

ВІДГУК РЕЦЕНЗЕНТА

доктора технічних наук, професора

Нечипоренко Аліни Сергіївни

на дисертаційну роботу

Гребенюка Микити Олександровича на тему

„Дослідження та розробка методів рекомендаційної системи з використанням алгоритмів нечіткої логіки та генетичного алгоритму“,

подану на здобуття ступеня доктора філософії

за спеціальністю 122 Комп'ютерні науки, галузь знань 12 Інформаційні технології

1. Актуальність теми дослідження

Дисертаційна робота присвячена актуальній науково-практичній задачі підвищення ефективності рекомендаційних систем в умовах розрідженості даних та проблеми холодного старту.

Стрімке зростання обсягів цифрового контенту, розвиток електронної комерції, медіасервісів та освітніх платформ зумовлюють необхідність створення ефективних методів персоналізованого відбору інформації. Одним із найбільш поширених підходів до побудови рекомендаційних систем є колаборативна фільтрація. Водночас ефективність класичних методів колаборативної фільтрації істотно знижується за умов високої розрідженості матриці оцінок, недостатньої кількості взаємодій користувача із системою та відсутності спільно оцінених об'єктів. Особливо гостро ці обмеження проявляються для нових користувачів у межах так званої проблеми холодного старту.

У дисертаційній роботі для вирішення зазначених проблем запропоновано комплексний підхід, що поєднує компактне гібридне представлення профілю користувача, апарат нечіткої логіки та методи генетичної і коеволюційної оптимізації. Такий підхід дозволяє використовувати не лише історію оцінювання

об'єктів, але й додаткові характеристики користувача та забезпечує адаптивне налаштування параметрів рекомендаційної моделі.

Метою дисертаційної роботи є підвищення якості персоналізованих рекомендацій в умовах розрідженості матриці оцінок шляхом розроблення методів і моделей гібридної рекомендаційної системи на основі нечіткої логіки та коєволюційної оптимізації структурно-параметричних характеристик гібридного профілю користувача.

Сформульована мета відповідає актуальним напрямам розвитку рекомендаційних систем, машинного навчання, нечітких систем та еволюційних обчислень. Тому актуальність обраної теми дисертаційного дослідження не викликає сумнівів.

2. Ступінь обґрунтованості наукових положень і висновків

Дисертація має наступну структуру: вступ, чотири розділи, висновки, перелік джерел посилань, додатки.

У вступі обґрунтовано актуальність теми дисертаційного дослідження, сформульовано мету та завдання, визначено об'єкт, предмет і методи дослідження, а також представлено наукову новизну та практичне значення отриманих результатів. Окремо наведено відомості щодо апробації результатів дослідження та їх висвітлення у наукових публікаціях.

У першому розділі виконано аналіз сучасного стану проблеми побудови персоналізованих рекомендаційних систем. Розглянуто основні підходи до формування рекомендацій, зокрема методи колаборативної та контентної фільтрації, проаналізовано їх переваги та обмеження. Особливу увагу приділено проблемам розрідженості матриці оцінок та холодного старту, що істотно обмежують ефективність класичних методів колаборативної фільтрації. Проаналізовано можливості використання компактних гібридних моделей користувача, нечіткої логіки та еволюційних методів оптимізації для подолання зазначених обмежень. На основі проведеного аналізу обґрунтовано постановку загальної наукової задачі та сформульовано завдання дослідження.

Другий розділ дослідження присвячено розробленню та теоретичному обґрунтуванню методів формування компактного гібридного профілю користувача та визначення подібності між профілями. Удосконалено показник жанрової зацікавленості користувача шляхом введення штрафного коефіцієнта негативного досвіду, що дозволяє враховувати не лише позитивні, але й негативні оцінки об'єктів певного жанру та розрізняти стани байдужості й антипатії користувача до відповідного контенту. Удосконалений показник представлено як Індикатор Цікавості Жанру (GII). Окремий акцент зроблено на вдосконаленні метрики нечіткої відстані між гібридними профілями користувачів шляхом введення нормалізуючого коефіцієнта, що забезпечує масштабну інваріантність та коректну агрегацію різномірних атрибутів. Запропонована метрика дозволяє формувати множину сусідів навіть за відсутності спільно оцінених об'єктів, що є важливим для рекомендаційних систем в умовах розрідженості даних.

