

ПОРІВНЯННЯ ТА ВИЗНАЧЕННЯ НЕДОЛІКІВ РОБОТИ АЛГОРИТМІВ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

Джичка Ярослав

Науковий керівник: канд.техн.наук, доцент Бредіхін В.М.

Харківський національний університет міського господарства

імені О.М. Бекетова, м. Харків, Україна

У статті проілюстровано застосування математичних методів для побудови рекомендаційних систем. Досліджено основні підходи до створення рекомендаційної системи. Розглянуті основні види та алгоритми побудови рекомендаційних систем. Проведено аналіз роботи алгоритмів, що дозволило виявити їх загальні недоліки та переваги. Приділено увагу важливості машинного навчання і перераховані основні проблеми даного класу методів. Результати дослідження дають можливість оцінити точність, зрозумілість і корисність рекомендацій що лежить в основі для створення ідеальних рекомендаційних систем та сервісів.

Ключові слова: рекомендаційні системи, фільтрація на основі змісту, колаборативна фільтрація, гібридний підхід, аналіз даних.

Comparison and definition of work defects algorithms of recommendation systems

Y. Dzhichka

Supervisor: Candidate of Technical Sciences, Associate Professor Bredikhin V.M.

Kharkiv National University of Municipal Economy named after OM Beketova,

Kharkiv, Ukraine

The article illustrates the application of mathematical methods for the construction of recommendation systems. The main approaches to the creation of a recommendation system are studied. The main types and algorithms construction of recommendation systems are considered. The analysis of algorithms operation is carried out that allowed to reveal their general lacks and advantages. Attention is to the importance of machine learning and the main problems of this class of methods are listed. The results of the study provide an opportunity to assess the accuracy, clarity and usefulness of the recommendations that underlie the creation of ideal recommendation systems and services.

Keywords: recommendation systems, content-based filtering, collaborative filtering, hybrid approach, data analysis.

Постановка проблеми. На сьогоднішній день інформаційні технології ввійшли майже в усі сфери людського життя, це пов'язано з величезним обсягом інформації, з якої людина зустрічається щодня. В усьому світі

збільшується кількість додатків, які розробляються для промисловості, медицини, комерційних цілей. У багатьох сферах більшість обладнань і програм функціонують із застосуванням штучного інтелекту (ШІ) і кардинальним образом трансформують умови існування людини [1]. Рекомендаційні системи є одним із прикладів таких систем до яких можна віднести Кинопоиск, Spotify або Amazon.

Рекомендаційні системи використовуються для надання рекомендацій товарів або послуг, які дозволяють користувачам орієнтуватися в численній інформації в Інтернеті, а також поліпшують взаємини із клієнтами. Розробка подібних систем — це велика ніша інтелектуального аналізу даних.

Аналіз досліджень і публікацій. Перші рекомендаційні системи з'явилися наприкінці ХХ сторіччя. Рекомендаційний сервіс – це найефективніше застосування штучного інтелекту при створенні мобільних додатків. Однак такі системи мають ряд проблем, пов'язаних з надлишком даних і складністю моделювання поведінки деяких користувачів. Інші проблеми пов'язані з масштабністю і конфіденційністю зібраних даних.

Основна мета будь-якої рекомендаційної системи - видача рекомендацій на основі прогнозів. Тому для проектування такої системи використовують безліч різних алгоритмів, методів і їх варіацій, які допомагають працювати не тільки з кількісними даними а і з якісними показниками. У якості одного з інструментів для побудови рекомендаційних систем використовують теорему Байєса, а в якості методу, наприклад, Байєсові мережі довіри (Bayesian Belief Nets). Цей метод застосовується тоді, коли існує неповне розуміння предметної області та не вистачає знань, і завдання має випадковий характер.

Ще одним методом, заснованому на тому ж підході, є Ланцюги Маркова (Markovchains). З його допомогою можна вирішувати проблему вироблення рекомендацій як послідовну оптимізацію, а не як просте прогнозування. Найчастіше Ланцюги Маркова використовують при генерації текстів. Обидва розглянутих методи легкі в застосовні для розв'язку якісних завдань.

Виклад основного матеріалу. Існує три основні підходи до створення

рекомендаційних систем:

- аналіз змісту (content-based);
- колаборативна фільтрація (collaborative filtering);
- гібридний підхід, що поєднує два попередні [2].

Системи, побудовані на основі підходу content-based, рекомендують користувачу об'єкти, які схожі на ті, що користувач вже вжив або використовував. Основна проблема таких систем – відсутність даних про нових користувачів що має назву «холодний старт».

Системи колаборативної фільтрації для рекомендацій використовують історію оцінок як конкретного, так і інших користувачів. Такий вид рекомендаційних систем часто дає кращий результат.

Гібридні системи поєднують алгоритми двох попередніх підходів і використовуються для підвищення точності результату та ефективності системи.

Для моделювання роботи рекомендаційної системи використовуються різні алгоритми, у числі яких:

- алгоритм розрахунків коефіцієнта кореляції Пірсона;
- алгоритм кластеризації;
- user-based алгоритм;
- item-based алгоритм;
- алгоритм спільної фільтрації;
- алгоритм фільтрації змісту;
- алгоритм SVD [3].

Кожний з алгоритмів має як переваги, так і недоліки.

Алгоритм розрахунків коефіцієнта кореляції Пірсона дозволяє виявити подібності між користувачами системи, за допомогою визначення лінійної залежності між елементами, і він є досить популярним алгоритмом у сфері колаборативної фільтрації. Використовують даний алгоритм для порівняння об'єктів, але для характеристики користувачів даний алгоритм не підходить.

