

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)

Кафедра _____ Інформаційних управляючих систем _____
(повна назва)

АТЕСТАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти _____ другий (магістерський) _____

Дослідження методів формування рекомендацій в проектах
побудови рекомендаційних систем в соціальних мережах.
(тема)

Виконав:
студент 2 курсу, групи УПГІТм-18-1
_____ Бондар Д.В. _____
(прізвище, ініціали)

Спеціальність _____ 122 Комп'ютерні _____
науки _____
(код і повна назва спеціальності)

Тип програми _____ освітньо-професійна _____
(освітньо-професійна або освітньо-наукова)

Освітня програма _____ Управління проектами _____
_____ в галузі інформаційних технологій _____
(повна назва освітньої програми)

Керівник _____ (проф. Чалий С.Ф.) _____
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____
(підпис)

_____ Петров К.Е. _____
(прізвище, ініціали)

2019р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Інформаційних управляючих систем
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Управління проєктами в галузі інформаційних технологій
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«_____» 20__ р.

ЗАВДАННЯ
НА АТЕСТАЦІЙНУ РОБОТУ

студентові Бондарю Денису Владиславовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів формування рекомендацій в проєктах побудови рекомендаційних систем в соціальних мережах

затверджена наказом по університету від _____ 20__ р. № _____

2. Термін подання студентом роботи до екзаменаційної комісії _____ 20__ р.

3. Вихідні дані до роботи Науково-технічні публікації та інтернет джерела з тематики _____
атестаційної роботи _____

4. Перелік питань, що потрібно опрацювати в роботі Аналіз процесу формування рекомендацій, аналіз методів та підходів аналіз обробки рекомендацій щодо людей, аналіз існуючих підходів до побудови рекомендаційних систем для соціальних мереж, удосконалення методу рекомендації за урахуванням попередньої фільтрації за демографічними атрибутами, Експериментальна перевірка удосконаленого методу рекомендації

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз літератури та інтернет-джерел		
2	Постановка задачі		
3	Обробка матеріалу		
4	Аналіз характеристик рекомендаційних систем		
5	Дослідження методів побудови рекомендацій в соціальних мережах		
6	Удосконалення методу колаборативної фільтрації з урахуванням демографічної фільтрації		
7	Апробація результатів дослідження на прикладі		
8	Підготовка презентації		
9	Перевірка на плагіат		
10	Нормконтроль		
11	Захист		

Дата видачі завдання _____ 20__ р.

Студент _____
(підпис)

Керівник роботи _____ проф. каф. ІУС Чалий С.Ф. _____
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до магістерської атестаційної роботи містить: 81 с., 4 розділи, 28 рис., 4 табл, 32 джерела.

У роботі виконано огляд методів рекомендацій та підходів до рекомендацій в соціальних мережах. Проаналізовані методи рекомендацій в соціальних мережах в режимі онлайн. На підставі проведеного аналізу запропоновано удосконалений метод побудови рекомендацій в соціальних мережах в режимі онлайн шляхом введення додаткового етапу попередньої фільтрації за демографічною ознакою.

В ході дослідження отримані такі результати: визначені існуючі види персоналізацій; визначені існуючі методи персоналізацій в соціальних мережах; визначені існуючі рекомендаційні системи; зроблена класифікація рекомендаційних систем та опис удосконаленого методу; проведена експериментальна перевірка по удосконаленому методу.

СОЦІАЛЬНІ МЕРЕЖІ, ПЕРСОНАЛІЗАЦІЯ В СОЦІАЛЬНИХ МЕРЕЖАХ, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, КОЛЛАБОРАТИВНА ФІЛЬТРАЦІЯ, ГІБРИДНИЙ МЕТОД, ДЕМОГРАФІЧНА ФІЛЬТРАЦІЯ

ABSTRACT

Explanatory Note to master certification work contains 81 pages, 4 sections, 28 pictures, 4 tables, 32 sources.

The paper gives review of the methods of recommendations in online social networks. On the basis of the analysis, an improved method of building recommendations in social networks online was made by introducing an additional stage of pre-filtering on demographic grounds is proposed.

The study obtained the following results: identified existing types of personalization; identified existing methods of personalization in social networks; identify existing recommender systems; made classification of recommender systems and a description of the improved method; experimental validation on an improved method.

SOCIAL NETWORKS, PERSONALIZATION IN SOCIAL NETWORKS, RECOMMENDATION SYSTEMS, COLLABORATIVE FILTERING, HYBRID METHOD, DEMOGRAPHIC FILTERING

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УСЛОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ ТА ТЕРМІНІВ	8
ВСТУП	9
1. АНАЛІЗ ХАРАКТЕРИСТИК РЕКОМЕНДАЦІЙНИХ СИСТЕМ	10
1.1. Структуризація рекомендаційних систем	10
1.1.1. Визначення рекомендаційних систем	10
1.1.2. Цілі рекомендаційних систем	13
1.1.3. Категорії рекомендаційних систем.....	15
1.2. Дослідження методів побудови рекомендацій з урахуванням властивостей користувача та контенту	21
1.2.1. Спрощені підходи.....	26
1.2.2. Демографічна фільтрація.....	27
1.2.3. Колаборативна фільтрація.....	28
1.2.4. Фільтрація на основі контенту.....	30
1.2.5. Гібридний метод.....	31
1.3. Порівняльний аналіз методів побудови рекомендаційних систем в соціальних мережах	32
1.4. Постановка задачі дослідження	34
2. ДОСЛІДЖЕННЯ МЕТОДІВ ПОБУДОВИ РЕКОМЕНДАЦІЙ В СОЦІАЛЬНИХ МЕРЕЖАХ	35
2.1. Методи побудови рекомендацій в соціальних мережах.....	35
2.1.1. Класи колаборативної фільтрації	35
2.1.2. Розробка удосконаленого підходу колаборативної фільтрації з урахуванням демографічної фільтрації	38
2.1.3. Демографічний компонент	41
2.1.4. Компонент інтереси	46
2.1.5. Компонент базової інформації.....	48
2.1.6. Компонент діяльності	49
2.1.7. Компонент відносин.....	50
2.2. Удосконалення методу колаборативної фільтрації в соціальних мережах в режимі онлайн.....	51
2.2.1. Вимоги до рекомендаційної системи	51

2.2.2. Рекомендаційна система на основі колаборативної фільтрації з додаванням демографічної фільтрації.....	53
3. РОЗРОБКА ПРОЕКТУ ПОБУДОВИ РЕКОМЕНДАЦІЙ В СОЦІАЛЬНИХ МЕРЕЖАХ В РЕЖИМІ ОНЛАЙН.....	58
3.1. Розробка проекту побудови рекомендацій в соціальних мережах в режимі онлайн.....	58
3.1.1. Розробка проекту побудови рекомендацій в соціальних мережах в режимі онлайн.....	65
3.2. Детальний опис удосконаленого методу колаборативної фільтрації в режимі онлайн.....	69
3.2.1. Етап попередньої демографічної фільтрації.....	69
3.2.2. Етап знаходження повної східності між користувачами.....	70
3.2.3. Активність користувача.....	72
3.2.4. Соціальна фільтрація.....	73
3.2.5. Результат.....	75
4. СТРУКТУРА РЕКОМЕНДАЦІЙНИХ СИСТЕМ.....	76
4.1. Структура зберігання елементів в рекомендаційній системі.....	76
4.2. Експериментальна перевірка удосконаленого методу.....	77
ВИСНОВКИ.....	81
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ.....	82
ДОДАТОК А.....	86

ПЕРЕЛІК СКОРОЧЕНЬ, УСЛОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ ТА ТЕРМІНІВ

ІС – інформаційна система;

РС – рекомендаційна система

КФ – колаборативна фільтрація;

AUC – Area under ROC curve;

RS – Рекомендаційна система

ROC – Receiver operating characteristic;

VAN – Value Added Networks.

CBF – Фільтрація на основі контенту

ВСТУП

Ідея як онлайн-соціальних мереж, так і рекомендаційних систем була розроблена багатьма дослідниками. Всі існуючі методи рекомендацій пропонують людям продукти або послуги. Вся ідея застосування рекомендаційних систем в соціальних мережах є інноваційною, тому виникло багато нових проблем. Побудова такої структури вимагає інтеграції двох окремих галузей знання: інформатики та соціології. Це робить роботу ще більш цікавою, оскільки дозволяє створювати міждисциплінарні групи, в яких люди з різним досвідом можуть співпрацювати один з одним.

З одного боку, коли ми рекомендуємо одну людину іншому, ми повинні мати глибокі знання про алгоритми, які можуть бути використані і вибрати відповідний. Як правило, обраний метод повинен бути адаптований до конкретних потреб системи. Крім того, це звичайна практика, що творці РС інтегрують багато різних методів рекомендації, тому що деякі методи доповнюють інші. Іншими словами, одна техніка зазвичай справляється з недоліком іншої, і цей зв'язок є двонаправленим.

Проте, потрібні не тільки знання про рекомендаційні системи. З іншого боку, ми повинні знати соціальну поведінку людей, які створюють мережу, і знати їх уподобання. У рекомендаційній системі для онлайн-соціальної мережі повинні враховуватися не тільки потреби і переваги людини, яка очікує якихось пропозицій, але і очікування людини, яка буде запропонована. Цей фактор робить всю проблему ще більш складною, але і більш захоплюючою.

1. АНАЛІЗ ХАРАКТЕРИСТИК РЕКОМЕНДАЦІЙНИХ СИСТЕМ

1.1. Структуризація рекомендаційних систем

Рекомендаційні системи відносяться до області досліджень, що інтенсивно розвивається, оскільки вона допомагає досягти більш високої віддачі від інвестицій в інтернет-магазини і забезпечити більш високу якість послуг. Запропонована в даній пояснювальній записці система не рекомендує типові споживчі товари, продавані в електронній комерції, такі як книги, компакт-диски або інші речі. Проте, основні визначення і класифікація класичних рекомендаційних систем будуть представлені в цьому розділі. Необхідно побудувати загальне уявлення про РС, їх цілі, а також їх переваги і недоліки, щоб зрозуміти всю концепцію РС для онлайн-соціальної мережі.

1.1.1. Визначення рекомендаційних систем

Існує безліч визначень рекомендаційних систем. Один з перших був представлений Резніком і Хелом в 1997 році. Вони стверджують, що "в типовій рекомендаційній системі люди надають рекомендації в якості вхідних даних, які потім система агрегує і направляє відповідним одержувачам".

Ці системи зазвичай визначаються з точки зору їх функціональності як системи або агенти, які пропонують продукти користувачам, які купують продукти на сайтах електронної комерції. Рекомендаційні системи допомагають споживачеві прийняти рішення, що купити[2].

Рекомендаційні системи можна класифікувати за рівнем персоналізації на неперсоналізовані та персоналізовані методи. Перші методи не враховують характеристики і переваги клієнтів, в той час як другі тісно залежать від профілю користувача. Окремі методи, перераховані на рисунку 1.1.



Рис 1.1 Рівні персоналізації в рекомендаційних системах

Прикладом неперсоналізованого методу є рекомендація, яка передбачає, що продукти, які були найкраще оцінені в минулому усіма клієнтами в середньому ("кращий рейтинг") або кількість їх копій, які були продані, є найбільшим ("краща покупка"). Для створення такого роду рекомендацій зазвичай використовуються статистичні методи. Більш того, в іншому варіанті не персоналізованого підходу можуть бути рекомендовані навіть нові предмети, наприклад, книги, опубліковані в минулому місяці. Цей вид рекомендацій залежить від політики електронної комерції і відноситься до методів, де не потрібно великого розрахунку. Головна особливість цих пропозицій полягає в тому, що вони однакові для всіх клієнтів[3]. Як правило, користувачеві легко і зручно знайти один пункт зі списку найпопулярніших.

Це схоже на ситуацію, коли хтось заходить в книжковий магазин і знаходить там полицю з бестселерами.

Проте деякі дослідження стверджують, що рекомендаційні системи-це тільки ті, які виробляють персоналізовані рекомендації. Іншими словами, результатом роботи цих систем є індивідуалізована рекомендація, яка допомагає направляти одного користувача до продуктів або послуги, які задовольняють їх особливі потреби. В результаті вони справляються з інформаційним перевантаженням краще, ніж неперсоніфіковані методи, що дозволяють знаходити і купувати потрібні предмети з великої кількості можливих варіантів.

Персоналізована рекомендація ґрунтується або на демографічній інформації про користувачів, або на аналізі минулого поведінки користувача з метою прогнозування його майбутньої поведінки (колаборативна і контентна фільтрація).

Більш того, персоналізація може бути або постійною, або ефемерною. Постійна персоналізація, заснована на поведінці попередніх користувачів, дозволяє створити унікальний список продуктів для кожного користувача. Вимога, яку повинно бути виконано в цій ситуації, полягає в тому, що клієнти повинні увійти в систему, щоб створити профіль користувача для кожного з них. В постійній персоналізованій рекомендації кожна людина на веб-сайті бачить різні рекомендації, тому що вони напряму залежать від особистих даних користувачів. Рекомендації засновані на інформації, отриманої з відповідей на опитування, історії покупок, рейтингів товарів і т.д. Профіль користувача не потрібен в ефемерною персоналізації. У цьому випадку рекомендації створюються відповідно до поведінки користувачів під час поточного сеансу, їх навігацією і вибором. У цьому методі рекомендації однакові для всіх користувачів[4].

Одне з формальних визначень проблеми рекомендацій було сформульовано що проблема рекомендації полягає в тому, щоб передбачити,

як користувачі будуть оцінювати продукти, які вони ще не бачили. Коли система оцінює рейтинги, вона може рекомендувати користувачеві елементи з найвищим рейтингом. Формулювання проблеми рекомендації можна представити у вигляді:

$$\forall c \in C, s'_c = \arg \max_{s \in S} u(c, s). \quad (1.1)$$

де:

C – Вибірка користувачів

S – Вибірка усіх предметів які можуть бути рекомендовані

s'_c – Продукт, який ще не був помічений користувачем c і має найвищий оціночний рейтинг з усіх товарів, не оцінених користувачем c

$u(c, s)$ – Вимірювання називається функцією корисності, яка обчислює корисність елемента s для користувача c

1.1.2. Цілі рекомендаційних систем

Рекомендаційні системи стали важливою і майже невід'ємною частиною сучасних веб-сайтів, більше того, переважна їх кількість застосовується в електронній комерції. Джефф Безос, генеральний директор Amazon.com, сказав: "Якщо б у мене було 3 мільйони клієнтів в Інтернеті, я повинен був би мати 3 мільйони магазинів в Інтернеті". Чому люди вважають, що Персоналізація та рекомендації є важливою частиною електронної комерції?

Мета цих систем-допомогти потенційним покупцям вибрати підходящий продукт для покупки, щоб їх можна було розглядати як системи підтримки прийняття рішень[5]. З іншого боку, вони служать в якості маркетингової допомоги для магазинів електронної комерції, тому що вони підвищують привабливість пропозиції.

Основні цілі рекомендаційних систем полягають в наступному:

- Справлятися з інформаційним перевантаженням
- Допомогати всім клієнтам (нових, частим і нечастим) приймати рішення про те, які продукти купувати, які новини читати далі , який фільм варто подивитися і т. д.
- Конвертувати спостерігачів в покупців
- Побудувати довіру через співтовариство і підтримувати лояльність клієнтів
- Закликати клієнтів прийти назад
- Для підвищення продажів електронної комерції і перехресних продажів

Перші два пункти показують, чому РС важливі з точки зору споживача. Перш за все, вони є дуже корисним інструментом, що допомагає впоратися з інформаційним перевантаженням. Рекомендаційні системи дозволяють вибрати невелику підмножину елементів з мільйонів продуктів, які, очевидно, відповідають потребам і перевагам користувачів. Хоча майже неможливо точно передбачити потреби користувачів, такий набір пропозицій допомагає обмежити кількість варіантів[6]. Крім того, обмежуючи кількість пропонованих продуктів, такі системи допомагають людям приймати рішення, які товари купувати, які новини читати далі або який фільм варто подивитися, набагато швидше, ніж при звичайному перегляді.

