

Є. В. Мелешко

Центральноукраїнський національний технічний університет, Кропивницький, Україна

МЕТОДИ ОЦІНКИ ЯКОСТІ РОБОТИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

Предметом вивчення у статті є процес оцінки якості роботи рекомендаційних систем. **Метою** є виявлення найважливіших показників якості роботи рекомендаційних систем та визначення методів їх оцінювання. **Завдання:** дослідити методи оцінки якості роботи рекомендаційних систем, дослідити показники якості роботи рекомендаційних систем. Отримані такі **результати:** Розглянуто основні та додаткові показники якості роботи рекомендаційних систем. Проведено дослідження їх важливості з погляду оцінки різних властивостей списку рекомендацій. Визначено напрямки подальших досліджень для розробки методів оцінки якості роботи рекомендаційних систем. **Висновки.** Основними показниками якості роботи рекомендаційних систем є точність прогнозування оцінок користувача та покриття простору об'єктів і покриття простору користувачів. Додатковими показниками якості роботи рекомендаційних систем є різноманітність, новизна, неочікуваність, робастність, приватність, ризик тощо. Показники різноманітність, новизна, неочікуваність дозволяють оцінити якість структури рекомендацій та ймовірність появи проблеми бульбашки фільтрів. Показники робастність, приватність, ризик дозволяють оцінити інформаційну безпеку рекомендаційної системи та її користувачів. Для кожного окремого веб-ресурсу чи додатку можна скласти свій набір додаткових показників якості роботи рекомендаційної системи, але розглянуті додаткові показники будуть актуальними практично для будь-якого випадку. Також важливими можуть бути наступні показники: впевненість, довіра, адаптивність, масштабованість, пропускна здатність, корисність, тощо. Оскільки не існує загальноприйнятих способів вимірювання та методів оцінки таких показників якості роботи рекомендаційних систем, як різноманітність, новизна, неочікуваність, приватність, ризик, робастність, – науково-практична задача розробки таких методів є актуальною.

Ключові слова: рекомендаційні системи, показники якості, точність роботи, якість роботи, покриття, бульбашка фільтрів, інформаційна безпека, аналіз даних, тестування програмного забезпечення.

Вступ

Важливою задачею при побудові рекомендаційних систем (РС) є оцінка якості їх роботи. В найпростішому випадку для дослідження якості роботи рекомендаційної системи застосовують такі показники як точність прогнозування вподобань користувача та покриття каталогу товарів, в той же час досить важливими з погляду якості роботи системи можуть бути й інші показники, наприклад, приватність, робастність, різноманітність, новизна, тощо.

Методів побудови рекомендаційних систем на сьогоднішній день досить багато [1-3]. Для різних типів вхідних даних та різних видів веб-ресурсів чи додатків слід підбирати й різні методи створення рекомендацій. Але необхідно мати чіткі критерії, за якими здійснювати такий вибір. Тому науково-практична задача визначення якості роботи рекомендаційних систем є актуальною.

Метою даної роботи є виявлення найважливіших показників якості роботи рекомендаційних систем та дослідження методів їх оцінювання.

Основна частина

Перевірку роботи рекомендаційної системи можна здійснити, використовуючи наступні дані:

1. За допомогою тестової вибірки, підготовленої заздалегідь.

2. За допомогою моніторингу роботи РС та збору даних у реальному часі.

У першому варіанті перевірки усі наявні дані слід розділити на навчачою та тестову вибірку. Тестову вибірку не можна використовувати при навчанні рекомендаційної системи та при генерації списку рекомендацій. Тестова вибірка буде використана для порівняння даних з неї з даними зі списку

рекомендацій. Для перевірки точності прогнозування вподобань при знаходженні однакових об'єктів у тестовій вибірці та у списку рекомендацій, їх прогнозовані та реально поставлені користувачем оцінки порівнюються. Таку перевірку легко організувати, однак її результати не досить інформативні, оскільки збігів у цих двох наборах даних буде досить мало.

