

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА ТЕХНОЛОГІЙ

Кафедра Комп'ютерних інформаційних технологій

ДОПУСТИТИ ДО ЗАХИСТУ
Завідувач випускової кафедри
Аліна САВЧЕНКО.
«_____» _____ 2023р.

КВАЛІФІКАЦІЙНА РОБОТА
(ДИПЛОМНИЙ ПРОЄКТ, ПОЯСНЮВАЛЬНА ЗАПИСКА)
ВИПУСКНИКА ОСВІТНЬОГО СТУПЕНЯ “БАКАЛАВР”
ЗА ОСВІТНЬО-ПРОФЕСІЙНОЮ ПРОГРАМОЮ
“ІНФОРМАЦІЙНІ УПРАВЛЯЮЧІ СИСТЕМИ ТА ТЕХНОЛОГІЇ”

Тема: “Застосування нейронних мереж в системах рекомендацій
на прикладі Netflix”

Виконавець: студентка групи УС-411Б Кічапіна Марія Олександрівна

Керівник: к.т.н., доцент Харченко Олександр Григорович

Нормоконтролер: _____ Олександр ШЕВЧЕНКО

Київ – 2023

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет комп'ютерних наук та технологій

Кафедра Комп'ютерних інформаційних технологій

Галузь знань, спеціальність, освітньо-професійна програма: 12 “Інформаційні технології”, 122 “Комп'ютерні науки”, “Інформаційні управляючі системи та технології”

ЗАТВЕРДЖУЮ

Завідувач випускової кафедри

_____ Аліна САВЧЕНКО

« _____ » _____ 2023р.

ЗАВДАННЯ

на виконання кваліфікаційної роботи студентки

Кічапіної Марії Олександрівни

- 1. Тема роботи:** «Застосування нейронних мереж в системах рекомендацій на прикладі Netflix» затверджена наказом ректора № 623/ст. від 01.05.2023р.
- 2. Термін виконання роботи:** 15.05.2023 – 25.06.2023
- 3. Вихідні дані до роботи:** статті «Глибоке навчання для систем рекомендацій: вивчення на прикладі Netflix», «Рекомендаційні системи та їхні етичні проблеми».
- 4. Зміст пояснювальної записки:** історія створення нейронних мереж, види та класифікація нейронних мереж, типи рекомендаційних систем, проблеми «холодного старту» у рекомендаційних системах, етичні питання систем рекомендацій, історія розвитку Netflix, його архітектура системи персоналізації та рекомендацій, проблематика Netflix.
- 5. Перелік обов'язкового графічного матеріалу:** рисунки нейронної мережі з різною топологією зв'язків, вигляду сторінок на Netflix в різні періоду його розвитку, а також рисунки вигляду загальної схеми і більш детальних схем архітектури рекомендаційних систем Netflix.

6. Календарний план-графік

№ п/п	Завдання	Термін виконання	Підпис керівника
1.	Аналіз предметної області дослідження.	15.05.2023 – 17.05.2023	
2.	Збір, аналіз і опрацювання інформації за тематикою дипломного проекту.	18.05.2023 – 21.05.2023	
3.	Дослідження історії створення, розвитку нейромереж та видів нейромереж і сфер їхнього застосування.	22.05.2023	
4.	Класифікування нейронних мережі і дослідження способів навчання нейронних мереж.	23.05.2023 – 25.05.2023	
5.	Вивчення типів рекомендаційних систем та їхніх проблем.	26.05.2023 – 28.05.2023	
6.	Дослідження історії розвитку Netflix.	29.05.2023 – 31.05.2023	
7.	Вивчення архітектури системи персоналізації та рекомендацій, а також проблематики.	01.06.2023 – 06.06.2023	
8.	Створення презентації, доповіді та підготовка до захисту дипломної роботи.	07.06.2023 – 13.06.2023	
9.	Підготовка матеріалів дипломної роботи для передачі секретарю ДЕК (папка, конверт, диск із файлом диплому, рецензія, відгук).	14.06.2023 – 15.06.2023	

7. Дата видачі завдання: «15» травня 2023 р.

Керівник дипломної роботи _____ Олександр ХАРЧЕНКО

Завдання прийняв до виконання _____ Марія КІЧАПНА

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи «Застосування нейронних мереж в системах рекомендацій на прикладі Netflix» складається зі вступу, трьох розділів, висновку, списку бібліографічних посилань і містить 59 сторінок та 9 рисунків. Список бібліографічних посилань складається з 9 найменувань.

Мета дипломного проєкту: дослідити ефективність систем рекомендацій на прикладі Netflix.

Для досягнення поставленої мети необхідно вирішити такі **завдання:**

- проаналізувати дані наукових досліджень з рекомендаційних систем;
- дослідити типи рекомендаційних систем та їхні проблеми;
- проаналізувати архітектуру системи персоналізації та рекомендацій Netflix.

Об'єкт дослідження: нейронні мережі в системах рекомендацій.

Предмет дослідження: системи рекомендацій стримінгової платформи Netflix.

Метод дослідження: аналіз і порівняння різних типів систем рекомендацій.

Результат проєкту:

- розглянуто історія створення та розвитку нейромереж;
- наведено види нейромереж, їхня класифікація та сфера застосування;
- досліджено типи, етичні питання, проблеми рекомендаційних систем;
- проаналізовано архітектуру системи персоналізації та рекомендацій Netflix.

Розглянуті підходи в розробці системі рекомендацій: спільна фільтрація, фільтрація на основі вмісту та гібридні підходи.

- виявлено, що гібридний підхід є найкращим в розробці рекомендаційних систем.

НЕЙРОМЕРЕЖІ, СИСТЕМИ РЕКОМЕНДАЦІЙ, ХОЛОДНИЙ СТАРТ, ТИПИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ, NETFLIX, СИСТЕМИ РЕКОМЕНДАЦІЙ NETFLIX, АРХІТЕКТУРА СИСТЕМИ ПЕРСОНАЛІЗАЦІЇ ТА РЕКОМЕНДАЦІЙ NETFLIX.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ.....	7
ВСТУП.....	8
РОЗДІЛ 1. ПРИНЦИПИ ТЕХНОЛОГІЙ НЕЙРОННИХ МЕРЕЖ	10
1.1. Історія створення нейронних мереж	10
1.2. Види та класифікація нейронних мереж.....	12
1.3. Правила навчання нейронних мереж	14
1.3.1. Контрольоване навчання	15
1.3.2. Неконтрольоване навчання	17
1.3.3. Оцінка та правила навчання.....	17
1.4. Висновок до 1 розділу.....	20
РОЗДІЛ 2. СИСТЕМИ РЕКОМЕНДАЦІЙ.....	22
2.1. Типи рекомендаційних систем.....	22
2.1.1. Спільна фільтрація.....	23
2.1.2. Фільтрація на основі вмісту	25
2.1.3. Гібридні підходи	27
2.2. Проблеми «холодного старту» у рекомендаційних системах	29
2.3. Етичні питання систем рекомендацій	31
2.3.1. Неприйнятний вміст	33
2.3.2. Конфіденційність	34
2.3.3. Автономія та особиста ідентичність	35
2.3.4. Непрозорість	36
2.3.5. Справедливість	36

2.3.6. Соціальні наслідки	38
2.4. Висновок до 2 розділу	38
РОЗДІЛ 3. СИСТЕМА РЕКОМЕНДАЦІЙ В NETFLIX	40
3.1. Ключові віхи в історії розвитку Netflix.....	40
3.2. Архітектура системи персоналізації та рекомендацій Netflix	44
3.2.1. Онлайнові, ніарлайнові та офлайнові обчислення	46
3.2.2. Сигнали і моделі.....	49
3.2.3. Розподілення подій і даних	50
3.3. Проблематика	52
3.4. Висновок до 3 розділу.....	54
ВИСНОВКИ.....	56
СПИСОК БІБЛІОГРАФІЧНИХ ПОСИЛАНЬ ВИКОРИСТАНИХ ДЖЕРЕЛ.....	58

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

НМ	Нейронна мережа
РС	Рекомендаційна система
ТЧ-ЗЧД	Термін частота-зворотна частота документа
ОВМ	Опорні векторні машини
ПХС	Проблема холодного старту
ППХС	Проблема постійного холодного старту
КППХС	Проблема постійного холодного старту, яка може виникати з користувачами
ОППХС	Проблема постійного холодного старту, яка може виникати з об'єктами

ВСТУП

Системи рекомендацій стали необхідною складовою багатьох платформ та сервісів у сучасному цифровому світі. Вони допомагають користувачам знаходити та відкривати новий контент, товари, музику, фільми та інші об'єкти інтересу, що сприяє персоналізованому користувацькому досвіду.

Рекомендаційні системи використовують різні алгоритми та техніки, включаючи машинне навчання та нейронні мережі, для аналізу користувацьких даних та генерації рекомендацій на основі індивідуальних уподобань та контекстуальної інформації.

Однією з ключових проблем, які вирішують системи рекомендацій, є величезна кількість доступних варіантів. У сучасному цифровому середовищі користувачі перевантажені надзвичайною різноманітністю продуктів, послуг і контенту. Рекомендаційні системи покращують взаємодію сервісів надання послуг з користувачами, пропонуючи їм предмети, продукти або вміст, які відповідають їхнім інтересам і вподобанням. Окрім цього, системи рекомендацій діють як інтелектуальні фільтри, звужуючи вибір і пропонуючи найбільш доречні варіанти на основі індивідуальних характеристик користувача.

Платформи та сервіси, що використовують системи рекомендацій, охоплюють широкий спектр галузей, від електронної комерції та соціальних медіа до стрімінгових платформ та новинних веб-сайтів.

Платформи, такі як Facebook, Instagram, YouTube, TikTok, використовують системи рекомендацій для пропозицій контенту на основі інтересів, попередніх переглядів, рейтингів, поведінки користувачів та соціальних зв'язків.

Стрімінгові платформи, такі як Netflix, Hulu, Disney+, Amazon Prime Video, використовують системи рекомендацій для пропозицій фільмів, серіалів та інших відео на основі переглядів, оцінок, жанрів, популярності та інших факторів. Це допомагає користувачам знайти новий контент, який їм сподобається.

Наразі Netflix знаходиться на передовій досліджень систем рекомендацій. Компанія продовжує інвестувати в дослідження передових методів, таких як посилене глибоке навчання, нейронні мережі графів і оптимізація з різними цілями, щоб вдосконалити свої рекомендації та підвищити задоволеність користувачів.

РОЗДІЛ 1

ПРИНЦИПИ ТЕХНОЛОГІЙ НЕЙРОННИХ МЕРЕЖ

1.1. Історія створення нейронних мереж

Концепція нейронних мереж базується на ідеї, що мозок — це складна мережа взаємопов'язаних нейронів, які спілкуються один з одним за допомогою електричних і хімічних сигналів. У 1940-х роках дослідники почали досліджувати можливість моделювання цієї складної мережі за допомогою математичних функцій з метою створення машини, яка могла б вивчати та обробляти інформацію подібно до людського мозку.

Одними з перших піонерів у цій галузі були Уоррен МакКаллох, нейрофізіолог, і Уолтер Піттс, логік, які співпрацювали над статтею під назвою «Логічне обчислення ідей, властивих нервовій діяльності», опублікованій у 1943 році. У статті введено концепцію штучний нейрон, який був математичною функцією, яка могла приймати дані від інших нейронів і виробляти вихідний сигнал на основі набору вагових коефіцієнтів і порогів. Нейрон Мак-Каллоха-Піттса, як стало відомо, був першою спробою змоделювати поведінку біологічного нейрона за допомогою математичної функції.

У наступні десятиліття дослідники продовжували досліджувати потенціал нейронних мереж, але прогрес був повільним через обмеження обчислювальної технології, доступної на той час. Однак у 1950-х і 1960-х роках Френк Розенблатт, психолог і комп'ютерний вчений, розробив новий тип нейронної мережі, відомий як перцептрон.

Кафедра КІТ				НАУ 23 11 57 000 ПЗ			
Виконавець	Кічапіна М.О.			ПРИНЦИПИ ТЕХНОЛОГІЙ НЕЙРОННИХ МЕРЕЖ	Літера	Аркуш	Аркушів
Керівник	Харченко О.Г.					10	12
Консультант					УС-411Б 122		
Н.Контроль	Шевченко О.П.						

Перцептрон був типом нейронної мережі, яка могла навчитися класифікувати зображення, призначаючи ваги кожному пікселю зображення та порівнюючи результати з набором попередньо позначених зображень. Перцептрон був першою нейронною мережею, успішно застосованою до проблеми реального світу, і викликав велике хвилювання серед дослідницького співтовариства.

Однак незабаром було виявлено, що перцептрон має обмеження, і він може класифікувати лише лінійно роздільні зображення, що означало, що він не міг розрізнити більш складні зображення, які не були лінійно розділеними. Це призвело до зниження інтересу до нейронних мереж у 1970-х роках.

У 1980-х відбулося відновлення інтересу до нейронних мереж, частково завдяки розробці нових алгоритмів навчання та прогресу в обчислювальній потужності. Одним із ключових проривів стала розробка зворотного поширення, методу навчання нейронних мереж, який представили в 1986 році Девід Румельхарт, Джеффри Хінтон і Рональд Вільямс. Зворотне поширення дозволило нейронним мережам вивчати більш складні завдання, і це проклало шлях до розвитку багатошарових нейронних мереж, які мали кілька шарів нейронів, які могли навчитися розпізнавати більш складні шаблони.