Інші з перерахованих вище пунктів показують, що РС можна розглядати як маркетингові інструменти, тому що вони підвищують продажі електронної

комерції. Як уже згадувалося раніше, ці системи можуть допомогти людям знайти продукти, які вони хочуть мати. В результаті це полегшує перетворення людей, які тільки дивляться на покупців. Коли споживачі купують речі, які рекомендовані системою, додаткові предмети можуть бути запропоновані для того, щоб збільшити перехресний продаж. Це призводить до створення і підтримки лояльності клієнтів, більш того, це спонукає клієнта повертатися в майбутньому. В Інтернеті та електронної комерції, де число конкурентів дуже велике, ця особливість є найважливішою перевагою рекомендаційних систем.

Мета всіх цілей, які були вказані вище, - задовольнити клієнта. Причина проста. Дослідження показують, що зберегти поточного клієнта набагато дешевше, ніж знайти нового. Крім того, незадоволений клієнт має тенденцію скаржитися на продукт або послугу в два рази більшій кількості людей, ніж задоволені клієнти будуть говорити позитивні речі про послугу або продукт.

Крім того, РС повинні бути якомога більш високою ефективністю, щоб збільшити їх ROI (рентабельність інвестицій). Однак ці рекомендації не тільки існувати, але і повинні бути актуальними. Проблема, яка може виникнути, - це занадто велика кількість хибно позитивних рекомендацій, які визначаються як пропозиції, створені для користувачів, хоча вони їм не підходять. На закінчення відзначимо, що цілі рекомендацій можуть бути досягнуті тільки в тому випадку, якщо сформовані пропозиції є релевантними[6].

1.1.3. Категорії рекомендаційних систем

Існує безліч категорій рекомендаційних систем. Їх можна розділити в залежності від того, чи є створена рекомендація персоналізованою чи ні.

Прикладом персоналізованого методу є колаборативна фільтрація, в той час як прикладом неперсоніфікованого методу є статистичний аналіз (див. рис 1.1)



Рис 1.2 Приклад категорій рекомендаційних систем

У деяких дослідженнях виділяються три основні категорії РС, де всі вони є персоналізованими, а саме: колаборативна фільтрація, фільтрація на основі контенту і гібридні методи (рис.1.3). Адомавічюс і Тужилін стверджують, що ці три категорії є найбільш популярними і значущими методами рекомендацій. Однак вони вказують на недоліки цих методів і пропонують можливі поліпшення.

Інше дослідження показує, що РС можуть бути розділені на різні групи за різних критеріїв. Робін Берк виділив п'ять прийомів рекомендації (рис. 2.3) залежно від типу бази та вхідних даних, а також алгоритму, який використовується для створення пропозицій[7].

Вихідні дані - це інформація, якою володіє система до початку процесу рекомендації, в той час як вхідні дані дозволяють створювати рекомендації для конкретного користувача. Вхідні дані надаються користувачами і безпосередньо пов'язані з користувачем, для якого генеруються рекомендації. Вихідні дані є основою, що дозволяє виділити наступні методи рекомендації: колаборативні, засновані на змісті, демографічні, засновані на корисності і, нарешті, засновані на знаннях.



Рис 1.3 Приклад категорій рекомендаційних систем

У колаборативної фільтрації фоновими даними є всі рейтинги товарів, якими володіє магазин, складені усіма клієнтами магазину. Вхідні дані-це інформація про рейтинги товарів в магазині, складених однією людиною, для якого система створює рекомендацію. Нарешті, алгоритм ідентифікує тих користувачів, які схожі на користувача, для якого підготовлено пропозицію, і рекомендує ці продукти, які були високо оцінені визнаними аналогічними користувачами[8].

Інші перераховані вище методи характеризуються аналогічним чином, і ці характеристики представлені в таблиці 1.1

Таблиця 1.1 Характеристики методів рекомендаційних систем

Метод	Довідкові дані	Вхідні дані	Метод
Колаборативний	Рейтинги з $C1$ о предметах в $S2$	Рейтинги з $c3$ о предметах в S	Ідентифікувати користувачів в C схожих на s та екстраполювати їх рейтинги з $s4$

На основі контенту	Особливості елементів в S	Рейтинги користувачів с предметів в S	Створення класифікатора, який відповідає вимогам користувача с рейтингова поведінка і використовувати його на s
Демографічний	Демографічна інформація про компанію C і їх рейтинги предметів в S	Демографічна інформація про с	Визначити користувачів, які є демографічно схожими на с і екстраполювати з їх оцінок s
Утилітарний	Особливості елементів в S	Функція корисності над елементами в St що описує переваги с	Застосувати функцію до елементів і визначити ранг s
Заснований на знаннях	Особливості елементів в S. Знання того, як ці елементи задовольняють потреби користувача	Опис потреб або інтересів користувача с	Визначити відповідність між s та потреб с користувача

Ще одне дослідження, яке варто згадати, - це таксономія рекомендаційних агентів, запропонована Микелем Монтанером, Беатріс Лопес і Луїсом де ла Роза. У своїх дослідженнях автори виділяють два основних підходи до проблеми РС: просторовий і функціональний. Крім того, з функціональної точки зору вони створюють вісім вимірювань, які є основою для подальшої класифікації рекомендаційних агентів. П'ять з них стосуються створення та обслуговування профілів, а три-експлуатації профілів

користувачів. Хоча створення і підтримка профілю є дуже важливими частинами рекомендаційних систем, вони виходять за рамки даної магістерської дисертації. Передбачається, що вихідними даними для методу рекомендацій є відповідний профіль користувача. На рис. 3.3 представлені розміри, що характеризують RS.

Фільтрація інформації, зіставлення елементів профілю користувача і зіставлення профілів користувачів - це три основних виміри використання профілю. Що стосується методів фільтрації інформації, то найбільш важливими методами є демографічна, контент-орієнтована і колаборативна фільтрація. Мета зіставлення профілю користувача і елемента полягає в тому, щоб порівняти уявлення профілю користувача (наприклад, інтереси користувача) і опис елемента і в результаті вибрати елементи, що мають відношення до конкретного клієнта. Приклади таких методів представлені на малюнку 1.3. Останнім виділене вимір-це зіставлення профілів користувачів, яке дозволяє знаходити схожих користувачів або групи користувачів.



Рисунок 1.4 Приклад категорій рекомендаційних систем

Шафер розглядає та аналізує не тільки методи рекомендацій, але і, подібно Берку, вхідні дані, які доставляються цільовим клієнтом (тим, для кого

створюється рекомендація) та іншими клієнтами. Ці дані служать вихідним матеріалом для методу рекомендацій (рис.1.4). В результаті застосування відповідної методики цільовий клієнт отримує пропозицію, який товар купити. Однак результатом застосування рекомендаційного методу є не тільки пропозиції, але і рейтинги і прогнози. Рейтинги зазвичай використовуються, коли кількість клієнтів невелика і користувачі знають один одного. У такому випадку може бути корисно відобразити індивідуальні рейтинги інших клієнтів. Кілька РС забезпечують прогнозування оцінок, які користувач, ймовірно, дав би продукту[9].

Згідно Шафера можна виділити наступні типи рекомендаційних систем: необроблений пошук (так звана “нульова рекомендацією”), вибір вручну, статистичне підсумовування, атрибутивна кореляція, кореляція між елементами (також звана фільтрація на основі контенту) і кореляція між користувачами (також звана колаборативна фільтрація) (рис.1.5)



Рисунок 1.5 Приклад категорій рекомендаційних систем

Існує багато підходів до РС, і вони, перераховані вище, створюють лише невелику частину існуючих класифікацій. В наступній частині перераховуються і коротко описуються найбільш поширені і широко застосовувані методи рекомендації в соціальних мережах.

1.2. Дослідження методів побудови рекомендацій з урахуванням властивостей користувача та контенту

Соціальні мережі (СМ) останнім часом стають все більш важливим елементом інформаційного суспільства. Проблема соціальних мереж дуже широка, і її концепція розробляється в багатьох областях досліджень, таких як корпоративні партнерські мережі, мережі співпраці вчених, мережі кіноакторів і т. д. Величезну різноманітність і безліч можливих областей застосування соціальних мереж обумовлюють те, що вони стали предметом багатьох досліджень. Оскільки відносини з мережі, їх зміст та якість відображають соціальну поведінку індивідів, дослідження їх може бути корисним при кількісній і якісній оцінці людських відносин в епоху інформаційного суспільства. Основна ідея соціальної мережі проста. Це набір акторів, тобто група людей або організацій, які є вузлами мережі, і зв'язків, які пов'язують вузли одним або кількома відносинами. Соціальна мережа вказує на те, як актори пов'язані між собою. Зв'язок між акторами може підтримуватися у відповідності з одним або кількома відносинами більш того, мережа дає єго (фокусною акторам) доступ не тільки до їх змін (людям, які безпосередньо пов'язані з єго), але і до змін їх змін (також званих "друзями моїх друзів")[11].

Вузли соціальної мережі не є незалежними істотами. Деякі характеристики, що описують членів мережі, можуть бути визначені (наприклад, демографічні та процентні дані про людей). Однак жоден з методів СНР не відбирає індивідуумів самостійно. Актори пов'язані відносинами, які характеризуються змістом, спрямованістю і силою. Зміст вказує на ресурс, який обмінюється, наприклад, в комп'ютерному зв'язку (СМС) інформація може розглядатися як ресурс. Напрямок визначає, чи є

ставлення спрямованим або неорієнтованим. Відносини між співробітниками та їх керівником є спрямованими. Перший працює на керівника, і це одне з відносин між співробітниками і босом. Останній виплачує заробітну плату або зарплату співробітникам, і це ще одне спрямоване ставлення між босом і співробітниками. Дружба, як правило, ненаправлена, проте вона може бути незбалансованою. Це означає, що одна людина може визначити дружбу з другою людиною як сильну, в той час як друга людина може стверджувати, що ця дружба слабка. Останній з перерахованих характеристик відносин є їх сила. Існує безліч способів визначити, чи є ці відносини сильними або слабкими, наприклад, вказати частоту, з якою актори спілкуються один з одним, важливість обміну інформацією та обсяг соціального капіталу, що передається від одного об'єкта до іншого.

Кожна соціальна мережа може бути візуалізована як точками і лініями, так і матрицями. Теоретичні основи, що дозволяють будувати ці уявлення соціальних мереж, виведені з теорії графів. Графічний приклад соціальної мережі, яку також називають соціограмою, представлений на малюнку 3.1. Соціограми складаються з вузлів, з'єднаних ребрами. Один вид відносин (симплекс) або кілька типів відносин (мультиплекс) можуть бути представлені набором точок і ліній. Крім того, графіки можуть бути спрямованими або неорієнтованими. Перший з них явно показує, як джерело і мета знаходяться в певному відношенні. В іншому випадку це неорієнтований граф. Графіки також дозволяють представити силу зв'язків між членами. Бінарні графіки показують, чи є зв'язок між акторами існує чи ні. Підписані графіки представляють негативний, позитивний або нульовий зв'язок. Порядкові графіки впорядковують відносини у відповідності з їх силою-найсильніша зв'язок, така сама сильна і т. д. Порядкові графіки можна розглядати як підмножину зважених графіків, які також можуть використовуватися для представлення зв'язків між членами спільноти. У такому разі кожному відношенню, представленому дугою графіка, присвоюється значення, яке

відноситься до сили відносини між людьми. Всі ці особливості графів можуть бути застосовані в теорії соціальних мереж і представлені в графічному вигляді[12].

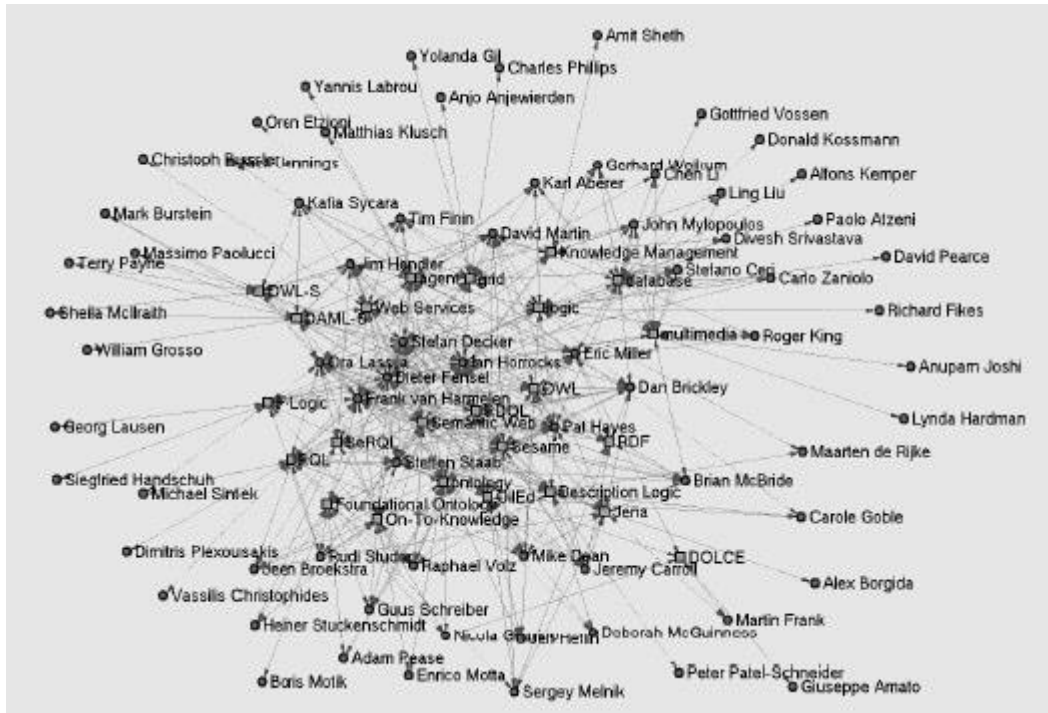


Рис 1.6 Члени семантичної веб спільноти та їх зв'язок з дослідницькими інтересами

Соціальні мережі можна розділити на кілька груп за різними критеріями. Соціальні мережі можуть бути: спеціалізовані (наприклад, знайомства або ділові мережі, мережі друзів, випускників, клуби розваг), непрямими (інтернет–комунікатори, адресні книги, електронна пошта), загальними (наприклад, співавтори наукових робіт, спів організатори заходів), локальними (наприклад, люди, що живуть по сусідству), сімейними, мережами співробітників, мережами гіперпосилання (посилання між домашніми сторінками) і т. д.

Щоб упорядкувати ці різні види мереж, їх можна класифікувати в залежності від типу відносин, які пов'язують двох людей. У цьому випадку

можна виділити ділові та соціальні зв'язки (Рисунок 1.7). Перші містять соціальні мережі, що складаються з людей, які пов'язані один з одним через те, що вони роблять разом, але одночасно вони не розділяють своє приватне життя. Їх можна назвати професійними мережами. Компанія-співробітники створюють соціальну мережу співробітників може бути хорошим прикладом. Крім того, люди, які разом організують, наприклад, конференцію або інший захід, створюють соціальну мережу співорганізаторів[14]. Ці люди пов'язані, тому що вони працюють разом, і їх співпраця зазвичай приносить якийсь результат, наприклад, стаття, конференція, книга і т. д.



Рис 1.7 Поділ соціальних мереж на основі типу відносин

З іншого боку, соціальні відносини вказують на зв'язки з емоційним фоном. Родичі-це група людей, з якими ми є сім'єю; проте, зазвичай люди не знаходяться в контактi з кожним членом своєї сім'ї.

Слід підкреслити, що зв'язок між двома людьми зазвичай являє собою поєднання різних видів відносин, які можуть відрізнятися за силою.

Рекомендаційні системи для соціальних мереж відрізняються від типових видів рекомендаційних рішень тим, що вони пропонують раціональних людей іншим, а не неживі предмети. Після вибору рекомендації одна людина ініціалізує відносини з іншим, і останній може відповісти або позитивно, або негативно на запрошення. Така взаємодія неможлива з продуктами, контентом або послугами. Більш того, зв'язок між людьми

двонаправлена на протипагу відносинам між людиною і об'єктом. Отже, мета такої системи рекомендацій полягає в тому, щоб знайти для поточного користувача іншої людини, який також міг би відреагувати сприятливим чином. Речі не володіють вільною волею і не можуть відмовитися від продажу. Навпаки, одна людина може відмовитися підтримувати відносини з іншою людиною. З цієї причини рекомендаційні системи для соціальних мереж повинні враховувати переваги і людські обмеження людей, які будуть запропоновані, наприклад, найбільша кількість відносин, які може підтримувати один чоловік. На закінчення, рекомендаційна система повинна пропонувати користувачеві тільки тих членів мережі, які потенційно є хорошими друзями або колегами для цього користувача. В результаті в мережі можуть з'явитися нові відносини поточного користувача[18].

