МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ
ФАКУЛЬТЕТ КОМП'ЮТЕРНО-ІНФОРМАЦІЙНИХ СИСТЕМ
І ПРОГРАМНОЇ ІНЖЕНЕРІЇ

ДАВИДОВ АНДРІЙ ОЛЕКСАНДРОВИЧ

УДК 004.4

МЕТОДИ І ЗАСОБИ КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЇ ПРИ ПОБУДОВІ РЕКОМЕНДАЦІЙНИХ СЕРВІСІВ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

123 «Комп'ютерна інженерія»

Автореферат
дипломної роботи на здобуття освітнього ступеня «магістр»

Тернопіль 2018

Роботу виконано на кафедрі комп'ютерних систем та мереж Тернопільського національного технічного університету імені Івана Пулюя Міністерства освіти і науки України

Керівник роботи: кандидат технічних наук, доцент кафедри комп'ютерних систем та мереж
Яцишин Василь Володимирович,
Тернопільський національний технічний університет імені Івана Пулюя,

Рецензент: кандидат технічних наук, доцент кафедри автоматизації технологічних процесів та виробництв
Коноваленко Ігор Володимирович,
Тернопільський національний технічний університет імені Івана Пулюя

Захист відбудеться 26 грудня 2018 р. о $9^{.00}$ годині на засіданні екзаменаційної комісії №34 у Тернопільському національному технічному університеті імені Івана Пулюя за адресою: 46001, м. Тернопіль, вул. Руська, 56, навчальний корпус №1, ауд. 603

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми роботи. Сучасні технології проектування інформаційних систем, розробки програмних та програмно-технічних засобів характеризується необхідністю опрацювання великої кількості інформації, що призвело до стрімкого розвитку таких галузей як Big Data, Data Mining, Text Mining, прикладних систем штучного інтелекту. Широкої популярності серед таких систем набувають рекомендаційні системи, оскільки дають змогу значно розширити сферу діяльності бізнесу шляхом стимулювання продаж, формування адекватних пропозицій для кінцевого користувача та підвищення ефективності роботи підприємства в цілому. Для проектування рекомендаційних систем, систем машинного навчання та інших «smart» систем важливим є вибір та побудова математичних методів і засобів, що є фундаментальним аспектом ефективності алгоритмів опрацювання даних та одержання точних і достовірних результатів.

Дослідженню методів побудови рекомендаційних сервісів присвячено ряд наукових праць як вітчизняних, так і закордонних вчених. Зокрема серед українських вчених та вчених СНД варто виділити праці Плескач В.Л., Воронцова К.В., Коршунова А.В., Шумейко А.А. та ряду інших. Серед закордонних вчених, які зробили значний внесок у розвиток прогнозування рекомендацій варто виділити праці G. Karypis, Resnick P., J. Ben Shafer та ін.

Дослідження і практичну реалізацію рекомендаційних сервісів для задач електронної комерції проводять такі компанії як Yahoo!, Google, Yandex, Amazon. Український бізнес також розвиває напрям побудови рекомендаційних систем, зокрема, пошуковик Мета, електронна площадка Rozetka, Prom, OLX.

Хоча в наукових працях досягнуто значних результатів щодо побудови рекомендаційних сервісів, однак значення прогнозу все ж потребують оптимізації.

Побудова рекомендаційних систем є актуальною задачею практично для усіх сфер діяльності. При їх проектуванні характерним є різна природа і походження даних. При цьому дані є слабоструктурованими або апріорі невідомими, присутні дефекти, що призводить до опрацювання недостовірної інформації та як наслідок недостовірних і не точних результатів.

Тому актуальною задачею у сфері розробки рекомендаційних систем для електронної комерції є побудова ефективних алгоритмів для формування пропозицій кінцевим користувачам та відповідного розробка засобів їх підтримки

Метою роботи є дослідження та обґрунтування методів і засобів колаборативної фільтрації для стимулювання крос-продаж в системах електронної комерції.