У 1990-х роках нейронні мережі застосовувалися для широкого кола програм, включаючи розпізнавання мови, обробку зображень і фінансове моделювання. Однак обмеження нейронних мереж, такі як відсутність інтерпретації та схильність до надмірного пристосування, означали, що вони не отримали широкого застосування в промисловості.

В останні роки відновився інтерес до нейронних мереж завдяки розробці нових архітектур, таких як згортові нейронні мережі та рекурентні нейронні мережі, а також прогресу в таких областях, як обробка природної мови та комп'ютерне бачення. Сьогодні нейронні мережі широко використовуються в промисловості та дослідженнях, і вони змінили такі галузі, як розпізнавання зображень і мови, автономні транспортні засоби та відкриття ліків.

1.2. Види та класифікація нейронних мереж

Нейронні мережі призначені виконувати безліч завдань, щоб покращити та полегшити життя людині. Нейромережі використовуються у багатьох галузях та сферах діяльності людини, а зокрема, у медицині, економіці, обробці інформації, зв'язку та безпеці охоронних систем тощо. Якщо підсумувати функції нейронних мереж, то це буде: асоціативна пам'ять, стиснення даних, апроксимація функцій, розпізнавання та класифікація, оптимізаційні задачі, керування складними процесами та прогнозування.

Загалом нейромережа — це сукупність нейронів, що зв'язані певним чином. В залежності від способу поєднання нейронів, функцій нейроелементів, а також механізмів та напрямків розповсюдження сигналів по мережі виділяють різні види нейромереж.

Класифікація нейронних мереж:

- за типом вхідної інформації: аналогові та бінарні;
- за типом функції активації нейронів: неперервні, дискретні та дискретно-неперервні;
- за типом графа міжнейронних зв'язків виділяють: мережі без циклів та мережі з циклами, які в свою чергу поділяються на рівноважні мережі з циклами та мережі з обмеженими циклами;
- за типом структур нейронів виділяють: гомогенні нейромережі та гетерогенні;
- за кількістю шарів нейронів виділяють: одношарові та багатшарові нейромережі;
- та типом дискримінантної функції нейромережі поділяють на зважені та без ваг зв'язків;
- за принципом синтезу виділяють: мережі, що навчаються та мережі, що конструюються;
- за топологією зв'язків виділяють (див. рис. 1.1.): повнозв'язні, неповнозв'язні та слабозв'язні мережі;

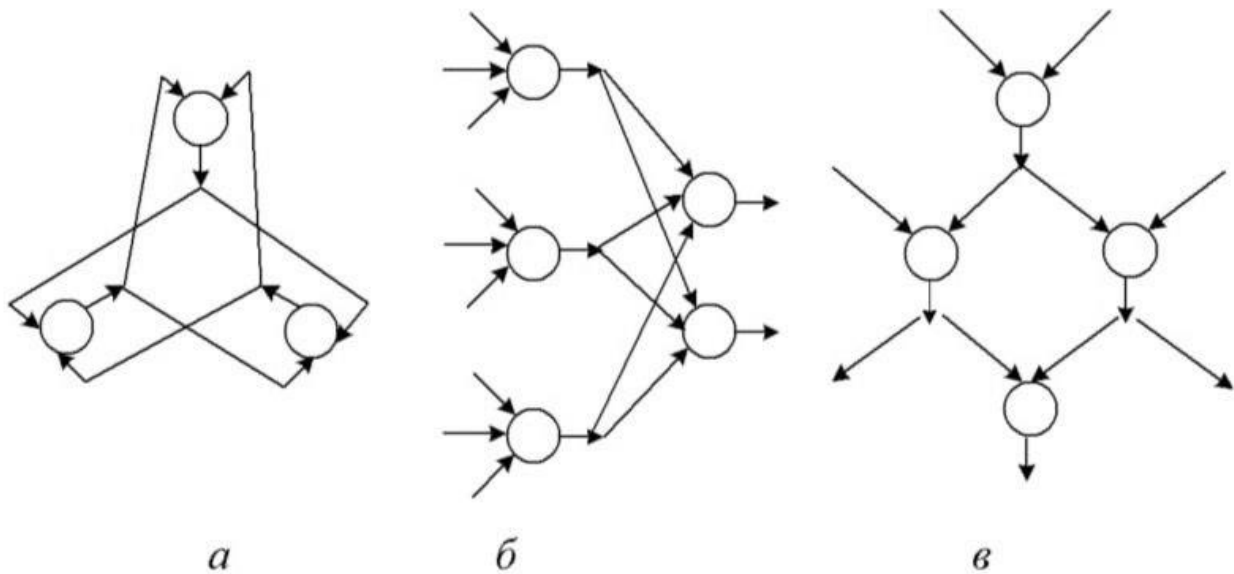


Рис. 1.1. НМ з різною топологією зв'язків: *a* - повнозв'язна мережа; *б* – шарувата мережа; *в* – слабозв'язана мережа

- за характером зв'язків виділяють: мережі прямого поширення та рекурентні мережі, що поділяються на релаксаційні та шаруваті мережі із зворотними зв'язками. Шаруваті мережі із зворотними зв'язками поділяють на шарувато-циклічні, шаруватоповнозв'язні, повнозв'язно-шаруваті, рециркуляційні та мережі з боковими зворотними зв'язками;

- за характером поділу зв'язків виділяють: немонотонні та монотонні мережі;

- за типом методу навчання: мережі з динамічними зв'язками і ітеративним навчанням, що засновано на принципі корекції помилок та мережі з фіксованими зв'язками і неітеративним навчанням;

- за характером навчання виділяють: мережі з контрольованим, неконтрольованим та змішаним навчанням;

- за методом навчання виділяють мережі: з неітеративним навчанням, з методом зворотного поширення помилки, з конкурентним навчанням (правило Хебба), з гібридним навчанням тощо;

- за типом часу функціонування виділяють мережі: з неперервним часом, з дискретним асинхронним часом та з дискретним часом, що функціонує синхронно;

- у залежності від врахування попереднього стану мережі виділяють мережі: статичні, нечіткі та нетрадиційні.

1.3. Правила навчання нейронних мереж

Найбільш важливою властивістю нейронних мереж є їхня здатність вчитися на даних про конкретну предметну галузь та підвищувати продуктивність за рахунок навчання. Продуктивність покращується з часом відповідно до певних правил.

Процес навчання можна розглядати як визначення архітектури мережі та ітеративне коригування ваг синаптичних сполук для ефективного виконання конкретного завдання.

Під час навчання нейронної мережі передбачається така послідовність подій:

- нейронна мережа отримує сигнали із зовнішнього середовища;
- змінено вільні параметри нейронної мережі;
- після зміни внутрішньої структури нейронна мережа по-різному реагує на вхідний сигнал.

Правила навчання нейронної мережі називають алгоритмом навчання. Немає загального алгоритму навчання, придатного всім архітектур нейронних мереж. Алгоритми навчання різняться за способом налаштування синаптичних вагів нейронів.

Нейронна мережа регулює ваги з'єднань відповідно до доступного навчального набору. Можливість вчитися на прикладах робить їх більш привабливими, ніж системи, що функціонують за системою правил, сформульованих експертами.

Для процесу навчання необхідна модель довкілля, у якій працює нейронна мережа, тобто інформація, необхідна для розв'язання завдання. По-друге, потрібно визначити, як модифікувати вагові параметри мережі. Алгоритм навчання відноситься до процесу використання правил навчання для коригування ваг.

Існують три загальні парадигми навчання:

- навчання під наглядом (з учителем);
- навчання без нагляду (самотужки);
- змішане навчання.

У першому випадку нейронна мережа має правильну відповідь (вихід мережі) для кожного вхідного прикладу. Вагові коефіцієнти налаштовуються таким чином, щоб мережа надала відповідь, якомога ближчу до відомої правильної відповіді.

Щоб навчатися без супервізора, вам не потрібно знати правильні відповіді для кожного прикладу у вашому навчальному зразку. У цьому випадку розкривається внутрішня структура даних і кореляції між вибірками в навчальному наборі, що дозволяє розподіляти вибірки за категоріями.

У змішаному навчанні деякі ваги визначаються шляхом навчання з учителем, а інші – шляхом самостійного навчання.

1.3.1. Контрольоване навчання

Більшість рішень походять від нейронних мереж із контрольованим навчанням, де поточний результат постійно порівнюється з бажаним результатом. Вагові коефіцієнти спочатку встановлюються випадковим чином, але коригуються під час наступних ітерацій, щоб точно відповідати бажаним і фактичним результатам. Створений метод навчання має на меті мінімізувати помилки продуктивності всіх елементів обробки, що з'являються з часом, шляхом постійної зміни синаптичних ваг до досягнення прийнятної точності мережі.

Нейронні мережі необхідно навчити за допомогою навчання під наглядом, перш ніж їх можна буде використовувати. Етап навчання може зайняти деякий час. Особливо для прототипів систем з недостатньою обчислювальною потужністю навчання може тривати годинами.

Навчання вважається завершеним, коли нейронна мережа досягає заданого користувачем рівня ефективності. Цей рівень означає, що мережа досягла бажаної

статистичної точності, оскільки виробляє бажаний вихід для даної вхідної послідовності.

Після тренування суглобові обважнювачі фіксуються для подальшого використання. Деякі типи мереж дозволяють безперервне навчання під час використання зі значно повільнішими темпами навчання, що допомагає мережі адаптуватися до умов, що повільно змінюються.

Навчальний набір має бути достатньо великим, щоб містити всю інформацію, необхідну для визначення важливих характеристик і взаємозв'язків. Крім того, навчальні приклади повинні містити різні дані. Якщо мережа навчається лише на одному прикладі, ваги, ретельно встановлені для цього прикладу, радикально змінюються під час навчання на наступному прикладі. Вивчивши наступний приклад, попередній можна легко забути. Таким чином, система повинна вивчити їх усі разом і знайти оптимальні ваги для всього набору прикладів.

Наприклад, під час навчання розпізнавача десятизначного растрового зображення, представленого двадцятьма екземплярами кожної цифри, недоцільно представляти послідовно всі екземпляри цифри "сім". Краще спочатку надати мережі один спосіб представлення всіх чисел, а потім надати їй другий спосіб. Ключовим компонентом успішної роботи мережі є представлення та кодування вхідних і вихідних даних. Оскільки штучні мережі працюють лише з числовими вхідними даними, нам потрібно перетворити необроблені дані із зовнішнього середовища.

Крім того, потрібне масштабування, себто нормалізація даних відповідно до діапазону всіх значень. Попередня обробка зовнішніх даних, отриманих від датчиків, у машинну форму є звичайною та легко доступною на стандартних комп'ютерах.

Якщо після навчання з учителем нейронна мережа ефективно обробила дані з навчальної вибірки, то її ефективність при обробці даних, які не використовуються для навчання, стає важливою. Якщо результати тестового набору незадовільні, продовжуйте навчання. Тестування використовується для

забезпечення не тільки запам'ятовування даних з навчальної вибірки, але й створення загальних образів, які можуть міститися в даних.

1.3.2. Неконтрольоване навчання

Неконтрольоване навчання може стати великою перевагою в майбутньому, оскільки комп'ютери можуть вчитися самі по собі в істинно роботизованому значенні. В даний час неконтрольоване навчання використовується в мережах, відомих як карти, що самоорганізуються, які мають досить обмежене застосування, але демонструють перспективність самоконтрольованого навчання.

Мережа не використовує зовнішні впливи для коригування ваг, пошуку закономірностей чи тенденцій у вхідному сигналі або внутрішнього моніторингу продуктивності шляхом коригування відповідно до функції навчання. Навіть без повідомлення про правильні чи неправильні дії мережа повинна мати власну організаційну інформацію, вбудовану в топологію мережі та засвоєні правила.

Алгоритми неконтрольованого навчання спрямовані на знаходження близькості між групами нейронів, які працюють разом. Коли зовнішній сигнал активує вузол у групі нейронів, загальна дія групи збільшується. Так само зменшення зовнішніх сигналів у групі має гальмівний вплив на всю групу. Конкуренція між нейронами є основою навчання. Навчання конкуруючих нейронів посилює відповіді певних груп на певні сигнали. Це пов'язує групи одна з одною та з зворотним зв'язком. Під час змагання змінюються лише ваги нейронів-переможців.

1.3.3. Оцінка та правила навчання

Оцінка ефективності навчання нейронної мережі залежить від кількох керованих факторів. Теорія навчання враховує три фундаментальні властивості, пов'язані з навчанням — це ємність, складність вибірки та обчислювальна складність:

- ємність вказує кількість зразків, які мережа може зберігати, і обмеження рішень, які можуть бути сформовані для зразків;
- складність вибірки визначає кількість навчальних вибірок, необхідних для досягнення узагальнення мережі;
- обчислювальна складність безпосередньо пов'язана з продуктивністю процесора комп'ютера.

Існує багато правил навчання для нейронних мереж, але більшість із них є варіантами правила Гебба — найвідомішого та найстаршого правила навчання. Ось кілька основних правил навчання.

Правило Гебба

Опис правил можна знайти в книзі «Організація поведінки» (Дональд Олдінг Гебб, 1949).

