

УПРАВЛІННЯ ПОШТОВОЮ РОЗСИЛКОЮ НА ОСНОВІ МЕТОДІВ ФІЛЬТРАЦІЇ

Д.т.н. В.В. Безкоровайний, М.С. Кротких, Харківський національний університет радіоелектроніки

Виконано огляд методів формування рекомендацій для управління поштовою розсилкою, що використовують методи фільтрації. На основі аналізу переваг та недоліків існуючих методів запропонована модифікація методу формування переліку рекомендацій для формування розсилок. Проведено експериментальне дослідження запропонованої модифікації методу, методу сингулярного розкладання, методів, що орієнтовані на користувачів та об'єкти, з використанням різних методів визначення подібності векторів уподобань, яке виявило його перевагу за показником точності.

Выполнен обзор методов формирования рекомендаций для управления почтовой рассылкой, использующих методы фильтрации. На основе анализа преимуществ и недостатков существующих методов предложена модификация метода формирования перечня рекомендаций для формирования рассылок. Проведено экспериментальное исследование предложенной модификации метода, метода сингулярного разложения, методов, ориентированных на пользователей и объекты, использующих разных методов вычисления подобия векторов предпочтений, которое выявило его преимущества по показателю точности.

The methods of forming recommendations for mailing list management using filtering methods are reviewed. Based on an analysis of the advantages and disadvantages of existing methods, a modification of the method of forming a list of recommendations for the formation of mailings is proposed. An experimental study of the proposed modification of the method, the method of singular decomposition, methods aimed at users and objects using different methods for calculating the similarity of preferences vectors, which revealed its advantages in terms of accuracy was carried out.

Ключові слова: поштова розсилка, управління, рекомендаційні системи, колаборативна фільтрація, контентна фільтрація.

Вступ

Обсяги доступної для користувачів комп'ютерних мереж інформації стрімко зростають, набагато перевищуючи можливості людини з її переробки. З метою створення комфортних умов для користувачів мережевих сервісів створюються так звані рекомендаційні системи, які знаходять все більш широке розповсюдження у професійній діяльності, в Інтернет-комерції, у побуті [1–5]. Рекомендаційні системи являють собою спеціальний тип техніки фільтрації інформації, що дозволяє надавати користувачеві об'єкти відповідні його інтересам.

Зазвичай рекомендаційні системи використовуються в додатках Інтернет або адаптованих під споживача веб-сайтах. Це робить актуальними завдання створення інформаційно-аналітичних систем, які б збирали інформацію щодо уподобань кожного користувача та формувати для них відповідні пропозиції,

зменшуючи при цьому обсяги поштових розсилок.

Для формування рекомендацій користувачам мереж на практиці використовуються методи фільтрації, що засновані на вмісті, знаннях, контексті та технології колаборативної фільтрації [6]. Кожен з методів має певні переваги та недоліки.

Метою роботи є підвищення ефективності методів управління поштовою розсилкою в Інтернет-сервісах за рахунок використання у рекомендаційних системах сучасних процедур фільтрації.

Для досягнення мети необхідно:

- виконати аналітичний аналіз сучасних методів фільтрації інформації;
- удосконалити найбільш ефективний з існуючих методів фільтрації, за допомогою якого можна було б здійснювати в автоматичному режимі створення та відправлення актуальної інформації користувачеві;
- провести порівняльне експериментальне дослідження ефективності методів управління поштовою розсилкою на основі процедур фільтрації.

Методи розв'язання задачі фільтрації

У сучасних системах управління поштовою розсилкою для генерації рекомендацій використовують декілька підходів і методів до розв'язання задачі фільтрації [6]: фільтрація, що заснована на вмісті; стохастичний метод формування рекомендацій; фільтрація, що заснована на знаннях; фільтрація, що заснована на контексті; колаборативна фільтрація.

Фільтрація, заснована на вмісті. Рекомендаційні системи з тематичною фільтрацією відбирають об'єкти, що ще не мали оцінки користувача, але які найбільш схожі на ті, що викликали його зацікавленість [7, 8].