Більш інформативним буде другий варіант, оскільки при моніторингу у реальному часі одразу буде видно як користувач реагує на кожну рекомендацію. Можна буде перевірити кожен з прогнозів в тій чи іншій формі – виявити факт обрання/необрання об'єктів зі списку рекомендацій користувачем, а при виставленні оцінки, порівняти оцінку з прогнозованою. Але даний спосіб більш складно реалізувати, необхідно створити систему моніторингу веб-ресурсу, бажано мати доступ до нього на рівні адміністратора.

Розглянемо показники якості роботи рекомендаційної системи та способи їх оцінки.

Найголовнішим показником якості роботи рекомендаційної системи є *точність прогнозування вподобань (Prediction Accuracy)*. Для перевірки точності прогнозування вподобань користувачів порівнюють два вектори [4, 5]:

1) вектор $\hat{R} = (\hat{r}_1, \hat{r}_2, \dots, \hat{r}_n)$, що містить список прогнозованих оцінок користувача, впорядкований по спаданню за величиною оцінок;

2) вектор $R = (r_1, r_2, \dots, r_n)$, що містить справжні оцінки користувача, невідомі системі на етапі формування списку рекомендацій.

Порівняння даних векторів можна здійснити різними способами.

Вимірювання точності прогнозу. Якщо користувачі виставляють оцінки об'єктам, точність про-

гнозування можна визначити за допомогою середньоквадратичної помилки (1) або середньої абсолютної помилки (2):

$$RMSE = \sqrt{\frac{1}{|\tau|} \sum_{(u,i) \in \tau} (\hat{r}_{ui} - r_{ui})^2}, \quad (1)$$

$$MAE = \frac{1}{|\tau|} \sum_{(u,i) \in \tau} |\hat{r}_{ui} - r_{ui}|, \quad (2)$$

Де \hat{r}_{ui} – прогнозовані рейтинги для тестового набору даних τ пар користувач-об’єкт (u, i) , r_{ui} – справжні рейтинги.

Система генерує прогнозовані рейтинги \hat{r}_{ui} для тестового набору даних τ пар користувач-об’єкт (u, i) , для яких відомі справжні рейтинги r_{ui} . Реальні та прогнозовані рейтинги порівнюються.

Прогнозування використання. Оцінювати точність роботи рекомендаційної системи можна завдяки реакції користувача на об’єкти наведені у списку рекомендацій. В такому разі інформація для оцінки роботи рекомендаційної системи з’являється в процесі її використання.

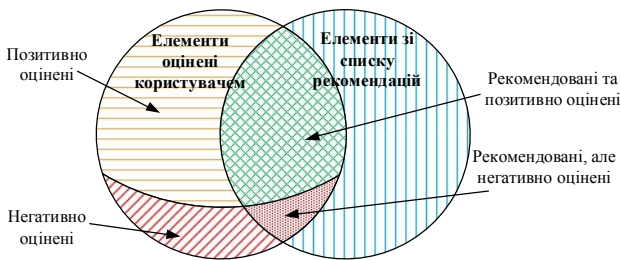


Рис. 1. Можливі варіанти реакції користувача на об’єкти у списку рекомендацій

Наведемо можливі результати прогнозу рекомендаційної системи (табл. 1).

Таблиця 1 – Класифікація можливих результатів рекомендацій

	Рекомендували	Не рекомендували
Позитивно оцінено	True-Positive (tp)	False-Negative (fn)
Негативно оцінено	False-Positive (fp)	True-Negative (tn)

Як видно з таблиці, можливі чотири результати роботи рекомендаційної системи:

- tp – результати, в яких позитивний прогноз виявився вірним;
- tn – результати, в яких негативний прогноз виявився вірним;
- fp – результати, в яких позитивний прогноз виявився помилковим;
- fn – результати, в яких негативний прогноз виявився помилковим.

Можна підрахувати кількість подій, що відповідають кожній з комірок, та на основі одержаних значень оцінити точність роботи рекомендаційної системи [5]:

$$Precision = \frac{tp}{tp + fp}, \quad (3)$$

$$Recall (True Positive Rate) = \frac{tp}{tp + fn}, \quad (4)$$

$$False Positive Rate = \frac{fp}{fp + tn}, \quad (5)$$

Міри ранжирування. Важливою задачею є ранжування об’єктів у списку рекомендацій. За допомогою даної міри можна визначити наскільки вірно рекомендаційна система впорядкувала елементи у списку рекомендацій.