З іншого боку, власники соціальної мережі можуть проводити свою власну політику. Отже, їх першою метою може бути створення найширшої соціальної мережі, з максимально можливою кількістю зв'язків між людьми. Ще однією можливою метою було б створення мережі, в якій відносини відображають сильне схожість між людьми і, отже, мережа складається з багатьох близьких гетерогенних груп. Ця еволюція соціальної мережі може бути стимульована різними видами рекомендацій. В обох випадках мета полягає в досягненні спільноти, в якому зв'язки між людьми є постійними. Для досягнення обраної мети власників конкретної соціальної мережі можуть бути адаптовані і використані концепції, відомі як *bonding* і *bridging*.

Загальна мета пропонованої нижче структури полягає в тому, щоб забезпечити можливість коригування рекомендацій з урахуванням профілю конкретного користувача, а також загальної політики мережі.

Класифікація, що лежить в основі подальших описів рекомендаційних систем, представлена на рисунку 1.8



Рис 1.8 Класифікація рекомендаційних систем з урахуванням властивостей користувача та контенту.

Методи рекомендації діляться на інтелектуальну фільтрацію і спрощений підхід. Інтелектуальна фільтрація складається з чотирьох основних категорій: демографічна, контент-орієнтована, колаборативна фільтрація і гібридний метод.

1.2.1. Спрощені підходи

До спрощених підходів відносяться методи статистичного аналізу і нерозрахункового підходу. Обидва вони не вимагають складних обчислень і є неперсоніфікованими методами рекомендації.

Прикладами нерозрахункових методів є системи ручного вибору і загальних анотацій (рис.2.6). Перша з них також називається "людською

рекомендацією" і заснована на пропозиціях експертів. Наприклад, в магазині, в якому продаються фільми, критиків можна вважати експертами. Загальні системи анотацій дозволяють обмінюватися думками між клієнтами. Вони можуть представити свою думку про предмети, які вони придбали, і таким чином допомогти іншим людям вибрати продукти, які є високоякісними[19].

Статистичний аналіз, на відміну від нерозрахункових методів, вимагає розрахунків, які, тим не менш, не дуже складні. Система надає рейтинги продуктів, які засновані на статистичних факторах. Деякі з цих факторів: кількість проданих одиниць кожного з продуктів і середній рейтинг продукту, представлений клієнтами, які вже купили цей продукт. Статистичні дані розраховуються в контексті всієї спільноти.

Перевага полягає в наступному:

Ці методи не вимагають складних обчислень і це дає можливість створювати рекомендації в режимі онлайн

Недоліки полягають в наступному:

Рекомендації однакові для всіх користувачів, і, отже, пропозиції занадто загальні і не персоналізовані. В результаті неможливо надати рекомендації, що відповідають унікальним перевагам деяких користувачів

1.2.2. Демографічна фільтрація

У методі демографічної фільтрації (DF) користувачі діляться на демографічні класи з точки зору їх особистих атрибутів. Ці класи служать вхідними даними для процесу рекомендацій. Мета цього процесу- знайти класи людей, яким подобається певний продукт. Якщо людям з класу С подобається

продукт s і є людина c (цей користувач належить до класу C), який ще не бачив продукт s , то цей продукт можна рекомендувати людині c .

Клієнти надають персональні дані опитувань, які вони заповнюють у процесі реєстрації або можуть бути вилучені з історії покупок.

Перевага полягає в наступному:

Цей метод може не вимагати збору складних даних, таких як історія покупок користувачів і рейтинги

Недоліки полягають в наступному:

Класифікація може бути занадто загальною, і це призводить до втрати індивідуальності користувачів. В результаті рекомендації носять занадто загальний характер. Цей метод використовує дані, надані користувачами. Ці дані можуть бути неповними або не відповідають дійсності

Класифікація створюється відповідно до інтересів клієнтів, які мають тенденцію змінюватися з плином часу. Демографічна фільтрація не підтримує прийняття профілю користувача до змін

1.2.3. Колаборативна фільтрація

Метод, який є найбільш зрілим і найбільш широко використовується для РС є колаборативна фільтрація (КФ). Він покладається тільки на думки, явно виражені користувачами по пунктах. Система рекомендує цільовому клієнту продукти (або людей), які були оцінені в плюсі іншими людьми, рейтинги яких аналогічні рейтингам цільового клієнта-користувача.

Вимога, яка повинна бути виконана, полягає в тому, що клієнт повинен увійти в систему, щоб створити для нього профіль користувача. Подання профілю користувача може бути вектором продуктів і рейтингів, які були

присвоєні конкретних елементів. Вектор змінюється, коли користувач оцінює елемент . Рейтинги можуть мати логічне значення (клієнтам подобається або не подобається товар) або значення з більш широкою шкали ,наприклад, з діапазону $[-5, 5]$.

Переваги полягають в наступному:

Хоча це найстаріший метод рекомендації, він все ще дуже ефективний

Цей метод не вимагає надання уявлення об'єкта, яке може бути легко прочитано комп'ютерами

Недоліки полягають в наступному:

Проблеми з новими користувачами (холодний старт), а також нові продукти (рання проблема), які ще не були оцінені користувачами. Багато людей повинні оцінити продукти, перш ніж система буде ефективною

Розрідженість даних-це проблема, яка може виникнути, коли є багато предметів для оцінки; набір предметів змінюється дуже часто, і кількість клієнтів відносно невелика . Це ускладнює пошук користувачів, схожих на цільову

Труднощі виявлення " непередбачуваних " користувачів з рідкісними уподобаннями і незвичайними думками про продукти

У традиційних системах колаборативної фільтрації обсяг роботи збільшується одночасно з кількістю учасників і елементів у системі. Метод обчислення досить складний, тому він зазвичай виконується в автономному режимі. Крім того, це дорого, тому що це вимагає збору великої кількості даних, щоб бути ефективним

1.2.4. Фільтрація на основі контенту

Фільтрація на основі контенту (CBF) використовує опис елементів, які раніше були переглянуті або придбані клієнтом і оцінені їм позитивно. Система рекомендує споживачам товари, схожі на ті, які їм подобалися в минулому. Профіль користувача створюється на основі функцій, які зустрічаються в елементах, позитивно оцінених користувачем, і містить смаки, потреби і переваги користувачів. Рекомендації на основі контенту зазвичай застосовуються для пропозиції документів, веб-сторінок, повідомлень новин та інших текстових елементів.

Перевага полягає в наступному:

Тільки аналіз предметів, які бачив або купив один незалежний користувач, повинен бути виконаний. На відміну від CF, ця техніка не така складна

Недоліки полягають в наступному:

Надспеціалізація - коли система рекомендує елементи, аналогічні тим, які були високо оцінені клієнтом, може виникнути ефект надспеціалізації. Це означає, що елементи, які пропонуються Користувачеві, будуть дуже схожі, і клієнтові може бути нудно від безперервного перегляду документів з вікон контентом

Обмежений контент-аналіз-в CBF кожен елемент описується функціями. Не завжди вдається створити достатній набір функцій. Витяг інформації з текстового документа щодо легко в порівнянні з іншими типами документів (графічними, аудіо-або відеодокументами). Більш того, якщо два документи містять одні і ті ж терміни і в результаті мають один і той же набір ознак, то їх

неможливо розрізнити. В цьому випадку система не здатна відрізнити добре написаний документ від погано написаного

Нова проблема користувача-метод стає ефективним, коли користувач оцінює достатню кількість елементів. Надійна рекомендація може бути створена тільки тоді, коли система має точні знання про переваги і потреби користувачів.

1.2.5. Гібридний метод

Гібридний підхід до рекомендації поєднує в собі контент-орієнтовану і колаборативну фільтрацію. Основна мета гібридних методів полягає в тому, щоб уникнути вад двох раніше перерахованих методів. Існує безліч різних способів комбінування контентної і колаборативної фільтрації. Найбільш відомими є:

Реалізувати обидва методи окремо і об'єднати результати цих методів

Додайте деякі характеристики на основі контенту в спільну фільтрацію

Додайте деякі характеристики спільної роботи в фільтрацію на основі вмісту

Розробити одну модель, яка застосовувала б як контент-орієнтовані, так і колаборативні характеристики.

Ці два підходи доповнюють один одного і сприяють ефективності один одного.

1.3. Порівняльний аналіз методів побудови рекомендаційних систем в соціальних мережах

Для того, щоб вибрати найбільш потрібний метод рекомендацій потрібно визначити критерії для методів рекомендаційних систем.

Були вибрані такі критерії як:

Швидкість, який потрібен для того щоб успішно використовувати рекомендації в режимі онлайн потрібна велика швидкість обробки інформації для видачі рекомендації.

Точність, який є одним з ключових характеристик рекомендаційних систем. Для успіху використання рекомендаційних систем потрібно щоб вона найбільш вцілено генерувала рекомендації кінцевому користувачу.

Направленість, який потрібен щоб відобразити можливості використання даного підходу рекомендацій в різних сферах де потрібна рекомендація.

Вартість обчислення, який тісно зв'язан зі швидкістю та є одним з важливим критерієм тому що на сьогодні більшість магазинів, соціальних мереж використовують хмарні технології в яких у більшу частину випадків кошторис за використання ресурсів складається з використовуваного місця, трафіку та використаного процесорного часу.

Використання в соціальних мережах який описує в яких випадках метод може бути використаний при побудові рекомендацій в соціальних мережах. Цей критерій був доданий до порівняння з причини того, що основною тезою магістерської роботи є дослідження побудови рекомендацій для соціальних мереж.

Для критеріїв як Швидкість, Точність та Вартість обчислення будуть використовуватися спрощений діапазон оцінок Високий, Середній, Низький.

Для критерія Направленність буде використовуватися діапазон оцінок як: Всенаправленна, Вузька.

Таблиця 1.2 Порівняльна таблиця методів рекомендаційних систем

	Швидкість	Точність	Направленість	Вартість обчислення	Використання в соціальних мережах
Статистичний аналіз	Висока	Низька	Всенаправленна	Низька	Рекомендація популярного контенту.
Нерозрахунковий підхід	Низька	Низька	Всенаправленна	Низька	Рекомендації типу «Вибір редакції»
Колаборативна фільтрація	Середня	Висока	Всенаправленна	Середня	Рекомендація інших людей.
Фільтрація на основі контенту	Середня	Висока	Вузька	Висока	Рекомендація подібного медіа контенту
Гібридний метод	Середня	Висока	Всенаправленна	Середня	Рекомендація на основі декількох методів
Демографічна фільтрація	Середня	Низька	Всенаправленна	Низька	Рекомендація залежить від персональних атрибутів людини.

1.4. Постановка задачі дослідження

Об'єктом дослідження в рамках магістерської атестаційної роботи є процес побудови рекомендацій в соціальних мережах в режимі онлайн.

Предметом дослідження являються методи рекомендацій які використовуються у рекомендаційних системах в режимі онлайн.

Метою даної роботи є підвищення ефективності методу колаборативної фільтрації для рекомендацій які використовуються у соціальних мережах в режимі онлайн.

Для досягнення мети, необхідно вирішити наступні питання:

- дослідити характеристики рекомендаційних систем
- дослідити рекомендаційні системи які призначені для соціальних мереж.
- зробити порівняльний аналіз методів побудови рекомендаційних систем для соціальних мереж.
- створити проект з розробки рекомендаційної системи для соціальних мереж в режимі онлайн.
- удосконалення методу колаборативної фільтрації для рекомендацій в режимі онлайн для соціальної мережі.

2. ДОСЛІДЖЕННЯ МЕТОДІВ ПОБУДОВИ РЕКОМЕНДАЦІЙ В СОЦІАЛЬНИХ МЕРЕЖАХ

2.1. Методи побудови рекомендацій в соціальних мережах

2.1.1. Класи колаборативної фільтрації

Існує два класи колаборативної фільтрації:

- колаборативна фільтрація на основі пам'яті;
- колаборативна фільтрація на основі моделей

Колаборативна фільтрація на основі пам'яті використовує інформацію об оцінках користувачів та дані про користувача щоб обчислити та знайти схожість між користувачами або предметами. Типовими прикладами такого підходу є CF на базі сусідства та рекомендації, що ґрунтуються на позиціях / користувачах. Наприклад, у підходах, що базуються на значеннях оцінок користувача u , що надаються елементу.

$$r_{u,i} = \text{aggr}_{u'} \in U^{R_{u'},i} \quad (2.1)$$

Де U відноситься до вибірки N користувачів які найбільш схожі до користувача u який оцінив i .

Деякі приклади функцій агрегації:

$$r_{u,i} = \frac{1}{N} \sum_{u' \in U} r_{u',i} \quad (2.2)$$

$$r_{u,i} = k \sum_{u' \in U} \text{simul}(u, u') r_{u',i} \quad (2.3)$$

$$r_{u,i} = \bar{r}_u + k \sum_{u' \in U} \text{simul}(u, u') (r_{u',i} - \bar{r}_{u'}) \quad (2.4)$$

Де k – нормуючий множник який визначається як

$$k = 1 / \sum_{u' \in U} |\text{simul}(u, u')| \quad (2.5)$$

\bar{r}_u є середня оцінка користувача u для всіх виробів, оцінених і.

Алгоритм, заснований на сусідстві, обчислює подібність двох користувачів або виробів, виробляє прогноз для користувача, приймаючи середнє зважене всіх оцінок. Обчислення схожості між виробами або користувачами є важливою частиною цього підходу. Багаторазові заходи, такі як кореляції Пірсона і схожість, заснована на скалярному добутку, використовуються для цього.

Схожість двох користувачів X , Y через кореляцію Пірсона визначається як

$$\text{simil}(x, y) = \frac{\sum_{i \in I_{xy}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in I_{xy}} (r_{x,i} - \bar{r}_x)^2 \sum_{i \in I_{xy}} (r_{y,i} - \bar{r}_y)^2}} \quad (2.6)$$

де I_{xy} - це набір елементів, оцінених як користувачем x , так і користувачем y .

$$\text{simil}(x, y) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \times \|\vec{y}\|} = \frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_x} r_{x,i}^2} \sqrt{\sum_{i \in I_y} r_{y,i}^2}} \quad (2.7)$$

Підхід, заснований на скалярному добутку визначає скалярний добуток між двома користувачами x і y , як:

Заснований на користувачеві алгоритм найкращий з N рекомендації використовує засновану на подібності векторну модель для визначення K — більшості подібних користувачів до активного користувача. Після того, як знайдені найбільш схожі користувачі, їх відповідні матриці агрегуються для визначення рекомендованого набору елементів. Популярний метод, знаходження схожих користувачів — Locality-sensitive hashing який реалізує механізм пошуку найближчих сусідів у лінійному часі.

Переваги цього підходу включають в себе: очікуваність результатів, що є важливим аспектом рекомендаційних систем; просте створення і використання; просте полегшення нових даних; добра масштабованість зі співавторами рейтингових пунктів.

Є також кілька недоліків при такому підході. Його продуктивність знижується, коли дані становляться розрідженими, що трапляється часто з виробами, пов'язаними з мережею. Це ускладнює масштабованість такого підходу і створює проблеми з великими наборами даних. Хоча він може ефективно обробляти нових користувачів, тому що спирається на структури даних, додавання нових елементів стає більш складним, що, як правило, спирається уявленням про конкретну складову векторного простору. Додавання нових елементів вимагає включення нового пункту і повторного включення всіх елементів у структуру.

Колаборативна фільтрація на основі моделей використовує різні алгоритми інтелектуального аналізу даних, машинного навчання для прогнозування рейтингу користувачів нерейтингових елементів. Існує безліч алгоритмів CF на основі моделей. Байєсівські мережі, моделі кластеризації, латентні семантичні моделі, такі як сингулярна декомпозиція, імовірнісний латентний семантичний аналіз, множинний мультиплікативний фактор, латентний розподіл Діріхле і Марковські моделі прийняття рішень.