Рекомендаційні системи, що використовують цей метод, виконують аналіз множин описів об'єктів, які отримали оцінки користувача раніше, створюють профіль (модель) інтересів і переваг, заснованих на особливостях цих об'єктів. Рекомендаційний процес у своїй основі складається із зіставлення атрибутів профілю користувача й атрибутів об'єкта. Результатом є оцінка релевантності, що показує рівень потенційної зацікавленості користувача об'єктом. Рекомендаційний процес у таких системах здійснюється в три етапи, кожен з яких забезпечується окремим компонентом: аналізатором контенту, аналізатором профілю та фільтром. Результатом є оцінка релевантності, що служить для побудови рангового списку потенційно цікавих активному користувачеві об'єктів [7].

Перший етап проводиться аналізатором контенту, який отримує ключові слова з неструктурованого тексту і виробляє структуроване уявлення об'єкта, що зберігається в базі даних «Об'єкти».

Для побудови та оновлення профілю активного користувача u_a , його реакція на об'єкти збирається і зберігається в базі даних «Зворотній зв'язок». Запис зворотного зв'язку користувача може здійснюватися

двома різними техніками: явним і неявним збором даних.

Володіючи великою кількістю об'єктів з відповідними їм рейтингами, виробленими користувачем u_a , $TR_a = \{I_k, r_k\}$ аналізатор профілю застосовує алгоритми навчання для отримання прогнозу моделі – профілю користувача, що зберігається в базі «Профілі» для подальшого використання фільтром.

Фільтр, отримуючи на вхід новий об'єкт, застосовує певну стратегію для визначення релевантності цього об'єкта користувачеві, зіставляючи профіль користувача і характеристику об'єкта, а потім формує список рекомендацій для користувача u_a .

Для виявлення ключових слів використовуються різні алгоритми, серед яких отримав найбільш широке поширення алгоритм, що використовує статистичну міру $TF-IDF$ [8].

Міра $TF-IDF$ використовується для оцінки важливості слова в контексті документа, що є частиною колекції документів. Вага деякого слова пропорційна кількості вживань цього слова в документі і обернено пропорційна частоті вживання слова в інших документах колекції.

TF (*term frequency*) – це нормалізована частота слова t у документі d :

$$TF(t, d) = n_i / \sum_{i=1}^k n_i, \quad (1)$$

де n_i – кількість входжень слова t в документ; знаменник – загальна кількість слів у документі d .

IDF (*inverse document frequency*) – інверсія частоти, з якою деяке слово зустрічається в документах колекції.

$$IDF(t, D) = \log \frac{|D|}{|t_i \subset d_i|}, \quad (2)$$

де $|D|$ – кількість документів в колекції; $|t_i \subset d_i|$ – кількість документів, в яких зустрічається слово.

З урахуванням (1) і (2) міра $TF-IDF$ визначається співвідношенням:

$$TFIDF(t, d, D) = TF(t, d) \times IDF(t, D). \quad (3)$$

Для порівняння двох текстів (об'єктів), їх можна подати у вигляді векторів в евклідовому багатовимірному просторі. Кожному вимірюванню відповідає слово, а значення компонент вектора дорівнює $TF-IDF$. Визначити релевантність двох текстів по відношенню один до одного можна порахувавши косинус між векторами, що їх подають [8].

Колаборативна фільтрація. Колаборативна фільтрація виробляє рекомендації, засновані на моделі попередньої поведінки користувача. Ця модель може бути побудована виключно на основі поведінки конкретного користувача або, що більш ефективно, з урахуванням поведінки інших користувачів з подібними характеристиками. У тих випадках, коли колаборативна фільтрація бере до уваги реакцію інших користувачів, вона використовує знання про групу (*group knowledge*) для вироблення рекомендацій на основі подібності користувачів. Рекомендації базуються на автоматичній взаємодії множини користувачів і виділенні (методом фільтрації) тих користувачів, які демонструють схожі

переваги або шаблони поведінки [9].

Інший спосіб розгляду цих відносини заснований на їх схожості та відмінності. Схожість визначають, за якими ознаками (за допомогою відповідного алгоритму) слід групувати користувачів, що мають схожі інтереси. Відмінності – це можливості, які можуть бути використані для вироблення рекомендацій – наприклад, за допомогою застосування фільтра популярності [10].

Алгоритм методу колаборативної фільтрації:

1. Для кожного користувача u обчислити, наскільки його інтереси збігаються з інтересами користувача a ;

2. Визначити множину користувачів, найбільш близьких за вподобаннями до користувача a ;

3. Визначити оцінку на основі оцінок об'єкта і користувачів, найбільш близьких за вподобаннями з ним.

