

УДК 004.4

Бублик В. В., Горностаї М. П.

ВПРОВАДЖЕННЯ СИСТЕМ НАДАННЯ РЕКОМЕНДАЦІЙ В ЕЛЕКТРОННЕ НАВЧАННЯ

Розглянуто системи надання рекомендацій, які доцільно використовувати для систем електронного навчання. Досліджено підходи до побудови профілю користувача для різних типів систем електронного навчання та систем надання рекомендацій, запропоновано алгоритм побудови профілю псевдо-активного користувача, що базується на характерних рисах.

Вступ. Персоналізація в Інтернеті

Термін «перевантаження інформацією» – синонімічний терміну «Інтернет», якщо брати до уваги величезний обсяг інформації, що існує в електронному форматі в Інтернеті, та нездатність людей сприймати всю цю інформацію. Ця проблема стає дедалі актуальнішою, якщо враховувати швидкі темпи надходження електронної інформації [1].

Протягом останнього десятиліття технології пошуку інформації розвивались швидкими темпами, з'явилися різні типи пошукових систем. Пошукові системи виконують завдання пошуку та роблять великі обсяги інформації доступними для широкого кола користувачів. Результати пошуку є корисними для користувача, якщо він точно визначив параметри пошуку. Проте часто пошукові системи повертають значно більше інформації, ніж користувач міг би обробити, тому для відбору потрібної інформації користувачеві потрібно витратити багато додаткових зусиль. Потреба забезпечити користувача саме необхідною йому інформацією привела до розробки різноманітних технологій фільтрування інформації, що будують профіль користувача (чи збирають інтереси користувача) та фільтрують великі обсяги інформації, надаючи користувачеві тільки ті елементи, які можуть викликати в нього інтерес [1]. Такі системи допомагають обійти обмеження, пов'язані з перевантаженням інформацією сучасного Інтернету, і їх називають системами надання рекомендацій.

Сфера застосування систем надання рекомендацій дуже широка – практично це будь-який сайт із великою кількістю елементів. Сьогодні існують реалізації систем надання рекомендацій для сайтів, що пропонують користувачеві новини, музику чи фільми, та для сайтів електронної комерції. Ці системи довели свою ефективність, оскільки вони створюють зручний інтерфейс для

користувача та сприяють тому, що він неодноразово повертатиметься саме на цей сайт. Новим етапом у розвитку систем персоналізації є впровадження їх у навчальні системи.

Освіта посідає центральне місце в сучасному суспільстві у зв'язку з необхідністю навчання протягом усього життя. Освітні заклади та різні компанії сьогодні стикаються з проблемою зробити свої навчальні ресурси доступними в будь-який час. Електронні навчальні системи надають для цього чудову можливість. Системи електронного навчання за роки свого існування довели свою доцільність та широко застосовуються. Електронне навчання також успішно використовується як доповнення до традиційних методів навчання, а також для розгортання нових форм дистанційної освіти.

У багатьох навчальних закладах України зроблено певні кроки у розвитку та впровадженні дистанційних технологій у навчальний процес, тому постає питання розроблення інтерфейсу, який допоможе користувачам працювати ефективно із безліччю навчальних курсів, що і є завданням персоналізації.

Мета пропонованого дослідження – розгляд систем персоналізації, які можна реалізувати в електронному навчанні. Зроблено огляд систем надання рекомендацій (перший розділ), розглянуто підходи до формування профілю користувача (другий розділ), гібридні системи надання рекомендацій (третій розділ), системи надання рекомендацій, що базуються на характерних рисах елементів (четвертий розділ), та запропоновано алгоритм формування профілю за характерними рисами елементів (п'ятий розділ).

1. Системи надання рекомендацій для електронного навчання

Сучасні літературні джерела виокремлюють три основні типи систем надання рекомендацій –

системи, засновані на сукупності правил; системи, що базуються на фільтруванні змісту, та системи сумісного фільтрування. Детальний опис кожного типу систем наведено в [2].