Завдяки цьому підходу методи зменшення розмірності в основному використовуються в якості додаткового методу для підвищення надійності і

точності підходу, заснованого на пам'яті. У цьому сенсі такі методи, як сингулярна декомпозиція, аналіз головних компонент, відомі як латентні факторні моделі, стискають матрицю елементів в мале подання в термінах латентних факторів. Однією з переваг використання цього підходу є те, що замість матриці з високою розмірністю, що містить велику кількість пропущених значень, ми будемо мати справу з набагато меншою матрицею в просторі з більш низькою розмірністю. Скорочене подання може бути використано для алгоритмів околиці, заснованих на користувачів чи елементах, які представлені в попередньому розділі. У цієї парадигми є кілька переваг. Він обробляє розрідженість вихідної матриці краще, ніж на основі пам'яті. Крім того, порівняння подібності на результуючій матриці набагато більш масштабоване, особливо при роботі з великими розрідженими наборами даних.

2.1.2. Розробка удосконаленого підходу колаборативної фільтрації з урахуванням демографічної фільтрації

Вимоги, які були перераховані, допомагають побудувати концепцію рекомендації, оскільки вони підкреслюють найбільш важливі аспекти системи. На цьому рівні розвитку будуть розглянуті в основному функціональні вимоги. Загальний процес рекомендації, мета якого полягає в поданні пропозицій користувачеві x , повинен містити такі елементи, як:

Моніторинг поведінки користувачів в соціальній мережі

Збір і підготовка даних

Розрахунок і періодичний перерахунок рейтингових списків для учасника x

Вибір k -найближчих сусідів для користувача x

Надання рекомендації користувачеві x

Періодичний перерахунок W_{ij} на основі зворотного зв'язку користувача

x

Всі перераховані кроки однакові для всіх людей з усієї спільноти. Деякі з них займають більше часу, ніж інші. Згідно з цим критерієм, це означає час, який необхідно для виконання конкретної задачі, їх можна розділити на виконуються в режимі онлайн і оффлайн (рис. 2.1).

Одним з найважливіших елементів при створенні спеціальних рекомендацій є використання відповідних даних. По-перше, він повинен бути сучасним; в іншому випадку підготовлена пропозиція не має відношення до справи. По-друге, дані, на основі яких створюється рекомендація, повинні бути важливі з точки зору процесу рекомендації. Це означає, що інформація, яка важлива з точки зору людини, коли він вибирає нових друзів, повинна бути зібрана. Немає сенсу використовувати дані, які не хвилюють користувача. Це забезпечить марну рекомендацію і матиме негативні наслідки, такі як невдоволення користувачів США, що може призвести до відставки від служби. У цьому розділі представлені дані, які будуть використовуватися при підготовці пропозицій[20].

Для забезпечення найвищої якості створюваних пропозицій необхідно визначити основні джерела даних, корисних для рекомендацій. У соціальних мережах доступні не тільки дані, що описують користувача як окремої людини, такі як демографічні дані та інтереси, але також точний опис відносин користувача і його діяльності може бути отримано з історії користувача.



Рис 2.1 Розподіл категорії особистої інформації користувача в соціальній мережі.

Для приведення в порядок складних характеристик учасників соціальних мереж створюється профіль (рис.2.1). Він складається з двох основних частин: статичної і динамічної. Користувачі самі доставляють перше, в той час як останнє контролюється і збирається системою. На основі аналізу існуючих онлайн-спільнот Профіль користувача був розбитий на набір компонентів. Крім того, кожен з компонентів складається з декількох окремих атрибутів (рис. 2.1). До статичних елементів відносяться демографічні, процентні і краці компоненти. Інформація для цих компонентів отримана з опитувань, які члени мережі заповнюють в процесі реєстрації і після цього підтримують себе. Пошук, дія та зв'язок є динамічними компонентами профілю користувача. Кожен з цих компонентів буде описаний в наступній частині цього розділу.

Крім того, кожному з атрибутів користувач може присвоїти рівень конфіденційності. Дані про Користувача можуть бути показані тільки друзям користувача, а також друзям друзів цієї людини або, нарешті, всім членам соціальної мережі. Пропонована система рекомендацій дозволяє створювати

пропозиції на основі всіх наявних даних про користувача, не показуючи цю інформацію іншим користувачам.

Описане вище поділ профілю на компоненти і далі в атрибутах робить створену структуру рекомендацій дуже підтримуваною. Дуже легко додавати нові або видаляти старі компоненти і атрибути.

2.1.3. Демографічний компонент

Демографічний компонент містить інформацію, яку користувачі надають самі, заповнюючи спеціальні форми, наприклад опитування. Кожна онлайн соціальна мережа вимагає від своїх членів входу в систему та надання хоча б основної інформації про користувачів (наприклад, ім'я і адресу електронної пошти). Однак, які елементи необхідні, тісно залежить від вимог конкретної мережі. Хоча, є багато різних соціальних мереж які проводять аналогічні обстеження в процесі реєстрації. Користувачі самі підтримують всі ці дані.

Профіль у віртуальному співтоваристві MySpace розділений на кілька елементів. Демографічні дані містяться в трьох з них: ім'я (рис 2.2), основна інформація (рисунок 2.3), а також спосіб життя (рис 2.4).



Рис 2.2 Елемент Name в соціальній мережі

Елемент Name містить інформацію про ім'я та прізвище користувача, а також ім'я, яке відображається іншим членам мережі, Ім'я Користувача MySpace і URL Myspace. Це важливі дані з точки зору окремого користувача, оскільки вони дозволяють зберегти особу. Проте це марно з рекомендаційної точки зору. Причина в тому, що люди підбираються не на основі їх Імен і прізвищ, а на основі їх характеристик і поведінки. Елемент базової інформації складається з декількох пунктів, які відносяться до базової демографічної інформації (рис.2.3), такий як стать, дата народження, рід занять і місце розташування. Більш того, учасник визначає тут, кого вони шукають. Можливі варіанти залежать від функціональності мережі, і це може бути, наприклад, дружба, дата або ділові контакти.

Interests & Personality	Background & Lifestyle	Schools	Companies	Networking	Name	Basic Info	Profile Songs
<p>Gender: <input type="radio"/> Male <input checked="" type="radio"/> Female</p> <p>Date Of Birth: Oct / 28 / 1982</p> <p>Occupation: Student</p> <p>City: Wroclaw</p> <p>State/Region: N/A (US State) [] (Non-US region) []</p> <p>Country: Poland</p> <p>Zip/Postal Code: []</p> <p>Ethnicity: White / Caucasian</p> <p>Body Type: No Answer</p> <p>Height: 0 Feet 0 Inches <input checked="" type="checkbox"/> No Answer</p> <p>I would like to make space for:</p> <p><input type="checkbox"/> Dating</p> <p><input type="checkbox"/> Serious Relationships</p> <p><input checked="" type="checkbox"/> Friends</p> <p><input type="checkbox"/> Networking</p>							

Рис 2.3 Елемент «Базова інформація» в мережі MySpace

Нарешті, розділ спосіб життя надає інформацію про таких демографічних даних, як сімейний стан, сексуальна орієнтація, освіта, рідне місто і т. д. (Рисунок 4.5). Такий поділ на три елементи, представлених на малюнках, характерно для мережі MySpace. Наприклад, мережа Friendster надає опитування про демографічні дані у вигляді одного аркуша.

Interests & Personality	Background & Lifestyle	Schools	Companies	Networking	Name	Basic Info	Profile Songs
<p>Marital Status: <input type="radio"/> Swinger <input checked="" type="radio"/> In a Relationship <input type="radio"/> Single <input type="radio"/> Divorced <input type="radio"/> Married</p> <p>Sexual Orientation: <input type="radio"/> Bi <input type="radio"/> Lesbian <input checked="" type="radio"/> Straight <input type="radio"/> Not Sure <input type="radio"/> No Answer</p> <p>Hometown: <input type="text" value="Radomsko"/></p> <p>Religion: <input type="text" value="Catholic"/></p> <p>Smoker: <input type="radio"/> Yes <input checked="" type="radio"/> No <input type="radio"/> No Answer</p> <p>Drinker: <input checked="" type="radio"/> Yes <input type="radio"/> No <input type="radio"/> No Answer</p> <p>Children: <input type="radio"/> I don't want kids <input checked="" type="radio"/> Someday <input type="radio"/> Undecided <input type="radio"/> Love kids, but not for me <input type="radio"/> Proud parent <input type="radio"/> No Answer</p> <p>Education: <input checked="" type="radio"/> High school <input type="radio"/> Some college <input type="radio"/> In college <input type="radio"/> College graduate <input type="radio"/> Grad / professional school <input type="radio"/> Post grad <input type="radio"/> No Answer</p> <p>Income: <input type="text" value="Less than \$30,000"/></p>							

Рис 2.4 Розділ «Спосіб життя» в мережі MySpace

Для вилучення демографічних даних, які будуть корисні з точки зору рекомендацій, було проведено дослідження декількох онлайн-співтовариств. Приклад MySpace network представлений тут, щоб забезпечити загальне уявлення профілів користувачів. Метою проведеного дослідження було створення профілю користувача, який можна було б використовувати в рекомендаційному процесі в різних соціальних мережах.

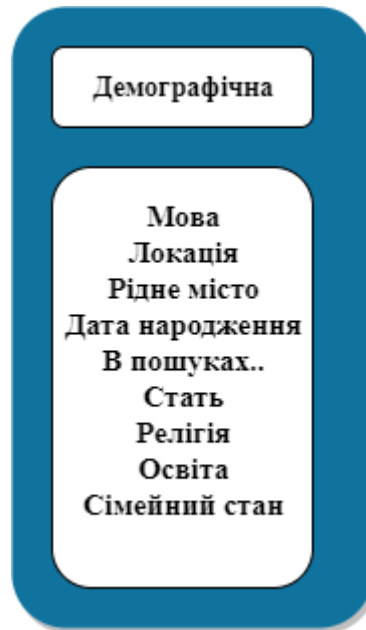


Рис 2.5 Демографічний компонент даних про користувача

Демографічні дані, які будуть зібрані і проаналізовані в подальшій частині цієї тези, представлені на рисунку 2.5 Мова є ключовим елементом, тому що коли двоє людей хочуть спілкуватися один з одним, то принаймні одна мова для цих людей повинна бути однаковою. Місце-це місце проживання, а рідне місто-місце походження. Обидва ці атрибути є ієрархічними. Це означає, що кожна людина приїжджає з певної країни, округу або району та міста. Пошук позначає мета запрошення нових друзів, наприклад дружба, шлюб, професійні контакти і т. д. Це важливий елемент з точки зору рекомендації, тому що якщо ми зіставимо двох осіб, то обидва вони повинні мати одну і ту ж причину приєднання до мережі. Сексуальна орієнтація, стать, професія, релігія, освіта і воєнний стан є категоріальними атрибутами. Наприклад, одна людина може мати тільки один військовий статус і мати тільки одну дату народження в той час[21].

Ці перераховані атрибути будуть використовуватися в подальших дослідженнях для визначення подібності між двома членами мережі.

2.1.4. Компонент інтереси

Інформація по компоненту, що цікавить, аналогічно демографічному компоненту, надається самими користувачами шляхом заповнення спеціальних форм. Більш того, користувачі самі підтримують всі ці дані. Компонент інтересу може включати в себе безліч атрибутів. Приклади таких атрибутів можна знайти в будь-якій соціальній мережі.

Багато аналізи були зроблені для того, щоб класифікувати величезний і різноманітний набір інтересів, які можуть мати люди. Одна з дуже цікавих рішень, яке може бути застосоване для складання процентної рекомендації, представлено Ст. Автори пропонують вилучити з профілів соціальних мереж нормалізоване подання шляхом зіставлення випадково заданих ключових слів і фраз в формальну онтологію дескрипторів. Після цього нормалізовані профілі з метаданими доповнюються для полегшення встановлення з'єднання. Нарешті, метод машинного навчання застосовується для вивчення ваг семантичної пов'язаності між кожною парою дескрипторів. Результатом цього процесу є семантична мережа, вузли якої є дескрипторами, а зважені ребра представляють силу семантичної зв'язаності. Дана онтологія може бути використана в пропонованій в даній дисертації структурі, оскільки багато соціальних мереж дозволяють вводити текстові дані про інтереси¹¹. Наприклад, MySpace надає закладку "інтереси і особистість", яка дозволяє користувачам писати про музику, яку вони слухають, фільмах, які вони дивляться, книжки, які вони читають, і т. д. (Рисунок 2.6).

Крім того, мережа інші соціальні мережі дають можливість користувачам представити свій інтерес в іншому форматі. Після дослідження декількох віртуальних спільнот був створений компонент інтересу, який

представлений на рисунку 2.6. Кожен користувач може мати свої улюблені книги, письменників, групи, музикантів, фільми і т. д

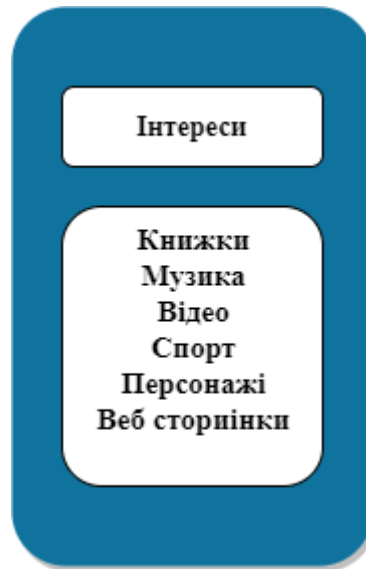


Рис 2.6 Компонент інтересів користувача.

Хоча всі демографічні дані відносяться до категоріальним змінним, більшість атрибутів інтересу ієрархичні, наприклад книги можна розділити на кілька категорій: фентезі, драма, кримінал і т. д. (Рис. 2.6). Далі, фантастичні книги можна розділити по авторам: Пратчетт, Кінг та ін Ієрархії для книг виглядає наступним чином: Категорія автор - назва. Інші ієрархічні атрибути: Музика, Фільми і спорт. Ієрархія музики: Категорія performer Виконавець (група, музикант або співак) назва, в той час як ієрархія фільмів побудована наступним чином: Категорія-назва. Нарешті, спортивна ієрархія: категорія (зима чи літо) місце, де відбувається спорт (вода, повітря, земля) назва виду спорту. Веб-сайти та особи, чиїм шанувальником є користувач, відносяться до категоріальних атрибутів[25].

Для обчислення схожості між членами мережі будуть використовуватися як ієрархічні, так і категоріальні атрибути. Проблема, яка може виникнути тут, - це потенційна відсутність даних, оскільки вся інформація про інтерес є необов'язковою. Це означає, що користувач може надати цю інформацію, але це не потрібно.

2.1.5. Компонент базової інформації

Користувач також передає базову інформацію, яка є прямою підказкою для системи про людей, яких шукає користувач. Користувачі самі підтримують всі ці дані. Атрибути цього компонента представлені на малюнку 4.11. Хоча ні Friendster, ні MySpace не надають окремі закладки для цього компонента, цей елемент доданий у інших соціальних мережах як частина профілю користувача. Це важливий компонент, тому що він може дати прямий натяк, в яких користувачах х він зацікавлений.

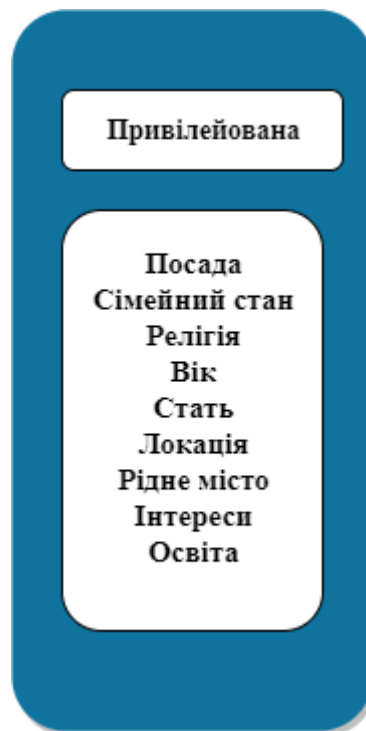


Рис 2.7 Компонент базової інформації

Слід підкреслити, що ці атрибути містять ті ж атрибути, що і демографічні компоненти і компонент інтересів. Різниця полягає в тому, що елементи з базового компонента описують людину, яку шукає користувач х, в той час як демографічні та компонент інтересу надають опис характеристики користувача х.