Оцінку ступеня близькості характеристик користувачів можна здійснювати через значення [7, 9]: косинуса кута між векторами їхніх характеристик; коефіцієнт кореляції Пірсона; міру близькості Дайса. Після визначення множини K найближчих за уподобаннями користувачів обчислюється значення оцінки, яку поставив би користувач a об'єкту i [10]. Вона являє собою середнє по всім користувачам з множини найближчих за уподобаннями K . При цьому, чим ближче користувач u за уподобаннями до користувача a , тим більший вплив він має на прогнозу оцінку.

Цей метод дозволяє прогнозувати оцінки для об'єктів, які користувач мережі ще не оцінював. Для того, щоб зробити рекомендацію для даного користувача, досить обчислити оцінку для всіх ще не оцінених об'єктів і обрати об'єкт з найбільшою оцінкою [8].

Практична реалізація методу колаборативної фільтрації може здійснюватися з використанням різних алгоритмів [9, 10].

Фільтрація, що заснована на контексті.

Контекстна фільтрація формує рекомендацію на основі поведінки користувача. У рамках цього підходу може використовуватись ретроспективна інформація щодо уподобань користувача. Фільтрація, що заснована на контексті, може використовувати ретроспективну інформацію для виявлення подібного контенту і рекомендувати його користувачеві. Цей контент може бути визначений у ручному режимі або автоматично на основі інших методів подібності [6].

Колаборативна фільтрація, що заснована на сусідстві користувачів. Методи колаборативної фільтрації цієї групи використовують для передбачення рейтингу користувача для нового об'єкта дані про оцінки цього об'єкта іншими користувачами, з найбільш близькими уподобаннями («найближчими сусідами») [6].

Якщо для кожного користувача мережі $u \in U$ встановлено значення, яке визначає подібність між перевагами щодо об'єкта i та k його «найближчих сусідів» $N_i(u)$, то рейтинг r_{ui} може бути розрахований як середній рейтинг даний i цими сусідами:

$$r_{ui} = \frac{1}{|N_i(u)|} \sum_{v \in N_i(u)} r_{vi}. \quad (4)$$

Однак, такий підхід не враховує того, що:

– сусіди можуть мати різний ступінь подібності в своїх перевагах;

– користувачі можуть використовувати різні

значення оцінки для позначення однакового рівня задоволеності об'єктом.

Для вирішення першої проблеми можна використовувати вагові коефіцієнти користувачів, позначаючи ступінь близькості профілю переваг одного користувача до іншого.

Для вирішенням другої проблеми можна використовувати нормалізацію рейтингів користувачів $h(r_{vi})$ [11]:

$$r_{ui} = h^{-1} \frac{\sum_{v \in N_i(u)} w_{uv} \cdot h(r_{vi})}{\sum_{v \in N_i(u)} |w_{uv}|} \quad (5)$$

Метод обчислення оцінки нового для користувача об'єкта, де прогнозовані рейтинги обчислюються як середньозважене оцінок сусідів, по суті, вирішує проблему регресії. Класифікація заснована на сусідстві, з іншого боку, знаходить найбільш ймовірний рейтинг відданий користувачем u до об'єкта i шляхом голосування найближчих сусідів користувача u за рейтинг.

Колаборативна фільтрація, що заснована на сусідстві об'єктів. Методи фільтрації цієї групи використовують рейтинги схожих об'єктів. Визначаючи k «найближчих сусідів» $N_i(u)$ як найбільш схожі об'єкти, оцінені користувачем u по відношенню до i , рейтинг розраховується як зважене середнє рейтингів відданих u до об'єктів $w_{ij}r_{uj}$:

$$r_{ui} = \frac{\sum_{j \in N_u(i)} w_{ij} \cdot r_{uj}}{\sum_{j \in N_u(i)} |w_{ij}|} \quad (6)$$

Різниця в індивідуальних призначених для користувача підходах до оцінювання може бути врахована введенням нормалізації h :

$$r_{ui} = h^{-1} \frac{\sum_{j \in N_u(i)} w_{ij} \cdot h(r_{uj})}{\sum_{j \in N_u(i)} |w_{ij}|} \quad (7)$$

Колаборативна фільтрація, що заснована на аналізі моделі даних. Методи, засновані на аналізі моделі даних, припускають, що спочатку за сукупністю оцінок формується описова модель вподобань користувачів, об'єктів і взаємозв'язків між ними, а потім формуються рекомендації на підставі отриманої моделі. Процес формування рекомендацій розбивається на два етапи: навчання моделі у відкладеному режимі і досить просте обчислення рекомендацій на основі існуючої моделі в реальному часі [9]. Такі алгоритми можуть бути засновані на імовірнісному підході, кластерному аналізі, а також на аналізі прихованих чинників [12].

