

Кучерук Володимир Юрійович
Глушко Михайло Васильович

МЕТОДИ ОЦІНКИ ЯКОСТІ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

Вінницький національний технічний університет

Анотація

У даній статті описуються деякі типи рекомендаційних систем і методи їх оцінки. Рекомендаційні системи є невід'ємною частиною сервісів, які надають користувачеві якийсь вибір серед безлічі альтернатив. Користуючись будь-яким сервісом, більшість людей так і не замислюються про те, що ж їм надає все нові і нові рекомендації. Дані системи істотно полегшують використання таких сервісів, так як вони надають рекомендації на основі інформації про користувача. Особливу увагу в статті приділено методам оцінки якості рекомендаційних систем. Оцінка рекомендаційних систем необхідна для того, щоб зрозуміти, наскільки ефективна та чи інша система. Ефективність системи дуже важлива, так як, чим ефективніше вона, тим краще і якісніше рекомендації вона надає. Стаття розділена на дві смислові частини: в першій говориться про те, що таке рекомендаційна система, які види рекомендаційних систем існують, також описані переваги та недоліки кожного з видів, у другій описуються методи оцінки рекомендаційних систем. У другій смисловій частині говориться про два методи оцінки рекомендаційних систем: метод оцінки якості рекомендацій і метод оцінки якості роботи алгоритмів.

Ключові слова: рекомендаційна система, методи оцінки, фільтрація, колаборативна.

Abstract

This article describes some types of recommendation systems and methods for their evaluation. Recommendation systems are an integral part of services that provide the user with a choice among a variety of alternatives. Using a service, most people do not even think about what gives them more and more new recommendations. These systems greatly facilitate the use of such services, since they provide recommendations based on user information. Particular attention is paid in the article to methods for assessing the quality of reference systems. Evaluation of advisory systems is necessary in order to understand how effective this or that system is. The effectiveness of the system is very important, since the more effective it is, the better and more qualitative recommendations it provides. The article is divided into two semantic parts: the first one deals with what is the recommendation system, which types of recommendation systems exist, the advantages and disadvantages of each species are also described, the second describes methods for evaluating the recommendation systems. In the second semantic part, two methods for evaluating recommendatory systems are described: the method for assessing the quality of recommendations and the method for evaluating the quality of work of algorithms.

Keywords: advisory system, assessment methods, filtering, collaborative.

Що таке рекомендаційні системи? Кожна рекомендаційна система спрямована на взаємодію з користувачем, якому надається величезна кількість об'єктів, серед яких він повинен зробити свій вибір. Користувачеві може не вистачати знань або досвіду, або того й іншого, щоб вибрати те, що відповідає його потребам. Користувач явно або побічно надає системі інформацію про свої уподобання. Таким чином, рекомендаційна система представляється у вигляді системи (програми), що використовує певний алгоритм фільтрації та наявну інформацію про потреби користувача, щоб рекомендувати йому набір об'єктів, які він вважає найбільш корисними для себе [1]. Рекомендаційні системи

використовуються в середовищі електронної комерції [2], пошукових системах, системах електронної освіти [3].

Рекомендаційні системи класифікують за способом відбору необхідного користувачеві матеріалу [4]. В основному застосовується два базові підходи: колаборативна фільтрація і контентна фільтрація. Також існує гібридна фільтрація, яка поєднує в собі як колаборативну, так і контентну фільтрацію.

У рекомендаційних системах, які використовують контентну фільтрацію (фільтрація по вмісту), користувачі не залежать від інших користувачів системи [5]. Для формування рекомендацій системі необхідний профіль користувача з інформацією про його інтересах. У профілі в певній формі зберігається інформація про визначні місця. Система також містить інформацію про всі предмети, які вона може рекомендувати. Така система використовує опис об'єктів в профілі користувача, знаходить схожі об'єкти в своїй базі даних, а потім рекомендує їх йому. Застосування фільтрації такого роду дуже доречно, коли користувач має чітко визначені, конкретні інтереси і шукає схожі рекомендації. Перевага контентної фільтрації полягає в тому, що для початку надання рекомендацій не потрібно велику кількість зареєстрованих користувачів, тобто рекомендації не залежать від інших користувачів системи. Основним обмеженням даного методу є неможливість системи з таким видом фільтрації рекомендувати нові об'єкти, які не відповідають інтересам користувача.