2.1.6. Компонент діяльності

Двома найбільш складними компонентами є діяльність і взаємини. Перший з них вимірює активність користувача в співтоваристві. Активність користувачів може бути виміряна на основі частоти їх входу в систему, частоти зміни інформації в фотоальбомах і блогах, а також частоти відправки електронних листів та участі в різних видах чатів в мережі. Активність конкретного користувача змінюється з плином часу. Це означає, що кожна людина може бути більш активним в один період і менш активним в інший[27].

Залежно від функціональності, яку надає конкретна мережа, користувачі можуть використовувати один з наступних елементів: фотоальбоми, кіноальбоми, чати, електронні листи, блоги, форуми, відправка і отримання файлів. На основі дослідження, проведеного в різних соціальних мережах, був створений компонент активності, який представлений на малюнку 2.8

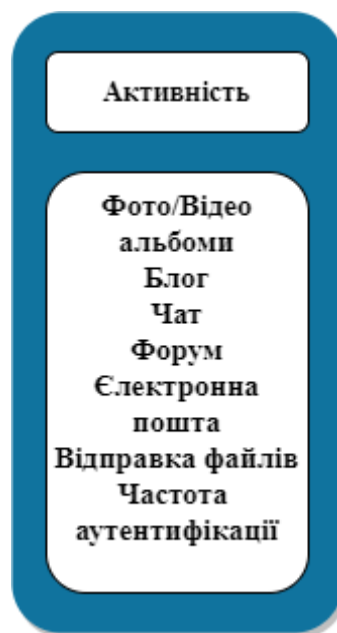


Рис 2.8 Компонент активності користувача.

2.1.7. Компонент відносин

Другим з найбільш складних компонентів є компонент відносин, який описує кількість і тривалість відносин користувача, а також деякі інші особливості, що характеризують ці відносини. На основі функціональності, що надається віртуальними спільнотами, можна виділити кілька характеристик відносин. Люди можуть відправляти і отримувати запрошення від усіх членів спільноти, а також брати участь в різних чатах. Крім того, люди можуть залишати коментарі про чужі блоги, фото-і кіноальбоми, а також ділитися своєю думкою на форумі

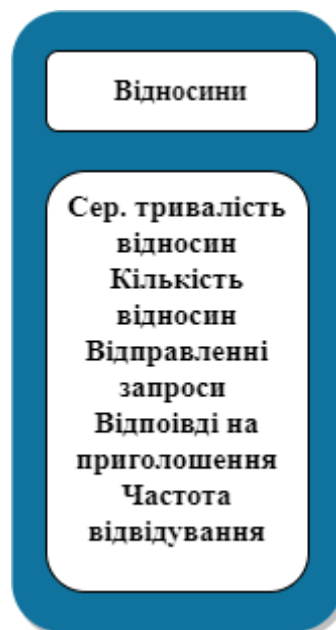


Рис 2.9 Елемент відносин в соціальній мережі.

Всі перераховані атрибути зібрані в компонент відносини, що є останньою частиною профілю користувача (рис. 2.9). Ці атрибути відстежуються системою, і їх значення періодично оновлюються, завдяки наданню рекомендацій, які будуть актуальними.

2.2. Удосконалення методу колаборативної фільтрації в соціальних мережах в режимі онлайн.

2.2.1. Вимоги до рекомендаційної системи

Перш ніж будувати рекомендаційну структуру, необхідно підготувати вимоги і передумови системи. Метою пропонованого в даній роботі рішення є надання персоналізованих рекомендацій. Це вимагає, щоб системи відповідали певним функціональним і якісним вимогам.

Функціональні вимоги наступні:

- Належна підготовка даних
- Створення персоналізованого методу рекомендації
- Звести до мінімуму кількість хибнопозитивних рекомендацій

Правильні дані про користувачів повинні бути зібрані і проаналізовані. Однак збір даних здійснюється не тільки один раз. Дані про користувачів змінюються з плином часу, і це призводить до того, що вони повинні постійно контролюватися. Крім того, для створення відповідної рекомендації необхідно провести достатню підготовку даних, яка включає класифікацію даних з кількох компонентів, що створює Профіль користувача (див. розділ 4.4). Ключовим елементом є опис, який показує, як впоратися з порожніми даними. Відсутність даних є великою проблемою, оскільки обмежує можливість створення відповідних рекомендацій[27].

Не тільки правильні дані визначають якість рекомендації, але і метод, який використовується для дослідження подібності між членами спільноти. Наступною вимогою є створення функції, що дозволяє виміряти схожість між користувачами. На основі цієї функції рекомендується людина, яка підходить поточному користувачеві. Ця функція подібності повинна враховувати

демографічні та процентні дані, а також вимірювання активності і сили відносин, підтримуваних учасниками соціальних мереж. Ця функція повинна періодично перераховуватися і дозволяє точно визначити k-найближчий сусід людину, якій створюється рекомендація.

Пропоновані рамки повинні звести до мінімуму число хибнопозитивних рекомендацій. Це означає, що ми не можемо рекомендувати людей, які не відповідають профілю поточного користувача. Безпечніше не рекомендувати людину, яка підходить поточному користувачеві (помилково негативні рекомендації), ніж давати безліч хибнопозитивних пропозицій, які можуть бути джерелом роздратування і незадоволеності людини, що отримала рекомендацію.

Вимоги до якості, які повинні бути виконані на найвищому рівні за структурою:

- Продуктивність - розрахунки, особливо зроблені в Інтернеті, не повинні тривати довго, тому що користувачі будуть роздратовані, якщо їм доведеться чекати рекомендацій. Таким чином, потрібна висока продуктивність.
- Ремонтпридатність-покращення в профілі користувача (наприклад, розширення профілю користувача шляхом додавання нових компонентів і атрибутів) і в методі рекомендації повинні бути легко впроваджені. Це допоможе забезпечити високий рівень чергової вимоги до якості для даної системи, тобто доступність. Крім того, рівень адаптивності системи повинен дозволяти застосовувати рекомендаційні рамки в різних соціальних мережах.
- Доступність - система повинна бути доступна протягом більшої частини часу, який користувачі проводять в мережі. Це призведе до того, що учасник буде звикати мати рекомендації поруч. Це допомагає створити зв'язок між користувачем і системою і призводить до підвищення лояльності користувачів.

- Швидкість – Система повинна швидко обробляти запити на рекомендації в режимі онлайн задля забезпечення плавної роботи сервісу

2.2.2. Рекомендаційна система на основі колаборативної фільтрації з додаванням демографічної фільтрації.

На основі зібраної інформації може бути побудована рекомендаційна структура, що дозволяє створювати пропозиції для соціальної мережі. Основною метою системи є надання найбільш актуальних рекомендацій користувачам. Крім того, аналізуючи кілька різних джерел даних, метод дозволяє новому користувачеві приєднатися до мережі і задовольнити його очікування[30].

Загальний вигляд процесу рекомендацій для соціальної мережі представлений на малюнку 2.10. Мета цього процесу-знайти користувача у (або групу таких користувачів), який може бути рекомендований користувачеві x . це означає, що люди, які позитивно дадуть відповідь на запрошення користувача x , повинні бути ідентифіковані.

Цей метод можна розділити на етапи:

На першому етапі зібрати інформацію про користувачів і провести достатню підготовку даних. Всі необхідні дані повинні бути розміщені в репозиторії. Таким чином, профіль користувача визначається для того, щоб визначити, які дані необхідні для створення відповідної рекомендації.

Підготовка даних є першим і одним з найбільш важливих елементів процесу підготовки рекомендацій. В цілому, в інтелектуальному аналізі даних ця дія складається з набору наступних етапів:

- Доступ до даних
- Аудит даних
- Розширення і збагачення даних
- Визначення структури даних
- Моделювання даних

На другому етапі виконується так звана попередня демографічна фільтрація. Цей етап включає в себе набір правил, за якими кожен користувач у повинен бути відфільтрований. Таку фільтрацію можна описати формулами:

If ($x_{\text{language}} \neq y_{\text{language}}$) то відкинути персону у з рекомендації

If ($x_{\text{city}} \neq y_{\text{city}}$) то відкинути персону у з рекомендації

If ($x_{\text{looking for}} \neq y_{\text{looking for}}$) то відкинути персону у з рекомендації

Якщо Користувач у не проходить через демографічну фільтрацію відхиляється від подальшого аналізу. Далі система обчислює кінцеву функцію подібності $r(x \rightarrow y)$ для кожної пари користувачів[22].

Цей етап був доданий навмисно тому що цей фільтр виконує значну фільтрацію користувачів за короткий проміжок часу і дозволяє швидко відсіяти користувачів, які будуть нерелевантні для швидкого пошуку.

На третьому етапі знаходиться повний список релевантних користувачів. Спочатку, за допомогою колаборативної фільтрації знаходиться повна схожість, аналізуються статистичні дані, такі як інтереси та базова інформація. На цьому етапі відсутня фільтрація по демографічним атрибутам тому що вона була зроблена на попередньому етапі і вхідна кількість даних зменшена, що дозволяє швидше обробити вхідні дані і зробити більш швидко рекомендацію. Аналізується компліментарність ініціації відносин та активність користувача. Також проводиться простий аналіз міцності стосунків який полягає в алгоритмі PageRank і досліджує не тільки відносини користувача, який буде потенційно рекомендований, але і його друзів і так далі[22].

Підсумкове подобання складається з чотирьох основних елементів:

- Фільтрація на основі демографічних атрибутів.
- Пряма схожість, отримана безпосередньо з статичних атрибутів користувачів
- Загальний показник активності кандидата на рекомендацію (користувач у)
- Аналіз динамічних характеристик користувача у.

Аналіз статичних атрибутів найчастіше описується формулою прямої подібності:

$$f_{language}(x, y) = \begin{cases} 1 & \text{if at least one language is the same} \\ 0 & \text{if none of the language is the same} \end{cases} \quad (2.8)$$

На прикладі атрибута мова.

Функції прямої подібності для статусу, пошуку, професії, релігії та сексуальної орієнтації мають ту ж форму, що і у формулі 1. Значення цих прямих функцій дорівнює 1, якщо атрибути для обох користувачів мають однакове значення, і 0 в іншому випадку[29].

Для безперервних атрибутів, таких як вік, може бути використаний простий, нормалізована зворотня різниця, наприклад $f_{age}(10,10)=1$ і $f_{age}(1,100)=0$, де макс. вік = 100 р. і мін. вік = 1. Функція прямої подібності для віку атрибута визначається наступним чином.

$$f_{age}(x, y) = 1 - \left| \frac{x_{age} - y_{age}}{m - l} \right| \quad (2.9)$$

де:

x_{age} – користувач для кого генерується рекомендація

y_{age} – користувач який порівнюється з користувачем x

l – мінімальні значення атрибута вік в системі(пр. 0)

m – максимальне значення атрибута вік в системі(пр. 100)

До кожного атрибуту додається вага, яка виражає важливість цього атрибуту при видачі рекомендації.

Існує проблема, яка полягає в тому, що кінцева функція подібності $r(x \rightarrow y)$ може мати значення 0. В цьому випадку всі обчислення повторюються з модифікацією кроку, званого попередньою фільтрацією. Беруться сусідні, більш схожі демографічні показники та процес повторюється знову. Причина досить проста, рекомендаційна система не може нічого або нікого рекомендувати. Краще запропонувати людину, яка може не відповідати перевагам користувача, ніж не рекомендувати нікого.

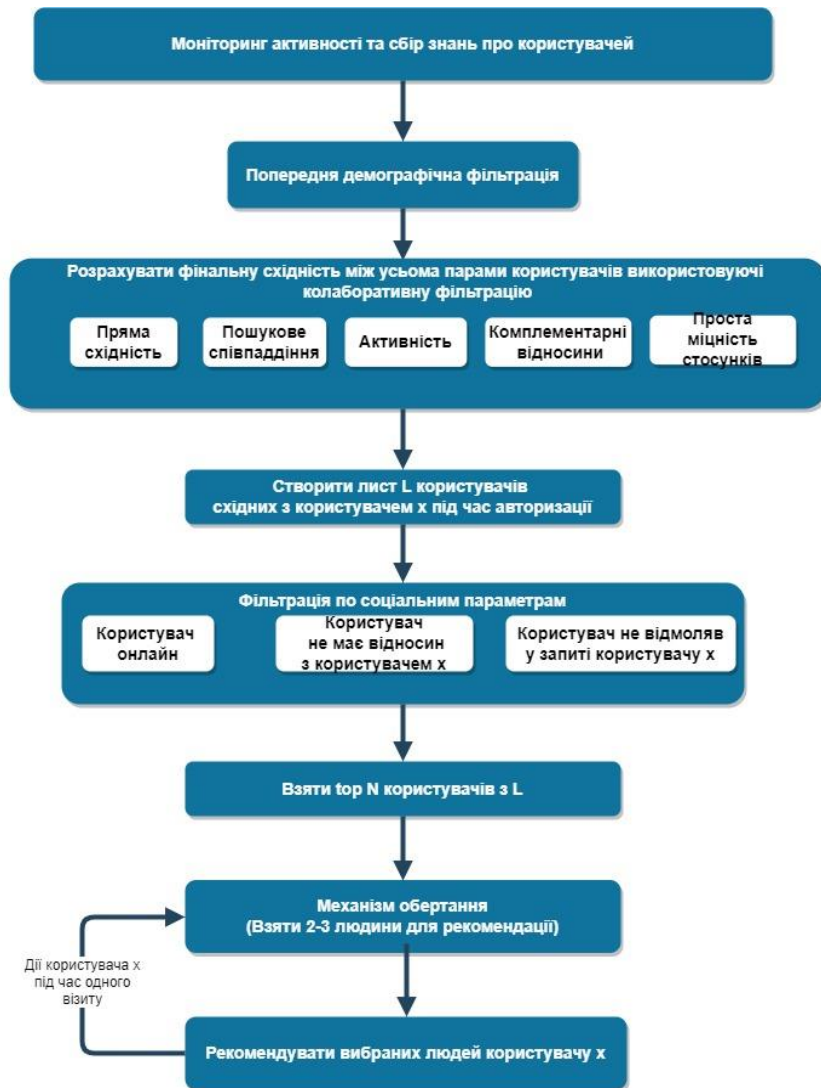


Рис 2.10 Процес рекомендацій для користувача в соціальній мережі.

Коли користувач x входить в систему, соціальна фільтрація списку L виконується для того, щоб видавати найбільш актуальні рекомендаційні дані.

З-за обмежень ефективності, тільки топ N найбільш підходящих користувачів фіксуються для рекомендації користувачу x під час їх перебування в мережі. Зрештою, тільки 2-3 обраних людини запропоновані до персони x .

Роторний механізм використаний, для того щоб запобігти повторень людей бути рекомендованим до споживача x весь час.

3. РОЗРОБКА ПРОЕКТУ ПОБУДОВИ РЕКОМЕНДАЦІЙ В СОЦІАЛЬНИХ МЕРЕЖАХ В РЕЖИМІ ОНЛАЙН

3.1. Розробка проекту побудови рекомендацій в соціальних мережах в режимі онлайн

Як кожне масштабне функціональне удосконалення в рамках існуючої онлайн системи яка належить компанії, потрібно створити процеси, за допомогою яких досягнути поставленої мети. Для цього буде створений проект з розробки та впровадження рекомендаційної системи для соціальної мережі.

Розробка проекту можна розуміти як проміжок часу від появи проекту до моменту його завершення і складається з окремих фаз стадій і етапів. Кожен проект поділяють на три фази:

- Передінвестиційну
- Інвестиційну
- Експлуатаційну

Кожна з цих фаз може розбиватися на стадії

Передінвестиційна фаза має в своєму складі такі стадії:

- преідентифікація;
- ідентифікація;
- підготовка;
- розробка та експертиза;
- детальне проектування.