«Якщо нейрон отримує дані від іншого нейрона, і обидва дуже активні (математично мають однаковий знак), вага між ними повинна бути збільшена».

Якщо на k -му кроці навчання одночасно збуджуються два нейрони з виходами (x_j, y_i) , вага синаптичного зв'язку між ними збільшується, в іншому випадку – зменшується.

$$\Delta W_{ij}(k) = r x_j(k) y_i(k), \quad (1.1)$$

де r – коефіцієнт швидкості навчання.

Можна використовувати і "з супервізором" і "без супервізора".

Правило Хопфілда

Подібний до правила Гебба, за винятком того, що він визначає величину підсилення або ослаблення. «Якщо вихідний і вхідний сигнали нейрона активні або неактивні одночасно, збільште вагу з'єднання за допомогою оцінки навчання. В іншому випадку зменшіть вагу шляхом оцінки навчання».

Правило "дельта"

Це правило є ще однією модифікацією правила Гебба і одним із найбільш часто використовуваних правил. Це правило засноване на простій ідеї безперервної

зміни синаптичних ваг, щоб зменшити різницю ("дельта") між бажаним і фактичним вихідними значеннями нейрона.

$$\Delta W_{ij} = x_j (d_i - y_i) \quad (1.2)$$

Це правило мінімізує середню квадратичну помилку мережі. У правилі «дельта» помилка, отримана на вихідному рівні, перетворюється на похідну передавальної функції, яка послідовно шар за шаром передається на попередній рівень для корекції синаптичних ваг. Процес зворотного поширення мережевих помилок триває, доки не буде досягнуто першого рівня. Добре відома парадигма Подача прямого зворотного поширення успадкувала свою назву від цього методу обчислення помилок.

Порушення у наборі вхідних даних є важливим при застосуванні правила «дельта». За умови добре впорядкованого або структурованого представлення навчальної множини результати мережі можуть не збігатися з бажаною точністю, і мережа вважається непридатною для навчання.

Правило градієнтного спуску

Це правило подібне до правила "дельта" тим, що воно використовує похідну функції передачі для виправлення помилки дельта перед застосуванням її до ваг з'єднання. Константа пропорційності для оцінки навчання додається до кінцевого коефіцієнта варіації, що діє на вагові коефіцієнти. Хоча процес навчання дуже повільно наближається до стабільності, це правило широко поширене і часто використовується.

Показано, що різні оцінки навчання для різних рівнів мережі допомагають прискорити конвергенцію процесу навчання. Рівні ближче до виходу мають нижчі бали навчання, ніж рівні ближче до входу.

Правило змагання

На відміну від закону Гебба, де кілька вихідних нейронів можуть спрацьовувати одночасно, у конкурентному навчанні вихідні нейрони змагаються між собою за активацію. Це явище відоме як правило "переможець отримує все".

Подібне навчання відбувається в біологічних нейронних мережах. Навчання через змагання дозволяє кластеризувати вхідні дані: подібні приклади згруповані мережами за співвідношенням і представлені одним елементом.

Під час навчання синаптичні ваги нейронів-переможців змінюються. Ефект цього правила досягається шляхом зміни шаблону, що зберігається в мережі (вектора синаптичних ваг нейронів-переможців), щоб він був подібний до прикладу вхідних даних. Нейрон із найбільшим вихідним сигналом оголошується переможцем і має здатність гальмувати конкурентів і збуджувати сусідні нейрони. Вихід нейрона-переможця використовується, і тільки цей нейрон та його сусіди можуть регулювати ваги з'єднань.

$$\Delta W_{ij}(k+1) = W_{ij}(k) + r [x_j - W_{ij}(k)] \quad (1.3)$$

Під час навчання розмір сусідніх зон може змінюватися. Звичайна парадигма повинна починатися з великого домену визначення сусідства та звужуватися в процесі навчання. Мережа Кохонена моделює розподіл вхідних даних, оскільки виграшний елемент визначається найбільшою відповідністю вхідному шаблону. Це правило використовується в самоорганізованих картах.

1.4. Висновок до 1 розділу

У першому розділі було досліджено нейронні мережі, їхні етапи розвитку, класифікацію та правила навчання. Однією з головних переваг нейронних мереж є їхня здатність навчатися на основі даних. Вони можуть обробляти великі обсяги даних і вивчати складні шаблони, які люди можуть не легко виявити. Нейронні мережі мають широкий спектр застосувань, включаючи розпізнавання зображень і мови, обробку природної мови та автономні транспортні засоби. Вони використовуються в комп'ютерному зорі для класифікації зображень, виявлення об'єктів і розпізнавання обличчя. У системах розпізнавання мови нейронні мережі використовуються для аналізу та розуміння людської мови. В автономних

транспортних засобах вони використовуються для виявлення та класифікації об'єктів у навколишньому середовищі, таких як інші транспортні засоби, пішоходи та дорожні знаки.

Незважаючи на численні переваги, нейронні мережі також мають деякі обмеження. Одним із головних обмежень є їхня природа «чорної скриньки», що означає, що може бути важко зрозуміти, як мережа приймає рішення. Це може ускладнити виявлення та виправлення помилок у системі. Нейронні мережі також потребують великої кількості даних для ефективного навчання, що може бути дорогим і трудомістким. Нарешті, нейронні мережі можуть бути сприйнятливі до переобладнання, коли мережа стає занадто спеціалізованою для навчальних даних і не може узагальнювати нові дані.

Загалом, нейронні мережі — це потужна технологія з широким спектром застосувань. Оскільки технології продовжують розвиватися, ми можемо очікувати ще більш захоплюючих застосувань нейронних мереж у майбутньому. Однак важливо також враховувати їхнє обмеження та продовжувати розробляти шляхи вирішення цих проблем.

РОЗДІЛ 2

СИСТЕМИ РЕКОМЕНДАЦІЙ

2.1. Типи рекомендаційних систем

Рекомендаційна система (РС) – це система, що використовується для прогнозування індивідуальних вподобань користувачів веб-сайту на основі зібраної про них інформації для створення їм рекомендацій з метою збільшення загального інтересу до об'єктів на веб-сайті (товарів, фільмів, новин, тощо).

Системи рекомендацій відіграють вирішальну роль у сучасну цифрову епоху, надаючи персоналізовані пропозиції користувачам. Ці системи аналізують дані користувачів, уподобання та поведінку, щоб створити актуальні та індивідуальні рекомендації для широкого спектру програм, включаючи електронну комерцію, соціальні мережі, стримінгове передавання вмісту тощо.

Основна мета рекомендаційних систем полягає в тому, щоб покращити взаємодію з користувачами, пропонуючи їм предмети, продукти або вміст, які відповідають їхнім інтересам і вподобанням. Використовуючи алгоритми машинного навчання та методи аналізу даних, ці системи допомагають користувачам відкривати нові та відповідні варіанти, зменшують перевантаження інформацією та спрощують процес прийняття рішень.

Однією з ключових проблем, які вирішують системи рекомендацій, є величезна кількість доступних варіантів. У сучасному цифровому середовищі користувачі завалені надзвичайною різноманітністю продуктів, послуг і контенту.

Кафедра КІТ				НАУ 23 11 57 000 ПЗ			
Виконавець	Кічапіна М.О.			СИСТЕМИ РЕКОМЕНДАЦІЙ	Літера	Аркуш	Аркушів
Керівник	Харченко О.Г.					22	18
Консультант					<i>УС-411Б 122</i>		
Н.Контроль	Шевченко О.П.						

Системи рекомендацій діють як інтелектуальні фільтри, звужуючи вибір і пропонуючи найбільш доречні варіанти на основі індивідуальних характеристик користувача.

Існують наступні типи систем рекомендацій, які зазвичай використовуються:

- спільна фільтрація;
- фільтрація на основі вмісту;
- гібридні підходи.

2.1.1. Спільна фільтрація

Спільна фільтрація є популярною технікою в системах рекомендацій, яка використовує взаємодію користувачів і схожість між користувачами для створення персоналізованих рекомендацій. Вона базується на припущенні, що користувачі, які мали подібні переваги чи поведінку в минулому, ймовірно, матимуть подібні переваги в майбутньому. Ця техніка аналізує дані про взаємодію між користувачем і товаром, такі як оцінки, відгуки або історію покупок, щоб визначити моделі та зв'язки між користувачами та елементами.

Спільна фільтрація на основі користувачів: у цьому підході рекомендації створюються на основі подібності між користувачами. Він ідентифікує користувачів зі схожими смаками та вподобаннями та пропонує елементи, які подобалися цим схожим користувачам або з якими вони взаємодіяли в минулому. Спільна фільтрація на основі користувачів може бути ефективною в ситуаціях, коли користувачі довіряють рекомендаціям однодумців.

Спільна фільтрація на основі елементів: цей підхід зосереджується на подібності між елементами, а не на користувачах. Він визначає продукти, які часто подобаються або споживаються користувачами разом, і рекомендує продукти, подібні до тих, які вже віддають перевагу певний користувач. Спільна фільтрація на основі елементів корисна, коли зв'язки та схожість елементів стабільніші, ніж уподобання користувача.

Спільна фільтрація може сприяти випадковим відкриттям, рекомендуючи елементи, які користувачі, можливо, не знайшли самостійно. Використовуючи колективну мудрість користувачів, він може виявити приховані зв'язки та представити користувачам нові та несподівані варіанти.

Спільна фільтрація може обробляти великі набори даних і враховувати зростаючу кількість користувачів і елементів. Оскільки він покладається на взаємодію користувача, а не на явні функції предметів, він може добре масштабуватися для обробки великих обсягів даних і підтримки рекомендацій у реальному часі.

Спільне фільтрування може пом'якшити проблему холодного старту, коли дані для нових користувачів або елементів обмежені або відсутні. Порівнюючи поведінку та вподобання користувачів із подібними користувачами чи елементами, спільна фільтрація може генерувати рекомендації, навіть якщо явних даних мало.

Розрідженість даних: спільна фільтрація може зіткнутися з труднощами під час роботи з розрідженими даними, коли користувачі надають обмежені оцінки або взаємодії. Для вирішення цієї проблеми можна застосувати такі методи, як матрична факторізація, методи на основі сусідства та обчислення даних.

Масштабованість і рекомендації в режимі реального часу: зі збільшенням кількості користувачів і елементів обчислювальна складність спільної фільтрації може зростати. Для забезпечення масштабованості та швидкості реагування в реальному часі часто потрібні вдосконалені алгоритми та методи розподілених обчислень.

Проблема холодного старту для нових користувачів і елементів: спільна фільтрація покладається на історичні взаємодії користувачів, які можуть бути недоступні для нових користувачів або елементів. Для вирішення проблеми холодного старту можна використовувати такі методи, як популярність предмета, фільтрація на основі вмісту або гібридні підходи.

Спільна фільтрація — це потужна техніка в системах рекомендацій, яка використовує колективну мудрість користувачів для надання персоналізованих і відповідних рекомендацій. Аналізуючи взаємодію користувачів і подібності,

спільне фільтрування дозволяє робити випадкові відкриття, вирішує проблему холодного старту та добре масштабується для обробки великих наборів даних. Однак такі проблеми, як розрідженість даних, масштабованість і проблема холодного старту для нових користувачів і елементів, повинні бути ретельно розглянуті, щоб використовувати весь потенціал спільної фільтрації для надання точних і ефективних рекомендацій.

2.1.2. Фільтрація на основі вмісту

Системи рекомендацій на основі вмісту є відомим підходом у системах рекомендацій, які використовують атрибути предметів і профілі користувачів для створення персоналізованих рекомендацій.

Системи рекомендацій на основі вмісту аналізують атрибути елементів, такі як описи, метадані або характеристики, щоб зрозуміти характеристики елементів. Порівнюючи ці атрибути елемента з уподобаннями або профілями користувача, фільтрація на основі вмісту визначає елементи, які точно відповідають інтересам користувача, і відповідно рекомендує їх. Цей підхід фокусується на внутрішніх властивостях елементів, а не на взаємодії користувача з елементом.

Системи рекомендацій на основі вмісту покладаються на представлення елементів на основі їхніх атрибутів. Це передбачає виділення релевантних ознак або дескрипторів, які фіксують характеристики елементів, такі як жанр, ключові слова або текстова інформація.

Профілі користувачів створюються на основі вподобань користувача, історії або явних відгуків. Ці профілі представляють інтереси та вподобання користувача, які потім зіставляються з атрибутами предмета для створення персоналізованих рекомендацій.

Методи, які застосовуються в системах рекомендацій на основі вмісту

Модель векторного простору: у цій техніці елементи та профілі користувачів представлені у вигляді векторів у багатовимірному просторі. Подібність між елементами та профілями користувачів обчислюється за допомогою показників

подібності, наприклад косинусної подібності, щоб визначити релевантність і придатність елементів для рекомендації.

Термін частота-зворотна частота документа (ТЧ-ЗЧД): ТЧ-ЗЧД є загальноживаним методом для вимірювання важливості термінів у документі. Він призначає ваги термінам на основі їх частоти в документі та їх рідкості в усій колекції. ТЧ-ЗЧД використовується для представлення елементів і профілів користувачів і обчислення подібності між ними.