Міри ранжирування – використання заздалегідь відомих рейтингів. Як міру ранжування можна використати Normalized Distance-based Performance Measure (NDPM) [6]. Якщо у нас є справжні рейтинги об’єктів r_{ui} , одержані в результаті дій користувача, та рейтинги, згенеровані рекомендаційною системою \hat{r}_{ui} для n_u об’єктів i користувачу u , то можна одержати:

$$C^+ = \sum_{ij} \text{sgn}(r_{ui} - r_{uj}) \text{sgn}(\hat{r}_{ui} - \hat{r}_{uj}), \quad (6)$$

$$C^- = \sum_{ij} \text{sgn}(r_{ui} - r_{uj}) \text{sgn}(\hat{r}_{uij} - \hat{r}_{uij}), \quad (7)$$

$$C^u = \sum_{ij} \text{sgn}^2(r_{ui} - r_{uj}), \quad (8)$$

$$C^s = \sum_{ij} \text{sgn}^2(\hat{r}_{uij} - \hat{r}_{uij}), \quad (9)$$

$$C^{u0} = C^u - (C^+ + C^-), \quad (10)$$

де суми перевищують $\frac{1}{2}n_u(n_u - 1)$ пар об’єктів. Таким чином, C^u – це кількість пар об’єктів, для яких справжні рейтинги встановлюють впорядкування один відносно одного, тоді як C^+ та C^- – це кількість пар об’єктів, для яких рекомендаційна система встановила правильний порядок і неправильний порядок відповідно. C^{u0} – це кількість пар, для яких немає впорядкування в справжньому рейтингу, але рекомендаційна система встановлює для них деяке впорядкування. NDPM одержується наступним чином:

$$NDPM = \frac{C^- + 0.5C^{u0}}{C^u}, \quad (11)$$

Таким чином, міра NDPM дає найкращу оцінку 0 – для систем, які правильно передбачають кожне впорядкування пар об’єктів. А найгірша оцінка 1 – призначена системам, що суперечать усім вірним впорядкуванням пар об’єктів.

Міри ранжирування – на основі корисності. Популярною альтернативою попереднього методу є припущення, що корисність переліку рекомендацій є сумарною, що визначається сумою корисності окремих рекомендацій. Корисність кожної рекомендації – це корисність рекомендованого об’єкту помножена на коефіцієнт зменшення, який залежить від позиції об’єкту в переліку рекомендацій. Одним

з прикладів такої корисності є ймовірність того, що користувач буде дотримуватися рекомендації в даній позиції у списку. Зазвичай передбачається, що користувачі проглядають списки рекомендацій від початку до кінця, з урахуванням того, що переваги рекомендацій значно знижуються до кінця списку.

В багатьох додатках список рекомендацій далеко не самий основний спосіб пошуку об'єктів, він містить невелику кількість елементів, а в крайніх випадках може містити тільки один елемент. В таких системах користувач зазвичай переглядає тільки невелику частину списку рекомендацій – тільки декілька пунктів на початку. В таких випадках цінність рекомендацій знижується дуже швидко відносно позицій елементів у списку. Для таких додатків можна використати метрику R-Score [5, 7].

Метрика R-Score передбачає, що корисність рекомендацій знижується експоненціально вниз по впорядкованому списку рекомендацій, щоб отримати наступний бал для кожного користувача u :

$$R_u = \sum_u \sum_j \frac{\max(r_{uj} - d, 0)}{(j-1)/(2\alpha-1)}, \quad (12)$$

де i_j – це об'єкт в j -й позиції, r_{ui} – рейтинг користувача i , d – випадки неважливого голосування, а α – період напіврозпаду, який контролює експоненціальне зниження значення позицій у ранжовому списку. У випадку задач прогнозування рейтингів r_{ui} – це рейтинг, наданий користувачем для кожного елемента (наприклад, 4 зірки), а d – це неважливе голосування (наприклад, 3 зірки), і алгоритм отримує кредити лише за позиції рейтингу вище неважливого, вищого за d (наприклад, 4 або 5 зірок). У задачі передбачення використання r_{ui} звичайно дорівнює 1, якщо \mathbb{I} вибирає i , та 0 в іншому випадку, якщо $d = 0$.

