, стать користувача, рід занять користувача, zip-код користувача, назва фільму та жанр фільму.

На рис. 5 представлено результати тестування удосконаленого методу розрахунку коефіцієнтів подібності за формулами (6) і (7). Тестування запропонованого методу проводилось на тестовому наборі MovLens. По осі абсцис відкладено розмір області k -найближчих, у межах якої здійснюється прогнозування рекомендацій. По осі ординат відкладено середню абсолютну похибку прогнозування (MAE).

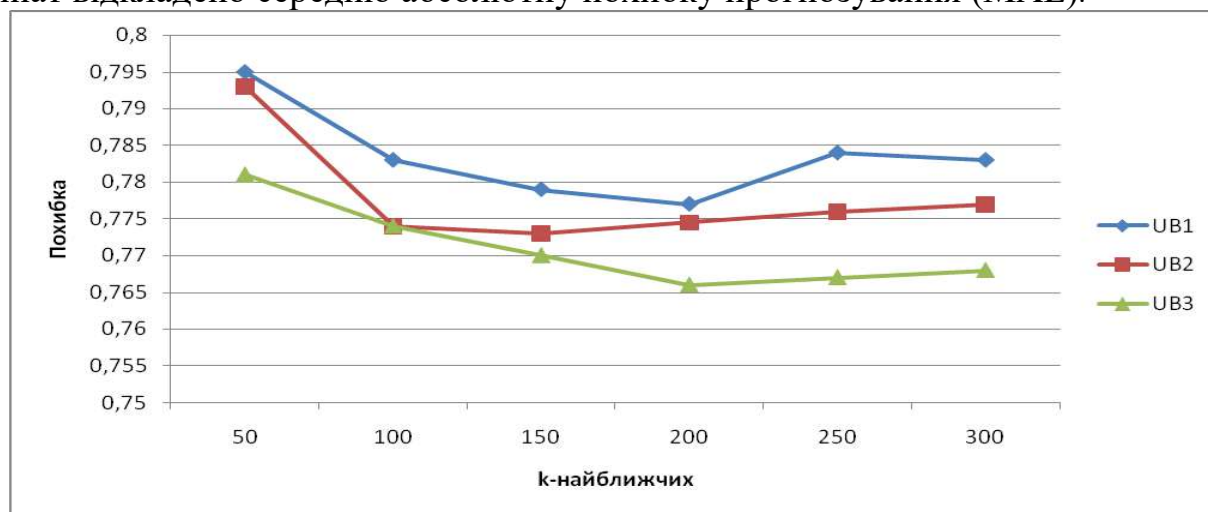


Рис. 5. Залежність похибки прогнозування рекомендацій для тестового набору MovLens в залежності від розміру області k -найближчих:

UB1 – міра подібності на основі оберненої евклідової відстані (5); UB2 – модифікована міра подібності, домінуюче значення демографічної подібності (11); UB3 – модифікована міра подібності, домінуюче значення рейтингової подібності (10)

На рис. 6 представлено результати тестування гібридного методу кластеризації для пошуку груп подібних користувачів. Тестування проводилось для тестового набору Movielens для кількості кластерів $|G|=8$. Коефіцієнт розрідженості розраховувався за формулою (15) і дорівнював 0,9576. Зміна коефіцієнту розрідженості відбувалася шляхом поетапної зміни кількості рейтингових оцінок у матриці користувач-предмет набору даних Movielens. По осі ординат відкладено середню абсолютну похибку прогнозування (MAE). Наведені тестові результати підтверджують ефективність розроблених моделей і методів.