Алгоритми машинного навчання: системи рекомендацій на основі вмісту можуть використовувати алгоритми машинного навчання, такі як дерева рішень, опорні векторні машини (ОВМ) або нейронні мережі, щоб вивчати закономірності та зв'язки між атрибутами елементів і уподобаннями користувача. Ці моделі можуть фіксувати складні взаємодії та давати точніші рекомендації.

Незалежність від історії користувача: системи рекомендацій на основі вмісту не залежать від явної взаємодії користувача з елементами. Вони можуть надавати рекомендації навіть для нових користувачів з обмеженою історією або без неї, що робить їх корисними для сценаріїв холодного старту.

Можливість рекомендувати нішеві предмети: системи на основі вмісту чудово рекомендують нішеві або менш популярні предмети. Враховуючи атрибути предметів і зіставляючи їх із уподобаннями користувачів, ці системи можуть надавати користувачам унікальні та персоналізовані рекомендації.

Прозорість і пояснення: рекомендації на основі вмісту базуються на атрибутах елементів, які користувачі можуть легко зрозуміти та інтерпретувати. Користувачі можуть бачити, чому їм рекомендовано певні елементи, що підвищує прозорість і довіру користувачів.

Системи рекомендацій на основі вмісту можуть страждати від надмірної спеціалізації, коли користувачам рекомендуються лише подібні предмети та втрачаються різноманітні варіанти. Щоб пом'якшити це, можна застосувати такі методи, як алгоритми диверсифікації або гібридні підходи.

Системи, що базуються на вмісті, можуть мати проблеми з рекомендаціями елементів, які виходять за межі відомих уподобань користувача. Щоб вирішити цю

проблему, можна використати гібридні підходи, що поєднують методи фільтрації на основі вмісту та спільні методи фільтрації, щоб додати випадковість у рекомендації.

Проблема холодного старту для нових елементів: системи на основі вмісту можуть зіткнутися з проблемами під час рекомендації нових або нещодавно доданих елементів, оскільки їм бракує достатніх атрибутів або відгуків користувачів. Такі методи, як популярність товару або гібридні підходи, можуть допомогти вирішити проблему холодного старту.

Системи рекомендацій на основі вмісту пропонують потужний підхід для надання персоналізованих рекомендацій на основі атрибутів предметів і профілів користувачів. Використовуючи внутрішні властивості елементів і узгоджуючи їх із уподобаннями користувачів, ці системи можуть досягти результатів.

2.1.3. Гібридні підходи

У світі систем рекомендацій гібридні підходи привернули значну увагу завдяки своїй здатності використовувати сильні сторони багатьох методів. Гібридні системи рекомендацій поєднують спільну фільтрацію та методи фільтрації на основі вмісту, щоб подолати обмеження індивідуальних підходів і надати більш точні та різноманітні рекомендації користувачам, а також запобігти проблемі холодного запуску.

Спільна фільтрація покладається на взаємодію між користувачами та схожість між користувачами для надання рекомендацій, тоді як фільтрація на основі вмісту використовує характеристики елементів і профілі користувачів для створення пропозицій. Поєднуючи ці підходи, гібридні системи можуть надавати персоналізовані рекомендації навіть у сценаріях із розрідженими або неповними даними.

Використовуючи спільну фільтрацію та фільтрацію на основі вмісту, гібридні системи можуть пом'якшити обмеження кожного методу. Спільна фільтрація може усунути ефект «мудрості натовпу», тоді як фільтрація на основі

вмісту може впоратися з проблемою холодного запуску та надати більш детальні рекомендації на основі атрибутів предмета.

Гібридні системи мають можливість пропонувати різноманітні рекомендації, враховуючи поведінку користувачів і характеристики предметів. Це веде до більш повного охоплення простору предметів і надає користувачам ширший спектр варіантів, потенційно покращуючи випадкові відкриття.

Гібридні системи більш стійкі до розрідженості даних і нових записів користувачів або елементів. Вони можуть надавати рекомендації на основі спільної фільтрації, коли є достатньо даних користувача, і перемикаються на фільтрацію на основі вмісту, коли даних користувачів мало або для нових елементів без попередніх взаємодій.

Вибір відповідної стратегії комбінування має вирішальне значення в гібридних системах. Це передбачає визначення того, як результати спільної фільтрації та фільтрації на основі вмісту зважуються або комбінуються для створення остаточних рекомендацій. Різні методи злиття, такі як перемикання, комбінація функцій і гібридизація, можуть бути використані на основі конкретних вимог програми.

Гібридні системи вимагають інтеграції та обробки даних про взаємодію між користувачем і елементом і даних про вміст елемента. Це вимагає ефективного зберігання даних, попередньої обробки та методів вилучення функцій, щоб забезпечити безперебійне функціонування системи.

Інтеграція кількох алгоритмів і джерел даних може призвести до підвищення складності системи. Належне проектування та реалізація необхідні для керування обчислювальними ресурсами та забезпечення масштабованості та оперативності реагування в режимі реального часу гібридної системи рекомендацій.

Прикладом успішних гібридних систем рекомендацій є стримінгова платформа Netflix. Механізм рекомендацій популярної платформи Netflix поєднує спільну фільтрацію з фільтрацією на основі вмісту. Він враховує оцінки та взаємодію користувачів (спільно), а також атрибути та жанри елементів (на основі вмісту), щоб пропонувати користувачам персоналізовані фільми та телешоу.

Amazon використовує гібридну систему рекомендацій, яка поєднує спільну фільтрацію та фільтрацію на основі вмісту. Він враховує історію покупок користувачів, поведінку веб-переглядача (співпраця), атрибути продукту й описи (на основі вмісту), щоб надавати персоналізовані рекомендації щодо продуктів.

Гібридні системи рекомендацій пропонують потужний підхід для підвищення точності рекомендацій, збільшення різноманітності та усунення обмежень окремих методів. Завдяки поєднанню спільної фільтрації та фільтрації на основі вмісту ці системи надають користувачам більш персоналізовані та релевантні рекомендації. Однак такі проблеми, як поєднання алгоритмів, інтеграція даних і складність системи, повинні ретельно розглядатися під час розробки та впровадження гібридних систем рекомендацій. Загалом гібридні системи продовжують залишатися перспективним напрямком для підвищення ефективності механізмів рекомендацій у різних сферах.

2.2. Проблеми «холодного старту» у рекомендаційних системах

Проблема холодного старту виникає при додаванні нового об'єкту чи користувача до системи. До появи відповідних новому об'єкту чи користувачу оцінок відсутня інформація для формування рекомендацій.

Проблема бульбашки фільтрів виникає у рекомендаційних та пошукових системах, коли алгоритм видачі інформації веб-сайту вибірково підбирає дані, враховуючи яку інформацію користувач хотів би бачити, базуючись на інформації про його вподобання і, в результаті, користувачі відділяються від інформації, яка їх не цікавить або їм не подобається, фактично ізолюючи їх у власних бульбашках. Одна з головних проблем РС – проблема холодного старту (ПХС). Вона виникає тоді, коли в системі з'являються нові елементи – або нові користувачі, історія вподобань яких порожня, або нові об'єкти, у яких ще немає оцінок та/або набору ознак. У багатьох реальних системах ПХС може набувати характеру циклічної проблеми для вже відомих користувачів або об'єктів. Наприклад, якщо частина користувачів змінює свої інтереси. Дана проблема отримала назву проблеми

постійного холодного старту (ППХС). Як і ПХС, проблема постійного холодного старту може виникати з користувачами (КППХС) та з об'єктами (ОППХС). КППХС виникає для користувачів, що змінюють свої вподобання, або рідко з'являються у системі та рідко оцінюють нові об'єкти. ОППХС виникає при наявності об'єктів, властивості яких можуть змінитися з часом. Для вирішення проблеми холодного старту, як правило, застосовують наступні підходи:

- гібридизація РС з поєднанням контентної та колаборативної фільтрації;
- використання контексту, в якому створюються та надаються рекомендації (демографічні дані, час та дата, тощо).

Однак всі ці способи не підходять в разі проблеми постійного холодного старту, оскільки припускають, що після того, як користувач став «відомим», він залишається таким необмежену кількість часу, а об'єкти рекомендацій не можуть змінювати свої властивості. Для рішення даної проблеми треба не тільки прогнозувати вподобання користувачів, а й відслідковувати та прогнозувати зміну їх вподобань, а також враховувати можливість зміни властивостей об'єктів рекомендацій. Дану проблему на сьогоднішній день намагаються вирішувати методами машинного навчання, що підвищують адаптивність системи до постійних змін. Не менш важливою проблемою РС є проблема бульбашки фільтрів. Класичні РС пропонують користувачам об'єкти, виходячи лише з їх попередніх вподобань. Отже, користувач потрапляє у інформаційне середовище, в якому спостерігає лише обмежену кількість однотипних об'єктів. Наслідки, викликані бульбашкою фільтрів:

- користувач не одержує альтернативну інформацію, яка може бути йому корисною (наприклад, види об'єктів, про які він зовсім не знає, але які ефективніше вирішать задачі його пошуку);
- у користувача формується викривлена точка зору на інформаційне середовище, так як він не бачить картини в цілому (напр., при рекомендації новин);

- користувач може втратити інтерес до списку рекомендацій, так як йому весь час пропонують однотипні об'єкти (наприклад, втратить інтерес до прослуховування онлайн радіо з однотипним набором пісень).

Так як усі РС як основну метрику якості своєї роботи використовують точність прогнозування вподобань користувачів, а формальна постановка задачі прогнозу оцінок виглядає наступним чином:

$$d(R,V) \rightarrow \min, \quad (2.1)$$

де $R=(r_1, r_2, \dots, r_n)$ – вектор, що містить список прогнозованих оцінок користувача, впорядкований по спаданню за величиною оцінок, $V=(v_1, v_2, \dots, v_n)$ – вектор, що містить справжні оцінки користувача, невідомі системі на етапі формування списку рекомендацій, то для всіх РС проблема бульбашки фільтрів є актуальною, так як якісна рекомендаційна система повинна створювати рекомендації максимально схожі на попередні вподобання користувача.

2.3. Етичні питання систем рекомендацій

Питання про те, які моральні принципи можуть бути правильними, глибоко спірне і обговорюється у філософії.

Цінність деяких наслідків часто вимірюється в термінах корисності, яку вони містять. Таким чином, логічно припустити, що будь-який аспект системи рекомендацій, який може негативно вплинути на корисність будь-якої з її зацікавлених сторін, або ризик спричинити такий негативний вплив, становить функцію, яка є етично значущою.

Хоча концепція корисності може бути введена в експлуатацію з використанням кількісних показників, права, як правило, приймаються для забезпечення якісних обмежень на дії. Думаючи про дії та наслідки, ми можемо визначити два способи, в яких рекомендаційна система може мати етичний вплив.

По-перше, його операції можуть вплинути (негативно) на корисність будь-якого з її зацікавлених сторін і/або порушувати їхні права.

По-друге, ці два види етичного впливу можуть бути миттєвими — наприклад, рекомендація може бути неточною, що призведе до зниження корисності для користувача — або вони можуть наражати відповідні сторони на майбутні ризики. Етика нав'язування ризиків є предметом дедалі більшої кількості філософської літератури, яка висвітлює, як більшість видів діяльності передбачає накладання ризиків. Наприклад, у випадку системи рекомендацій ризики можуть полягати в тому, що користувачі можуть порушувати конфіденційність зовнішніми особами або потрапляти в потенційно невідповідний або шкідливий вміст. Подібність до такого роду ризиків може вважатися помилкою, навіть якщо фактично нема ніяких негативних наслідків.

Враховуючи попередній аналіз, тепер ми можемо класифікувати етичні проблеми, викликані системами рекомендацій, за двома вимірами:

- чи негативно впливає (вказана характеристика) система рекомендацій на корисність деяких із її зацікавлених сторін або, натомість, становить порушення прав, яке не обов'язково вимірюється з точки зору корисності;
- і чи є негативний вплив негайною шкодою чи наражає відповідну сторону на майбутній ризик шкоди або порушення прав.

Література, що вирішує етичні проблеми, пов'язані з РС, малодосліджена, з обговоренням конкретних питань, часто пов'язаних з конкретним екземпляром РС, і фрагментована через дисциплінарні розбіжності. Через мультидисциплінарний, порівняльний мета-аналіз, було визначено шість наступних основних областей етичних проблем.

2.3.1. Неприйнятний вміст

Лише кілька досліджень на сьогоднішній день прямо стосуються етики РС як окремого питання. Попередня робота над питанням етичних рекомендацій більше зосереджена на змісті рекомендацій і пропонує способи фільтрації пунктів,

рекомендованих системою, на основі культурних і етичних уподобань. Особливо актуальними є чотири дослідження.

У першому дослідженні Камаль Суалі, Абделлатіф Ель Афія і Рдуан Файзі розглядають питання РС, які не відповідають культурі, і пропонують «етичну базу даних», побудовану на основі того, що вважається загальноприйнятими культурними нормами регіону, які діють як фільтр для рекомендацій.

Тіффані Я Тан і Піната Віното у другому дослідженні застосували більш динамічний підхід до проблеми, пропонуючи двошарову РС, що включає регульований користувачем «етичний фільтр», який відсіює елементи, які можна рекомендувати на основі визначених етичних уподобань користувача.