Фільтрація, заснована на знаннях. Методи цієї групи використовуються для ситуацій, коли в системі управління розсилками не має потрібного обсягу інформації щодо здійснюваних користувачем транзакцій. Вони поділяються на дві основні групи[6]:

- використання жорстких обмежень;
- вибір близьких об'єктів.

Обидва методи містять у собі реалізацію ідеї, що користувач формує вимоги до об'єкту, а система

намагається знайти необхідний йому об'єкт. Відмінність методів полягає у тому, що перші зорієнтовані на пошук об'єктів, які точно відповідають усім вимогам користувача, а другі зорієнтовані на пошук об'єктів з характеристиками, близькими до вимог.

Якщо користувач вказує занадто жорсткі вимоги щодо об'єктів пошуку, то може виявитися що об'єкти з такими характеристиками відсутні в системі. У таких випадках необхідно в автоматичному режимі послідовно послаблювати початкові обмеження.

Для методу вибору подібних об'єктів характерним є формування рекомендацій, які в повному обсязі задовольняють вимогам. Міра близькості між вимогами і властивостями обчислюється за формулою [13]:

$$SIM(p, N_r) = \frac{\sum_{r=1}^{N_r} w_r \cdot sim(p, r)}{\sum_{r=1}^{N_r} w_r} \quad (8)$$

де p – об'єкт; r – вимога; N_r – кількість вимог користувача; w_r – вага вимоги r ; $sim(p, r)$ – міра близькості об'єкта p до вимоги r .

При цьому, чисельна міра близькості об'єкта p до вимоги r до вимоги може бути одного з трьох типів:

– «більше – краще»:

$$sim(p, r) = \frac{u_r(p) - \min(r)}{\max(r) - \min(r)}; \quad (9)$$

– «менше – краще»:

$$sim(p, r) = \frac{\max(r) - u_r(p)}{\max(r) - \min(r)}; \quad (10)$$

– «ближче – краще»:

$$sim(p, r) = 1 - \frac{|u_r(p) - f_r|}{\max(r) - \min(r)}, \quad (11)$$

де $\max(r)$, $\min(r)$ – максимальні і мінімальні значення властивості відповідно; f_r – точне значення вимоги r користувача; $u_r(p)$ – значення відповідної властивості r у об'єкта p .

Модифікація методу управління розсилкою.

Для підвищення ефективності технологій фільтрації пропонується для створення множини з n рекомендацій розсилки включати лише рекомендації щодо тих об'єктів, категорія яких у даний час найбільше відповідає користувачеві. Для цього необхідно ранжувати множину об'єктів за балами та датами виставлення рейтингу певної категорії. Якщо таких об'єктів менше ніж n , тоді необхідно включати до розсилки об'єкти, що мають найбільшу кількість високих оцінок за час після останньої розсилки.

Повідомлення такої поштової розсилки буде мати структуру, що включатиме:

- перелік рекомендованих об'єктів, що мають високий рейтинг і відносяться до категорій, яким підписувач нещодавно ставив високі бали рейтингу;

– перелік рекомендованих матеріалів, що мають найбільший рейтинг;

– перелік рекомендованих об'єктів, які були рекомендовані як найкращі за період після останньої розсилки.

Для формування рекомендацій визначається кількість матеріалів n , що будуть представлені у розсилці. Вони залежать від того, з яким інтервалом на веб-ресурсі відбувається публікація нової інформації: меншій частоті публікації відповідає менший обсяг рекомендацій та навпаки. Після цього визначається інтервал, з яким буде відбуватися розсилка.

Метод формування переліку з n рекомендацій включає такі етапи.

