

Мельничук В.І.

студент,

*Чернівецький національний університет
імені Юрія Федьковича*

АНАЛІЗ АЛГОРИТМІВ НАВЧАННЯ ДЛЯ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

Рекомендаційні системи в Інтернеті з'явилися близько 20 років тому. Але бурхливий розвиток в цій області розпочався приблизно 5-10 років тому, коли відбулося змагання Netflix Prize, якій дуже важливо було підвищити якість рекомендацій, а передбачення вимірювалися за допомогою середньо-квадратичного відхилення [1]. Сьогодні, рекомендаційні системи змінюють способи взаємодії Web-сайтів зі своїми користувачами. Не використовуючи статичної інформації, рекомендаційні системи виконують збільшення ступеня інтерактивності з метою розширення призначених для користувача можливостей. Також рекомендаційні системи виконують незалежне формування рекомендацій кожному конкретному користувачеві, ґрунтуючись на конкретній історії дій, а також, урахувавши реакцію інших користувачів [1].

Задачею є аналіз алгоритмів навчання для рекомендаційних систем.

Більшість рекомендаційних систем передбачає застосування колаборативної фільтрації (collaborative filtering) або контентної фільтрації (content-based filtering), чи в іншому випадку – застосування гібридних підходів [1; 2].

Для колаборативної фільтрації характерне вироблення рекомендацій, заснованих на моделі попередньої поведінки користувача [2]. Таку модель можна побудувати, ґрунтуючись виключно на поведінці певного користувача або при обліку поведінки подібних користувачів, що мають подібні характеристики. У разі колаборативної фільтрації, коли береться до уваги характер поведінки інших подібних користувачів, відбувається використання знань про групу (group knowledge) з метою вироблення певних рекомендацій на основі порівняння таких користувачів. В цілому рекомендації відбуваються в режимі автоматичної співпраці множини подібних користувачів з виділенням (метод фільтрації) таких користувачів, яким характерна демонстрація подібних переваг або шаблонів поведінки.

Контентній фільтрації характерне формування рекомендацій, що ґрунтуються на поведінці користувача [1; 3]. Наприклад, при цьому підході може бути використана ретроспективна інформація про перегляди (перелік блогів, які читаються користувачем і характеристика таких блогів). Сам контент можна визначити, використовуючи ручний режим або витягти автоматично за допомогою інших методів подібності.

Гібридним підходам, яким характерне поєднання колаборативної і контентної фільтрації, властиво підвищення ефективності (і складності) рекомендаційних систем [1]. Об'єднання результатів колаборативної і контентної фільтрації потенційно дозволяє підвищити точність рекомендації. Крім того, гібридний підхід може бути корисний, якщо застосування

коллаборативної фільтрації починається при значній розрідженості даних. Гібридний підхід дозволяє спочатку зважувати результати згідно контентної фільтрації, а потім зміщувати ці ваги у напрямку до коллаборативної фільтрації (в міру «визрівання» доступного набору даних по конкретному користувачеві). Слід зазначити, що в рекомендаційних механізмах можуть бути використані найрізноманітніші алгоритми, багато з яких прийшли з області машинного навчання [3]. Так, наприклад, за допомогою кореляції Пірсона, може бути точно обчислено схожість між двома користувачами (і їх атрибутами, наприклад, такими як статті, прочитані в колекції блогів). Цей алгоритм вимірює лінійну залежність між двома змінними (або користувачами) як функцію їх атрибутів. Однак він не обчислює цю міру за всією сукупністю користувачів. Цю сукупність необхідно попередньо відфільтрувати до близьких елементів на основі високорівневих показників подібності, таких як, наприклад, читання подібних блогів.