Родрігес і Воткінс, проводячи третє дослідження застосовують більш абстрактний підхід до проблеми етичних рекомендацій, пропонуючи бачення успішної РС, метою якого є «створення суспільств, у яких індивіди відчують задоволення через глибоку участь у світі». Автори передбачають, що це може бути досягнуто завдяки використанню взаємопов'язаних структур великих даних. Нарешті, Парашакіс, провівши четверте дослідження, надає один із найдетальніших звітів. Зосереджуючись на програмах електронної комерції, Парашакіс припускає, що існує п'ять етичних проблемних областей:

- практики профілювання користувачів;
- публікація даних;
- розробка алгоритму;
- дизайн інтерфейсу користувача;
- онлайн-експерименти або А/В-тестування, тобто практика викриття обраних груп користувачів модифікаціям алгоритму, з метою збору відгуків про ефективність кожної версії з відповідей користувачів.

Ризики, які він ідентифікує, пов'язані з порушеннями конфіденційності користувача (наприклад, через витоки даних або шляхом збору даних за відсутності явної згоди), порушення анонімності, маніпулювання поведінкою та упередженість у рекомендаціях, наданих користувачеві, цензура контенту, вплив побічних ефектів

та нерівне ставлення до А/В-тестування з недостатньою обізнаністю користувачів, що призводить до відсутності довіри. Рішення, висунуті в Парашакіс, обертаються навколо підходу до дизайну, орієнтованого на користувача, впроваджуючи регульовані інструменти для користувачів, щоб контролювати явно спосіб, яким РС використовує свої особисті дані, фільтрувати маркетингові упередження або цензуру вмісту, а також відмовитися від онлайн-експериментів.

2.3.2. Конфіденційність

Конфіденційність користувачів є однією з головних проблем для систем рекомендацій. Це можна вважати неминучим, враховуючи, що більшість комерційно успішних систем рекомендацій базуються на гібридних або спільних методах фільтрації та працюють шляхом побудови моделей своїх користувачів для створення персоналізованих рекомендацій.

Ризики конфіденційності виникають щонайменше в чотири етапи. По-перше, вони можуть виникати, коли дані збираються або надаються без явної згоди користувача.

По-друге, коли набори даних зберігаються, існує додатковий ризик того, що вони можуть потрапити до зовнішніх агентів або стати предметом спроб деанонізації. На обох етапах порушення конфіденційності наражають користувачів на ризики, які можуть призвести до втрати користі (наприклад, якщо окремі користувачі стають ціллю зловмисних агентів у результаті) або до порушення прав (наприклад, якщо використовується особиста інформація користувачів способами, які загрожують їхній індивідуальній автономії).

По-третє, незалежно від того, наскільки безпечно збираються та зберігаються дані, питання конфіденційності також виникають на етапі висновків, які система може (дозволити) зробити з даних. Користувачі можуть не знати про природу цих висновків і можуть заперечувати проти такого використання їхніх особистих даних, якщо вони будуть краще проінформовані.

Ризики конфіденційності стосуються не лише збору даних, тому що, наприклад, зовнішній агент, який дотримується рекомендацій, які система генерує для певного користувача, може отримати певну конфіденційну інформацію про користувача. Розширити поняття інформованої згоди на непрямі висновки з рекомендацій користувачів видається складним. Нарешті, є ще одна тонка, але важлива системна проблема щодо конфіденційності, яка виникає на етапі спільної фільтрації: система може побудувати модель користувача на основі даних, які вона збрала про взаємодію інших користувачів.

Іншими словами, поки достатньо користувачів взаємодіє та ділиться своїми даними з системою, система може створити досить точний профіль навіть для тих користувачів, про яких у неї менше даних. Це вказує на те, що для окремих користувачів може бути неможливо повністю захиститися від тих видів висновків, які система може зробити про них. Це може бути позитивною рисою в деяких сферах, як-от медичні дослідження, але також може виявитися проблематичним в інших сферах, як-от підбір персоналу чи фінанси.

2.3.3. Автономія та особиста ідентичність

Системи рекомендацій можуть посягати на автономію окремих користувачів, надаючи рекомендації, які спонукають користувачів у певному напрямку, намагаючись «прив'язати» їх до певного типу вмісту або обмежуючи діапазон опцій, до яких вони піддаються. Ці втручання можуть варіюватися від доброякісних (надання можливості індивідуальним діям і підтримка кращого прийняття рішень шляхом відфільтрування нерелевантних варіантів), до сумнівних (переконання, підштовхування) і, можливо, шкідливих (маніпулятивних і примусових).

2.3.4. Непрозорість

Теоретично, пояснення того, як генеруються персоналізовані рекомендації для окремих користувачів, може допомогти зменшити ризик посягання на їх автономію, надаючи їм доступ до причин, чому система «думає», що деякі параметри є актуальними для них. Це також допомогло б підвищити прозорість алгоритмічних рішень щодо того, як класифікувати та моделювати користувачів, таким чином допомагаючи захиститися від упередженості.

Розробка та оцінка пояснень для систем рекомендацій може приймати різні форми залежно від конкретних програм. Як повідомили Тінтарев і Мастхофф, у кількох дослідженнях використовувався орієнтований на користувача підхід до метрик оцінки, включаючи метрики для оцінки пояснень рекомендацій. Те, що вважається хорошим поясненням, залежить від кількох критеріїв: мета рекомендації для користувача; чи пояснення точно відповідає механізму, за допомогою якого генерується рекомендація; чи покращує це прозорість і контрольованість системи; і чи допомагає це користувачеві приймати рішення ефективніше (наприклад, швидше) і ефективніше, напр. з точки зору підвищення задоволеності.

2.3.5. Справедливість

Справедливість у алгоритмічному прийнятті рішень є широкомасштабною проблемою, яка ускладнюється існуванням кількох понять справедливості, які не всі є взаємно сумісними. У контексті систем рекомендацій кілька статей, визначених у цьому огляді, стосуються питання рекомендацій, які можуть відтворювати соціальні упередження. Вони можуть бути синтезовані навколо двох підходів.

З одного боку, Яо та Хуанг розглядають кілька можливих джерел несправедливості у спільній фільтрації та вводять чотири нові показники для їх усунення, вимірюючи відстань між рекомендаціями, наданими системою різним

групам користувачів. Зосереджуючись на методах спільної фільтрації, вони зазначають, що ці методи припускають, що відсутні рейтинги (тобто ті, які системі потрібно вивести зі статистичних даних, щоб передбачити переваги користувача) розподіляються випадковим чином. Однак це припущення про випадковість створює потенційне джерело упередженості в прогнозах системи, оскільки добре задокументовано, що базові вподобання користувачів часто відрізняються від вибіркового рейтингу, оскільки на останні впливають соціальні фактори, які можуть бути упередженими. Слідом за Яо з Хуангом, Фарнаді разом з іншими також ідентифікують два основних джерела упередженості в рекомендаційних системах з двома проблемними моделями збору даних, а саме упередженість спостереження, яка є результатом циклів зворотного зв'язку, створених рекомендаціями системи конкретним групам користувачів, і дисбаланс населення, коли доступні дані системі відображають існуючі соціальні моделі, виражаючи упередженість щодо деяких груп. Вони пропонують підхід імовірнісного програмування, щоб пом'якшити упередженість системи щодо захищених соціальних груп.

З іншого боку, Берк пропонує розглядати справедливість у системах рекомендацій як багатосторонню концепцію. Грунтуючись на цьому підході, він зосереджується на трьох поняттях чесних рекомендацій, враховуючи точку зору або користувача/споживача (С-чесність), чия зацікавленість полягає в отриманні найбільш відповідних рекомендацій; або постачальник (Р-чесність), який зацікавлений у тому, щоб їхні власні продукти чи послуги були рекомендовані потенційно зацікавленим користувачам; або, нарешті, комбінація двох (СР-чесність). Ця таксономія дозволяє розробнику системи рекомендацій визначити, як рекомендації системи впливають на конкуруючі інтереси різних сторін, і, отже, розробити архітектуру системи, яка може ефективно виступати посередником між цими інтересами.

В обох підходах питання справедливості пов'язане з вибором правильного рівня абстракції для конкретного застосування системи рекомендацій. З огляду на те, що концепція справедливості тісно пов'язана із соціальним контекстом, у якому

система збирає свої дані та дає рекомендації, поширення того самого підходу на будь-яке застосування систем рекомендацій може бути нежиттєздатним.

2.3.6. Соціальні наслідки

Наслідком деяких систем рекомендацій, який часто обговорюється, є їх трансформаційний вплив на суспільство. Зокрема, системи рекомендацій новин і фільтри соціальних мереж, за своєю природою, ризикують ізолювати користувачів від впливу різних точок зору, створюючи самозміцнювані упередження та «фільтрувальні бульбашки», які шкодять нормальному функціонуванню публічних дебатів, групове обговорення та демократичні інститути загалом. Ця особливість систем рекомендацій може мати негативний вплив на соціальну корисність. Відносно недавній, але тривожний приклад – поширення пропаганди проти вакцин, яку пов'язують зі зниженням колективного імунітету.

2.4. Висновок до 2 розділу

Отже, системи рекомендацій відіграють вирішальну роль у покращенні взаємодії з користувачами та покращенні бізнес-результатів у різних сферах. Проведено аналіз різних типів систем рекомендацій, включаючи спільну фільтрацію, фільтрацію на основі вмісту та гібридні підходи. Вирішення проблеми холодного старту в системах рекомендацій є важливим для ефективного обслуговування нових користувачів та елементів. У цьому розділі було розглянуто методи рекомендації на основі контенту, знань та гібридні методи. Застосовуючи ці стратегії, спеціалісти-практики можуть подолати проблеми, пов'язані з обмеженими або відсутніми даними користувачів, забезпечуючи точні та персоналізовані рекомендації з самого початку.

Етичні міркування є важливими в системах рекомендацій, оскільки вони можуть вплинути на довіру користувачів, справедливість і конфіденційність. Розглянуто потенційні упередження, проблеми справедливості та проблеми

конфіденційності, які виникають при розробці та впровадженні рекомендаційних систем. У роботі систем рекомендацій важливі прозорість, підзвітність та згода користувачів у забезпеченні етичних рекомендацій.

РОЗДІЛ 3

СИСТЕМА РЕКОМЕНДАЦІЙ В NETFLIX

3.1. Ключові віхи в історії розвитку Netflix

Система рекомендацій у Netflix значно розвинулася за ці роки, перейшовши від простих алгоритмів до більш складних методів, керованих машинним навчанням і нейронними мережами. На початку 2000-х коли Netflix вперше запусився, його система рекомендацій, яка тоді мала назву Cinematch, в основному спиралася на методи спільної фільтрації. Спільна фільтрація аналізує поведінку та вподобання користувачів, щоб знаходити шаблони та давати рекомендації. Cinematch використовувала оцінки, надані користувачами, щоб рекомендувати фільми на основі подібних оцінок інших користувачів.

У 2006 році Netflix запусив конкурс (див. рис. 2.1.), відомий як Netflix Prize, пропонуючи значний грошовий приз кожному, хто покращить алгоритм рекомендацій принаймні на 10 % (див. рис. 2.2.). Конкурс залучив тисячі учасників, які розробили інноваційні алгоритми, включно з використанням методів машинного навчання, для підвищення точності рекомендацій.

Кафедра КІТ				НАУ 23 11 57 000 ПЗ			
Виконавець	Кічапіна М.О.			СИСТЕМА РЕКОМЕНДАЦІЙ В NETFLIX	Літера	Аркуш	Аркушів
Керівник	Харченко О.Г.					40	16
Консультант					<i>УС-411Б 122</i>		
Н.Контроль	Шевченко О.П.						



Рис. 2.1. Вигляд сторінки на Netflix з оголошенням про конкурс



Рис. 2.2. Переможці конкурсу з грошовим призом

Наприкінці 2000-х в результаті конкурсу Netflix Prize Netflix запровадив гібридну систему рекомендацій, яка поєднала спільне фільтрування та фільтрування на основі вмісту. Фільтрування на основі вмісту передбачає аналіз

атрибутів фільмів (наприклад, жанр, акторський склад, режисер) і зіставлення їх із уподобаннями користувача.

Netflix почав інтегрувати нейронні мережі у свою систему рекомендацій на початку 2010-х років. Першим помітним прикладом було впровадження глибоких нейронних мереж для підвищення точності вибору заставок для відео, оптимізації оформлення, що відображається користувачам (див. рис. 2.3, 2.4.).