Кластеризація дає можливість сформулювати категорії, кластери осіб, що

мають загальні «сучасні переваги», схожі смаки. За допомогою кластерного аналізу можливо досягти тільки приблизної точності в вирішенні завдання. Метод кластеризації простий, але має свої недоліки, наприклад, не враховується специфіка нового користувача.

User-based алгоритм допомагає позбутися деяких проблем кластерного аналізу. Алгоритм дозволяє розрахувати середню оцінку користувача за рахунок підсумку середньої оцінки конкретного користувача та середньої зваженої величини оцінок усіх інших користувачів. Недолік даного алгоритму полягає в тому, що абсолютно нові об'єкти важко піддаються рекомендаціям.

Алгоритм Item-based це алгоритм на основі елементів. Проблема нових елементів залишається, як і в алгоритмі user-based. Також слід відзначити, що рекомендації при використанні такого методу часто залишаються тривіальними. Алгоритм Item-based і user-based мають низьку ефективність.

Розподіл даних за кластерами дає і алгоритм спільної фільтрації. В основі даного алгоритму лежить матриця корисності, у стовпцях якої перебувають предмети рекомендацій, а в рядках - користувачі. Рекомендації визначаються на основі рейтингів - значень у клітинках даної матриці. Недолік алгоритму полягає в тому, що не завжди матриця корисності може бути повністю заповнена тому що дані розріджені. Для боротьби з виділеною особливістю використовують даний алгоритм разом з методом кластеризації, але за рахунок цього точність рекомендацій для конкретного користувача знижується.

Алгоритм фільтрації змісту є симетричним попередньому й полягає у вивченні системою якихось параметрів контенту та користувача, які потім служать підставою для рекомендації. Складність збору певного набору даних - основна проблема даного алгоритму. Як і алгоритм спільної фільтрації, алгоритм фільтрації змісту не є універсальним.

За рахунок використання матриць в алгоритмі SVD, дані стають наочними й зрозумілими, а завдяки властивостям і особливостям самих матриць, точність отриманих рекомендацій досить висока. Алгоритм є простим,

але з його допомогою стає можливим виявити приховані ознаки об'єктів і інтереси користувачів. Головним недоліком алгоритму є відсутність на перших етапах матриці оцінок, яка зазнає використовується в подальшому розкладанні, причому таке розкладання є не єдиним.

На рисунку 1 показані групи основних недоліків, які мають більшість алгоритмів для побудови рекомендаційної системи.

Недоліки	Алгоритм кореляції Пірсона;	Алгоритм кластеризації	user-based алгоритм	item-based алгоритм	Алгоритм спільної фільтрації	Алгоритм фільтрації змісту	SVD алгоритм
Низька ефективність	+	+	+	+	-	+	-
Неточність результатів	+	+	-	-	+	+	-
Відсутність універсальності	+	+	+	+	+	+	-
Недолік інформації/даних	+	+	+	+	+	+	+
Проблема «Холодного старту»	-	+	+	+	+	+	+
Складність реалізації	-	-	+	+	+	+	+

Рис. 1. Недоліки алгоритмів рекомендаційних систем [4]

Виходячи з аналізу перерахованих вище алгоритмів, можна зробити висновок про наявність загальних проблеми та недоліків:

- можливість рекомендувати зовсім новий об'єкт (або нового користувача) відсутня проблема «холодного» старту;
- тривіальність рекомендацій;
- недолік інформації для одержання якісного прогнозу.

Щоб упоратися з деякими недоліками при моделюванні рекомендаційних систем і одержати модель пророкування, слід звернутися до машинного навчання (Machinelearning). Методи машинного навчання дозволяють перейти до більш глибокого аналізу даних, що дозволяє одержати раніше недоступну аналітичну інформацію, аналізувати поведінку користувачів і об'єктів, тим самим можливим стає одержання рекомендацій для нового користувача, поступово позбуваючись від проблеми недоліку даних.

Однак і ці системи мають недоліки. Важливе значення для завдань

машинного навчання має формування безлічі навчальних даних. Щоб одержати виграш, необхідно створювати навчальну базу великих розмірів. Але при цьому виникає час відведений на навчання зростає.

Ще одна проблема, це процес перенавчання. Зв'язане це з використанням надмірно складних моделей. Перенавчання спостерігається завжди, коли є оптимізація параметрів по кінцевій (свідомо неповної) вибірці [5].

Висновки. Джерел даних про об'єкти та користувачів з кожним днем стає усе більше. Завдяки прогресу в інформаційних технологіях можна стежити за фізичним станом об'єкта та настроєм користувача, його місцем розташування. Актуальність рекомендаційних систем з кожним днем збільшується. Основне завдання таких систем - одержати відповідну реакцію користувача, та його лояльність. Рекомендаційні системи можуть вирішувати проблему вибору користувача, відкривати йому нові обрії. Точність, зрозумілість і корисність рекомендацій лежить в основі для створення ідеальних рекомендаційних систем і сервісів.

Список літератури

1. Adomavicius G. and Kwon Y. "New recommendation techniques for multicriteria rating systems" in IEEE Intelligent Systems, 22, p. 48–55, May 2017.
2. Burke R. "Hybrid Recommender Systems: Survey and Experiments" in UMUAI 12 (4), pp. 331- 370, 2020.
3. Dehuri S. "Intelligent Techniques in Recommendation Systems: Contextual Advancements and New Methods", IGI Global, 2021, p. 51.
4. Хабрахабр. Рекомендательные системы: "You can (not) advise": [Electronic resource] – Access mode: <https://habrahabr.ru/post/176549/>. [Дата звернення: 21.10.2021].
5. Ю.В. Парфененко, А.А. Ковтун, А.А. Вербицька. Рекомендаційна інформаційна система для пошуку відеоматеріалів // Вісник КрНУ імені Михайла Остроградського. Випуск 5/2019 (118) С. 97-102