Отримані для кожного користувача результати агрегуються за допомогою:

$$R = 100 \frac{\sum_u R_u}{\sum_u R_u^*}, \quad (13)$$

де R_u^* – оцінка найкращого рейтингу для користувача u .

До сучасних рекомендаційних систем висувається значно більше вимог, ніж просто прогнозування оцінок користувача. Крім точності формування та ранжування списку рекомендацій до рекомендаційної системи може висуватися багато інших вимог, наприклад: покриття, різноманітність, новизна, приватність, робастність до атак, адаптивність, масштабованість, пропускна здатність, тощо. Покращення даних показників рекомендаційної системи може знизити її точність прогнозування оцінок, але підвищити загальну якість роботи. Розглянемо основні з них.

Важливим показником якості роботи РС є покриття (Coverage), існує декілька його видів:

– Покриття каталогу (покриття простору об'єктів) – може визначатися як відсоток усіх еле-

ментів, які можуть бути рекомендовані, даний показник дозволяє виявити об'єкти, які нікому не рекомендуються.

– Покриття простору користувачів – може характеризуватися часткою користувачів або взаємодій користувачів, для яких система може рекомендувати об'єкти (у багатьох системах рекомендації можуть не надаватися для користувачів, про яких зібрано мало даних, через низьку впевненість у точності прогнозів), якщо рекомендації слід надавати всім користувачам у системі, то необхідно йти на компроміс між покриттям та точністю.

– Різноманітність збуту – міра неоднорідності вибору різних об'єктів користувачами зі списку рекомендацій, для її визначення можна використовувати різні індекси, зокрема, коефіцієнт Джині (14) або ентропію Шеннона (15):

$$G = \frac{1}{n-1} \sum_{j=1}^n (2j-n-1)p(i_j), \quad (14)$$

де i_1, \dots, i_n – це список об'єктів, впорядкованих за збільшенням частоти їх вибору користувачем або частоти їх появи у списку рекомендацій $p(i)$. Індекс $G = 0$, коли всі елементи вибираються однаково часто, а $G = 1$ – коли завжди вибирається один елемент.

$$H = -\sum_{i=1}^n p(i) \log p(i), \quad (15)$$

де ентропія дорівнює 0, коли завжди обирається або рекомендується один об'єкт, та $\log n$, коли n об'єктів вибирається чи рекомендується однаково часто.

Точність прогнозування вподобань і покриття є настільки важливими показниками якості роботи рекомендаційних систем, що їх слід віднести до основних.

Розглянемо й інші показники, за допомогою яких можна оцінити якість роботи рекомендаційної системи за різними критеріями.

Усі сучасні рекомендаційні системи схильні до проблеми бульбашки фільтрів, що виникає, коли алгоритм формування списку рекомендацій підбирає інформацію, яку користувач хотів би бачити, і, в результаті, користувачі відділяються від інформації, яка їх не цікавить або їм не подобається, фактично ізолюючи їх у власних «бульбашках». Для вирішення даної проблеми до рекомендаційних систем висуваються наступні вимоги – список рекомендацій повинен володіти наступними властивостями:

– *Різноманітність рекомендацій (Diversity)*. Це міра схожості елементів списку. Елементи у рекомендаційній системі не повинні бути майже однаковими, вони повинні містити різнотипні об'єкти (наприклад, фільми різних жанрів, а не тільки одного жанру, чи однієї трилогії). Для визначення схожості елементів можна застосовувати різні коефіцієнти подоби (коефіцієнт кореляції Пірсона, косинусну міру, евклідову відстань, відстать Хеммінга тощо), за допомогою яких попарно порівнювати елементи списку, після визначення рівня схожості між окремими елементами можна буде оцінити різноманітність списку рекомендацій в цілому.