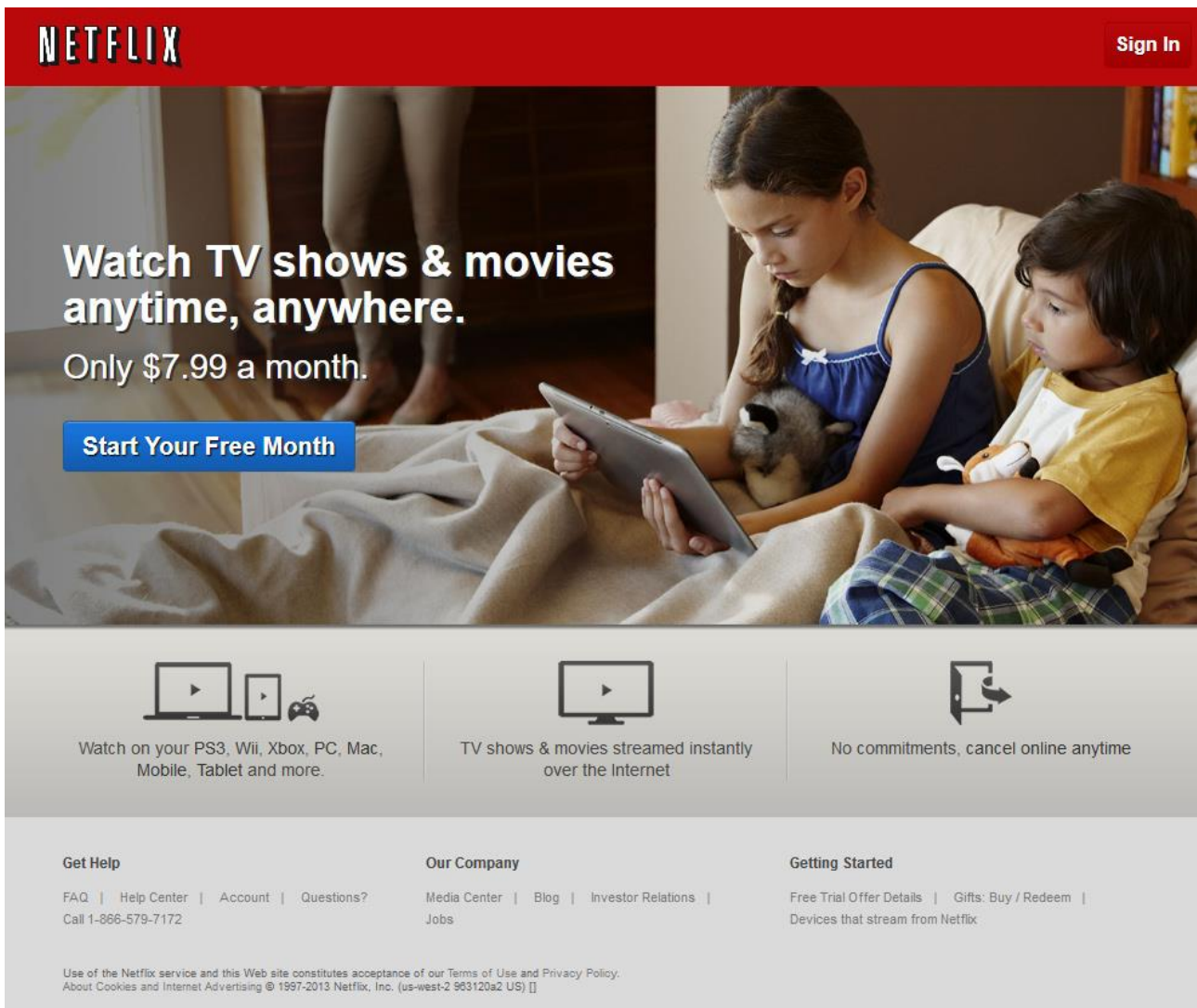


Рис. 2.3. Вигляд сторінки на Netflix на початку 2010-х років

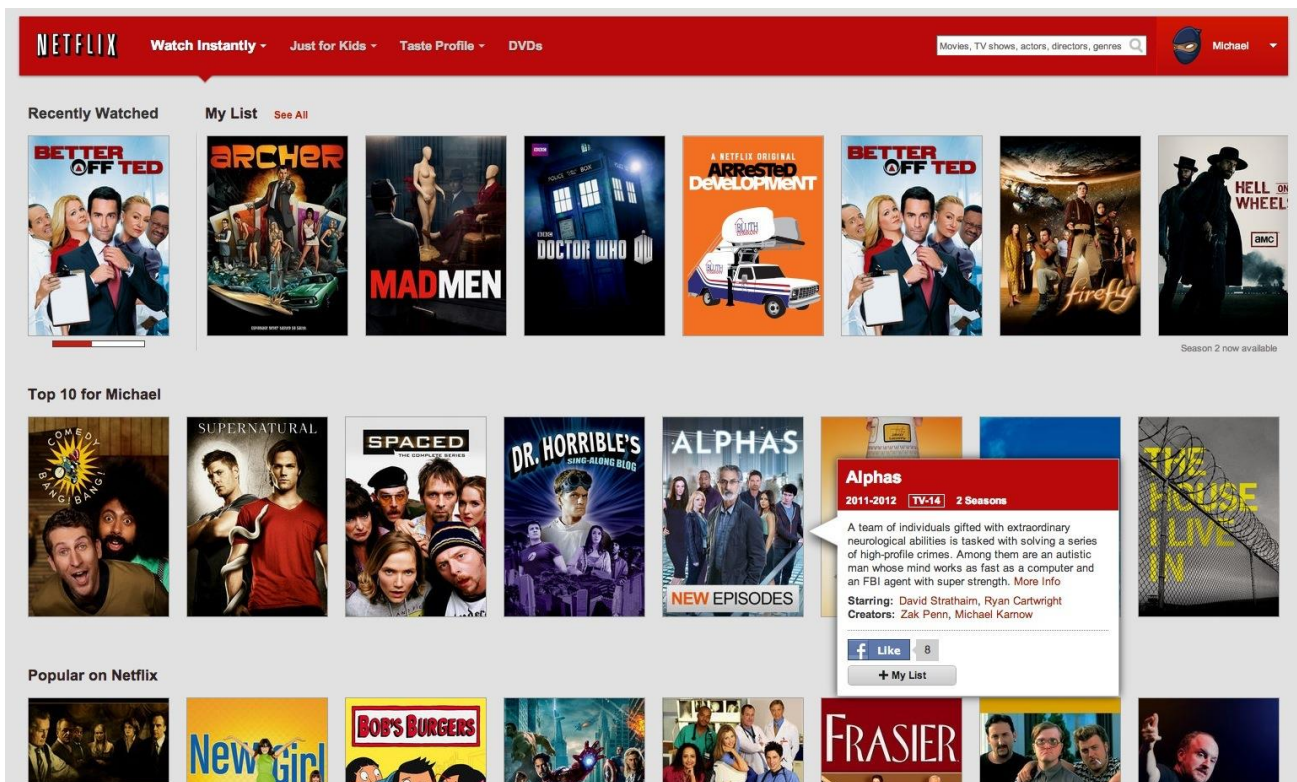


Рис. 2.4. Вигляд найпопулярнішої стримінгової платформи у 2012-2013 роках

У 2014 році Netflix зробив значний зсув у напрямку використання методів глибокого навчання для персоналізації. Моделі глибокого навчання, такі як глибокі автокодери, використовувалися для захоплення складних моделей поведінки користувачів і функцій фільму. Ці моделі дозволили Netflix генерувати більш персоналізовані рекомендації на основі історії переглядів користувачів, оцінок і контекстної інформації.

У 2016 році Netflix почав досліджувати використання так званих бандитських алгоритмів, зокрема контекстних бандитів, щоб оптимізувати свою систему рекомендацій. Такі алгоритми дозволяють приймати рішення в реальному часі шляхом постійного вивчення та адаптації рекомендацій на основі взаємодії користувача. Контекстні бандити враховують контекстну інформацію, таку як час доби, тип пристрою та демографічні показники користувачів, щоб надавати більш індивідуальні рекомендації.

Починаючи з 2017 року Netflix представив посилені методи навчання, щоб ще більше вдосконалити свою систему рекомендацій. Посилене навчання дає змогу системі навчатися на основі відгуків і відповідним чином оптимізувати рекомендації. Використовуючи таке навчання, Netflix може постійно покращувати актуальність і ефективність своїх рекомендацій.

Наразі Netflix залишається на передовій досліджень систем рекомендацій. Компанія продовжує інвестувати в дослідження та розробки, досліджуючи передові методи, такі як посилене глибоке навчання, нейронні мережі графів і оптимізацію з різними цілями, щоб вдосконалити свої рекомендації та підвищити задоволеність користувачів.

Важливо зазначити, що система рекомендацій Netflix є результатом комбінації алгоритмів і методів. Точні деталі їхньої поточної системи є конфіденційними та ретельно охороняються. Однак еволюція рекомендацій Netflix демонструє поступовий перехід до більш складного машинного навчання та підходів на основі нейронної мережі для надання високоперсоналізованих рекомендацій своїм користувачам.

3.2. Архітектура системи персоналізації та рекомендацій Netflix

Netflix розробили архітектуру рекомендаційних систем, яка:

- обробляє величезні обсяги вже існуючих даних;
- чуйно реагує на користувальницькі взаємодії із системою;
- дає можливість легко експериментувати з новими підходами до рекомендацій.

Основні компоненти архітектури містять декілька алгоритмів машинного навчання.

Загальна схема архітектури рекомендаційних систем Netflix виглядає наступним чином (див. рис. 2.5.):

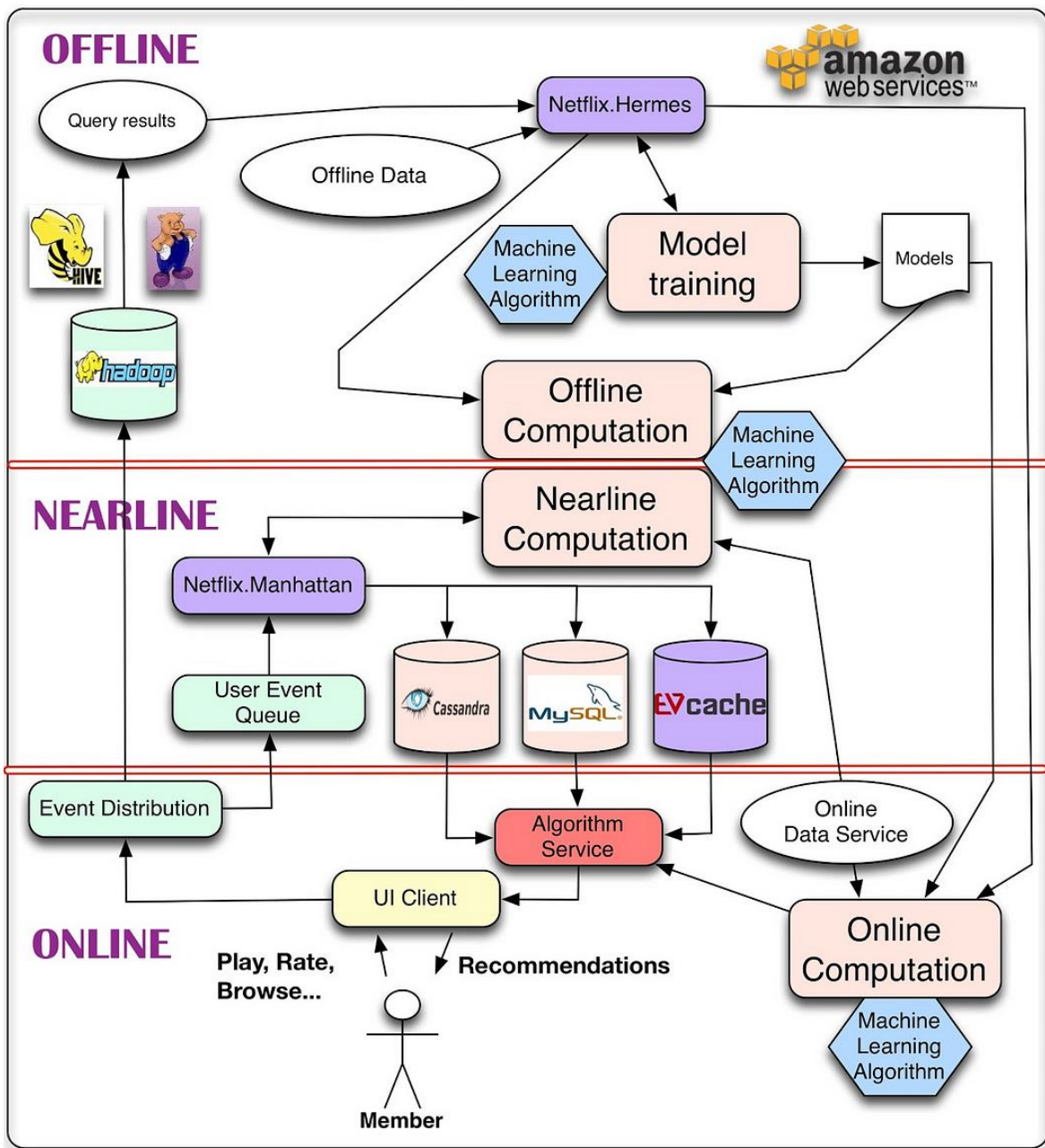


Рис. 2.5. Вигляд загальної схеми архітектури рекомендаційних систем

Netflix

Найпростіше, що можна робити з даними — зберігати їх, щоб потім спокійно обробляти в офлайн. За цю частину діаграми відповідає зона «Офлайн-вакансії». Тим не менш, обчислення можуть вироблятися не тільки офлайн, але ще і ніарлайново і онлайн. Онлайн-обчислення краще відповідають на нещодавні події та взаємодію користувачів (наприклад, на сайт зараз було завантажено новий боевик або ви як користувач щойно залишили вподобайку на картинці з цуценям),

але вони повинні відповідати на запити в реальному часі, і це суттєво обмежує обчислювальну складність алгоритмів, які використовуються в онлайні. Кількість даних, які можна опрацювати в онлайні також обмежена вимогами по швидкості. В офлайнних обчисленнях набагато менше подібних обмежень, так як вони вираховуються батчами, в яких досить м'які таймінги. Тем не менш, результати таких обчислень можуть бути неактуальними для користувача, тому що нещодавні дані не були там зауважені (наприклад, у випадку якщо за останні 2 години ви різко зацікавилися вуличним балетом). Це і є однією з ключових проблем побудови архітектури персоналізації — потрібно сумістити онлайніві та офлайніві вчислення так, щоб вони працювали як одне ціле — гармонійно, швидко, точно. Обчислення типу ніарлайн — це компроміс між онлайніом і офлайніом. Тут є можливість виконувати обчислення онлайн, але від них не вимагається моментального виконання — наприклад, ви почали залишати вподобайки на котів у стрічці, а стрічка пропонує вам більше котів лише за 5 хвилин.