1. Якщо кількість матеріалів, що відносяться до категорій, якими підписувач цікавиться, дорівнює $i \leq n$, тоді до листа поштової розсилки буде введено i рекомендованих матеріалів a , з категорій, якими підписувач цікавився.

2. Якщо кількість матеріалів $a \leq n$, вважається, що кількості матеріалів не достатньо для формування розсилки. Тоді до переліку рекомендацій включаються об'єкти, що являються найбільш високо оцінюваними за останній час.

3. Якщо кількість матеріалів b , що відносяться до унікальних найпопулярніших об'яв, дорівнює $j \leq n - i$, тоді в лист поштової розсилки буде введено j рекомендованих матеріалів b найпопулярніших об'яв.

4. У разі якщо кількість матеріалів $b + a < n$, вважається, що кількості матеріалів не достатньо для формування розсилки. Тоді до переліку рекомендацій додаються c матеріалів, що відмічені адміністратором як найкращі за останній час. Якщо $c = k \leq n - i - j$, тоді до розсилки буде введено k рекомендованих матеріалів c .

При формуванні переліку рекомендацій повинні виконуватися такі умови:

– категорії матеріалів для етапів 2-3 повинні бути унікальними. Якщо на якомусь етапі виявлена об'ява, що повторювалась раніше, вона більше не буде використовуватися для уникнення повторення рекомендацій;

– кожен використаний для рекомендації матеріал повинен мати дату публікації не раніше зазначеного адміністратором інтервалу поштової розсилки щоб уникнути використання в якості рекомендації матеріалу, що був відправлений у попередній розсилці;

– порядок виводу матеріалів у розсилці здійснюється за датою публікації;

– у разі, якщо після етапів 1-3 не виявлено ні одного нового матеріалу, розсилка переноситься на наступний інтервал.

Порівняльна оцінка ефективності методів

За результатами експериментального дослідження проведено аналіз ефективності алгоритмів, що реалізують основні з розглянутих методів, і орієнтовані на користувача та об'єкт. У якості вхідних даних використовувалась інформація з бази даних MovieLens, що містить 100 000 рейтингів, що зробили 1000 користувачів стосовно до 1700 об'єктів. Дана вибірка містить перелік з ідентифікатором користувача, ідентифікатором об'єкта, який ним обрано. Перелік

об'єктів містить назву, ознаку та рейтинг, який виставив користувач (оцінку від 1 до 5 з кроком 1) у певний час [14].

Якість отримуваних рекомендацій щодо розсилок залежить від обраної метрики визначення схожості векторів об'єктів (наприклад, уподобань користувачів) і розмірів групи користувачів, між якими встановлюють подібність («сусідів»).

Для проведення експериментів був застосований *Apache Mahout*. Для визначення помилки рекомендацій обрано середнє абсолютне відхилення (*MAE* – *Mean Absolute Error*). Уся вибірка з бази даних рейтингів була поділена на тестову та навчальну у пропорції 1:9 відповідно.

За результатами експериментів з алгоритмом, орієнтованим на користувача (рис. 1), підтверджено, що збільшення кількості «сусідів» позитивно впливає на якість надання рекомендацій. Найкращий результат даний алгоритм продемонстрував з використанням евклідової відстані, визначеній між векторами вподобань користувачів у багатовимірному просторі з координатами-оцінками.

Проведені експерименти з алгоритмом, що орієнтований на об'єкти, зокрема, з підбором метрики обчислення подібності. Через те, що даний алгоритм не дозволяє корегувати кількість «сусідів», за результатами експериментів можна судити тільки про вибір тієї чи іншої метрики. Умови проведення експерименту і підібрані метрики аналогічні з попереднім експериментом (рис. 2). При цьому, підтверджується, що використання коефіцієнта Танімото, що більш застосовний до бульових даних, дозволяє отримати найкращий результат.