Інвестиційна фаза охоплює такий перелік робіт, які можна об'єднати в такі стадії:

- підготовка і проведення тендерів;
- інженерно-технічне проектування;

- виробничий маркетинг;
- навчання персоналу.
- Експлуатаційна фаза включає в себе такі стадії:
 - задача в експлуатацію (грань між інвестиційної та експлуатаційної фазами, тому може перебувати і в тій, і в іншій фазі);
- виробнича експлуатація;
- заміна та оновлення;
- розширення та інновації;
- заключна оцінка проекту.

Під час передінвестиційної фази проводяться дослідження щодо інвестиційних можливостей проекту, аналізують альтернативні варіанти і попередній вибір, а також рівень підготовки проекту та роблять висновки щодо проекту та рішення про його інвестування.

Предідентифікація - стадія визначення інвестиційних можливостей, яка зв'язана з отриманням інформації потенційними інвесторами про інвестиційні можливості, які виникають на різних рівнях - від сектора економіки до підприємства.

Дослідження інвестиційних можливостей ґрунтуються здебільшого на загальних оцінках, ніж на детальному аналізі.

Мета здійснення цих досліджень - швидко і без значних витрат оцінити найважливіші моменти інвестиційних можливостей.

На стадії ідентифікації складається перелік всіх можливих ідей, здатних забезпечити виконання цілей економічного розвитку.

Оскільки загальною метою будь-якого проекту є отримання додаткових вигод, відбір проектів здійснюється шляхом зіставлення можливих результатів від реалізації різних проектів.

При відборі ідей проекту аналітики натрапляють, з одного боку, на необхідність скорочення кількості ідей, які розглядаються, і, з іншого боку, в потреби детального підходу до відібраних ідей з метою більш ретельного

вивчення цих варіантів і відбору найбільш привабливого. Відповідальність, покладена на аналітика в процесі вибору проекту, вимагає від нього переконливої мотивації вибору або відхилення будь-якого варіанта проекту.

Досить поширеними критеріями відхилення ідей проектів є:

- висока вартість проекту в порівнянні з очікуваними доходами;
- відсутність зобов'язань з боку організацій або людей, які будуть отримувати вигоду від здійснення проекту;
- відсутність політичної підтримки;
- технологія, що не придатна для здійснення проекту;
- надмірний масштаб проекту, який не відповідає організаційним та управлінським можливостям;
- надмірний ризик проекту.

Як правило, при відборі ідей проектів кількість варіантів, які заслуговують детального вивчення, скорочується до двох-трьох, що дозволяє швидко і без великих витрат оцінити найважливіші моменти інвестиційних можливостей проекту на макро- і мікро-рівнях.

Метою проведення макроаналізу є розробка інвестиційної пропозиції та збір інформації для потенційних інвесторів. На цій самій стадії, окрім макроаналізу, необхідно провести мікроаналіз, метою якого є діагностика окремих суб'єктів, які мають потенціал та інвестиційну привабливість, яка дозволить сформулювати попередні цілі окремих інвестиційних пропозицій.

Стадія підготовки проекту ділиться на два етапи: попередня оцінка і додаткові дослідження.

Ідея проекту повинна бути детально розроблена на стадії ретельного дослідження. Тобто перш ніж вкладати великі кошти в детальну розробку проекту, треба оцінити його ідею за допомогою попереднього аналізу, в результаті якого формулюються основні положення:

- розглянуті всі можливі альтернативи проекту;
- проект може бути спрямований на детальну розробку;

- всі аспекти проекту мають важливе значення з точки зору його здійсненності і є потреба глибокого вивчення за допомогою функціональних досліджень;
- ідея проекту на підставі існуючої інформації може бути визначена як нежиттєздатна або недостатньо приваблива;

Мета підготовки проекту - ранжування і відбір існуючих варіантів проекту, які потрібно передати для детальної розробки, і його становлення. Це може відбуватися, якщо:

- визначені основні і альтернативні варіанти проекту;
- ідентифіковані основні організаційні та політичні проблеми і знайдені можливості їх вирішення;
- є результати приблизної оцінки очікуваних витрат і доходів;
- існує підтримка з боку політичної влади, яка отримує вигоду від проекту;
- є підстави і впевненість, що проект отримає надійне фінансування з внутрішніх джерел.

Підготовка проекту дає інформацію щодо таких напрямків дослідження, як:

- стратегія проекту або корпоративні стратегії, межі проекту;
- ринок продукту проекту і концепція маркетингу;
- місцезнаходження, ділянка і навколишнє середовище;
- проектування і технологія;
- організаційні та накладні витрати;
- трудові ресурси, зокрема управлінські кадри, витрати на оплату робочої сили, потреби в професійному навчанні і витрати на нього;
- графік здійснення проекту і складання бюджету.

Іноді добре виконані дослідження щодо підготовки проекту можуть служити достатнім його обґрунтуванням, проте якщо економічна сторона

проекту викличе сумніви, слід неодмінно провести додаткові дослідження за проектом[33].

На стадії розробки і експертизи готується вся інформація, необхідна для прийняття рішення про інвестування. Оскільки підготовка проекту - процес ітераційний, що передбачає повернення до вже оброблених етапів і уточнення перш сформульованих ідей, слід здійснювати вибір оптимального варіанта проекту шляхом зіставлення відібраних альтернатив по їх відносній вартості та дохідності і аналізувати з різним ступенем деталізації. На цій стадії визначаються здійсненність або достатня обґрунтованість проекту в цілому і по його окремими параметрами: технічна здійсненність, ринкова ефективність, інституційна прийнятність, соціальні аспекти, фінансова та економічна цінність. Всі ці дослідження виконуються з максимальною точністю шляхом підбору оптимальних характеристик, які повторюються, з урахуванням зворотних і взаємних зв'язків, включаючи визначення всіх факторів ризику та невизначеності[3].

Так, найбільш прийнятний пакет технічних рішень буде залежати від попиту на продукцію здійснюваного проекту, адміністративних можливостей організації, яка реалізує проект, а також від культурних традицій і соціальної поведінки кінцевих виробників або споживачів.

Аналіз здійснюється за такими критеріями:

- комерційним (або маркетинговим);
- технічним;
- інституційним;
- фінансовим;
- економічним.

Мета цієї стадії - дати детальну оцінку наявних проектів, вибрати найбільш ефективний, виробити пропозиції щодо фінансування.

Під експертизою проекту розуміють його оцінку зацікавленими або незалежними організаціями за формальними і неформальними критеріями.

Завданням експертизи є перевірка раціональності проекту, визначення доцільності його реалізації. Саме на цій стадії приймається остаточне рішення про прийняття або відхилення проекту. Зазвичай експертизу виконують за окремими складовими:

Комерційна експертиза ведеться для оцінки прийнятності, доступності і цінової привабливості.

Технічна експертиза, перш за все, повинна визначити переваги технічних пропозицій.

Фінансова експертиза передбачає оцінку фінансової можливості об'єкта, який здійснює реалізацію проекту, обґрунтованість фінансових прогнозів, здатність своєчасного забезпечення покриття платежів по позиках, фінансові наслідки проекту для інвесторів, замовників і підприємства, які реалізують проект.

Основною складовою проектування є вибір відповідної технології, а також планування придбання і освоєння цієї технології і відповідного ноу-хау.

Інвестиційна фаза проекту

Інвестиційна фаза, або фаза впровадження проекту, передбачає широкий спектр консультаційних та проектних робіт, головним чином в сфері управління проектом.

Переговори та укладання угод, пов'язані з правовими зобов'язаннями, які виникають при передачі технології а також при фінансуванні. На цій стадії здійснюється підписання угод між інвестором або підприємцем, з одного боку, і фінансовою установою, консультантами - з іншого. Стадія технічного проектування передбачає остаточний вибір технології та обладнання, здійснення всього діапазону робіт з планування і складання графіків.

Виробничий маркетинг безпосередньо аналізує готовність ринку до появи продукту проекту. Набір і навчання персоналу проводиться одночасно з етапами розробки проектів і може мати вирішальне значення для очікуваного підвищення продуктивності і ефективності роботи. Особливо важлива ця

стадія для великомасштабних проектів, які пов'язані із залученням значної кількості робочих, до яких пред'являються серйозні вимоги у професійній і кваліфікаційній підготовці. Навчання є невід'ємною частиною інвестиційної фази для проектів, які використовують нові технології, ноу-хау, які вимагають особливих навичок персоналу, а також нових прийомів роботи, які підвищують ефективність проектів взагалі.

Здача в експлуатацію, як правило, обмежені за часом, але технічно важливий період здійснення проекту. Він пов'язує інвестиційну фазу з експлуатаційною. Здача проекту в експлуатацію охоплює такі види робіт:

- передексплуатаційні перевірки;
- пробні пуски;
- експлуатаційні випробування;
- прийняття.

Найбільш простий і популярний метод поділу процесу реалізації проекту на певні дії і визначення тривалості кожного з них називається графіком Ганта. Якщо в передінвестиційній фазі головним критерієм оцінки її успішності є якість і надійність проекту, то на інвестиційній фазі першочергового значення набуває фактор часу, який пов'язаний з потребою утримати проект в межах прогнозованих даних, які отримані в обґрунтуванні.

Експлуатаційна фаза проекту

Експлуатаційна фаза повинна розглядатися як з точки зору короткострокових основ (які стосуються початку виробництва якщо можуть виникати проблеми, пов'язані із застосуванням технології через нестачу кваліфікованого персоналу) так і довгострокових основ (які стосуються обраної стратегії і загальних витрат на маркетинг, а також надходжень від продажів ці фактори безпосередньо пов'язані з прогнозом, зробленим в передінвестиційній фазі).

Якщо стратегії і перспективні оцінки виявляються помилковими, внесення будь-яких коректив буде не тільки складним, але і вкрай дорогим.

На цій фазі виділяють такі стадії:

- заміна та оновлення;
- розширення та інновації;
- заключна оцінка проекту.

Заклучна оцінка проекту ведеться після реалізації проекту та фактично є ретроспективним аналізом всіх виконаних за проектом робіт. Здійснення оцінки проекту, з одного боку, надає можливості встановити фактори успіху або причини провалу проекту (буде служити інформаційною базою для розробки наступних проектних рішень), з іншого боку, - визначити, наскільки ефективно використовувалися ресурси для досягнення поставлених проектних цілей.

Найбільш важливим моментом в оцінці проекту є збір інформації, який дозволяє на підставі узагальнення успішного або невдалого досвіду проектних рішень уникнути помилок при підготовці і розробці майбутніх проектів.

3.1.1. Розробка проекту побудови рекомендацій в соціальних мережах в режимі онлайн

Об'єктом розробки є рекомендаційна система для соціальної мережі.

Програмний продукт матиме ім'я Anylink і виконуватиме такі функції:

Рекомендація користувачів користувачам в режимі онлайн.

Головною метою цієї рекомендаційної системи буде висока швидкість та низька затримка під час виконання рекомендацій у режимі онлайн.

Призначення розробки

Призначенням розробки є потрібність у швидких рекомендаціях користувачам. Повільна рекомендація знизила відвідуваність сайту на 20% що привело до втрати популярності даного ресурсу.

Нова реалізація та покращена рекомендація повинна повисити відвідуваність ресурсу через покращення рекомендацій та збільшення швидкості рекомендацій.

Технічні потреби до програми або програмного продукту

Програмний продукт повинен виконувати наступні функції:

Аналіз існуючої бази користувачів та використання даних для рекомендації

- Рекомендації в режимі онлайн
- Потреби до надійності програмного продукту.
- Програмний продукт повинен стабільно працювати весь робочий період.
- Вірогідні помилки не повинні приводити к втраті даних.

Техніко економічні показники

Економічна ефективність розробки є достатньо значущою. Вона дозволить покращити якість та швидкість рекомендацій не підвищуючи розрахункові потужності. Це підвищить відвідуваність сайту та збереже кошти на розширення інфраструктури.

Розробка програмного продукту буде складатися з наступних задач і зображена на рис 3.1:

☑	▾ Розробка проекту	39 днів	1/11/18	25/12/18		
☑	▾ Планування	17.5 днів	1/11/18	26/11/18		
☑	▾ ▾ Визначення потреб	7 днів	1/11/18	9/11/18		
☑	Визначення бізнес цілей	4 днів	1/11/18	6/11/18		Business analyst, Manager
☑	Написання потреб до продукту	3 днів	7/11/18	9/11/18	4	Business analyst, Senior data analyst
☑	▾ ▾ Проектування Структури рекомендацій	13.5 днів	7/11/18	26/11/18		
☑	Аналіз існуючих алгоритмів рекомендацій	3.5 днів	7/11/18	12/11/18	4	Data analyst
☑	Вибір та перевірка існуючого методу рекомендацій	2 днів	12/11/18	14/11/18	7	Senior data analyst
☑	Удосконалення структури та алгоритму рекомендації	6 днів	14/11/18	22/11/18	8	Senior data analyst
☑	Симуляція рекомендацій з використанням алгоритму	1 день	22/11/18	23/11/18	9	Data analyst
☑	Презентація алгоритму	1 день	23/11/18	26/11/18	10	Data analyst, Manager
☑	▾ ▾ Розробка сервісу рекомендацій	14 днів	26/11/18	14/12/18		
☑	Розробка схеми зберігання даних для рекомендацій	2 днів	26/11/18	28/11/18	11	Senior Backend dev
☑	Імплементация алгоритму на платформі	2 днів	28/11/18	30/11/18	13	Senior data analyst, Senior Backend dev
☑	Розробка конекторів до сервісу	4 днів	30/11/18	6/12/18	14	Senior data analyst
☑	Інтеграція з існуючою системою	5 днів	30/11/18	7/12/18	14	Backend Dev, Senior Backend dev
☑	Тестування стабільності системи	1.5 днів	28/11/18	7/12/18	13	Backend Dev, Data analyst
☑	Інтеграційне тестування на тестовій вибірці	4 днів	28/11/18	4/12/18	13	Backend Dev
☑	Коригування алгоритму	6 днів	6/12/18	14/12/18	15	Senior Backend dev
☑	▾ ▾ Завершение	7.5 днів	14/12/18	25/12/18	19	
☑	Тестування на навчальній вибірці	3 днів	14/12/18	19/12/18	19	Testing engineer
☑	Опис документації	3.5 днів	19/12/18	25/12/18	21	Backend Dev, Business analyst, Data analyst
☑	Відладка алгоритму на великій вибірці	3 днів	19/12/18	24/12/18	21	Backend Dev, Senior data analyst
☑	Перевірка та виправлення помилок	2 днів	19/12/18	21/12/18	21	Testing engineer
☑	Аудит	1 день	21/12/18	24/12/18	24	Senior Backend dev
☑	Запуск системи в експлуатацію	1 день	24/12/18	25/12/18	25	Manager

Рис 3.1 Перелік задач для розробки рекомендаційної системи.

Також до плану була створена діаграма ганта з зображеними на ній завданнями.