Тренування моделі — це теж обчислення, у результаті яких ми отримуємо навчену на наявних даних модель, яка пізніше нам і видає рекомендації.

Ще одна частина архітектури — Система розподілу подій і даних. Вона описує, як події і дані різних типів оброблюються в системі. Схоже завдання — як опрацювати різні Сигнали и моделі. У кінці збирається середній Результат рекомендацій, який для нашого користувача буде виявляти інтерес.

3.2.1. Онлайніві, ніарлайніві та офлайніві обчислення

Онлайн-обчислення можуть швидко реагувати на недавні події та використовувати найсвіжіші дані. Наприклад, створення галереї фільмів-бойовиків, відсортованих для користувача з використанням його поточного контексту. Але онлайн-обчислення повинні відповідати Угоді про рівень доступності та часу відгуку, яка визначає максимальну затримку при відповіді на запити від клієнтських застосунків, тобто часу, поки користувач очікує появи

рекомендацій, не так багато. Це може ускладнити використання обчислювально складних алгоритмів (див. рис. 2.6.).

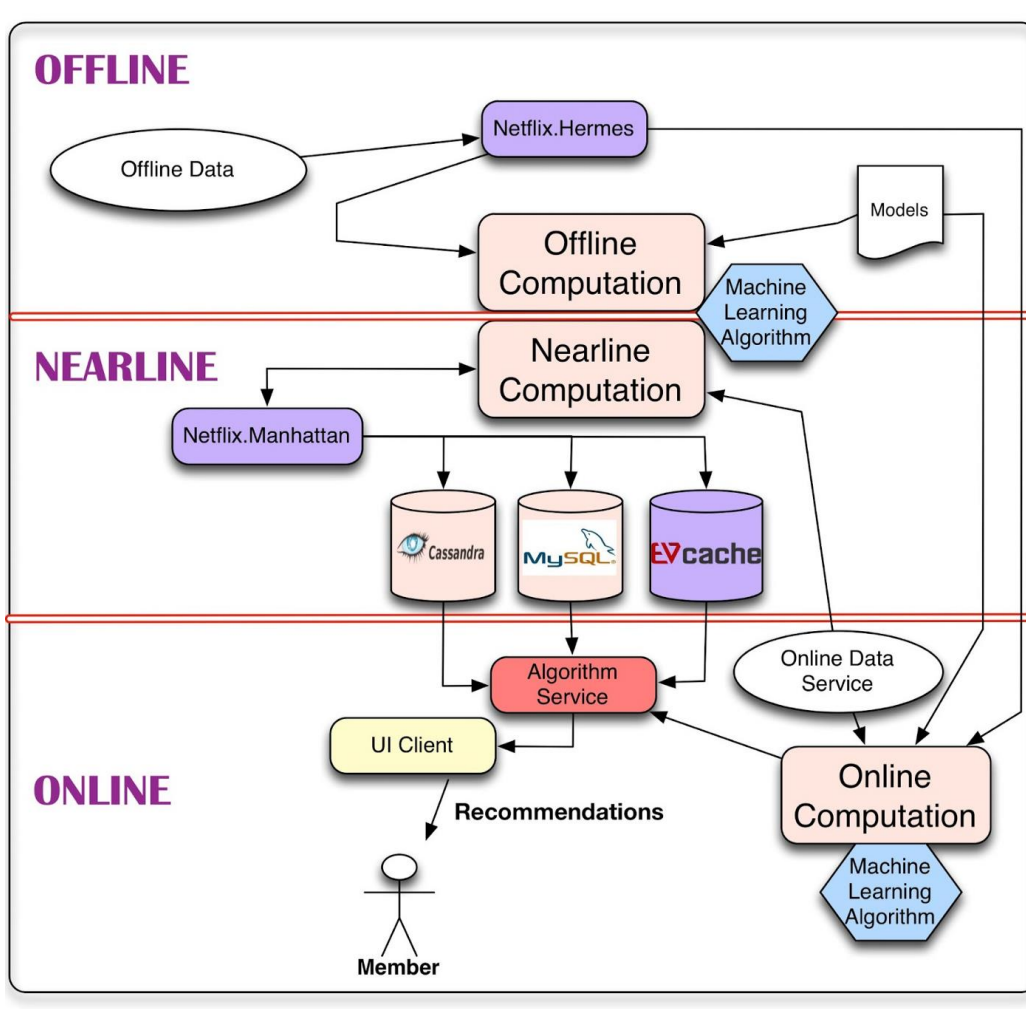


Рис. 2.6. Вигляд схеми онлайнного, ніарлайнного та офлайнного обчислення

Крім того, за деяких обставин онлайнні обчислення в чистому вигляді можуть не відповідати цій Угоді, тому завжди важливо продумати швидкий запасний механізм, наприклад повернення до попереднього, вже обчисленого результату. Обчислення в режимі онлайн також означають, що задіяні джерела даних також мають бути доступні в режимі онлайн, і це може потребувати додаткової інфраструктури.

З іншого боку, офлайнні обчислення дозволяють використовувати складніші алгоритми, і вони не мають таких жорстких обмежень на кількість

оброблюваних даних. Найпростіший приклад — періодичне узагальнення мільйонів переглядів, щоб створити базові загальні рекомендації найпопулярніших фільмів. В офлайн-системи можна простіше впроваджувати нові алгоритми без необхідності витратити купу часу на їхню оптимізацію. Netflix використовує офлайн-обчислення для підтримки швидкого експериментування: якщо новий експериментальний алгоритм виконується повільніше, можна просто розгорнути більше екземплярів Amazon EC2 для досягнення пропускної спроможності, необхідної для запуску експерименту, замість витратити дорогоцінний час інженерів на оптимізацію алгоритму, який у результаті може виявитися не таким цінним для бізнесу. Але офлайн-алгоритм не швидко реагуватиме на зміни контексту або нові дані. "Загальмованість" офлайн рекомендацій може призвести до "застою" рекомендацій, що, у свою чергу, може погіршити досвід користувача. Офлайн-обчислення також потребують інфраструктури для зберігання, обчислення та доступу до великих наборів попередньо обчислених результатів.

Ніарлайнові обчислення можна розглядати як компроміс між двома попередніми типами. У цьому випадку обчислення виконуються так само, як і в онлайн-підході, але від них більше не вимагається надавати результати відразу після обчислення, а натомість їх зберігають, роблять асинхронними. Це відкриває можливості для більш складної обробки подій. Прикладом може бути оновлення статусу фільму на "переглянутий" відразу після того, як користувач починає його дивитися. Результати можуть бути збережені у кеш або на бекенді. Ніарлайнові обчислення є полем для алгоритмів посиленого навчання.

У будь-якому випадку, не обов'язково вибирати якийсь із цих підходів, адже їх можна і потрібно поєднувати. Є багато способів збудувати таку суміщену архітектуру. Наприклад, використовуючи офлайн-обчислення як запасний аеродром при відмові онлайн-обчислень. Інший варіант — попередньо обчислити частину результату за допомогою офлайн-процесу і залишити менш обчислювально складні або контекстно-залежні частини алгоритмів для онлайн-обчислень.

Навіть моделювання може відбуватися у гібридному режимі офлайн/онлайн. Зазвичай, такий підхід не використовується для класифікації з учителем, він застосовується тільки в онлайні для класифікації нових вхідних даних. Однак такі підходи, як матрична факторизація, більш природно підходять для гібридного онлайн/офлайн моделювання: деякі етапи можуть бути попередньо обчислені в режимі офлайн, а інші можуть бути оновлені в режимі реального часу. Підходи без вчителя, такі як кластеризація, допускають офлайн-обчислення центрів кластерів та онлайн призначення кластерів. Ці приклади показують, що є можливість розділяти процес навчання моделі на частини — "важке", масштабне, потенційно досить складне навчання та більш легке навчання/оновлення під користувача.

3.2.2. Сигнали і моделі

Незалежно від того, який тип вичислених використовується, потрібно продумувати, як дана архітектура буде обробляти три типи даних — моделі, дані, сигнали. (див. рис. 2.7.).

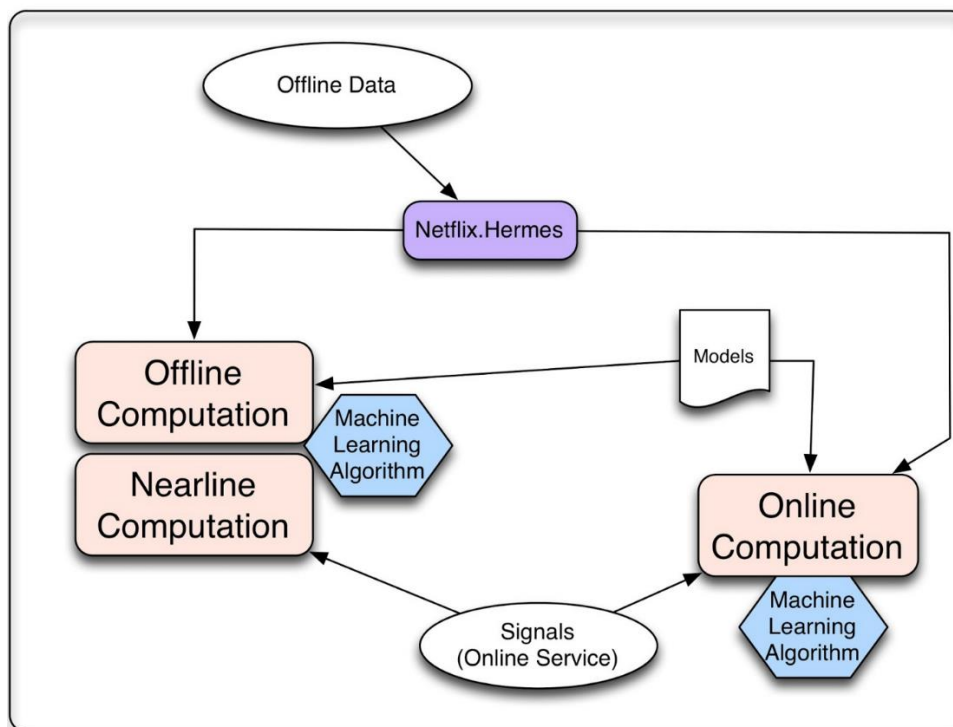


Рис. 2.7. Схема "Сигнали і моделі"

Моделі — це зазвичай невеликі файли з масами/параметрами, які були отримані в результаті офлайнового передобування. Дані — це попередньо оброблена інформація, яка зберігається в якійсь базі даних, наприклад метаданні фільму або рейтинги популярності. Під терміном «сигнал» маються на увазі дані про події, які приходять в систему — чиясь вподобайка, чийсь перегляд. Ці сигнали виходять з онлайн-сервісів, можуть представляти з себе інформацію користувача, наприклад, які фільми дивляться користувач, або ж контекстну інформацію, наприклад дані сесії, пристрої, дані та час.

3.2.3. Розподілення подій і даних

Мета компанії Netflix — перетворити дані користувацької взаємодії в інсайти для поліпшення досвіду користувача. Тому робиться все, щоб різні інтерфейси Netflix (розумні ТВ, планшенти, консолі і т.п.) збирали якомога більше даних користувача. Це можуть бути дані кліків, прокручування, переглядів або навіть вміст області видимості в момент часу (див. рис. 2.8.).

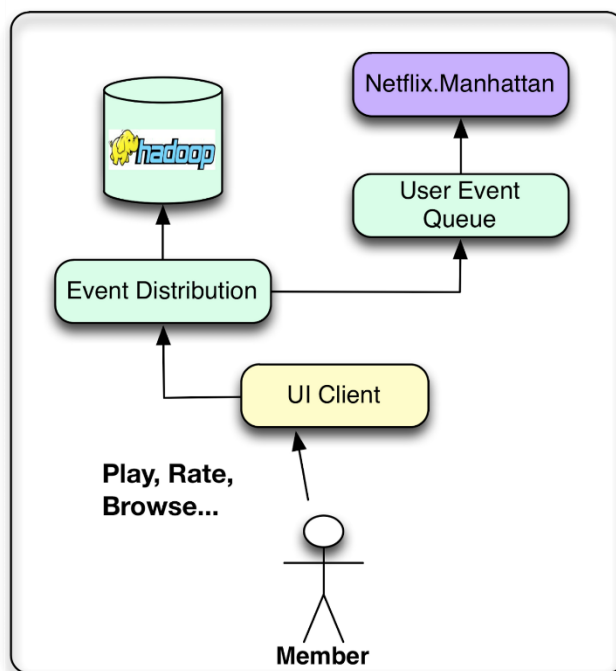


Рис. 2.8. Схема “Розподілення подій і даних”

Працівники компанії їх агрегують, систематизують та подають у моделі. Тут можна розмежувати події та дані, хоча кордон досить розмитий. Таким чином невеликі одиниці інформації, чутливої до часу, які необхідно обробляти з найменшою можливою затримкою називають подіями — вподобайка на фільмі, зміни поля “улюблений жанр” у профілі. Ці події запускають таку дію, наприклад, процес оновлення рекомендацій ніарлайново. Дані — це більш насичені інформаційні одиниці, які, можливо, потрібно обробити та зберегти для подальшого використання. Тут затримка не така важлива, як якість і кількість інформації. Звичайно, існують події користувача, які можна розглядати як події, так і дані і, отже, їх відправляють в обидва потоки.