Також до машинних методів належать алгоритми кластеризації, які є різновидом так званого «навчання без вчителя» (unsupervised learning), що дозволяє виявити структуру в рядах на перший погляд випадкових (або немаркованих) даних. У загальному випадку такий алгоритм базується на виявленні подібності між елементами (наприклад, між читачами блогу) за допомогою обчислення їх відстані від інших елементів в просторі ознак (feature space) (ознакою в просторі ознак може бути, наприклад, кількість прочитаних статей в наборі блогів). Кількістю незалежних ознак визначається розмірність просторових ознак. У випадку, коли елементи мають «близькість» один до одного, то вони можуть бути об'єднані в один кластер. Слід зазначити про існування множини алгоритмів кластеризації, найпростішим з яких є алгоритм k-середніх (k-means), який розділяє елементи на k-кластери.

Процес обчислення відстані між двома об'єктами може бути важким для візуалізації [4]. Один з поширених методів вирішення цього завдання полягає в тому, щоб розглядати кожен член кластера як багатовимірний вектор і обчислювати для нього «Евклідову відстань».

Крім того, слід згадати про існування множини інших існуючих різновидів кластеризації, до яких можна віднести теорію адаптивного резонансу (Adaptive Resonance Theory), нечітку кластеризацію методом C-середніх (Fuzzy C-means), вірогідну кластеризацію за допомогою EM-алгоритму (Expectation-Maximization) тощо.

Досліджуючи рекомендаційні механізми, можна відзначити можливість застосування багатьох алгоритмів, які і зараз успішно застосовуються. До таких належать:

- Байєсова мережа довіри (Bayesian Belief Net), яка може бути представлена, як орієнтований ациклічний граф, з ребрами, які представляють пов'язані ймовірності змінних;

- ланцюги Маркова (Markov chains), які мають такий же підхід, як у Байєсовій мережі довіри, але рішення проблеми вироблення рекомендацій виконується послідовною оптимізацією, а не простим прогнозуванням;

– класифікація за методом Роккі (Rocchio classification), яка заснована на векторній моделі і, в якій використовуються відгуки про релевантність вхідних елементів з метою підвищення точності рекомендаційної системи.

У цілому, можна відмітити про існування множини алгоритмів навчання для рекомендаційних систем, які застосовуються для машинного навчання, а також, які в будь-якому разі, потребують подальшого вдосконалення для отримання ще більшої точності надання рекомендацій.

Список використаних джерел:

1. Melville P., Mooney R., Nagarajan R. Content-Boosted Collaborative Filtering for Improved Recommendations. *National Conference on Artificial Intelligence : «AAAI-2002»* : materials (20-25 July 2016, Edmonton, Canada). Edmonton, Canada : AAAI, 2002. P. 187-192.
2. Linden G., Smith B., and York J. Item-to-Item Collaborative Filtering. Los Alamitos, CA, USA: IEEE Internet Computing, 2003. P. 76-80.
2. Офіційний блог для розробників компанії IBM. URL: <https://www.ibm.com/developerworks/ru/library/os-recommender1/> (дата звернення: 25.04.2019).
3. Мазурік О.Ю. Порівняльний аналіз моделей оцінювання в рекомендаційних системах. *Системний аналіз та інформаційні технології*. Київ : НТУУ «КПІ», 2016. С. 113.

Омельченко П.В.

студент,

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

АНАЛІЗ ПРОБЛЕМИ ГЕНЕРАЦІЇ БАЗИ ДАНИХ НА ОСНОВІ ОНТОЛОГІЇ ПРЕДМЕТНОЇ

Онтологія (за визначенням Н. Грубера) – специфікація концептуалізації [1]. Це спроба всеохоплюючого і детального формального опису певної області людських знань, за допомогою концептуальної схеми. Онтології придатні для розуміння людьми, але в той же час всебічний та формальний опис, придатний для автоматичної обробки спеціальними програмами. Онтології можуть використовуватись для:

- формування і фіксації загального знання певної предметної області або її частини;
- явної концептуалізації предметної області, що дозволяє описати семантику даних;
- опису певних компонентів, які можуть використовуватись як специфікація для певної задачі;
- розробки сховищ даних в якості високорівневого інтерфейсу;
- покращення взаємодії між програмними агентами та розробниками;
- уніфікації обміну даними;
- формалізації процесів специфікації.