Новизна рекомендацій (Novelty). Нові об'єкти у системі можуть ще не мати оцінок і не бути популярними, але вони можуть бути цікавими користувачам через свою новизну. В той же час нові об'єкти необхідно комусь рекомендувати, щоб вони не залишилися без уваги. новизна для користувача, поширеність продукту, частка його рейтингів. Якщо користувачу рекомендувати лише популярні об'єкти, скоріше за все він їх і так знає та обере без рекомендаційної системи, такі рекомендації не будуть містити для нього нової інформації. Це характеристика елемента у списку рекомендацій протилежна його популярності, і в найпростішому випадку може визначатися за формулою [8]:

$$\text{novelty}(i) = -\log_2 p(i), \quad (16)$$

де $p(i)$ – ймовірність того, що i потрапить у список рекомендацій (буде обрано).

Неочікуваність рекомендацій (Serendipity). Неочікуваність представляє собою деякий сюрприз у списку рекомендацій, несхожість на історію дій користувача. Не існує консенсусу у визначенні неочікуваності, однак більшість авторів вказує, що елемент, який має властивість неочікуваності, повинен бути важливим, новим та непрогнозованим для користувача [9]. Важливість для користувача виражається в його реакції на даний елемент після рекомендації, новизна виражається в тому, наскільки користувач знайомий з даним елементом. Елемент може бути незнайомим для користувача, якщо:

- 1) користувач ніколи не чув про даний елемент,
- 2) користувач чув про даний елемент, але ніколи не використовував,
- 3) користувач використовував даний елемент, але забув про це.

Елемент може бути непрогнозованим для користувача, якщо:

- 1) користувач не очікує, що цей елемент буде для нього актуальним,
- 2) користувач не очікує, що цей елемент буде рекомендований йому,
- 3) користувач не знайшов би цього елемента самостійно,
- 4) цей елемент значно відрізняється від елементів, які як правило, обирає користувач,
- 5) користувач не очікує даного елемента у списку рекомендацій, оскільки він переглядав інші види елементів.

Збільшення різноманітності, неочікуваності та новизни рекомендацій може знизити точність прогнозування та точність ранжування, в той же час може підвищитися покриття каталогу об'єктів, різноманітність збуту та частково вирішитися проблема бульбашки фільтрів.

На даний час не існує загальноприйнятих мір та методів оцінки різноманітності, новизни та неочікуваності списків рекомендацій.

Серед показників якості роботи рекомендаційних систем, що є важливими з погляду інформаційної безпеки можна виділити наступні:

1. *Приватність користувача (Privacy).* РС збирають велику кількість даних про користувачів, значну

частину яких користувачі охоче надають самі в обмін на корисні рекомендації.

Однак для більшості користувачів, важливо щоб їхні вподобання залишалися приватними, тобто, жодна третя сторона не могла використовувати РС, щоб дізнатися інформацію про них або їх вподобання. Дана загроза цілком реальна. Як один з прикладів можна навести скандальні ситуації з рекомендаціями друзів у Facebook, які виникали через експерименти з використанням даних геолокації, подібні рекомендації частково порушували приватні дані людей та давали інформацію третім особам про їх переміщення. Для забезпечення приватності користувача РС можна застосовувати наступні методи:

– Інформування користувачів про те, яку інформацію про них збирає РС, гнучкі налаштування параметрів конфіденційності.

– Анонімізація – інформація про користувача може частково видалятися або піддаватися обфускації (маскуванню) користувачем або власником РС.

– Рандомізація – дані користувача (напр., виставлені об'єктам оцінки) можуть бути частково зашумлені випадковими значеннями. Необхідний рівень шуму залежить від того, як часто дані будуть використовуватися, і передбачає балансування між точністю прогнозування та конфіденційністю користувача.

– Шифрування даних користувача, що зберігаються в базі даних РС.