У Netflix ніарлайновий потік подій управляється за допомогою внутрішньої платформи під назвою Manhattan. Manhattan — це розподілена обчислювальна система, яка займає центральне місце в архітектурі рекомендацій Netflix. Вона чимось схожа на Storm Твіттера, але вирішує інші проблеми та відповідає іншому набору внутрішніх вимог. На початкових етапах обробки управління потоком даних здійснюється в основному шляхом протоколювання через Chukwa Hadoop. Пізніше використовується Hermes, як реалізація патерну підписки (publish-subscribe) сервісів Netflix та обробників на повідомлення про події та зміни в даних.

Отже, мета машинного навчання в системі — надати персоналізовані рекомендації. Ці рекомендації можуть бути завантажені зі списку, який розраховується заздалегідь або ж генерується онлайнними алгоритмами. Звичайно, в хід йде і поєднання цих двох підходів, наприклад, можна вважати основні рекомендації офлайн, а потім додавати до них щось нове, обробляючи сигнали, що надходять з реального часу.

У Netflix зберігають офлайнові та проміжні результати в різних репозиторіях, таким чином вони доступні, коли надходить запит. В основному використовуються Cassandra, EVCache та MySQL як сховища даних. MySQL дозволяє зберігати структуровані реляційні дані, які можуть знадобитися для будь-якого майбутнього процесу. Уніфікованість запитів — безперечний плюс, але разом із цим плюсом приходять проблеми з масштабованістю у розподілених середовищах. Cassandra та

EVCache мають переваги сховищ типу “ключ-значення”. Cassandra — це добре відоме і стандартне рішення, коли потрібно розподілене та масштабоване сховище без SQL. Cassandra добре працює в деяких ситуаціях. Проте в тих випадках, коли відбуваються "інтенсивні" та постійні операції запису, EVCache підходить краще. Але питання все-таки полягає не в тому, де зберігати дані, а як побудувати таку архітектуру, яка б змогла поєднати конфліктуючі вимоги щодо складності запиту, затримки читання/запису та цілісності транзакцій.

Отже, Netflix вкладає багато зусиль для створення системи рекомендацій світового класу. При побудові такої системи важливо також думати про архітектуру програмного забезпечення, в якій вона буде розгорнута. Компанія хоче використовувати складні алгоритми машинного навчання, які можуть рости до довільної складності і можуть мати справу з величезними обсягами даних. Вони також працюють над архітектурою, яка забезпечує гнучкі інновації, де нові підходи можуть бути легко розроблені та підключені. Крім того, робиться все, щоб результати рекомендацій були свіжими та швидко реагували на нові дані та дії користувачів. Пошук оптимального варіанту між цими бажаннями не є тривіальним: він вимагає продуманого аналізу вимог, ретельного вибору технологій та стратегічного розкладання алгоритмів рекомендацій для досягнення найкращих результатів для членів команди.

3.3. Проблематика

Система рекомендацій є критично важливим компонентом успіху Netflix, оскільки вона відіграє важливу роль у пропонуванні персоналізованого контенту своїм передплатникам. Однак існує декілька практичних проблем, пов'язаних із системою рекомендацій.

Якість і кількість даних. Ефективність системи рекомендацій залежить від наявності високоякісних і різноманітних даних. Netflix збирає величезну кількість даних користувачів, включаючи історію переглядів, оцінки та взаємодії, щоб зрозуміти індивідуальні вподобання. Однак якість даних може бути низькою через

неповні або неточні введення користувачами, упереджені оцінки або обмежені дані для нових користувачів. Забезпечення точності та достатності даних є постійною проблемою.

Проблема холодного старту. Проблема холодного старту стосується завдання надання точних рекомендацій для нових користувачів або нещодавно випущеного вмісту. Коли дані для користувача обмежені або відсутні, стає важко зрозуміти його вподобання та надати відповідні рекомендації. Подібним чином для нового вмісту без значної історії переглядів важко точно передбачити його привабливість для користувачів.

Масштабованість і оновлення в режимі реального часу. Netflix обслуговує величезну базу передплатників, і його система рекомендацій повинна підтримувати масштабованість обробки даних і оновлення в реальному часі. Коли нові користувачі приєднуються, а існуючі користувачі взаємодіють із вмістом, система повинна постійно обробляти та оперативно оновлювати рекомендації. Підтримка систем із низькою затримкою та високою пропускнуою здатністю має вирішальне значення для забезпечення бездоганної взаємодії з користувачем.

Різноманітність і інтуїтивність. Хоча персоналізація важлива, існує проблема в тому, щоб збалансувати персоналізовані рекомендації з інтуїцією та надати користувачам новий і різноманітний вміст. Надмірний акцент на персоналізації може призвести до ефекту бульбашки фільтра, коли користувачам рекомендують лише подібний вміст, обмежуючи їх вплив на нові жанри, стилі чи точки зору. Досягнення балансу між персоналізованими рекомендаціями та пошуком різноманітного вмісту є проблемою.

Конфіденційність користувачів і етичні міркування: система рекомендацій базується на зборі та аналізі даних користувачів, що викликає занепокоєння щодо конфіденційності. Щоб забезпечити довіру користувачів, Netflix має дотримуватися суворої практики конфіденційності даних і відповідати нормам. Встановлення правильного балансу між персоналізованими рекомендаціями та дотриманням конфіденційності користувачів є серйозною проблемою.

Рекомендації щодо кількох профілів. Багато облікових записів Netflix мають кілька профілів користувачів, щоб обслуговувати різних членів сім'ї або окремих осіб, які мають один обліковий запис. Складним завданням є надання точних рекомендацій для кожного профілю, враховуючи переваги, що збігаються, і різноманітні звички перегляду.

Сезонність і тенденції вмісту. Система рекомендацій повинна адаптуватися до мінливих тенденцій і переваг сезонного вмісту. Наприклад, під час свят або особливих подій користувачі можуть мати різні уподобання щодо перегляду, і система повинна відповідним чином налаштувати рекомендації. Виявлення таких тимчасових змін і реагування на них є складним завданням.

3.4. Висновок до 3 розділу

Отже, Netflix виявився піонером в розвитку систем рекомендацій, і протягом років його рекомендаційна система пройшла значний шлях. Від початкових простих алгоритмів, які використовували спільне фільтрування, до складних методів, що базуються на машинному навчанні та нейронних мережах, Netflix безперервно здійснював інновації для поліпшення персоналізації рекомендацій.

Запровадження нейронних мереж дозволило Netflix значно покращити якість рекомендацій шляхом здійснення глибокого аналізу поведінки користувачів та властивостей фільмів. Шляхом використання глибокого навчання, контекстуальних алгоритмів та підсиленого навчання Netflix досяг високого рівня персоналізації та актуальності рекомендацій, що сприяє збільшенню задоволення користувачів та залученості.

Крім того, Netflix продовжує інвестувати в дослідження та розробки, щоб покращити свою рекомендаційну систему. Це означає, що компанія постійно вдосконалює свої алгоритми, досліджує нові техніки, такі як графові нейронні мережі та багатоцільова оптимізація, і активно експериментує з методами підсиленого навчання.

У майбутньому можна очікувати подальшого розвитку систем рекомендацій у Netflix. Це може включати ще більшу інтеграцію контекстуальних факторів, розширення використання нейронних мереж та удосконалення алгоритмів підсиленого навчання. Крім того, Netflix продовжуватиме займатися етичними аспектами рекомендаційних систем, забезпечуючи конфіденційність користувачів та борючись з можливим алгоритмічним упередженням.

В цілому, Netflix продемонстрував, що стежить за передовими технологіями та постійно розвивається у сфері систем рекомендацій. Його підхід до персоналізації та залучення користувачів служить прикладом для інших компаній, і майбутнє розширення системи рекомендацій Netflix може принести нові зрушення та покращення в цій галузі.

ВИСНОВКИ

У кваліфікаційній роботі за темою «Застосування нейронних мереж в системах рекомендацій на прикладі Netflix» проаналізовано дані наукових досліджень з рекомендаційних систем. Проаналізовані наступні типи систем рекомендацій: спільна фільтрація, фільтрація на основі вмісту та гібридні підходи.

Розкриті ключові виклики, з якими стикаються системи рекомендацій – це проблеми «холодного старту», «бульбашки фільтрів» і «мудрості натовпу». Проблема «холодного старту» виникає, коли з'являється новий користувач, про індивідуальні вподобання якого у сервісу немає жодних даних (наприклад, історії його покупок, переглядів тощо). Ще одним викликом для рекомендаційних систем є проблема «бульбашки фільтрів», яка виникає, коли алгоритми видачі інформації веб-сайту вибірково підбирають доречний контент для користувача, фактично ізолюючи його у «бульбашці». Розглянуто також проблема «мудрості натовпу», коли система пропонує контент користувачу, не враховуючи його ставлення до запропонованого контенту, базуючись тільки на оцінці і вподобаннях більшості інших користувачів.

У даній роботі класифіковано етичні проблеми, викликані системами рекомендацій, за двома вимірами – порушення прав та негативний вплив, пов'язаний з неприйнятним контентом тощо.

Проаналізовано архітектуру системи персоналізації та рекомендацій Netflix та виявлено, що дана архітектура обробляє величезні обсяги вже існуючих даних та чуйно реагує на користувальницькі взаємодії із системою. Доведено, що обчислення типу ніарлайн є вигіднішим. Виявлено, що механізм рекомендацій платформи Netflix – це гібридні підходи, які поєднують спільну фільтрацію з фільтрацією на основі вмісту. Він враховує оцінки та взаємодію користувачів, а також атрибути та жанри елементів, щоб пропонувати користувачам персоналізовані фільми та телешоу.

У даній кваліфікаційній роботі запропоновано при розробці цифрових платформ чи веб-сайтів використовувати гібридні підходи для покращення рекомендаційних систем.

СПИСОК БІБЛІОГРАФІЧНИХ ПОСИЛАНЬ ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кравченко В. Що таке нейронні мережі та як вони працюють? Класифікація штучних нейромереж [Електронний ресурс] - Режим доступу: <https://livingfo.com/shcho-take-nejronni-merezhi-ta-iak-vony-pratsiuiut/> (дата звернення 15.05.2023 р). – Назва з екрана.
2. Н. Steck [and others] Deep Learning for Recommender Systems: A Netflix Case Study / Н. Steck., L. Baltrunas, E. Elahi, D. Liang, Y. Raimond, J. Basilico // AI Magazine. – 2021. – № 3, Vol. 42. – P. 7-18.
3. Мелешко Є.В. Дослідження методів побудови рекомендаційних систем в мережі Інтернет / Є.В. Мелешко, Г.С. Семенов, В.Д. Хох. // Системи управління, навігації та зв'язку: зб. наук. праць. / ПНТУ ім. Ю. Кондратюка. – Полтава, 2018. – Вип. 1(47). – С. 131-136.
4. Bernardi L., Kamps J., Kiseleva J., Mueller M.J.I. The Continuous Cold Start Problem in e-Commerce Recommender Systems. – 2015. [Electronic resource] – Access mode: https://www.researchgate.net/publication/280773072_The_Continuous_Cold_Start_Problem_in_e-Commerce_Recommender_Systems (last access: 26.05.2023). – Title from the screen.
5. Silvia Milano, Mariarosaria Taddeo, Luciano Floridi. Recommender systems and their ethical challenges. – 2020. [Electronic resource] – Access mode: <https://link.springer.com/article/10.1007/s00146-020-00950-y#Tab2> (last access: 27.05.2023). – Title from the screen.
6. Carlos A Gómez-Uribe, Neil Hunt. The Netflix Recommender System: Algorithms, Business Value, and Innovation. – 2015. [Electronic resource] – Access mode: <https://dl.acm.org/doi/10.1145/2843948> (last access: 29.05.2023). – Title from the screen.

7. Ferrari Dacrema, M., P. Cremonesi, and D. Jannach. Are We Really Making much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches. – 2019. [Electronic resource] – Access mode: <https://arxiv.org/abs/1907.06902> (last access: 30.05.2023). – Title from the screen.

8. Dan Jackson. The Netflix Prize: How a \$1 Million Contest Changed Binge-Watching Forever. – 2017. [Electronic resource] – Access mode: <https://www.thrillist.com/entertainment/nation/the-netflix-prize> (last access: 31.05.2023). – Title from the screen.

9. Akanksha Bhatt. Neural Networks in Netflix. – 2021. [Electronic resource] – Access mode: <https://www.linkedin.com/pulse/neural-networks-netflix-akanksha-bhatt> (last access: 31.05.2023). – Title from the screen.