Третій розділ присвячено розробленню методів оптимізації параметрів гібридної рекомендаційної моделі з використанням генетичних та коеволюційних алгоритмів. Удосконалено метод генетичної оптимізації вагових коефіцієнтів GA-W шляхом розроблення композитної функції пристосованості з регуляризацією, яка одночасно враховує точність прогнозування та якість ранжування рекомендацій. Окремий акцент зроблено на розробленому коеволюційному генетичному алгоритмі CEGA, який забезпечує одночасну оптимізацію вагових коефіцієнтів ознак та параметрів функцій належності нечітких множин у межах єдиного еволюційного циклу. Для збереження семантичної цілісності нечітких множин у запропонованому алгоритмі передбачено механізми топологічного контролю мутації. Представлено математичне та алгоритмічне обґрунтування запропонованих методів, що дозволяє здійснювати структурно-параметричну адаптацію гібридного профілю до індивідуальних характеристик користувача.

У четвертому розділі представлено експериментальне дослідження розроблених методів на стандартному наборі даних MovieLens 100K. Проведено п'ятикратну крос-валідацію Монте-Карло для порівняння запропонованої нечіткої відстані з класичними мірами подібності, зокрема косинусною мірою та кореляцією Пірсона. Для статистичної верифікації отриманих результатів використано непараметричний критерій знакових рангів Вілкоксона, за результатами якого підтверджено статистично значущу перевагу запропонованої нечіткої відстані над розглянутими класичними мірами подібності. Окремо оцінено результати застосування алгоритмів GA-W та CEGA, проведено порівняльне багатокористувацьке тестування на стратифікованій вибірці та досліджено стійкість запропонованих методів в умовах холодного старту. Отримані експериментальні результати демонструють, що коеволюційний метод забезпечує покращення показників якості ранжування та зберігає стійкість в умовах суттєвого дефіциту даних.

Висновки дисертаційної роботи узгоджуються з поставленими завданнями, послідовно відображають основні теоретичні та експериментальні результати дослідження та підкреслюють їх наукову новизну і практичне значення. Обґрунтованість сформульованих наукових положень і висновків забезпечується використанням відповідного математичного апарату, проведенням експериментальної перевірки розроблених методів, порівнянням з базовими підходами та статистичною верифікацією отриманих результатів.

У підсумку, дисертація є завершеною науковою працею, що оформлена відповідно до вимог «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого постановою Кабінету Міністрів України від 12 січня 2022 р. № 44, та «Вимог до оформлення дисертації», затверджених наказом Міністерства освіти і науки України від 12.01.2017 № 40.

3. Достовірність та наукова новизна

Достовірність отриманих здобувачем Гребенюк М.О. результатів забезпечується коректним застосуванням математичного апарату, використанням загальнодоступного набору даних MovieLens 100K, проведенням серії експериментів, застосуванням перехресної валідації та статистичної верифікації отриманих результатів.

До найбільш вагомих наукових результатів дисертаційної роботи слід віднести такі.

По-перше, удосконалено показник жанрової зацікавленості користувача шляхом введення штрафного коефіцієнта негативного досвіду. На відміну від базового підходу, удосконалений Індикатор Цікавості Жанру (ГИ) враховує не лише позитивний, але й негативний досвід взаємодії користувача з контентом, що дозволяє краще розрізняти байдужість користувача до певного жанру та його негативне ставлення до нього.

По-друге, удосконалено метрику нечіткої відстані між гібридними профілями користувачів. Запропонована модифікація забезпечує масштабну інваріантність та коректніше об'єднання різнорідних характеристик профілю користувача, а також дозволяє формувати множину сусідів навіть за відсутності спільно оцінених об'єктів. Останнє є особливо важливим у контексті розрідженості даних та проблеми холодного старту.

По-третє, удосконалено метод генетичної оптимізації вагових коефіцієнтів гібридної моделі. Запропонована композитна функція пристосованості одночасно враховує точність прогнозування, якість ранжування та регуляризаційну складову.

Найбільш суттєвим елементом наукової новизни можна вважати запропонований автором коеволюційний генетичний алгоритм CEGA, який забезпечує одночасне налаштування вагових коефіцієнтів ознак та параметрів функцій належності нечітких множин у межах єдиного еволюційного циклу.