2. *Ризик для користувача (Risk).* В деяких випадках рекомендації можуть бути пов'язані з ризиком. Напр., якщо об'єктами в РС є акції, кредити, депозити, ліки, медичні послуги, політичні акції тощо. В таких випадках може бути необхідним врахування не тільки вподобань користувача при формуванні рекомендацій, а й інших факторів, врахування яких здатне мінімізувати ризик для користувача, що буде переглядати та обирати рекомендації.

3. *Робастність системи до атак (Robustness).* Здатність системи надавати адекватні рекомендації при появі некоректної інформації. Некоректна інформація може виникати при атаках на РС з метою збільшення рейтингу на певні об'єкти, напр., при створенні великої кількості фейкових аккаунтів та ботів, які виставляють високі оцінки певному об'єкту чи об'єктам. Побудова робастних РС базується на двох принципах:

- 1) виявлення спаму серед дій користувачів;
- 2) невраховування при побудові списку рекомендацій інформації користувачів, що поширюють спам.

На сьогоднішній день існують методи, які дозволяють виявити атаку на РС, використовуючи той факт, що розподіл коефіцієнтів подоби користувачів змінюється при наявності спам-користувачів в РС. Так як при атаці створюють не один фейковий профіль, а декілька схожих, такі спам-користувачі будуть мати незвично високу схожість, у порівнянні зі звичайними користувачами.

Однак, надійні методи виявлення атак на РС та протидії ним все ще залишаються активною областю досліджень.

Також, в залежності потреб певного веб-ресурсу чи додатку, можна виділити й інші показники якості роботи рекомендаційних систем, наприклад:

– *Впевненість (Confidence)*. Рекомендаційна система може додавати до своїх рекомендацій процент впевненості у них. Так наприклад, система може вказати для першої рекомендації прогнозовану оцінку користувача 5 балів з впевненістю на 95%, а для другої рекомендації – 5 балів з впевненістю на 89%. Користувач може враховувати даний параметр та обирати в першу чергу об'єкти з більшим значенням впевненості системи.

– *Довіра (Trust)*. Дана властивість характеризує наскільки користувач може довіряти даній рекомендаційній системі. Якщо рекомендаційна система запропонувала користувачу лише невідомі йому об'єкти, то користувач може засумніватися у правильності роботи системи, а якщо серед рекомендацій є певна кількість об'єктів, про які він знає і які йому подобаються, то рівень довіри до системи буде вищим. Тобто, деяка кількість об'єктів, про які відомо користувачу, може збільшити довіру до системи. Ще одним варіантом збільшення довіри до системи є пояснення до рекомендації. Наприклад, коли елемент зі списку рекомендацій подається у вигляді «Якщо Вам подобається X, то спробуйте Y». Також довіра користувача зростає, якщо рекомендації були йому корисні та відповідали його вподобанням, та знижується в протилежному випадку.

– *Адаптивність (Adaptivity)*. Реальна система рекомендацій може працювати у ситуації, коли колекції об'єктів, і/або інтереси користувачів швидко змінюються. Прикладом таких систем можуть бути рекомендаційні системи новинних сайтів, коли об'єкти (новини) цікаві тільки на проміжку певного часу. В таких системах з одного боку цікавість до певного об'єкту існує тільки певний час, з іншого боку деякі старі об'єкти можуть знову ставати дуже цікавими, якщо нові об'єкти у системі поновлюють інтерес до старих. В таких системах слід вибирати швидкі алгоритми, навіть якщо доводиться в деякій мірі жертвувати точністю. Інший тип адаптивності, це те, з якою швидкістю система адаптується до вподобань користувача або до змін в його профілі. Якщо дії користувача на сайті занадто повільно змінюють рекомендації, він може відмовитися оцінювати об'єкти, не отримавши зворотного зв'язку. Адаптивність алгоритмів можна оцінювати шляхом визначення різниці між списками рекомендацій до та після додавання нової інформації з профіля користувача, наприклад, за допомогою міри ентропії Шеннона.

– *Масштабованість (Scalability)*. Зі збільшенням об'ємів даних багато алгоритмів працюють значно повільніше або вимагають додаткових ресурсів,

таких як обчислювальні потужності або пам'ять. Тому важливо враховувати просторову та часову складність алгоритмів. В реальних системах може виникнути необхідність згодитися на меншу точність рекомендацій для одержання більш масштабованої системи.