Системи, засновані на сукупності правил, містять запрограмовані логічні правила, що пов'язують атрибути користувача (наприклад, мова, факультет, курс, сфера інтересів) з атрибутами системи (наприклад, мова інтерфейсу) чи елементами системи (навчальними курсами). Примітивні правила персоналізації (наприклад, автоматичний вибір мови інтерфейсу залежно від мови користувача) сьогодні є невід'ємною частиною інформаційних систем. Проте стандартний підхід до таких систем персоналізації – це реалізація правил на етапі реалізації самої інформаційної системи без подальшої їх зміни, що в разі систем електронного навчання може давати не-

якісну рекомендацію, оскільки кількість курсів та категорій курсів зазвичай постійно збільшується. Через зазначені обмеження в цій статті розглянуто системи, що базуються на фільтруванні змісту, та системи сумісного фільтрування.

Системи, що базуються на фільтруванні змісту, рекомендують користувачеві курси, схожі на ті, якими він уже зацікавився.

Системи сумісного фільтрування надають рекомендації, базуючись на порівнянні елементів профілю цільового користувача (наприклад, курсів) з елементами профілю інших користувачів (наприклад, курсами).

Для кращого розуміння реалізації систем надання рекомендацій на рис. 1 наведено модель – архітектуру інформаційної системи на рівні бізнес-логіки [3].

На цій схемі рівень 1 – рівень аналізу пове-

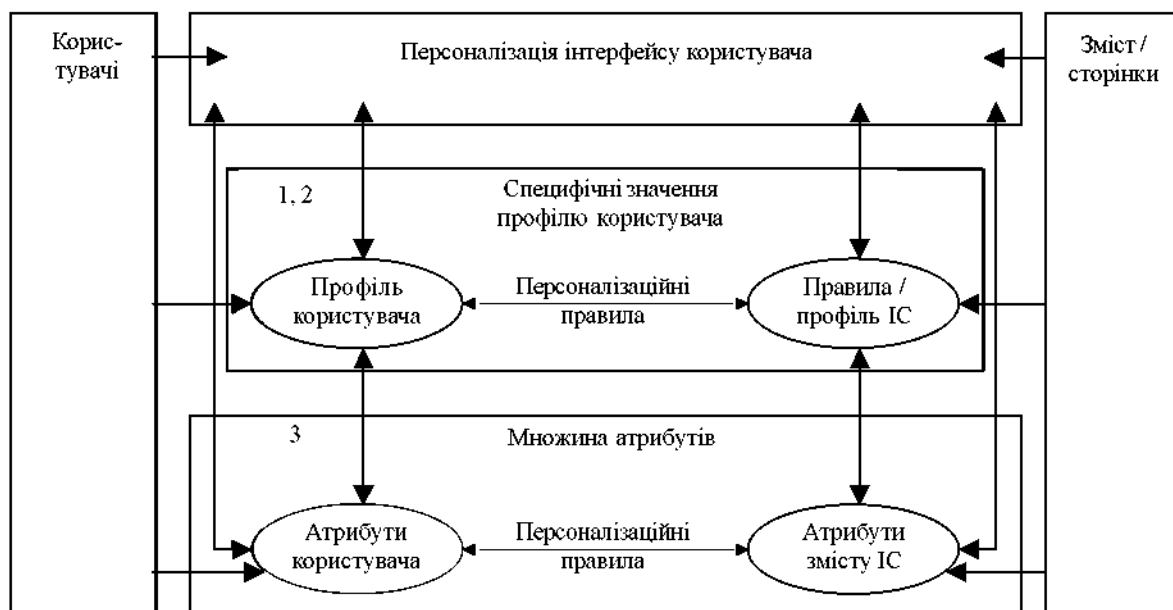


Рис. 1. Модель – архітектура інформаційної системи на рівні бізнес-логіки

дінки користувачів, 2 – рівень фільтрування змісту та 3 – рівень застосування сукупності правил.

Початковим кроком будь-якої системи надання рекомендацій є формування профілю користувача, яке розглянуто у наступних розділах дослідження. Огляд методів та алгоритмів персоналізації для кожної з систем залишимо поза межами цієї статті.

2. Формування профілю користувача

Поняття профілю користувача є центральним для систем надання рекомендацій, тому що рекомендації підраховуються, базуючись на елементах із профілю користувача.

Введемо такі позначення: U – це множина користувачів, $U = \{u_1, u_2, \dots, u_m\}$, $|U| = m$ та I – множина всіх можливих елементів оцінювання, $I = \{i_1, i_2, \dots, i_n\}$, $|I| = n$.