Для досягнення вказаної мети в роботі поставлено наступні **задачі**:

– аналіз наукових публікацій та стандартів в галузі побудови рекомендаційних систем;

– обґрунтування та формалізація метрик подібності об’єктів задач формування рекомендацій у системах електронної комерції ;

– обґрунтування методів колаборативної фільтрації та вибір оптимального при побудові рекомендаційних систем електронної комерції;

– побудова імітаційної моделі та розробка засобу формування

рекомендацій на основі колаборативної фільтрації для систем електронної комерції;

– апробація одержаних результатів щодо застосування в системах електронної комерції.

Об'єкт дослідження: процеси формування рекомендацій для систем електронної комерції.

Предмет дослідження: моделі, методи і засоби колаборативної фільтрації при побудові рекомендаційних систем.

Методи дослідження. Для вирішення поставлених задач використано наступні методи: аналіз та узагальнення – при проведенні аналізу існуючих моделей і методів побудови рекомендаційних систем; формалізації – при дослідженні метрик подібності об'єктів та розробці методу формування рекомендацій в системах електронної комерції; проектування та програмування – при розробці програмного засобу побудови рекомендацій на основі колаборативної фільтрації; експеримент та вимірювання – для апробації розробленого методу та засобу.

Наукова новизна одержаних результатів:

– уперше запропоновано метод формування рекомендацій із застосуванням метрик подібності та регуляризаторів, що дало змогу підвищити точність надання рекомендацій та враховувати розрідженість даних у вхідній матриці колаборативної фільтрації.

– набула подальшого розвитку технологія формування рекомендацій на основі методів колаборативної фільтрації, що дає змогу інтегрувати їх у системи електронної комерції.

Практичне значення одержаних результатів. Впровадження розробленого методу і побудованої імітаційної моделі на основі GraphLab Create дає змогу реалізувати рекомендаційні сервіси із виконанням як в хмарному сервісі EC2, так і при реалізації звичайних електронних магазинів не розміщених в cloud.

Апробація. Результати дослідження апробовано на міжнародній науково-технічній конференції «Фундаментальні та прикладні проблеми сучасних технологій» до 100 річчя з дня заснування НАН України та на вшанування пам'яті Івана Пулюя (100 річчя з дня смерті) 22-24 травня 2018 р. та VII міжнародній науково - технічній конференції молодих учених і студентів «Актуальні задачі сучасних технологій» (28-29 листопада 2018 р.) Тернопільського національного технічного університету імені Івана Пулюя у вигляді тез конференцій.

Структура роботи. Робота складається з розрахунково-пояснювальної записки та графічної частини. Розрахунково-пояснювальна записка складається з вступу, 6 розділів, висновків, списку використаної літератури та додатків. Обсяг роботи: розрахунково-пояснювальна записка – 137 арк. формату А4, графічна частина – 10 аркушів формату А1.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність дослідження методів і засобів колаборативної фільтрації при побудові рекомендаційних сервісів в системах електронної комерції, визначено мету і задачі дипломної роботи магістра, наведено наукову новизну і практичну цінність одержаних результатів.

У першому розділі дипломної роботи проведено аналіз літературних джерел та наукових публікацій в області побудови рекомендаційних сервісів та систем електронної комерції, проаналізовано особливості і класифікацію систем електронної комерції, визначено концептуальні моделі і застосування такого класу програмних систем та самої електронної комерції. Проведено аналіз платформ електронної комерції на предмет підтримки та реалізації рекомендаційних сервісів і встановлено, що у більшості з них існує можливість підтримки таких сервісів, однак вони є не гнучкими і складно адаптовуваними під конкретну реалізацію електронного магазину. Досліджено методи колаборативної фільтрації для побудови рекомендаційних сервісів і встановлено, що найбільш ефективними з точки зору практики є використання memory-based моделей із user-based і item-based підходів.

У другому розділі «Методи колаборативної фільтрації для реалізації рекомендаційних сервісів систем електронної комерції» формалізовано об'єкти предметної області для розв'язку задачі побудови рекомендаційних сервісів із застосуванням елементів теорії множин та математичної статистики, що дало змогу забезпечити необхідний рівень абстракції при реалізації алгоритмів колаборативної фільтрації незалежно від предметної області. Обґрунтовано застосування метрик подібності для визначення схожих товарів і користувачів, що дало змогу формувати персоніфіковані рекомендації із заданим рівнем точності. Запропоновано метод формування рекомендацій із застосуванням метрик подібності та регуляризаторів, що дало змогу підвищити точність надання рекомендацій та враховувати розрідженість даних у вхідній матриці колаборативної фільтрації.