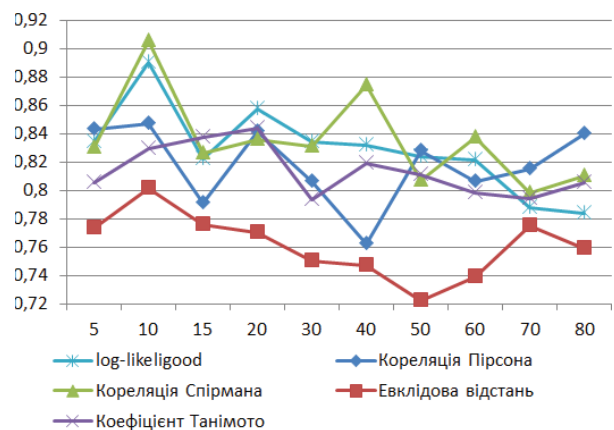


Рис. 1. Оцінки алгоритму, що орієнтований на користувача

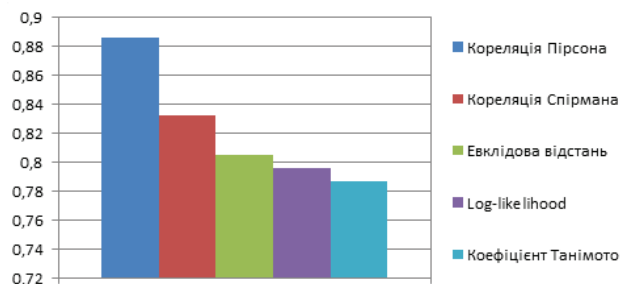


Рис. 2. Оцінки алгоритму, що орієнтований на об'єкти

При застосуванні алгоритму *SVD* (сингулярного розкладання) поряд з латентним фактором, що бере участь в обчисленні ваг для формування рекомендацій, важливе значення має вибір алгоритму факторизації (рис. 3) [15].

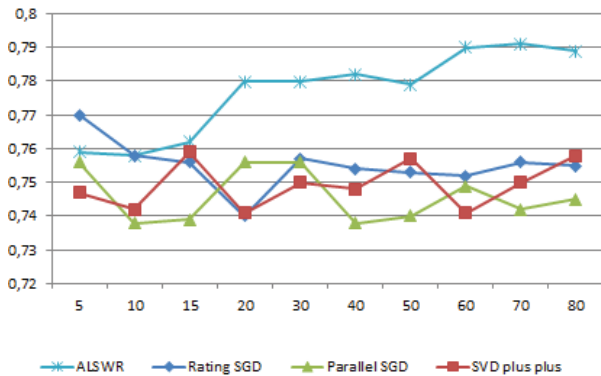


Рис. 3. Оцінки *SVD*-алгоритму з використанням різних методів факторизації

За результатами експериментів можна зауважити, що використання методів *Parallel SGD*, *Rating SGD* та *SVD plus plus* дозволяє отримувати подібні результати в межах 0.735-0.79, які є набагато кращими у порівнянні з іншими алгоритмами. Використання *ALS*-факторизатора продемонструвало найгірший результат навіть при збільшенні кількості латентних факторів через велике зашумлення, що пов'язане появою факторів у розрахунках, які не є важливими для формування моделі переваг користувачів.

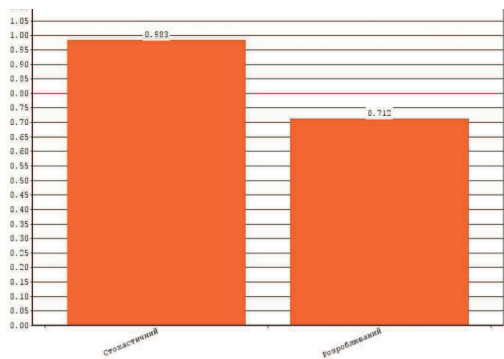


Рис. 4. Порівняльна оцінка запропонованої модифікації методу

Експериментальне порівняння роботи алгоритма випадкового формування рекомендацій та формування переліку рекомендацій за розробленим методом показало суттєву перевагу розробленої модифікації методу за показником точності (рис. 4).

Висновки

З метою підвищення ефективності методів управління поштовою розсилкою в рекомендаційних системах Інтернет-сервісів проаналізовано сучасні процедури фільтрації інформації. За результатами експериментальних досліджень найкращі показники точності рекомендацій продемонструвала запропонована модифікація методу формування рекомендацій для поштової розсилки. До числа методів, які продемонстрували високі показники ввійшли методи, що орієнтовані на користувача, в яких у якості метрики

подібності використовувалась евклідова відстань (алгоритм *SVD* з методами факторизації *rating SGD*, *parallel SGD*, *SVD plus plus*).