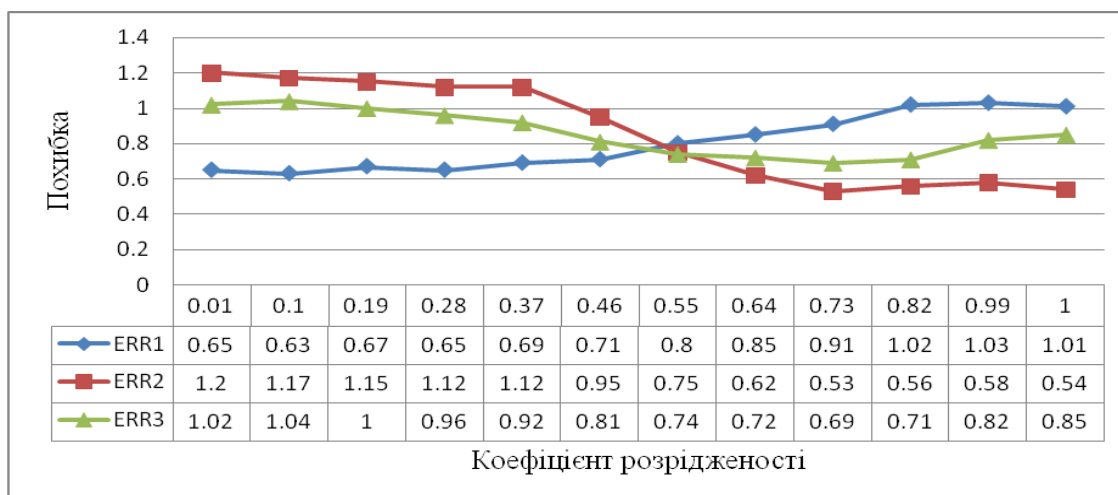


Рис. 6. Залежність похибки прогнозування рекомендацій для тестового набору Movielens у залежності від коефіцієнта розрідженості матриці користувач-предмет:

ERR1 – залежність похибки прогнозування від розрідженості матриці користувач-предмет для методу кластеризації числових профілів користувачів; ERR2 – залежність похибки прогнозування від розрідженості матриці користувач-предмет для методу двохетапної категоріально-числової кластеризації профілів користувачів; ERR3 – залежність похибки прогнозування від розрідженості матриці користувач-предмет для методу мішаної категоріально-числової кластеризації профілів користувачів.

ОСНОВНІ РЕЗУЛЬТАТИ ТА ВИСНОВКИ

У дисертаційній роботі розв'язано актуальне науково-прикладне завдання, яке полягає у розробленні моделей і методів для прогнозування рекомендацій інтернет-магазином для кінцевого користувача. При цьому отримано такі результати:

1. Виконано аналіз актуального стану моделей, методів, засобів і алгоритмів побудови рекомендаційних систем. Це дало можливість визначити переваги і недоліки сучасного математичного забезпечення рекомендаційних систем.

2. Розроблено формальну теоретико-множинну модель колаборативної рекомендаційної системи. Виведено цільову функцію, яку необхідно мінімізувати в процесі роботи колаборативної рекомендаційної системи.

3. Розроблено метод розрахунку коефіцієнтів подібності векторів характеристик користувачів і предметів, який, на відміну від існуючих, враховує демографічні і контентні характеристики користувачів і предметів та дозволяє розраховувати коефіцієнти подібності для нового користувача. Проведені експериментальні дослід-

ження, у порівнянні з існуючими методами розрахунку коефіцієнтів подібності на тестовій базі даних MovLens, показали збільшення точності прогнозування на 4%.

4. Розроблено метод прогнозування рекомендацій для користувачів інтернет-магазину, який використовує алгоритм пошуку асоціативних правил Apriori. Метод враховує рівень інтересу кожного користувача до відповідного предмета. Отримані рішення підтверджуються апробацією на тестовій базі даних MovLens.

5. Вперше на основі концепції застосування в одному методі категоріальної, мішаної і числової кластеризації розроблено метод пошуку груп користувачів, який адаптується до розрідженості матриці користувач-предмет. Отримані рішення підтверджуються апробацією на тестовій базі даних MovLens.

6. Отримав подальший розвиток метод кластеризації мішаних категоріально-числових даних, який, на відміну від існуючих, дозволяє автоматично визначати центри кластерів і використовується при виділенні груп користувачів за мішаними рейтинго-демографічними векторами профілів користувачів. Проведені експериментальні дослідження розробленого методу на тестовій базі даних MovLens показали збільшення точності прогнозування рекомендацій на 5% у порівнянні з застосуванням методів чіткої числової кластеризації.

7. За допомогою поділу тестової матриці користувач-предмет MovLens на тестову і прогнозовану частини проведено тестування точності методів і алгоритмів прогнозування рекомендацій

8. Розроблено програмне забезпечення, яке включає методи прогнозування рекомендацій, методи розрахунку подібності векторів профілів користувачів і предметів та методи тестування точності.

9. Результати дисертаційної роботи впроваджено в навчальний процес.

СПИСОК ОСНОВНИХ ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Лобур М.В., Шварц М.Є., Стех Ю.В. Моделі і методи прогнозування рекомендацій для колаборативних рекомендаційних систем // Вісник Національного університету «Львівська політехніка»: Інформаційні системи та мережі. Львів, 2018. № 901. С. 68-75.
2. Stekh Y., Lobur M., Shvarts M. Some methods for improving the accuracy of prediction recommendations // Вісник Національного університету «Львівська політехніка»: Комп'ютерні системи проектування. Теорія і практика. Львів, 2017. № 882. С. 46-49.
3. Лобур М.В., Стех Ю.В., Шварц М.Є. Метод і алгоритм прогнозування рекомендацій для спільнот користувачів // Квалілогія книги: Збірник наукових праць Української академії друкарства. Львів, 2017. № 1 (31). С. 88-93.
4. Лобур М.В., Стех Ю.В., Шварц М.Є. Побудова асоціативних правил для прогнозування рекомендацій в колаборативних рекомендаційних системах // Квалілогія книги: Збірник наукових праць Української академії друкарства. Львів, 2017. № 2 (32). С. 82-86.
5. Lobur M., Shvarts M., Stekh Y. Application of recommender systems in the design of complex microsystem devices // International Journal of Advanced Research in Computer Engineering & Technology. 2018. V. 7, № 9. P. 709-714.
6. Shvarts M., Lobur M., Stekh Y. Analysis of the effectiveness of similarity measures for recommendations systems // The Experience of Design and Application of CAD