У третьому розділі «Реалізація рекомендаційного сервісу для систем електронної комерції» проведено аналіз бібліотек і платформ машинного навчання та обґрунтовано використання платформи GraphLab Create як комплексного рішення для розв'язку задач машинного навчання, що дало змогу реалізовувати складні проекти з можливістю візуалізації та аналізу результатів моделювання при написанні мінімум коду і високим рівнем абстракції. Реалізовано імітаційну модель для формування рекомендацій товарів в електронних магазинах із застосуванням метрик подібності з регуляризаторами та без них, що дало змогу підтвердити ефективність застосування запропонованого методу (рекомендаційні систем з метриками подібності та регуляризаторами).

У четвертому розділі «Обґрунтування економічної ефективності» на основі результатів розрахунків техніко-економічних показників економічно обґрунтовано впровадження розробленого методу та реалізованої імітаційної моделі побудови рекомендацій у системах електронної комерції.

У п'ятому розділі «Охорона праці та безпека в надзвичайних ситуаціях» визначено вимоги з охорони праці та безпеки в надзвичайних ситуаціях щодо використання програмно реалізованої імітаційної моделі рекомендаційного сервісу, а також проаналізовано вимоги до організації робочого місця інженера з побудови рекомендаційних сервісів.

У шостому розділі «Екологія» досліджено питання електромагнітного забруднення довкілля, його вплив на людину, а також статистичну оцінку техногенних впливів.

У загальних висновках до дипломної роботи магістра наведено результати виконання розділів дипломної роботи магістра щодо побудови рекомендаційних сервісів в системах електронної комерції.

У додатках до пояснювальної записки наведено матеріали щодо апробації одержаних у дипломній роботі магістра результатів.

У графічній частині до дипломної роботи магістра наведено основні теоретичні і практичні результати дослідження щодо застосування методів колаборативної фільтрації при побудові рекомендаційних сервісів в системах електронної комерції.

ВИСНОВКИ

Основні наукові та практичні результати роботи полягають у наступному:

Проаналізовано особливості і класифікацію систем електронної комерції, визначено концептуальні моделі і застосування такого класу програмних систем та самої електронної комерції, що дало змогу обґрунтувати актуальність задач побудови рекомендаційних сервісів з подальшою їх інтеграцією в електронних магазинах.

Проведено аналіз платформ електронної комерції на предмет підтримки та реалізації рекомендаційних сервісів і встановлено, що у більшості з них існує можливість підтримки таких сервісів, однак вони є не гнучкими і складно адаптовуваними під конкретну реалізацію електронного магазину.

Досліджено методи колаборативної фільтрації для побудови рекомендаційних сервісів і встановлено, що найбільш ефективними з точки зору практики є використання memory-based моделей із user-based і item-based підходів.

Формалізовано об’єкти предметної області для розв’язку задачі побудови рекомендаційних сервісів із застосуванням елементів теорії множин та математичної статистики, що дало змогу забезпечити необхідний рівень абстракції при реалізації алгоритмів колаборативної фільтрації незалежно від предметної області.

Обґрунтовано застосування метрик подібності для визначення схожих товарів і користувачів, що дало змогу формувати персоніфіковані рекомендації із заданим рівнем точності.

Запропоновано метод формування рекомендацій із застосуванням метрик подібності та регуляризаторів, що дало змогу підвищити точність надання рекомендацій та враховувати розрідженість даних у вхідній матриці колаборативної фільтрації.

Обґрунтовано застосування інструментального засобу Jupyter Notebook, як середовища розробки систем машинного навчання, що дало змогу забезпечити ефективність та зручність виконання проектів шляхом використання web-браузерів і підтримки множини відкритих бібліотек для data science.

Проведено аналіз бібліотек і платформ машинного навчання та обґрунтовано використання платформи GraphLab Create як комплексного рішення для розв’язку задач машинного навчання, що дало змогу реалізовувати складні проекти з можливістю візуалізації та аналізу результатів моделювання при написанні мінімум коду і високим рівнем абстракції.