– *Пропускна здатність (Throughput)*. Це кількість рекомендацій, які система може надати в одиницю часу.

– *Корисність (Utility)*. Значення, що характеризує, яку вигоду від рекомендації отримує система та/або користувач.

При побудові рекомендаційної системи для конкретного веб-сайту чи додатку досить логічним кроком буде визначення списку показників якості, яким повинна задовольняти розроблювана РС, та вибір/розробка алгоритмів і методів її побудови на основі визначення у процесі їх тестування, наскільки вони задовольняють висунутим критеріям.

Висновки

У статті розглянуто основні показники якості роботи рекомендаційних систем та досліджено методи їх оцінки. Найголовнішими показниками якості роботи РС є точність прогнозування оцінок користувача та покриття простору об'єктів і простору користувачів.

Серед додаткових показників якості слід обов'язково звернути увагу на такі, що дозволяють оцінити наявність/відсутність проблеми бульбашки фільтрів та проблем з інформаційною безпекою.

Важливими є такі показники якості роботи РС як різноманітність, новизна та неочікуваність, які дозволяють оцінити структуру списку рекомендацій та ймовірність виникнення проблеми бульбашки фільтрів.

Суттєвими з погляду інформаційної безпеки є такі показники якості роботи РС як приватність, ризик та робастність. Вони дозволяють оцінити інформаційну безпеку користувача та інформаційну безпеку системи.

Безперечно, для кожного окремого веб-ресурсу чи додатку можна скласти свій набір додаткових показників якості роботи рекомендаційної системи, але розглянуті вище показники будуть актуальними практично для будь-якого випадку.

Також важливими можуть бути наступні показники якості: впевненість, довіра, адаптивність, масштабованість, пропускна здатність, корисність, тощо.

Оскільки не існує загальноприйнятих способів вимірювання та методів оцінки таких показників якості роботи рекомендаційних систем, як різноманітність, новизна та неочікуваність, приватність, ризик, робастність, – задача розробки таких методів є актуальною.

СПИСОК ЛІТЕРАТУРИ

1. Jones M. Recommender systems, Part 1. Introduction to approaches and algorithms. Learn about the concepts that underlie web recommendation engines / M. Jones – 2013. [Електронний ресурс] – Режим доступу: https://www.ibm.com/developerworks/opensource/library/os-recommender1/index.html?s_tact=105agx99&s_cmp=cp

2. Мелешко Є.В. Дослідження методів побудови рекомендаційних систем в мережі Інтернет / Є.В. Мелешко, Г.С. Семенов, В.Д. Хох. // Збірник наукових праць "Системи управління, навігації та зв'язку". Випуск 1(47). – Полтава: ПНТУ ім. Ю. Кондратюка. – 2018. – С. 131–136.
3. Сегаран Т. Программируем коллективный разум. – Пер. с англ. – СПб: Символ-Плюс, 2013. – 368 с.
4. Амелькин С.А. Оценка эффективности рекомендательных систем // Труды 14-й Всерос.НК «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL-2012, Переславль-Залесский, 2012. – С. 288-291.
5. Recommender Systems Handbook / Editors Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor. – 1st edition. – New York, NY, USA: Springer-Verlag New York, Inc., 2010. – 842 с.
6. Yao Y.Y. Measuring retrieval effectiveness based on user preference of documents / Yao Y.Y. // Journal of the American Society for Information Science. – 1995. – №46. – С. 133–145.
7. Breese, S., Heckerman, D., Kadie, C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. / In Proc. of the 14th Conference on Uncertainty in Artificial Intelligence, volume 461. – San Francisco, CA, 1998. – pages 43–52.
8. Castells, P., Vargas, S., Wang, J. Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance / P. Castells, S. Vargas, J. Wang. – 2011. [Електронний ресурс] – Режим доступу:
9. <https://www.semanticscholar.org/paper/Novelty-and-Diversity-Metrics-for-Recommender-and-Castells-Vargas/4ec6bd672aaaa075b42a751099eb9317857e6e0c>
10. Kaminskis M., Bridge D. Measuring Surprise in Recommender Systems. [Електронний ресурс] – Режим доступу: <https://www.insight-centre.org/sites/default/files/publications/recsys2014.pdf>.