- Systems in Microelectronics: Proc. of the 14th International Conference, Polyana-Svalyava (Zakarpattya), 21-25 February 2017. Lviv, 2017. P. 275-277.
7. Shvarts M., Lobur M., Stekh Y. Some trends in modern recommender systems // Perspective Technologies and Methods in MEMS Design: Proc. of the 13th International Conference, Polyana-Svalyava (Zakarpattya), 20-23 April 2017. Lviv, 2017. P. 167-169.
 8. Shvarts M., Lobur M., Stekh Y., Demkiv I. Some methods for predicting recommendations for MEMS designer communities // Perspective Technologies and Methods in MEMS Design: Proc. of the 14th International Conference, Polyana-Svalyava (Zakarpattya), 18-22 April 2018. Lviv, 2018. P. 196-199.
 9. Шварц М.Є., Стех Ю.В. Моделі і методи побудови рекомендаційних систем // Проблеми та перспективи розвитку економіки і підприємництва та комп'ютерних технологій в Україні: зб. тез XIII науково-практична конференції, м.Львів, 2017. Львів, 2017. С. 37-38.
 10. Лобур М., Стех Ю., Шварц М. Метод прогнозування рекомендацій з врахуванням інтересу спільноти користувачів // Комп'ютерне моделювання та програмне забезпечення інформаційних систем і технологій: зб. тез Третьої Всеукр. наук.-практ. конф., м.Рівне, 29-30 вересня 2017. Рівне, 2017. С. 135-137.
 11. Лобур М., Стех Ю., Шварц М. Використання демографічних характеристик користувачів при прогнозуванні рекомендацій // Комп'ютерне моделювання та програмне забезпечення інформаційних систем і технологій: зб. тез Третьої Всеукр. наук.-практ. конф., м.Рівне, 29-30 вересня 2017. Рівне, 2017. С. 138-139.
 12. Lobur M., Shvarts M., Stekh Y. The method of sequential clustering for predicting recommendations // CAD in Machinery Design-Implementation and Education Problems: Proc. of the XXV Polish-Ukrainian Conference: Bielsko-Biala, October 20-21. Bielsko-Biala, 2017. P. 19-20.
 13. Lobur M., Shvarts M., Stekh Y. The method and algorithm for increasing diversity in recommendation systems // CAD in Machinery Design-Implementation and Education Problems Issues: Proc. of the XXVI International Ukrainian-Polish Scientific and Technical Conference. Lviv, 2018. P. 110-114.
 14. Kosobutsky P., Kovalchuk A., Kuzmynykh M., Shvarts M. Geometric calculation of Pi using the Monte Carlo method // Perspective Technologies and Methods in MEMS Design: Proc. of the 12th International Conference, Polyana-Svalyava (Zakarpattya), 20-24 April 2016. Lviv, 2016. P. 167-169.

АНОТАЦІЇ

Шварц М.Є. Гібридні моделі і методи прогнозування рекомендацій для інтернет-магазину. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем. – Національний університет «Львівська політехніка» МОН України, Львів, 2019.

У дисертаційній роботі розв'язано науково-прикладне завдання розроблення гібридних моделей і методів прогнозування рекомендацій для інтернет-магазину. Удосконалено метод розрахунку коефіцієнтів подібності векторів для методу

зваженої суми, який на відміну від існуючих, використовує демографічні характеристики користувачів, що дозволяє підвищити точність прогнозування рекомендацій. Вперше на основі концепції застосування в одному методі категоріальної, мішаної і числової кластеризації розроблено метод пошуку груп користувачів, який адаптується до розрідженості матриці користувач-предмет. Отримав подальший розвиток метод мішаної кластеризації, який використовується для кластеризації категоріально-числових векторів профілів користувачів і, на відміну від існуючих, дає можливість автоматично вибирати центри кластерів і дозволяє зменшити час пошуку груп користувачів при високій точності виділення груп. Отримав подальший розвиток метод збільшення різноманітності рекомендованих предметів, який дозволяє врахувати оцінки подібних предметів в околі предметів активного користувача і вирішує проблему "довгого хвоста". Удосконалено метод надання рекомендацій користувачу рекомендаційної системи, який, на відміну від існуючих, базується на концепції асоціативних правил, враховує інтереси існуючих у системі користувачів, дає можливість надавати рекомендації новому користувачу, пропонувати користувачу супутні предмети. Для пошуку асоціативних правил метод використовує адаптивну стратегію ітеративного збільшення підтримки асоціативних правил. Проведено експериментальні дослідження розроблених моделей і методів.