Важливою складовою алгоритму є використання механізмів топологічного контролю мутації для збереження семантичної цілісності нечітких множин.

Таким чином, представлена робота містить нові та удосконалені наукові результати, що розвивають методологічну основу побудови гібридних рекомендаційних систем на основі нечіткої логіки та еволюційної оптимізації.

4. Повнота викладу в наукових публікаціях, зарахованих за темою дисертації, наукових положень, висновків і рекомендацій, сформульованих у дисертації

Основні результати дисертаційного дослідження викладено у 10 наукових працях, зокрема у 3 статтях у наукових фахових виданнях України категорії «Б», 2 статтях у міжнародних наукових збірниках, що індексуються наукометричними базами Scopus та/або Web of Science, а також у 5 тезах і матеріалах всеукраїнських наукових конференцій.

Тематика опублікованих праць відповідає основним напрямкам дисертаційного дослідження та відображає результати, пов'язані з розробленням і вдосконаленням гібридної моделі користувача, застосуванням нечіткої логіки, оптимізацією параметрів рекомендаційної системи та використанням генетичних алгоритмів.

Таким чином, наведені в дисертації основні наукові положення, результати, висновки та рекомендації достатньою мірою представлені в наукових публікаціях автора.

5. Значущість дослідження для науки і практики та шляхи його використання

Наукове значення дисертаційної роботи Гребенюка М.О. полягає у розвитку методів побудови гібридних рекомендаційних систем, орієнтованих на функціонування в умовах розріджених даних. Запропоновані автором методи поєднують переваги нечіткого представлення інформації з можливостями еволюційної оптимізації параметрів моделі. Особливий науковий інтерес

становить одночасна оптимізація ваг характеристик профілю користувача та параметрів функцій належності, що дозволяє розглядати рекомендаційну модель як адаптивну структурно-параметричну систему.

Практичне значення отриманих результатів полягає в можливості їх використання при розробленні рекомендаційних сервісів для електронної комерції, медіаплатформ та освітніх систем. Важливою перевагою запропонованих методів є можливість роботи за умов обмеженої кількості інформації про взаємодії користувача із системою.

Практична значущість роботи підтверджується впровадженням її результатів у виробничу діяльність компанії EFISCO Sp. z o.o. (Республіка Польща) у складі програмного компонента для побудови персоналізованих рекомендацій. Зокрема, удосконалений Індикатор Цікавості Жанру інтегровано до підсистеми побудови профілю користувача, а алгоритм SEGA використано як компонент адаптивної оптимізації.

Отримані результати можуть бути використані при подальшому створенні та вдосконаленні промислових рекомендаційних систем, особливо в задачах із високою розрідженістю даних та значною часткою нових користувачів.

6. Відсутність (наявність) порушення академічної доброчесності

За результатами ознайомлення з дисертаційною роботою Гребенюка М.О. підстав для висновку про наявність порушень академічної доброчесності не виявлено.

Наукові результати, що виносяться на захист, сформульовані із зазначенням характеру особистого внеску автора та відмежуванням запропонованих результатів від відомих підходів. Використані в роботі наукові результати інших авторів мають відповідні посилання.

Таким чином, за результатами рецензування ознак академічного плагіату, фабрикації або фальсифікації результатів дослідження не встановлено. Представлена до захисту робота відповідає чинним нормам законодавства України у сфері авторських і суміжних прав.

7. Дискусійні положення та зауваження

Позитивно оцінюючи отримані результати, слід зазначити, що дисертаційна робота містить окремі дискусійні положення та аспекти, які потребують додаткового обговорення.