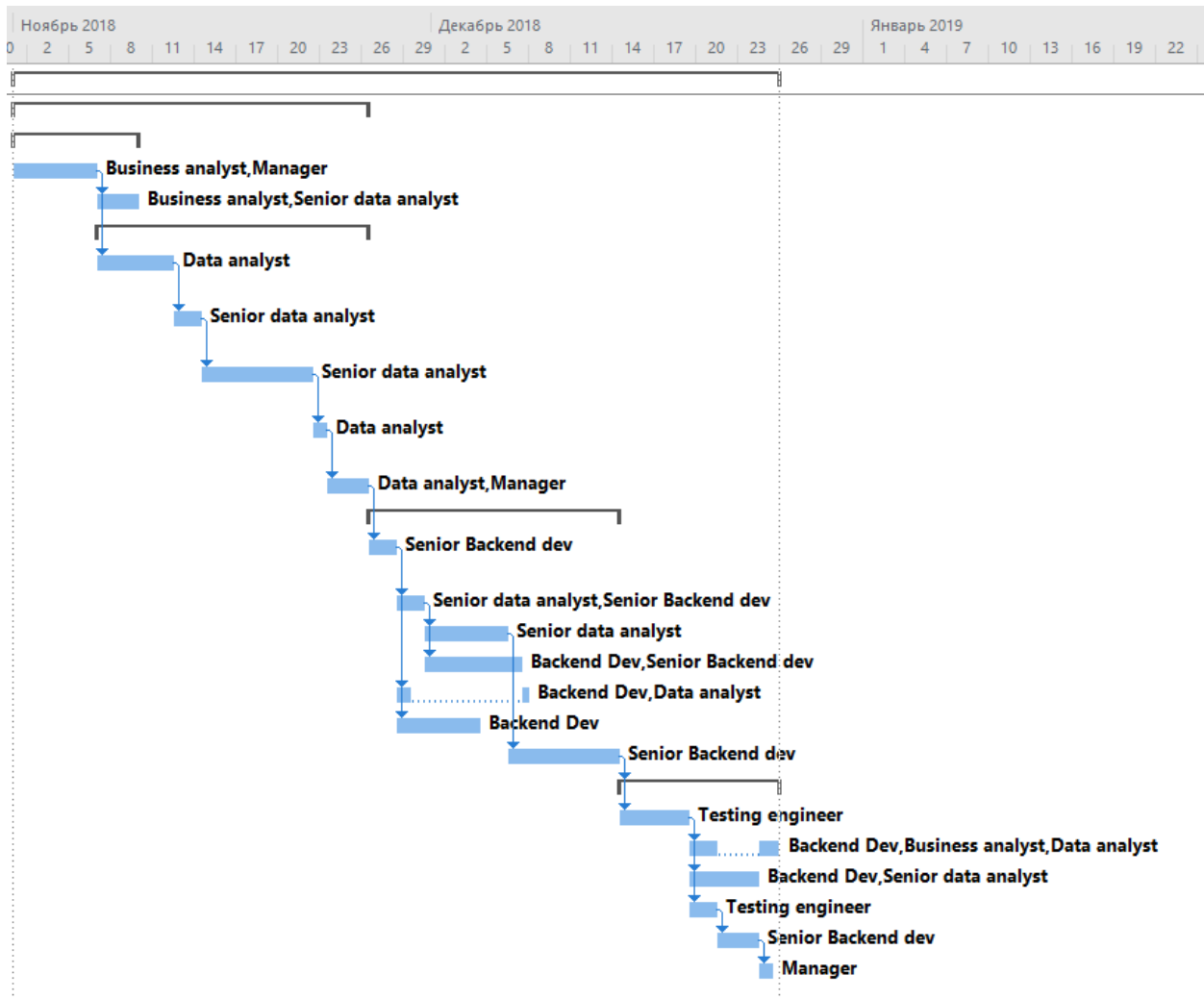


Рис 3.2 Діаграма ганта для проекту з розробки рекомендаційної системи для соц мережі.

Тривалість проекту по реалістичній оцінці складатиме 39 робочих днів. Використовується каскадна(waterfall) модель життєвого циклу проекту.

3.2. Детальний опис удосконаленого методу колаборативної фільтрації в режимі онлайн.

3.2.1. Етап попередньої демографічної фільтрації

Етап попередньої демографічної фільтрації необхідний для того, щоб відкинути людей, яких не можна рекомендувати. Він включає в себе набір правил, за якими кожен користувач повинен бути відкинутий за демографічним компонентом. У багатьох соціальних мережах такими вирішальними атрибутами є такі параметри як мова, соціальний статус та місто. Але в соціальних мережах, основною метою яких є знайомство з друзями, важливо підбирати людей, чия мета приєднання до мережі однакова. Наприклад, якщо людина шукає серйозних відносин, то система не може запропонувати людину, яка шукає тільки дружби.

Аналогічна ситуація виникає і в разі мовного атрибута. Якщо дві людини не говорять хоча б на одній спільній мові, то немає ніякої можливості, що вони можуть спілкуватися один з одним. Однак цей атрибут не має настільки стратегічного значення для професійних мереж. Наприклад, якщо є лише кілька вчених, які працюють над дуже важливим і унікальним рішенням, мовне обмеження може бути подолане. Вони можуть використовувати різні види перекладачів або звернутися за допомогою до перекладача.

В результаті правила, включені в попередню фільтрацію, будуть відрізнятися в залежності від функціональності соціальної мережі.

Результатом правильного вибору атрибутів для фільтрації є попередня фільтрація по невеликій кількості атрибутів яка опрацьовується швидше і дає перевагу в швидкості обробки рекомендацій.

Приклад правил для віртуальних спільнот, таких як Myspace, які аналізуються буде наступним:

Вхідні дані:

Множина користувачів $U = \{U_k\}; U_k = \{c, l, a, s\}$,

де: c – країна користувача;

l – мова користувача;

s – соціальний статус користувача;

a – вік користувача.

Обмеження за демографічною ознакою: $R = \{[a_1, a_2], C, L, S\}$

де C – множина країн;

L – множина допустимих мов;

S – множина статусів користувача в соц мережі;

$[a_1, a_2]$ – допустимий діапазон віку користувача.

Відбір користувачів виконується за наступних умов:

$$U^* = \{U_k: \forall U_k c \in C, l \in L, s \in S, a_1 \leq a \leq a_2\} \quad (1.10)$$

Після того, як люди, які не виконують правила фільтрації, були відхилені, можна виконати наступний крок процесу рекомендації. Остаточне схожість між користувачами розраховується для того, щоб знайти людей, які потенційно є кращими знайомими для користувача, який отримує пропозиції.

3.2.2. Етап знаходження повної східності між користувачами

Наступним етапом є знаходження повного співпадання між користувачами. Мета процесу рекомендації полягає в тому, щоб з'ясувати, чи

слід рекомендувати людину у людині X. Це досягається за допомогою кінцевої функції подібності $r(x \rightarrow y)$:

$$r(x \rightarrow y) = \alpha \cdot d(x, y) + \beta \cdot s(x) + \gamma \cdot c(x, y) + \delta \cdot a(y) + \lambda sr(y) \quad (1.11)$$

де:

$d(x, y)$ - пряме схожість між користувачем x і y , отримане в результаті порівняння всіх атрибутів з процентних і компоненту вподобань компонентів профілю користувача.

$s(x, y)$ - функція пошуку відповідності, яка відповідає атрибутам, отриманими з пошуку, виконаного x .

$c(x, y)$ - комплементарна функція ініціювання відносин описує соціальну поведінку користувачів у контексті встановлених відносин

$a(y)$ - активність користувача у розраховується на основі інформації, включеної в дію.

$\alpha, \beta, \gamma, \delta, \lambda$ - коефіцієнти вагомості, сума яких дорівнює 1. Вони використовуються для стимулювання і коригування еволюції соціальної мережі. Наприклад, якщо α низький і β високий, то метою мережі є побудова широкої мережі незалежно від взаємного прямого подібності користувачів. Значення цих факторів жорстко залежать від політики.

Так як сума коефіцієнтів $\alpha, \beta, \gamma, \delta$ і λ дорівнює 1, а значення всіх п'яти компонент знаходяться в діапазоні $[0,1]$, то значення кінцевої функції подібності $r(x \rightarrow y)$ належить діапазону $[0,1]$.

Пряма схожість $d(x, y)$ між двома користувачами x і y -це функція, яка порівнює пару атрибутів з демографічних, бажаних і зацікавлених компонентів, що характеризують обох цих членів мережі. Схожість пар атрибутів може бути досліджено, з одного боку, з допомогою прямого, індивідуального порівняння їх (див. розділ 4.8.1.1), наприклад, система робить порівняння між заняттями користувачів x і Y . З іншого боку, можна ввести ієрархію і вимірювання подібності між атрибутами на певному рівні (див.

розділ 4.8.1.2). Ієрархія побудована для компонента інтересу, а також атрибутів місця розташування і рідного міста

$$d(x, y) = \frac{w_1 \cdot f_1(x, y) + w_2 \cdot w_2 \cdot f_2(x, y) + \dots + w_n \cdot f_n(x, y)}{w_1^{max} \cdot f_1^{max}(x, y) + w_2^{max} \cdot (x, y) + \dots + w_n^{max} \cdot f_n^{max}(x, y)} \quad (1.12)$$

Де:

n – кількість атрибутів

$w_1 \dots w_n$ – ваги які прив'язані до атрибутів

$f_1(x, y), \dots, f_n(x, y)$ - функції подібності між x та y з урахування пар атрибутів $1, \dots, n$.

$w_1^{max} \dots w_n^{max}$ - максимальна вага яка може будти прив'язана до атрибуту

$f_1^{max}(x, y), \dots, f_n^{max}(x, y)$ – максимальне значення функцій $f_1(x, y), \dots, f_n(x, y)$

3.2.3. Активність користувача

Активність користувачів визначає, чи є їх поведінка більш активнішою ніж у інших, наприклад, їх оновлення фотографій/фільмів/блогів. Крім того, необхідно поважати загальний контекст діяльності в мережі, тобто всіх інших учасників, які є кращими і гіршими у своїй діяльності по оновленню. Крім того, користувач може бути більш активним в один період і менш активним в інший, і найстаріші періоди повинні чинити найменший вплив на кінцевий показник активності. Якщо мета полягає в тому, щоб знайти потенційних друзів користувача x , то діяльність всіх людей в співтоваристві повинна бути досліджена[14].

Приклади активності зображені на рис 3.3

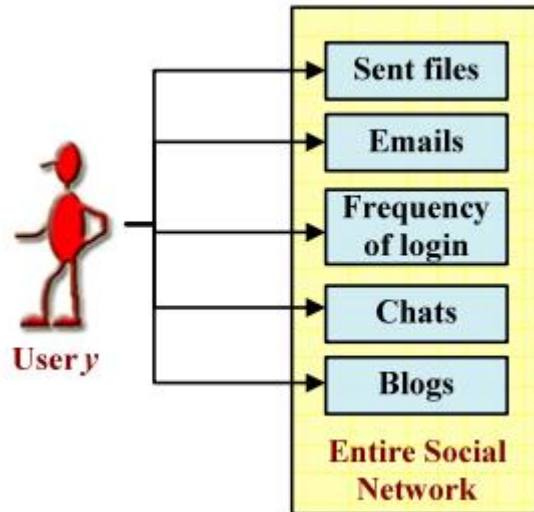


Рис 3.3 Соціальні активності користувача

Активність користувача $a(y)$ може бути описана як

$$a(y) = \frac{a_1(y) + a_2(y) + \dots + a_n(y)}{a_1^{\max}(y) + a_2^{\max}(y) + \dots + a_n^{\max}(y)} \quad (1.13)$$

де:

$a_1(y), \dots, a_n(y)$ - компонент функцій активності який описує частоту якихось дій(загрузка відео, підтримка альбому, частота входу в аккаунт)

$a_1^{\max}(y), \dots, a_n^{\max}(y)$ - максимальні значення функцій $a_1(y), \dots, a_n(y)$

n – кількість атрибутів

3.2.4. Соціальна фільтрація

Коли функція подібності вже обчислена і список L містить людей, які потенційно можуть бути рекомендовані користувачеві x , слід виконати процес соціальної фільтрації.

Соціальна фільтрація враховує соціальні характеристики людей і їх когнітивні обмеження. Він складається з шести елементів, і проти всіх них досліджується кожен член спільноти (рис. 3.4)

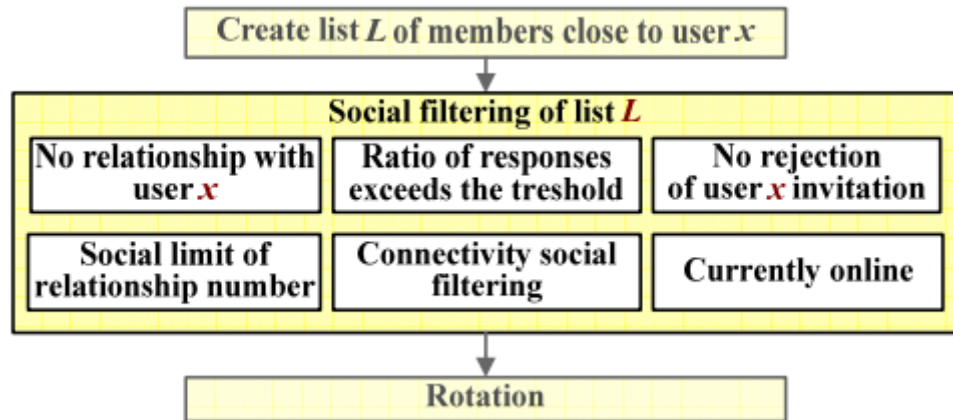


Рис 3.4 Фільтрація за соціальною активністю.

Перш за все, система рекомендацій перевіряє, чи особа, яка потенційно може бути рекомендована користувачеві x , минулого y відносинах з користувачем X . Якщо це так, то ця людина відхиляється набору людей, які можуть бути рекомендовані користувачеві X .

Другий елемент відповідає за розслідування, якщо відношення відповідей на запрошення користувача рекомендувати перевищує фіксований поріг. Це важливий елемент, тому що немає сенсу пропонувати людину, яка ніколи не відповідав на отримані запрошення. Цей поріг повинен встановлюватися в контексті всієї спільноти. Наприклад, алгоритм кластеризації може бути використаний для класифікації користувачів на певні групи респондентів, щоб оцінити, хто найчастіше відповідає на запрошення, а хто взагалі не відповідає. Соціальна фільтрація також дає можливість відхилити людей, які вже відповіли негативно на запрошення користувача X . Це цілком природне заняття, тому що якщо ми будемо рекомендувати такої людини, то це лише викличе роздратування користувача x і, як наслідок, цей користувач перестане користуватися рекомендаціями.

Крім того, деякі дослідження показали, що число стабільних відносин, які може підтримувати один чоловік становить близько 150. Це число, зване числом Данбара, визначає соціальну межу числа відносин. Інші вчені вказують, що ця соціальна межа набагато менше і дорівнює 20 людям. З досвіду людей, які є членами віртуальних спільнот, ця менша межа здається більш розумною. З цієї причини система не повинна рекомендувати членів, у яких вже було більше 20 відносин.

Після цього, в залежності від стратегії розвитку і цілі мережі, соціальна фільтрація підключення використовується для сприяння створенню зв'язків або всередині однієї групи, або між групами. Тут може бути використана концепція з'єднання і наведення мостів. Люди з тієї ж групи будуть рекомендовані, якщо сильні групи підтримуються, в той час як крос-групові рекомендації створюються для "вирівнювання" соціальної мережі.\

Крім того, система в рамках свого соціального фільтра просуває людей, які в даний час знаходяться в мережі, щоб забезпечити можливість швидкого відповіді на запрошення.

3.2.5. Результат

Створення рекомендаційної системи-дуже складне завдання. По-перше, був представлений збір відповідних даних та інтеграція цих даних. Після цього були описані елементи, які повинні бути прийняті до уваги. Мета використаних прикладів полягала в тому, щоб полегшити розуміння методу рекомендацій.

Дві основні частини методу-це розрахунок прямого подібності між користувачами і елементом соціальної фільтрації. Перший концентрується на обчисленні подібності між користувачами, їх активності і міцності їх відносин. Останнім враховує соціальні обмеження людини, такі як кількість знайомств, які може підтримувати один чоловік.

4. СТРУКТУРА РЕКОМЕНДАЦІЙНИХ СИСТЕМ

4.1. Структура зберігання елементів в рекомендаційній системі

Так як рекомендаційна система буде використовуватися для соціальної мережі. Схема зберігання даних про користувачів буде відображати тільки дані, які використовуються для рекомендацій. На рис 4.1 зображена базова схема даних про користувача в соціальній мережі.