Ключові слова: інтернет-магазин, прогнозування рекомендацій, кластеризація, демографічно-рейтинговий профіль користувача, рекомендаційна система.

Шварц М.Е. Гибридные модели и методы прогнозирования рекомендаций для интернет-магазина. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 01.05.03 - математическое и программное обеспечение вычислительных машин и систем. – Национальный университет «Львівська політехніка» МОН України, Львов, 2019.

В диссертационной работе решено научно-прикладное задание разработки гибридных моделей и методов прогнозирования рекомендаций для интернет-магазина. Усовершенствован метод расчета коэффициентов сходства векторов для метода взвешенной суммы, который в отличие от существующих, использует демографические характеристики пользователей и позволяет повысить точность прогнозирования рекомендаций. Впервые на основе концепции применения в одном методе категориальной, смешанной и числовой кластеризации разработан метод поиска групп пользователей, который адаптируется к разреженности матрицы пользователь-предмет. Получил дальнейшее развитие метод смешанной кластеризации, который позволяет использовать для кластеризации категориально-числовые векторы профилей пользователей и, в отличие от существующих, автоматически выбирать центры кластеров и позволяет уменьшить время поиска групп пользователей при высокой точности выделения групп. Получил дальнейшее развитие метод увеличения разнообразия рекомендованных предметов, который позволяет учесть оценки подобных предметов в окрестности предметов активного пользователя и решает проблему "длинного хвоста". Усовершенствован метод предоставления рекомендаций пользователю рекомендательной системы, который, в

отличие от существующих, базируется на концепции ассоциативных правил, учитывает интересы существующих в системе пользователей, позволяет предоставлять рекомендации новому пользователю и предлагать пользователю сопутствующие предметы. Для поиска ассоциативных правил метод использует адаптивную стратегию итеративного увеличения поддержки ассоциативных правил. Проведены экспериментальные исследования разработанных моделей и методов.

Ключевые слова: интернет-магазин, прогнозирование рекомендаций, кластеризация, демографическо-рейтинговый профиль пользователя, рекомендательная система.

Shvarts M.E. Hybrid models and methods for predicting recommendations for online store. – On the rights of the manuscript.

Dissertation for the degree of candidate of technical sciences in specialty 01.05.03 – Mathematical and software of computers and systems. – Lviv Polytechnic National University, Ministry of Education and Science of Ukraine, Lviv, 2019.

In the dissertation, the scientific and practical task of developing hybrid models and methods for predicting recommendations for users of an online store has been solved. The method for calculating vector similarity coefficients for the weighted sum method, which, unlike the existing ones, uses the demographic characteristics of users, improves the accuracy of forecasting recommendations, has been improved. For the first time, based on the concept of using categorical, mixed, and numerical clustering in a single method, a method of searching for user groups has been developed that adapts to the sparseness of the user-subject matrix. The mixed clustering method was further developed, which allows using categorical-numerical user profile vectors for clustering and, unlike existing ones, automatically select cluster centers and reduces the search time for user groups with high accuracy of group allocation. The method of increasing the variety of recommended items, which allows to take into account the estimates of such items in the vicinity of the items of the active user and solves the problem of the “long tail”, was further developed. The method of providing recommendations to the user of the recommendation system has been improved, which, unlike the existing ones, is based on the concept of association rules, takes into account the interests of the users existing in the system, allows to give recommendations to the new user, to offer the user related items. To search for association rules, the method uses an adaptive strategy of iterative increase in support for association rules. Experimental studies of the developed models and methods were carried out. A model and method for increasing the diversification of goods and services provided by an online store have been developed. With respect to online shopping, diversification means expanding the range of products and services that an online store offers to a potential buyer. The referral system offers the buyer top-N items, goods or services. However, the recommendation prediction algorithm predicts items with similar characteristics or high popularity. Diversification greatly helps to solve these problems and increase the profits of the online store.

Keywords: online store, predicting recommendations, clustering, demographic-rating profile of user, recommender system.

Підписано до друку 20.09.2019 р.
Формат 60×84/16. Папір офсетний.
Друк цифровий. Умовн. друк. арк. 0,9.
Наклад 100 прим. Зам. № 170

ТзОВ «Растр-7»
79005, м. Львів, вул. Кн.Романа, 9/1
тел./факс: (032) 235-52-05, e-mail: rastr.sim@gmail.com
Свідоцтво суб'єкта видавничої справи:
ЛВ №22 від 19.11.2002 р.