Профілем користувача $u \in U$ вважаємо вектор довжини n , що складається з упорядкованих пар: $u^{(n)} = \langle (i_1, s_u(i_1)), (i_2, s_u(i_2)), \dots, (i_n, s_u(i_n)) \rangle$, де $i_j \in I$, а $s_u(i_j)$ – функція, яка приписує елементу оцінювання i_j його вагу (чи рейтинг), що відповідає ступеню інтересу користувача u в елементі i_j .

Профілі всіх користувачів, що взаємодіють із системою, мають зберігатися протягом усього часу роботи системи. База даних всіх профілів може бути представлена як матриця розмірністю $m \times n$: $\Lambda = [s_{uk}(i_j)]_{m \times n}$, де $s_{uk}(i_j)$ – ступінь інтересу

користувача u_k до елемента i_j . Тому задача побудови профілю користувача зводиться до задачі визначення ваги (рейтингів) елементів.

До ваг може застосовуватись алгоритм нормалізації (у рамках одного профілю нові ваги отримуємо за формулою:

$$s'_u(i_j) = \frac{s_u(i_j) - m_u}{\max_{1 \leq i \leq n} \{s_u(i_i)\} - m_u}, \quad j = \overline{1, n},$$

де $m_u = \min_{1 \leq i \leq n} \{s_u(i_i)\}$, чи $s'_u(i_j) = \frac{s_u(i_j) - m(i_j)}{\max_{v \in U} \{s_v(i_j)\} - m(i_j)}$, $j = \overline{1, n}$ для усіх профілів у системі, де $m(i_j) = \min_{v \in U} \{s_v(i_j)\}$.

Розглянемо два підходи отримання інформації про рейтинги (ваги) елементів. Перший підхід базується на рейтингах, які користувачі призначають елементам у відповідь на явний запит системи, другий – на записях у системі про поведінку користувача. Приклади першого підходу – це відзначення рейтингів для фільмів, музики чи товарів за певною шкалою (наприклад, від 0 до 5). У рамках другого підходу [4] запропоновано використовувати функції переваг із бінарними значеннями, одиниця в яких означає, що файл було завантажено чи сторінка переглянута; запропоновано [2] визначати рейтинги залежно від часу, який користувач витратив на перегляд елемента, якщо така інформація є доступною в системі.

Більшість літературних джерел зазначають саме перший підхід, який уже досконало вивчено [4] і для якого запропоновано та реалізовано багато різних методів та підходів. Дослідження [4] свідчить, що зазвичай користувач не дуже схильний явно показувати свої переваги. Прочання вказати рейтинги елементів часто докучливе і його, за можливості, потрібно уникати. Проте також очевидно, що точна модель переваг та поведінки користувача не може бути побудована на підставі обмеженого обсягу інформації, отриманого внаслідок взаємодії користувача та інформаційної системи. Доведено [5], що користувачі бажають більшої взаємодії та наводять більше інформації про свої уподобання під час роботи саме із системою, яка надає їм більше переваг (у сенсі правильних рекомендацій).

Залежно від реалізації систем електронного навчання, вони можуть містити можливість оцінювати користувачем курси, на які він підписався. Якщо така можливість недоступна, то у формуванні профілю користувача рекомендуємо використовувати функцію переваг із бінарними значеннями, оскільки для предметної сфери навчання час, проведений користувачем на сторінці, не є визначальним, оскільки він може залежати від індивідуальних особливостей студентів.

3. Гібридні системи надання рекомендацій

Розглянемо системи, що базуються на фільтруванні змісту, та системи сумісного фільтрування. Кожна з цих систем має свої завдання та принципи роботи, тому для реалізації в системах електронного навчання потрібно обрати одну з них. Для систем електронного навчання реалізація системи, що базується на фільтруванні змісту, означає, що користувачеві завжди рекомендуватимуть курси з тієї самої предметної сфери, що й курси, на які він уже підписався. Це є доцільним, коли користувач має на меті досконало вивчити конкретні дисципліни, проте рекомендація не принесе користі, якщо користувач бажає вивчити кілька різних дисциплін (наприклад, у рамках навчального курсу його університету). Реалізація ж систем сумісного фільтрування дає змогу рекомендувати на основі схожості профілів користувачів. У цьому разі рекомендація складатиметься з елементів, які цільовий користувач ще не відвідав, проте які наявні у профілях схожих користувачів. Такі системи також мають обмеження, оскільки курси, на які ще ніхто не підписався, не можуть бути рекомендовані.