Для того, щоб система мала високоякісні рекомендації, необхідно постійно оновлювати інформацію про інтереси користувача. Але існує й інша проблема користувач завжди хоче мінімізувати час, що витрачається на взаємодію з системою, і неохоче ділиться інформацією для зворотного зв'язку з системою. Це типова проблема для всіх рекомендаційних систем, і оскільки рекомендації засновані виключно на інформації профілю користувача, тобто чим менше інформації користувач надає, тим менше відповідний набір рекомендацій користувач отримає. У такому випадку для вирішення цього завдання використовуються автоматичні системи збору інформації для підтримки профілю, але це ускладнює реалізацію системи. Цей метод фільтрації використовується в електронній комерції. У поєднанні з колаборативною фільтрації, використовується в таких областях, як електронна освіта, програмування і т. д.

Рекомендаційні системи, що використовують колаборативну фільтрацію, шукають користувачів зі спільними інтересами і рекомендують об'єкти, які були високо оцінені іншими користувачами [5, 6]. Інтереси зберігаються у вигляді графіка переваг, який показує, наскільки користувачеві подобається той чи інший предмет. Такої системи досить для оцінок з обраної тематики. Перевага цієї системи полягає в тому, що система повинна знати інтереси багатьох користувачів. Ця система не може створювати рекомендації, засновані тільки на користь однієї людини.

Головним недоліком колаборативної фільтрації є «холодний початок». Система не здатна ефективно генерувати рекомендації, якщо більша частина користувачів не вкажуть свої інтереси в профілі.

Гібридні системи зазвичай поєднують колаборативну фільтрацію і контентну фільтрацію. Це дозволяє вирішити ряд проблем, які виникають при використанні цих методів окремо. У гібридній системі інформація про інтереси користувачів представлена в профілі в двох видах як перелік властивостей певного об'єкта і як його оцінка користувачем. Ця характеристика є як перевагою, так і недоліком системи. Перевагою є повнота інформації, що дозволяє використовувати ефективні алгоритми фільтрації і формувати необхідні рекомендації. Недоліком є те, що користувач повинен ввести більше інформації в свій профіль, яка завжди дуже неохоче, або не робиться взагалі.

Якість рекомендаційних систем можна визначити, оцінивши якість роботи алгоритмів, якість рекомендацій.

Методика оцінки якості рекомендацій рекомендаційних систем.

Нехай вихідні дані: список користувачів $U = (u_1, \dots, u_m)$ і об'єкти $I = (i_1, \dots, i_n)$ представлені об'єктно-ознаковою таблицею. Кожному користувачеві в таблиці відповідає один рядок. В ході взаємодії користувачів з системою здійснюється заповнення (формування) даної таблиці. Бінарне відношення $T \subseteq U \times I$ показує, чи оцінив користувач $u \in U$ об'єкт $i \in I$, який можна розглядати як рейтинг $r_{u,i}$ об'єкта i у користувача u .

Для оцінки якості рекомендацій з точки зору точності і повноти існує кілька стандартних підходів [5]. Одним з таких підходів є метод змінного контролю [7]. Використання даного методу полягає у разі необхідності розділення вихідного набору множин U на дві підмножини: тестову U_{test} і навчальну $U_{training}$. Повнота і точність пошуку оцінюється на тестовій вибірці. Кожен елемент u тестової вибірки U_{test} розбивається на дві частини: на оцінені ознаки $I_{visible}$ і неоцінені I_{hidden} .

Алгоритм рекомендацій по оцінках користувачів використовує схожість об'єктів з тестової і навчальної вибірки для формування рекомендацій. Кожний користувач з U_{test} отримує рекомендації, як множина фіксованого розміру $r_n(u) = \{i_1, i_2, \dots, i_n\}$. Далі точність і повнота розраховуються наступним чином:

$$recall = \frac{|r_n(u) \cap u^I \cap I_{hidden}|}{|u^I \cap I_{hidden}|}, \quad (1)$$

$$precision = \frac{|r_n(u) \cap u^I \cap I_{hidden}|}{|r_n(u) \cap I_{hidden}|}, \quad (2)$$

де, u^I – сукупність всіх оцінених користувачем u об'єктів із I . Значення цих заходів обчислюються для кожного користувача і усереднюються. Експеримент проводиться для розбиття початкової множини, з якою відбувається вибір, на тестову і навчальну множину кілька разів. Значення заходів також усереднюються. В додатку до стандартних способів обчислення точності і повноти в статті [8] пропонуються нетрадиційні варіанти розкриття невизначеностей, що виникають при їх оцінці: $|u^I \cap I_{hidden}| = 0$, то $recall=1$. Якщо $|r_n(u) \cap I_{hidden}| = 0$ і $u^I = 0$, $toprecision = 1$, інакше $precision = 0$.