Розроблена модифікація методу формування рекомендацій поряд з іншими відомими методами фільтрації може успішно використовуватись для управління розсилками в інформаційних системах віддаленого продажу квитків на розважальні заходи, інтернет-магазинах, інформаційних сайтах тощо, що сприятиме підвищенню точності рекомендацій для користувачів комп'ютерних мереж.

СПИСОК ЛІТЕРАТУРИ:

1. Гомзин А.Г. Системы рекомендаций: обзор современных подходов / А. Г. Гомзин, А. В. Кориунов // Труды ИСП РАН. – 2012. – Т. 22, №3. – С. 401-417.
2. Архитектура рекомендательной системы, работающей на основе неявных пользовательских оценок: труды XIII Всероссийской научной конференции RCDL 2011. – Воронеж: Воронежский госуниверситет, 2011. – 76 с.
3. Recommender Systems Handbook / F. Ricci, L. Rokach, B. Shapira, P. B. Kantor. – New York: Springer, 2001. – 845 с.
4. Системы выработки рекомендаций / Business Data Analytics. – Режим доступа: [www / URL: http://www.businessdataanalytics.ru/RecommendationSystems.htm](http://www.businessdataanalytics.ru/RecommendationSystems.htm). – 01.02.2010 г. – Загл. с экрана.
5. Рынок удаленной продажи билетов России, 2008-2015, оценка J'son & Partners Consulting / Mskit.ru. Режим доступа: [www/ URL: http://mskit.ru/analytics/a86502](http://mskit.ru/analytics/a86502). – Загл. с экрана.
6. Рекомендательные системы. Часть 1. Введение в подходы и алгоритмы. Принципы работы рекомендательных механизмов Интернета / Ibm.com. Режим доступа: [www/ URL: https://www.ibm.com/developerworks/ru/library/os-recommender1/index.html](https://www.ibm.com/developerworks/ru/library/os-recommender1/index.html). – Загл. с экрана.
7. Melville P. Content-Boosted Collaborative Filtering for Improved Recommendations (англ.) / P. Melville, R. Mooney, R. Nagarajan // University of Texas, USA: Материалы конф. / AAAI-02, Austin, TX, USA, 2002. – 2002. – P. 187-192.
8. Кориунов А.Г. Системы рекомендаций: обзор современных подходов. / А.Г. Кориунов, А.В. Гомзин // Москва: Труды Института системного программирования РАН. 2012. – С. 20 – 22.
9. Бескоровайный В.В. Рекомендательная система на основе коллаборативной фильтрации / В.В. Бескоровайный, Ю.В. Мельничук // Информационные системы и технологии: материалы 4-й Международной науч.-техн. конф., Харьков, 21-27 сентября 2015г.: тезисы докладов. – X.: НТМТ, 2015. – С. 30-31.
10. Коллаборативная фильтрация / Intellect.ml. Режим доступа: [www/ URL: http://intellect.ml/kollaborativnaya-filtratsiya-4778](http://intellect.ml/kollaborativnaya-filtratsiya-4778). – Загл. с экрана.
11. Рекомендательные системы: SVD, часть I / Habrahbr.ru. – Режим доступа: [www / URL: https://habrahbr.ru/company/surfingbird/blog/139863](https://habrahbr.ru/company/surfingbird/blog/139863). – Загл. с экрана.
12. Zan Huang. Prediction Approach to Collaborative Filtering (англ.) / Z. Huang, L. Xin, C. Hsinchun // University of Arizona, USA: Материалы конф. / JCDL'05, Denver, Colorado, USA, June 7-11, 2005. – 2005.
13. Case-Based Recommendation: сб. науч. тр. / The Adaptive Web – Dublin: Springer Berlin Heidelberg, 2007. – 766 с.
14. SUMMARY & USAGE LICENSE [Электронный ресурс] / Files.grouplens. Режим доступа: [www/ URL: http://files.grouplens.org/datasets/movielens/ml-100k-README.txt](http://files.grouplens.org/datasets/movielens/ml-100k-README.txt). – Загл. с экрана.
15. Сингулярное разложение [Электронный ресурс] / Machine learning. Режим доступа: [www/ URL: http://www.machinelearning.ru/wiki/index.php?title=Сингулярное_разложение](http://www.machinelearning.ru/wiki/index.php?title=Сингулярное_разложение). – Загл. с экрана.