Щоб обійти обмеження кожної з систем та збільшити точність прогнозів, розроблено гібридні методи надання рекомендацій. Існує лінійна комбінація методів фільтрування змісту та сумісного фільтрування [6]. Фільтрування змісту застосовується для заповнення нульових елементів матриці переваг користувачів [5], далі застосовуються методи, що базуються на аналізі поведінки користувачів (сумісного фільтрування). Найпоширеніший підхід – це одночасне паралельне виконання обох методів із подальшим комбінуванням результатів [7]. Також поширеними є підходи, коли профіль користувача складається не з рейтингів елементів, а з елементів змісту оцінених сторінок (чи ключових слів, що повністю характеризують відвідані елементи). Такі підходи різняться методами визначення сусідів активного користувача [8, 9]. В інших прикладах гібридних систем враховують зміст елементів разом із близькістю користувачів [10] і використовують імовірнісні методи визначення наступного елемента, до якого звернеться користувач [4].

Для гібридних систем формування профілю користувача відбувається аналогічно формуванню профілю користувача для систем сумісного фільтрування та фільтрування змісту. Окремим випадком гібридних систем, як зазначено вище, є системи, профіль користувача в яких

складається з елементів змісту оцінених сторінок (чи ключових слів, що повністю характеризують відвідані елементи).

4. Надання рекомендацій, що базується на характерних рисах елементів

Розглянемо методи фільтрування, що базуються на семантичних рисах елементів та їх оцінюванні.

Семантичні, чи характерні риси елементів – це виділені розробником чи користувачем якості, що характеризують елемент. На відміну від ключових слів, які фактично пояснюють зміст елемента в найбільш можливій повній формі, семантичні риси елементів виражають певні якості елементів, що можуть прямо чи опосередковано стосуватися змісту елементів. Часто семантичні риси добирають таким чином, щоб вони розкривали основні фактори, що впливають на думку користувача під час вибору елемента. Множина ключових слів може бути підмножиною множини семантичних рис. Ці множини можуть перетинатися або не перетинатися.

Розглянемо приклади систем, що базуються на семантичних рисах елементів. Розроблено систему рекомендації ресторанів, у якій пропонується виділення специфічних рис елементів (наприклад, менш дорогий, більш тихий), а не оцінювання самих елементів [11]. У системі рекомендації музики [12] риси, що можуть бути оцінені користувачем, зафіксовано розробниками системи й запропоновано користувачеві для оцінювання. Інші системи дають користувачеві більше свободи – наприклад, користувачі під час оцінювання елементів можуть не лише давати оцінку, а й зазначити риси, що впливають на таку оцінку [13]. Також користувач може здійснювати пошук елементів за наявними в нього рисами. Користувачі можуть присвоювати елементам ярлики, що описують їхні якісні характеристики [14]. Кожен користувач може додавати семантичний рейтинг елементу, а для користування рекомендуються елементи з найбільшими рейтингами. Основна мета такої системи – зібрання ресурсів, що допомагають навчанню. Інші приклади подібних систем – Technorati.com (пошукова система, що явно використовує семантичні переваги користувачів) та Flickr.com (система розміщення фотографій, де користувачі можуть присвоювати фотографіям ярлики та виконувати пошук за ними).

Семантичні риси елементів можуть бути представлені як слова, що частіше повторюються в елементі. Також пропонується підраховувати ступінь суміжності між профілем користувача та

ключовими словами елементів і сукупностей елементів, а користувач сам може формувати свій профіль, задаючи ключові слова [6].

Вищенаведені підходи використовують інформацію не лише про переваги користувачів, а й про риси, належні елементам. Таку інформацію про риси елементів можна отримати автоматично з рейтингів елементів, використовуючи такі методи, як аналіз принципів компонентів [15] чи факторний аналіз [16].