Методика оцінки якості роботи алгоритмів.

Існує безліч метрик для оцінки якості алгоритмів. В основному це метрики для оцінки точності очікуваного і реального значення. Розглянемо деякі з них: MAE, RMSE.

MAE (Mean Absolute Error, переклад. Середня абсолютна помилка) - помилка оцінюється як різниця між розрахунковою і фактичною оцінкою по модулю:

$$MAE = \frac{\sum_{i \in n} |P_i - R_i|}{n} \quad (3)$$

RMSE (Root Mean Squared Error, переклад. Середня квадратична помилка) - помилка обчислюється як корінь з суми квадратів різниць між передбачуваним і дійсним значенням:

$$RMSE = \sqrt{\frac{\sum_{i \in n} |P_i - R_i|^2}{n}} \quad (4)$$

Інші характеристики рекомендаційних систем також можуть бути оцінені, але на основі всього переліку рекомендацій. Для цього необхідно мати інформацію про те, які елементи в списку рекомендацій відповідають запиту, а які ні. Виходячи з цього, характеристику точності можна представити у вигляді формули:

$$Accuracy = \frac{||T||}{L}, \quad (5)$$

де, L - довжина всього списку рекомендацій, T – множина відповідних рекомендацій, F – множина невідповідних елементів в списку - F .

Характеристику повноти, в свою чергу, можна представити у вигляді такої формули:

$$Recall = \frac{||T||}{||T|| + ||G||}, \quad (6)$$

де, G - безліч елементів, які повинні були бути порекомендованими, але не виявилися в списку. Обчислювальний експеримент У жовтні 2006 року Netflix, велика компанія з оренди DVD, проводила відкрите змагання на кращий алгоритм передбачення оцінки, яку глядач поставить фільму, на основі попередніх оцінок цього та інших глядачів. Головний приз становив \$1000000. У Netflix використовується алгоритм, заснований на лінійній регресії, який називається Cinematch. Тривіальний алгоритм, який ставить фільму середню оцінку, дає середньоквадратичне відхилення (СКВ) 1,0540. Алгоритм Cinematch, який використовує дані, надані учасникам змагання, дає СКВ 0,9525 на тестовій множині. Це означає, що він на 10% краще тривіального алгоритму. Звичайно, були суперечки з приводу того, чи дійсно зменшення помилки на 10% допоможе користувачам? На думку одного з учасників змагання, поліпшення помилки навіть на 0,01 істотно змінить список 10 рекомендованих фільмів.

Всім учасникам були надані 4 файли:

- Тренувальні дані, в якому були оцінки кожного клієнта, про кожному фільмі;
- Кваліфікаційні дані, які містили інформацію про те який користувач, який фільм, коли оцінив;
- Назви і роки виходу фільмів;
- Фрагмент тренувальної бази даних, розподілене так само, як кваліфікаційні дані.

Ніякої інформації про користувачів, в базах даних, не було надано учасникам з метою конфіденційності користувачів. Деякі кваліфікаційні та тренувальні дані були навмисно змінені в тих же цілях. Щороку кращій команді видавався приз в \$ 50000, якщо не було досягнуто головної мети. Майже через три роки (26.07.2009) змагання було припинено, перемога і приз були присвоєні команді «Bellkor's Pragmatic Chaos», яка покращила алгоритм на 10,09% і зменшила СКВ до 0,8554. Церемонія нагородження та оголошення результатів проводилися 21 квітня 2009 [9].