Реалізовано імітаційну модель для формування рекомендацій товарів в електронних магазинах із застосуванням метрик подібності з регуляризаторами та без них, що дало змогу підтвердити ефективність застосування запропонованого методу (рекомендаційні систем з метриками подібності з регялризаторами).

На основі розрахунків економічної ефективності щодо дослідження методів і засобів колаборативної фільтрації при розробці рекомендаційних сервісів систем електронної комерції обґрунтовано економічну доцільність запропонованих методів та імітаційної моделі.

Проаналізовано вимоги з охорони праці і техніки безпеки при реалізації та використанні рекомендаційних сервісів для систем електронної комерції, що дало змогу знизити шкідливий вплив негативних факторів при використанні комп'ютерної техніки на людину.

Досліджено питання електромагнітного забруднення довкілля, його вплив на людину, а також статистичну оцінку техногенних впливів.

СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ РОБОТИ

1. Давидов А.О. Методи колаборативної фільтрації при побудові систем кроспродаж/ А.О. Давидов, В.В. Яцишин, Д.О. Подолян// – Матеріали міжнародної науково-технічної конференції «Фундаментальні та прикладні проблеми сучасних технологій» до 100 річчя з дня заснування НАН України та на вшанування пам'яті Івана Пулюя (100 річчя з дня смерті) 22-24 травня 2018 р. – Тернопіль, ТНТУ – 2018 – с. 243 – 244

2. Подолян Д.О. Класифікація та препроцесинг текстових даних/ Д.О. Подолян, В.В. Яцишин, А.О. Давидов // – Матеріали VII міжнародної науково - технічної конференції молодих учених і студентів «Актуальні задачі сучасних технологій» (28-29 листопада 2018 р.) Тернопільського національного технічного університету імені Івана Пулюя у вигляді тез конференцій. – Тернопіль, ТНТУ – 2018 – с. 201

АНОТАЦІЯ

Давидов А.О. Методи і засоби колаборативної фільтрації при побудові рекомендаційних сервісів у системах електронної комерції.

Дипломна робота на здобуття освітнього ступеня магістра 123 – Комп'ютерна інженерія. – Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль 2018.

У дипломній роботі магістра проведено аналіз наукових публікацій та практик реалізації проектів щодо створення рекомендаційних систем, обґрунтовано та формалізовано метрики подібності об'єктів для розв'язку задач формування рекомендацій у системах електронної комерції, обґрунтовано методи колаборативної фільтрації та запропоновано метод побудови рекомендацій на основі метрик подібності і регуляризаторів. Крім того, побудовано імітаційну модель та розроблено засіб формування рекомендацій на основі колаборативної фільтрації для систем електронної комерції.

Реалізовано імітаційну модель для формування рекомендацій товарів в електронних магазинах із застосуванням метрик подібності з регуляризаторами та без них, що дало змогу підтвердити ефективність застосування запропонованого методу (рекомендаційні систем з метриками подібності та регуляризаторами).

Ключові слова: МЕТОД, КОЛАБОРАТИВНА ФІЛЬТРАЦІЯ, РЕКОМЕНДАЦІЯ, СЕРВІС, ЕЛЕКТРОННА КОМЕРЦІЯ.

ABSTRACT

Davydov A. O. Methods and tools of collaborative filtration at recommendation services development in e-commerce systems.

The diploma paper for obtaining the Master's degree 123 – Computer engineering – Ternopil Ivan Puluj National Technical University, Ternopil 2018.

The master's thesis analyzes scientific publications and practice of implementing projects on the creation of advisory systems, substantiates and formalizes the metrics of similarity of objects for solving the problems of forming recommendations in e-commerce systems, validates collaborative filtration methods and proposes a method of constructing recommendations based on similarity metrics and regularizers. In addition, an imitation model was constructed and a tool for forming recommendations based on collaborative filtering for e-commerce systems was developed.

An imitation model was implemented to formulate recommendations for goods in electronic stores using metrics of similarity to and without regularizers, which allowed to confirm the efficiency of the application of the proposed method (recommendation systems with similarity metrics and regularizers).

Keywords: METHOD, COLLABORATIVE FILTERING, RECOMMENDATION, SERVICE, E-COMMERCE.