5. Формування профілю псевдо-активного користувача, що базується на характерних рисах

Для збільшення точності рекомендацій та врахування всіх елементів системи в разі обчислення рекомендації, запропоновано ввести псевдоактивного користувача, профіль якого містить оцінки (ваги) для всіх елементів системи [5]. Вага елемента може визначатися як середнє арифметичне всіх оцінок даного елемента або виходячи з оцінок схожих елементів системи (тобто методом фільтрування змісту). Проте, користуючись таким підходом, неможливо визначити ваги для елементів, які ще не були оцінені ніким та для яких неможливо визначити ваги методом фільтрування змісту. Позначимо множину таких елементів V . Пропонується метод визначення ваги елементів із множини V , що базується на характерних рисах.

Розглянемо систему, для якої кожному елементу відповідає сукупність характерних рис, а множина ключових слів є підмножиною множини характерних рис, тобто кожен елемент визначається своїми характерними рисами. Наприклад, для системи електронного навчання така множина характерних рис може складатися з ключових слів, що характеризують курс, та семантичних рис, які описують галузь науки, до якої належить курс, та інші галузі, тематично близькі до курсу.

Позначимо F сукупність усіх характерних рис системи, $F = \{f_1, f_2, \dots, f_l\}$, де l – загальна кількість рис. F^j – риси, приписані елементу i_j . Кожна риса $f_k \in F$ може набувати різні значення. Позначимо f_k^j значення риси f_k для елемента i_j . Кожен елемент може бути представлено як вектор, що складається з характерних рис.

Сукупність значень рис для всіх елементів системи може бути представлена як матриця $\Phi = [f_k^j]_{n \times l}$ ($k = \overline{1, l}, j = \overline{1, n}$).

Розглянемо перетворення, яке дає змогу суттєво спростити підрахунок рейтингів за характерними рисами. Побудуємо граф, вершинами якого будуть елементи системи. Ребра будуємо за таким

правилом: дві вершини i_j та i_k з'єднані ребрами c , $c \in N \Leftrightarrow$ коли $\exists k_1, \dots, k_c: f_{k_1}^h = f_{k_1}^{j_2}, \dots, f_{k_c}^h = f_{k_c}^{j_2}$ і для $\forall k_j \in F \setminus \{f_{h_1}, \dots, f_{h_c}\} f_{k_j}^h \neq f_{k_j}^{j_2}$. Кількість ребер між двома елементами дорівнює кількості однакових значень відповідних рис. Далі для графа будується матриця інцидентності $H = [h_{j,k}]_{n \times n}$, у якій $i_{j,k}$ відповідає кількості ребер між вершинами i_j та i_k . Вага $s_u(i_j)$ цільового елемента i_j псевдоактивного користувача u обчислюється як сума добутків ваг інших елементів та відповідних значень із матриці інцидентності (що відповідають кількості ребер між цільовим елементом та елементом, який визначає доданок), розділена на суму значень із матриці інцидентності, що відповідає цільовому елементу, при цьому в обчисленні враховуються тільки ті елементи i_j , для яких

$$\text{ваги } s_u(i_j) \text{ визначені: } s_u(i_j) = \frac{\sum_{\substack{j=1 \\ j \neq i}}^n h'_{i,j} * s_u(i_j)}{\sum_{\substack{j=1 \\ j \neq i}}^n h'_{i,j}}, \quad \text{де}$$

$$h'_{i,j} = \begin{cases} h_{i,j}, & \text{якщо } s_u(i_j) \text{ визначене} \\ 0, & \text{інакше} \end{cases} \quad \text{і відповідно}$$

$h'_{i,j} * s_u(i_j) = 0$, якщо вага $s_u(i_j)$ невизначена.

На першому кроці алгоритму формуємо множину вершин V , для яких потрібно визначити ваги за характерними рисами (ваги всіх інших елементів оцінювання псевдоактивного користувача u , що не входять у множину V , вважаємо визначеними). На кожному наступному кроці алгоритму обчислюємо ваги для елементів із множини V та вилучаємо із множини V ті вер-

шини, для яких обчислення було успішним (що еквівалентно

Робота алгоритму зупиняється, якщо множина V порожня, або якщо вона не змінюється протягом двох послідовних кроків роботи алгоритму (тобто для деяких вершин підрахунок рейтингів неуспішний). Підрахунок рейтингу може бути неуспішним (вага залишається невизначеною) для певної вершини в разі, якщо в матриці інцидентності їй відповідають нулі, або якщо ця вершина з'єднана ребрами лише з вершинами, для яких також неможливий підрахунок ваг. У цьому разі доцільно призначити таким вершинам рейтинг за замовчуванням, який визначається, наприклад, як середнє арифметичне всіх рейтингів системи, щоб не виключати ці елементи в підрахунок рекомендації.