1. Обмежений перелік сучасних алгоритмів для порівняльного аналізу. На окремих базових етапах експериментального дослідження порівняння запропонованої нечіткої метрики здійснюється переважно з класичними memory-based підходами, зокрема косинусною мірою та кореляцією Пірсона. Для більш повного підтвердження конкурентоспроможності запропонованого підходу доцільним було б розширити порівняльний аналіз сучасними model-based рекомендаційними методами, наприклад Singular Value Decomposition Plus Plus (SVD++), Non-negative Matrix Factorization (NMF) або Neural Collaborative Filtering. Водночас слід відзначити, що в методології дисертації автор також використовує SVD, KNN Baseline та SlopeOne як базові алгоритми порівняння.
2. Недостатньо детально представлений аналіз внеску окремих компонентів запропонованої моделі. Доцільним було б провести розширене ablation study, яке дозволило б окремо оцінити вплив модифікованого Індикатору Цікавості Жанру (GII), запропонованої нечіткої метрики, оптимізації вагових коефіцієнтів та адаптації функцій належності на кінцеві показники рекомендаційної системи. Це дозволило б більш чітко визначити внесок кожного компонента у загальне покращення результатів.
3. Потребує додаткового дослідження вплив вибору типу та параметрів функцій належності. Оскільки функції належності є одним із ключових компонентів нечіткої моделі, доцільним було б детальніше проаналізувати, як різні їх типи, кількість і початкова конфігурація впливають на збіжність коеволюційного генетичного алгоритму SEGА та якість рекомендацій.

4. Експериментальна перевірка запропонованих методів проведена переважно на одному наборі даних MovieLens 100K. Використання стандартного набору даних забезпечує відтворюваність та можливість порівняння результатів, однак перевірка на додаткових наборах даних дозволила б краще оцінити здатність запропонованих методів до узагальнення.

5. Доцільною була б додаткова експериментальна перевірка на MovieLens 1M. Враховуючи більший обсяг цього набору та наявність демографічних характеристик користувачів, MovieLens 1M є логічним наступним етапом перевірки запропонованого гібридного підходу. Такий експеримент дозволив би додатково оцінити масштабованість методу, стабільність отриманих результатів та ефективність використання демографічних характеристик зі збільшенням кількості користувачів і взаємодій.

6. Потребує більш детальної оцінки обчислювальна складність запропонованих методів. Коеволюційна оптимізація вагових коефіцієнтів та параметрів функцій належності потенційно може потребувати значних обчислювальних ресурсів. У зв'язку з цим доцільним було б навести детальніший аналіз часу виконання, збіжності алгоритму та обчислювальної складності залежно від кількості користувачів, об'єктів, розміру популяції та кількості поколінь. Особливо важливим такий аналіз є з точки зору перспектив використання запропонованих методів на великих промислових наборах даних.

Зазначені зауваження мають переважно дискусійний та рекомендаційний характер, окреслюють перспективні напрями подальших досліджень і не знижують загальної наукової та практичної цінності дисертаційної роботи.

8. Загальна оцінка роботи, її відповідність встановленим вимогам

Дисертаційна робота Гребенюка М.О. на тему «Дослідження та розробка методів рекомендаційної системи з використанням алгоритмів нечіткої логіки та генетичного алгоритму» є завершеним науковим дослідженням, у якому розв'язано актуальну науково-практичну задачу підвищення якості персоналізованих рекомендацій в умовах розрідженості даних та проблеми

холодного старту.

У роботі отримано нові наукові результати, пов'язані з удосконаленням показника жанрової зацікавленості користувача, модифікацією нечіткої метрики відстані між профілями користувачів, генетичною оптимізацією вагових коефіцієнтів та розробленням коеволюційного генетичного алгоритму SEGA для одночасної оптимізації ваг ознак і параметрів функцій належності.

Структура дисертації є логічною та послідовною, поставлені в роботі завдання вирішені, а сформульовані висновки належним чином відображають отримані результати дослідження. Результати дисертації мають наукову новизну, теоретичне та практичне значення, пройшли експериментальну апробацію та статистичну верифікацію. Практична цінність результатів підтверджується їх впровадженням у діяльність компанії EFISCO Sp. z o.o.

Наведені в рецензії зауваження не впливають на загальну позитивну оцінку дисертаційного дослідження. За своїм змістом, рівнем наукової новизни, обґрунтованістю отриманих результатів, практичною значущістю та повнотою їх оприлюднення дисертаційна робота відповідає вимогам, що висуваються до дисертацій на здобуття ступеня доктора філософії, а її автор Гребенюк М.О. заслуговує на присудження наукового ступеня доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 122 Комп'ютерні науки.

Рецензент:

доктор технічних наук, професор,
професор кафедри комп'ютерного моделювання
та інтелектуальних технологій Харківського
національного університету радіоелектроніки

Аліна НЕЧИПОРЕНКО