З усього цього випливає, якість рекомендаційної системи залежить від якості алгоритмів. Якщо поліпшити якість алгоритмів, то покращиться якість рекомендацій, а це значить, що рекомендаційна система стане більш ефективною. У рекомендаційній системі компанії Netflix, використовується колаборативна фільтрація [10]. Незважаючи на всі недоліки такої фільтрації, якщо поліпшити її алгоритм, можна отримувати досить точні рекомендації. Проведемо експеримент, оцінимо якість рекомендацій рекомендаційної системи NetFlix, для цього реєструватимемося на сайті "КиноПоиск" [Www.kinopoisk.ru]. Для оцінки якості рекомендацій по повноті використовуємо формулу:

$$recall = \frac{|r_n(u) \cap u^I \cap I_{hidden}|}{|u^I \cap I_{hidden}|},$$

Так як ми тільки зареєструвались і нічого не замовляли і не шукали, то $u^I = 0$. Із цього слідує, що $|u^I \cap I_{hidden}| = 0$, а це означає що $recall = 1$.

Оцінимо якість рекомендаційної системи по точності. При першому вході в систему кількість рекомендацій не оцінених об'єктів, тобто об'єкти, які користувач не переглядав, дорівнюють кількості об'єктів в цій системі, тобто $|r_n(u)| = 107$ і $I_{hidden} = 107$. Із цього випливає, що $|r_n(u) \cap I_{hidden}| \neq 0$. Так як ми тільки що зареєструвались і нічого не замовляли і не шукали, то $u^I = 0$. Значить, $precision = 0$.

Якщо подивитися на результати оцінки якості рекомендаційної системи, що використовує колаборативну фільтрацію, то можна зрозуміти, що в ній є проблема "холодного початку". Система не здатна ефективно генерувати рекомендації, якщо більша частина користувачів не вкажуть свої інтереси в профілі. Тобто, в даному випадку постраждала точність, так як при створенні аккаунта не були зазначені інтереси користувача і схожих користувачів не виявилось.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Jannach D., Zanker M., Felfernig A. Friedrich G. Recommender Systems. An Introduction. New York: Cambridge University Press 32 Avenue of the Americas, 2011. 352 P.
2. Вахрушева М.Ю., Евдокимов И.В. Разработка программного обеспечения аналитических информационных систем // Труды Братского государственного университета. Серия: Экономика и управление. 2014. Т. 1. № 1. С. 196-199.
3. Кокташев В.В., Макеев В.В., Михалев А.С. Применение систем управления инцидентами для разработки программного обеспечения в образовании // Современные информационные технологии. 2017. №4 (30). С. 129-135.
4. Джонс Т.М. Рекомендательные системы. Часть 1. Введение в подходы и алгоритмы. [Электронный ресурс]: <https://www.ibm.com/developerworks/ru/library/os-recommender1/> (дата обращения: 11.05.2018).
5. Melville P., Sindhvani V. Recommender systems. Encyclopedia of Machine Learning. 2010.
6. Su X., Khoshgoftaar Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence. 2009.
7. Воронцов К.В. Комбинаторный подход к оценке качества обучаемых алгоритмов. Математические вопросы кибернетики / Под ред. О.Б. Лупанова. – М.: Физматлит, 2004. – Т. 13. – С.5-36.
8. Игнатов Д.И., Каминская С.Ю., Магизов Р.А. Метод скользящего контроля для оценки качества рекомендательных Интернет-сервисов // Труды 12й национальной конференции по искусственному интеллекту, М., Физматлит, Т. 1., С.183-191, 2010.
9. Netflix. [Электронный ресурс]: <https://www.netflix.com/ru/> (дата обращения: 11.05.2018).
10. В.Ю. Кучерук, М.В. Глушко. Покращення алгоритму "item to item" методу колаборативної фільтрації для розробки рекомендаційних систем на основі косинусної міри шляхом оцінки релевантності // ScienceRise. - 2018. - № 1. - С. 20-24.: http://nbuv.gov.ua/UJRN/text_2018_1_6

Кучерук Володимир Юрійович, д.т.н., професор. Завідувач кафедри МПА Факультету Комп'ютерних систем і автоматички. Вінницький національний технічний університет. Вінниця.

Глушко Михайло Васильович, студент-аспірант кафедри МПА Факультету Комп'ютерних систем і автоматички. Вінницький національний технічний університет. Вінниця.

Kucheruk Volodymyr Yuriiovich, Doctor of Technical Sciences, Professor. Head of the MPA Department of the Faculty of Computer Systems and Automation. Vinnytsia National Technical University. Vinnytsia.

Hlushko Mykhailo Vasyliovich, post-graduate student of the MPA department of the Faculty of Computer Systems and Automation. Vinnytsia National Technical University. Vinnytsia.