Висновки

Системи надання рекомендацій розвивалися протягом останнього десятиліття: було запропоновано різного типу алгоритми, огляд яких зроблено в цій статті. Існують практичні розробки та застосування таких систем у комерційній сфері; електронна освіта є новим напрямом їх застосування. Наведено опис методів, які можуть бути застосовані для систем електронного навчання. Детально розглянуто перший крок алгоритмів надання рекомендацій - побудову профілю користувача для різних систем надання рекомендацій. Запропоновано алгоритм побудови профілю псевдоактивного користувача, що базується на характерних рисах.

1. Anand S. S., Mobasher B. Intelligent techniques for web personalization // LNAI 3169.- Springer-Verlag, Berlin-Heidelberg; New York, 2005.- P. 1-36.
2. Mobasher B., Dai H., Luo T., Nakagawa M. Improving the effectiveness of collaborative filtering on anonymous web usage data // IJCAI 2001, workshop on intelligent techniques for web personalization (ITWP).- Seattle.- August 2001.- P. 53-60.
3. Instone K. An information architecture-based framework for personalization systems // Information architecture and personalization.- December 2000.- http://argus-acia.com/white_papers/personalization.html.
4. Wang J., Vries A., Reinders M. A user-item relevance model for log-based collaboration // 28th european conference on information retrieval research, ECIR 2006.- London, 2006.- P. 37-48.
5. Melville P., Mooney R., Nagarajan R. Content-booster collaborative filtering for improved recommendations // 18th national conference on artificial intelligence (AAAI).- Edmonton, 2002.- P. 187-192.
6. Claypool M., Gokhale A, Miranda T., Murnikov P, Netes D., Sartin M. Combining content-based and collaborative filters in an online newspaper // In proceedings of ACM SIGIR workshop on recommender systems - implementation and evaluation.- 1999.- August 19.
7. Cotter P., Smyth B. TV: Intelligent personalized TV guides // In twelfth conference on innovative applications of artificial intelligence.- Austin, 2000.- P. 957-964.
8. Pazzani M. J. A framework for collaborative, content-based and demographic filtering // Artificial intelligence review.- 1999.- Vol. 13, № 5-6.- P. 393-408.
9. Soboroff I., Nicholas C. Combining content and collaboration in text filtering // In proceedings of the IJCAI'99 workshop on machine learning in information filtering.- Stockholm, 1999.- P. 86-91.
10. Popescul A., Ungar L., Pennock D. M., Lawrence S. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments // In proceedings of the seventeenth conference on uncertainty in artificial intelligence.- August 2001.- P. 437-444.

11. *Burke R.* Hybrid recommender systems: Survey and experiments // *User modeling and user-adapted interaction.*— 2002.— Vol. 12, № 4.— P. 331–370.
12. *Anderson M., Ball M., Boley H., Greene S., Howse N., Lemire D., McGrath S.* RACOFI: A rule applying collaborative filtering system // *In proceedings of COLA'03.*— Halisak, 2003.
13. *Garden M., Dudek G.* Semantic feedback for hybrid recommendations in Recommendz // *In proceedings of the IEEE international conference on e-technology, e-commerce, and e-service (EEE05).*— Hongkong, 2005.— P. 754–759.
14. *Dron J.* CoFIND – an experiment in n-dimensional collaborative filtering // *Journal of network and computer applications.*— 2000.— Vol. 23, № 2.— P. 131–142.
15. *Goldberg K., Roeder T., Gupta D., Perkins C.* Eigentaste: A constant time collaborative filtering algorithm // *Information Retrieval.*— 2001.— Vol. 4, № 2.— P. 133–151.
16. *Canny J.* Collaborative filtering with privacy via factor analysis // *In proceedings of the 25th annual international ACM SIGIR conference on research and development in Information Retrieval.*— Tampere, 2002.— P. 238–245.

V. Boublik, M. Gornostay

DEVELOPING RECOMMENDER SYSTEMS FOR E-LEARNING

The article says about recommender systems which can be used in e-learning. Approaches for creating user profile are viewed for different e-learning systems and recommender systems, algorithm of pseudo-active user profile creation based on characteristic features is proposed.