Рецензент: д-р техн. наук, проф. С. Г. Семенов,
 Національний технічний університет «ХПІ», Харків;
 Received (Надійшла) 08.08.2018
 Accepted for publication (Прийнята до друку) 15.09.2018

Методы оценки качества работы рекомендательных систем

Е. В. Мелешко

Предметом изучения в статье является процесс оценки качества работы рекомендательных систем. **Целью** является выявление важнейших показателей качества работы рекомендательных систем и определение методов их оценивания. **Задача:** исследовать методы оценки качества работы рекомендательных систем, исследовать показатели качества работы рекомендательных систем. Получены следующие **результаты:** Рассмотрены основные и дополнительные показатели качества работы рекомендательных систем. Проведено исследование их важности с точки зрения оценки различных свойств списка рекомендаций. Определены направления дальнейших исследований для разработки методов оценки качества работы рекомендательных систем. **Выводы.** Основными показателями качества работы рекомендательных систем является точность прогнозирования оценок пользователя, а также покрытие пространства объектов и покрытие пространства пользователей. Дополнительными показателями качества работы рекомендательных систем является разнообразие, новизна, неожиданность, робастность, приватность, риск. Показатели разнообразие, новизна, неожиданность позволяют оценить качество структуры рекомендаций и вероятность появления проблемы пузыря фильтров. Показатели робастность, приватность, риск позволяют оценить информационную безопасность рекомендационной системы и ее пользователей. Для каждого отдельного веб-ресурса или приложения можно составить свой набор дополнительных показателей качества работы рекомендательной системы, но рассмотренные дополнительные показатели будут актуальны практически для любого случая. Также важными могут быть следующие показатели: уверенность, доверие, адаптивность, масштабируемость, пропускная способность, полезность и т.д. Поскольку не существует общепринятых способов измерения и методов оценки таких показателей качества работы рекомендационных систем, как разнообразие, новизна, неожиданность, приватность, риск, робастность, научно-практическая задача разработки таких методов является актуальной.

Ключевые слова: рекомендательные системы, показатели качества, точность работы, качество работы, покрытие, пузырь фильтров, информационная безопасность, анализ данных, тестирование программного обеспечения.

Quality assessment methods of work of recommendation systems

Yu. Meleshko

The **subject matter** of the article is the processes of assessing a quality of the work of a recommendation systems. The **goal** is to investigate the problems of modern recommendation systems and to find methods for their solution. The **tasks** to be solved are: to identify the most important a quality metrics of the work of a recommendation systems and research the methods for their assessment. The following **results** were obtained: the main and additional quality metrics of a recommendation systems work are considered. Their importance from the point of view of assessing the different properties of a list of recommendations have been studied. Further directions of research for the development of a quality assessing methods of the work of recommendation systems have been determined. **Conclusions.** The main quality metrics of the work of recommendation systems is the prediction accuracy of user's ratings, as well as covering the space of objects and covering the space of users. Additional quality metrics of recommendation systems are diversity, novelty, serendipity, robustness, privacy, risk. The metrics of diversity, novelty and serendipity allow us to assess the quality of a structure of recommendations and the probability of the filter bubble problem. Metrics of robustness, privacy and risk can assess the information security of the recommendation system and its users. For each individual web-resource or application, you can create your own set of additional quality metrics of the recommendation system, but the considered additional quality metrics will be relevant for almost any case. The following metrics can also be important: confidence, trust, adaptability, scalability, throughput, utility, etc. Since there are no generally accepted measure and methods for assessing such quality metrics of recommended systems, such as diversity, novelty, serendipity, privacy, risk, robustness, the scientific-practical task of developing such methods is actual.

Keywords: recommendation systems, quality metrics, accuracy of work, quality of work, coverage, filter bubble, information security, data analysis, software testing.