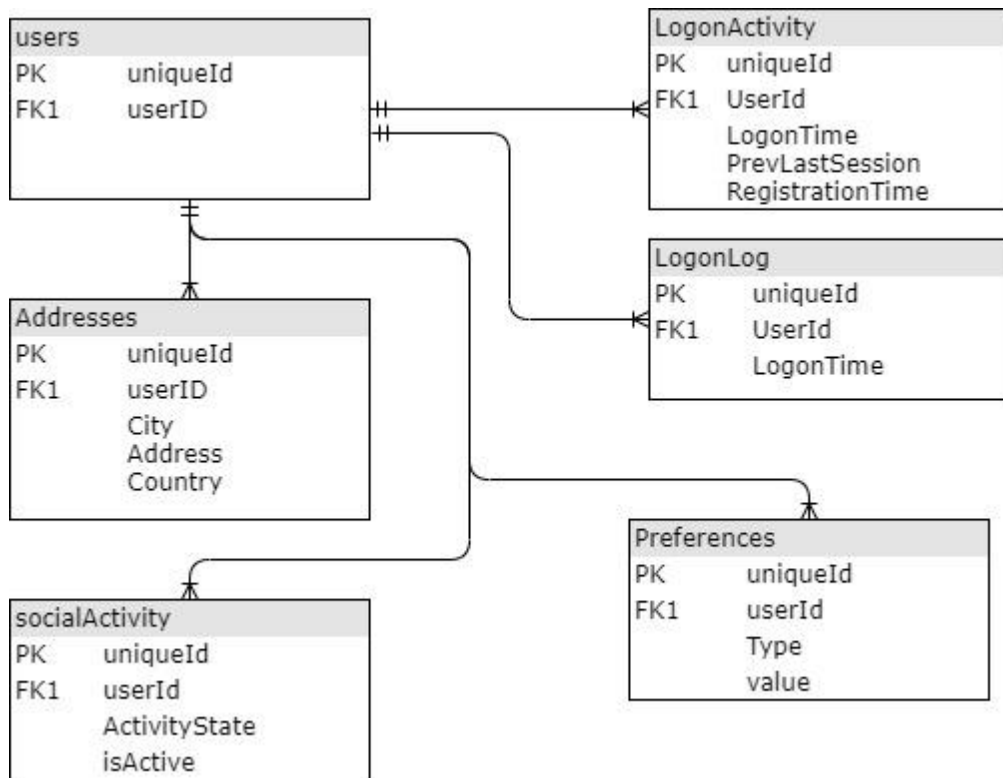


Рис 4.1 Схема зберігання атрибутів яка використовується в рекомендаційній системі.

4.2. Експериментальна перевірка удосконаленого методу

Метою експерименту є визначення показника AUC – площі, яка обмежена ROC кривою і віссю частки помилкових позитивних класифікацій та швидкості обробки рекомендацій при збереженні показника AUC. Чим вище показник AUC, тим якісніше класифікатор, при цьому значення 0,5 демонструє непридатність обраного методу класифікації.

Використовуючи формулу пошуку площу, обмежену ROC-кривою і віссю, буде проведено 3 пошуки показника AUC за допомогою мови програмування R в середовищі розробки R-studio. Пошук буде виконуватися для вибірки без урахування попередньої демографічної фільтрації та з додаванням попередньої демографічної фільтрації. Додатково буде розглянутий алгоритм без попередньої фільтрації (проблема, коли результат дорівнює 0).

Розглянемо базову ситуацію в соціальній мережі, коли користувач заходить на свою сторінку та відкриває лист своїх контактів. У багатьох соціальних мережах існує такий блок, як Рекомендовані контакти.

Для цього експерименту, користувач матиме такі поля.

Таблиця 4.1 – Вихідні дані експерименту

Поле	Значення
ПІБ	Іван Іванов Миколайович
Стать	Чоловік
Дата народження	12.01.1990
Місто	Харків
Мова	Українська, англійська

Активність авторизацій	Мапа (дата:кількість авторизацій)
Інтереси	Мапа(тип інтересу:перелік інтересів)
Соціальний статус	Вільний

Спочатку ми імпортуємо бібліотеки які нам потрібні для обробки даних користувача.

```
install.packages("recommenderlab")
install.packages("recosystem")
library(devtools)
library(recommenderlab)
library(recosystem)
library(SlopeOne)
library(SVDApproximation)
library(data.table)
library(RColorBrewer)
library(ggplot2)|
```

Рисунок 4.1.1 – Скріншот програмного коду

Потім ми беремо сгенерований датасет користувачів для рекомендацій зображений на рис 4.1.2.

id	name	surname	yearold	city	country	status	interests_music	interests_films
12	anna	ivanova	19.11.1990	Kiev	Ukraine	Searching for relations	[AM,Pink Floyd...]	[Tenet,interstellar...]
13	inna	sereda	11.11.1991	Kharkiv	Ukraine	Married	[TLSP,LSP,Ljay...]	[T-serries,TOFW...]
14	olena	vovchok	22.12.1992	Kharkiv	Ukraine	Divorced	[loboda,inna,Ljay...]	[Tenet,fast and furious...]
15	ivan	bogun	03.07.1993	Poltava	Ukraine	not in a search	[AM,Pink Floyd...]	[Tenet,inception...]
16	denys	kharko	27.01.1994	Odessa	Ukraine	Divorced	[loboda,inna,Ljay...]	[Peaky blinders,interstellar...]

Рисунок:4.1.2 – Приклад вхідної інформації для рекомендації

Тепер передамо дані до звичайного алгоритмів рекомендації

```
matrix<-data(users)
rec_users<-recommenderDefault(current_user,matrix)
preproc_users<-demographic_preprocess(current_user,matrix)
rec_users_adv<-recommenderDefault(current_user,preproc_users)
```

Рис 4.1.3 Запуск різних методів рекомендацій для аналізу.

Після цього, отримаємо рекомендації. Для цих трьох алгоритмів на невеликій вибірці результати будуть однаковими.

14	olena	vovchok	22.12.1992	Kharkiv	Ukraine	Divorced	[loboda,inna,Ljay...	[Tenet,fast and furious...]
13	inna	sereda	11.11.1991	Kharkiv	Ukraine	Married	[TLSP,LSP,Ljay..]	[T-serries,TOFW...]
12	anna	ivanova	19.11.1990	Kiev	Ukraine	Searching for relations	[AM,Pink Floyd...]	[Tenet,interstellar...]

Рис 4.1.4 Отримані рекомендації

Сгенеруємо вибірку розміром 10 тис. користувачів. Проведемо тіж самі операції але з вимірами по часу. Як результат отримаємо діаграму про затрачений час на обробку рекомендацій.

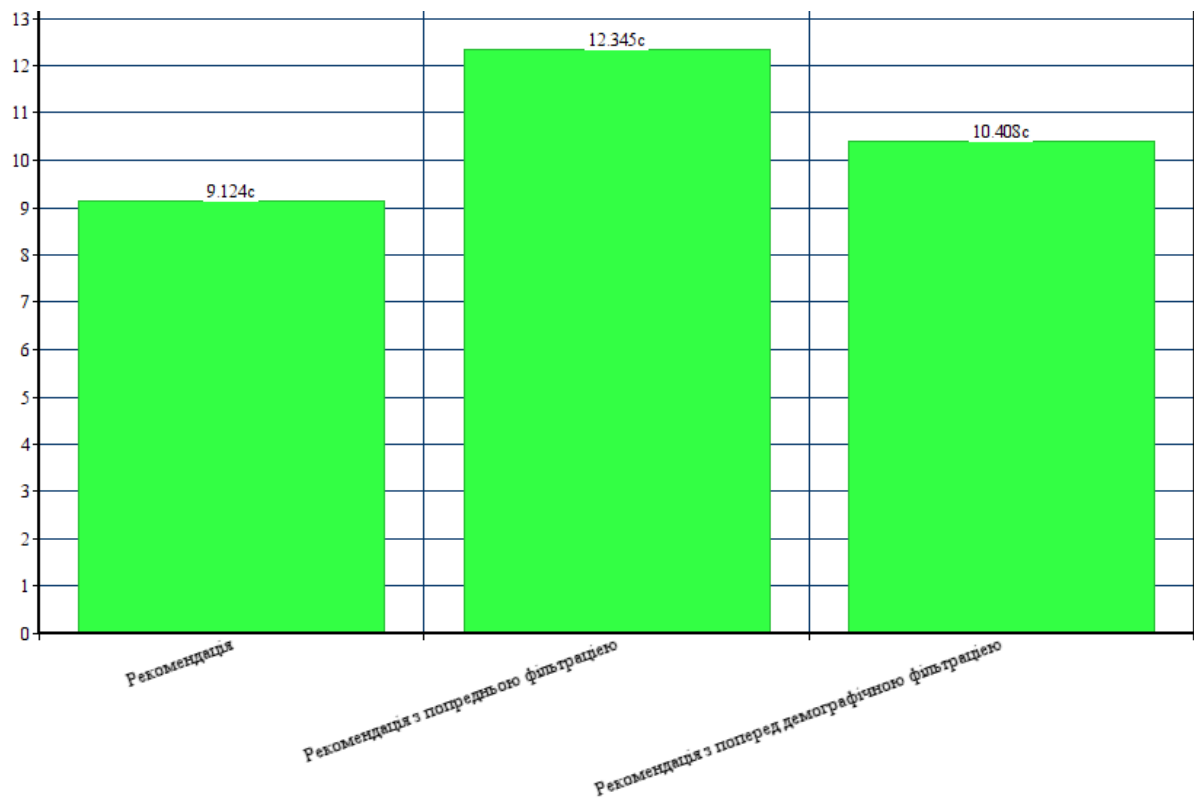


Рис 4.1.5 Діаграма швидкості обчислення рекомендацій на вибірці розміром 10 тис користувачів.

Після того, як були отримані результати рекомендацій та швидкості обчислення. Можна перевірити методи за допомогою показника ROC.

Як базовий еталонний алгоритм будемо вважати алгоритм з попередньою фільтрацією. Він є використовуваним у більшості соціальних мереж.

Результати пошуку AUC-показника представлені в таблиці 4.2

Таблиця 4.2 Значення AUC-показника для різних методів фільтрації

Метод	AUC
КФ без попередньої фільтрації	(0,862)
КФ з попередньою фільтрацією	(0,880)
КФ з попередньою демографічною фільтрацією	(0,871)

По результатам експерименту бачимо, що використання КФ з попередньою демографічною фільтрацією не сильно програє в якості формування рекомендацій і працює набагато швидше ніж КФ з попередньою фільтрацією.

ВИСНОВКИ

В результаті виконання магістерської роботи був виконаний огляд методів рекомендацій в соціальних мережах. Проаналізовано існуючі методи фільтрацій для рекомендаційних систем.

Була розглянута проблема рекомендацій для соціальних мереж. А саме колаборативної фільтрації з використанням додаткових попередньої фільтрації по демографічному компоненту, який прискорює процес рекомендації і дозволяє збільшити швидкість рекомендацій користувачам соціальних мереж.

В практичному аспекті метод дозволяє покращити показник швидкості від 12,145с. до 10,408 с.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Методичні вказівки щодо розробки та оформлення магістерської атестаційної роботи за спеціальністю 8.05010101 – Інформаційні управляючі системи та технології. Освітньо-кваліфікаційний рівень – магістр / Упоряд.: Левикін В.М., Міхнов Д.К., Саєнко В.І., Євланов М.В., Міхнова А.В., Керносів М.А. – Харків: ХНУРЕ, 2012. – 28 С.
2. Schafer, J. B., Konstan, J. and Riedi, J. (1999). Recommender systems in e-commerce. In Proceedings of the 1st ACM conference on Electronic commerce, pp. 158–166.
3. Adomavicius, G. and A. Tuzhilin, 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE T. Knowl. Data En., 17(6): 734-749.
4. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-adapted Interaction, Vol. 12, No. 4, pp. 331–370.
5. Melville, P. and Sindhvani, V. (2010). Recommender systems. Encyclopedia of Machine Learning, Vol. 1, pp. 829–838.
6. Blanco-Fernández, Y., Pazos-Arias, J. J., Gil-Solla, A., Ramos-Cabrera, M., López-Nores, M., García-Duque, J., Fernández-Vilas, A., Díaz-Redondo, R. P. and Bermejo-Muñoz, J. (2008). A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems. Knowledge-Based Systems, Vol. 21, No. 4, pp. 305-320.
7. Chen, T. and He, L. (2009). Collaborative filtering based on demographic attribute vector. In Future Computer and Communication, 2009. FCC'09. International Conference on, pp. 225–229.

8. Towle, B. and Quinn, C. (2000). Knowledge based recommender systems using explicit user models. In Proceedings of the AAAI Workshop on Knowledge-Based Electronic Markets, pp. 74–77.
9. Carrer-Neto, W., Hernández-Alcaraz, M. L., Valencia-García, R. and García-Sánchez, F. (2012). Social knowledge-based recommender system. Application to the movies domain. Expert Systems with Applications, Vol. 39, No. 12, pp. 10990–11000.
10. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання контексту в рекомендаційних системах. Науковий журнал «Проблеми інформаційних технологій», 2018, №. 1(023). С. 21-26..
11. Mobasher, B., Jin, X. and Zhou, Y. (2004). Semantically enhanced collaborative filtering on the web. In Web Mining: From Web to Semantic We, Springer, pp. 57–76.
12. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Інтеграція локальних контекстів споживачів в рекомендаційних системах на основі відношень еквівалентності, схожості та сумісності. Process mining Materials of the VII International Scientific Conference «Information-Control System and Technologies» 17th-18th September, 2018, Odessa. С.142-144.
13. Davoodi, E., Kianmehr, K. and Afsharchi, M. (2013). A semantic social network-based expert recommender system. Applied Intelligence, Vol. 39, No. 1, pp. 1–13.
14. Lee, D., Park, S. E., Kahng, M., Lee, S. and Lee, S. (2010). Exploiting contextual information from event logs for personalized recommendation. In Computer and Information Science 2010, Springer, pp. 121–139.
15. Vozalis, M. and Margaritis, K. G. (2004). Enhancing collaborative filtering with demographic data: The case of item-based filtering. In 4th International Conference on Intelligent Systems Design and Applications, pp. 361–366.

17. Ma, H., Zhou, T. C., Lyu, M. R. and King, I. (2011). Improving recommender systems by incorporating social contextual information. *ACM Transactions on Information Systems (TOIS)*, Vol. 29, No. 2, p. 9.
18. Wang, J., De Vries, A. P. and Reinders, M. J. (2006). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 501–508.
19. Billsus, D. and M.J. Pazzani, 1998. Learning collaborative information filters. *Proceeding of the 15th International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp: 46-54.
20. Billsus, D. and M.J. Pazzani, 2000. User modeling for adaptive news access. *User Mod. User-adapted Interac.*, 10(2-3): 147-180.
21. Claypool, M., P. Le, M. Wased and D. Brown, 2001. Implicit interest indicators. *Proceeding of International Conference on Intelligent User Interfaces*, pp: 33-40.
22. Pazzani, M.J. and D. Billsus, 2007. Content-based Recommendation Systems. In: Brusilovsky, P., A. Kobsa and W. Nejdl (Eds.), *the Adaptive Web*. LNCS 4321, Springer-Verlag, Berlin Heidelberg, pp: 325-341.
23. Lops, P., Degemmis, M. and Semeraro, G. (2007). Improving social filtering techniques through WordNet-Based user profiles. *User Modeling 2007*. Springer Berlin Heidelberg, pp. 268-277.
24. Liu, H., Hu, Z., Mian, A., Tian, H. and Zhu, X. (2014). A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems*, Vol. 56, pp. 156–166.
25. Sun, D., Luo, Z. and Zhang, F. (2011). A novel approach for collaborative filtering to alleviate the new item cold-start problem. In *Communications and Information Technologies (ISCIT)*, 2011 11th International Symposium on, IEEE, pp. 402–406.

26. Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D. and Sartin, M. (1999). Combining content-based and collaborative filters in an online newspaper. In Proceedings of ACM SIGIR workshop on recommender systems, Vol. 60.
27. Basu, C., Hirsh, H. and Cohen, W. (1998). Recommendation as classification: Using social and content-based information in recommendation. In Proceedings of the national conference on artificial intelligence, pp. 714–720.
28. De Campos, L. M., Fernández-Luna, J. M., Huete, J. F. and Rueda-Morales, M. A. (2010). Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. *International Journal of Approximate Reasoning*, Vol. 51, No. 7, pp. 785–799.
29. Tran, T. and Cohen, R. (2000). Hybrid recommender systems for electronic commerce. In Proc. Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, Technical Report WS-00-04, AAAI Press.
30. Ceylan, U. and Birturk, A. (2011). Combining Feature Weighting and Semantic Similarity Measure for a Hybrid Movie Recommender System. In The 5th SNA-KDD Workshop'11.
31. Sieg, A., Mobasher, B. and Burke, R. (2010). Improving the effectiveness of collaborative recommendation with ontology-based user profiles. In proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems, pp. 39–46.
32. Hu, B. and Zhou, Y. (2008). Content Semantic Similarity Boosted Collaborative Filtering. In Computational Intelligence and Security, 2008. CIS'08. International Conference on, Vol. 2, pp. 7–11.