

УДК 004.9+004.42+339.1

ПОБУДОВА АСОЦІАТИВНИХ ПРАВИЛ ДЛЯ ПРОГНОЗУВАННЯ РЕКОМЕНДАЦІЙ В КОЛАБОРАТИВНИХ РЕКОМЕНДАЦІЙНИХ СИСТЕМАХ

М. В. Лобур, Ю. В. Стех, М. Є. Шварц

*Національний університет «Львівська політехніка»,
вул. С. Бандери, 12, Львів, 79013, Україна*

Розроблено метод і алгоритм побудови асоціативних правил для прогнозування рекомендацій в колаборативних рекомендаційних системах. Метод використовує вектори профілів користувачів, які зберігаються в матриці користувач-предмет і використовує алгоритм Apriori для побудови асоціативних правил. Метод дає змогу прогнозувати рекомендації для нового користувача рекомендаційної системи і для активного користувача із заповненим вектором профілю.

Ключові слова: *рекомендаційна система, колаборативна фільтрація, асоціативні правила, прогнозування рекомендацій.*

Постановка проблеми. Рекомендаційні системи застосовуються здебільшого для пропозиції клієнтові в реальному часі продуктів (фільмів, книг, одягу), які, ймовірно, його зацікавлять [1, 2]. Особливо широко рекомендаційні системи використовуються в електронній комерції [3]. Застосування рекомендаційних систем поширюється останнім часом на стаціонарну роздрібну торгівлю, довідкові центри, пошук у програмному забезпеченні, наукових статей тощо. Це застосування характеризується наданням рекомендацій користувачам автоматично, на підставі вже вчинених дій (покупок, виставлених рейтингів, відвідувань тощо) і прийомом від них зворотного зв'язку (замовлення в магазинах, перехід по посиланнях і т. д.) [4, 5]. Рекомендаційні системи — один із важливих розділів інтелектуального аналізу даних — Data Mining [6]. Колаборативна фільтрація використовує відомі переваги (оцінки) групи користувачів для прогнозування невідомих переваг іншого користувача [7, 8]. Її основне допущення полягає у такому: ті, хто однаково оцінювали будь-які предмети в минулому, схильні давати схожі оцінки інших предметів у майбутньому. Наприклад, колаборативна фільтрація здатна прогнозувати, яка музика сподобається користувачеві, маючи неповний список його переваг (симпатій і антипатій). Прогнози складаються індивідуально для кожного користувача, хоча використовується інформація, зібрана від багатьох користувачів. Так колаборативна фільтрація відрізняється від більш простого підходу, який дає усереднену оцінку для кожного об'єкта інтересу, наприклад, що базується на кількості поданих за нього голосів. Сьогодні ведуться активні дослідження, що зумовлено наявністю ряду невирішених проблем в колаборативній фільтрації. До таких проблем відносять: підвищення точності прогнозування рекомендацій, введення інформації

про нового користувача і новий предмет в систему (проблема холодного старту), прогнозування предметів для нового користувача, проблема великої розмірності і великої розрідженості матриці користувач-предмет [4, 5].

Аналіз останніх досліджень та публікацій. Перші публікації з дослідження асоціативних прави з'явилися в 1993 — 1994 роках. Віттоді цей науковий напрям інтенсивно розвивався і став одним із провідних напрямків інтелектуального аналізу даних. Було розроблено низку алгоритмів побудови і дослідження асоціативних правил: Apriori [9], AprioriTID [10], FP-tree [11], Eclat [12]. Основні зусилля науковців направлені на: 1) зменшення числа звертань до бази даних транзакцій; 2) зменшення часу вибірки з бази даних транзакцій; 3) введення додаткових обмежень на структуру асоціативних правил; 4) розроблення паралельних алгоритмів пошуку асоціативних правил [13]. Застосування асоціативних правил у рекомендаційних системах направлене на підвищення точності прогнозування рекомендацій, вирішення проблеми нового користувача і нового предмета, підвищення рівня захисту рекомендаційної системи від зовнішніх атак [14].

Виклад основного матеріалу дослідження. Нехай задано множину векторів профілів користувачів \mathbf{U}_p

$$\mathbf{U}_p = (\mathbf{U}_{p_1}, \mathbf{U}_{p_2}, \dots, \mathbf{U}_{p_n}), \quad (1)$$

де $\mathbf{U}_{p_i} = (u_{p_{i1}}, u_{p_{i2}}, \dots, u_{p_{im}})$, $u_{p_{ij}}$ - j -та компонента вектора профілю i -того користувача, множину векторів профілів предметів \mathbf{I}_p

$$\mathbf{I}_p = (\mathbf{I}_{p_1}, \mathbf{I}_{p_2}, \dots, \mathbf{I}_{p_m}), \quad (2)$$

де $\mathbf{I}_{p_i} = (i_{p_{i1}}, i_{p_{i2}}, \dots, i_{p_{ij}})$, $i_{p_{ij}}$ - j -та компонента вектора профілю i -того предмета, матрицю користувач-предмет \mathbf{A} . Рядки матриці \mathbf{A} відповідають векторам профілів користувачів, стовпці матриці \mathbf{A} відповідають векторам профілів предметів. Елемент $r_{ij} > 0$ (рейтинг предмета), якщо користувач \mathbf{U}_{p_i} вибрав предмет \mathbf{I}_{p_j} . Елемент $r_{ij} = 0$, якщо користувач \mathbf{U}_{p_i} не вибрав предмет \mathbf{I}_{p_j} . Величина числового значення елементів матриці r_{ij} визначає рівень інтересу кожного користувача до відповідного предмета. Більше чисельне значення елемента r_{ij} відповідає більшому рівню інтересу користувача \mathbf{U}_{p_i} до предмета \mathbf{I}_{p_j} . У цій роботі асоціативні правила використовуються для вирішення таких задач:

1. Рекомендація предметів новому користувачу, який вперше звернувся до рекомендаційної системи і профіль якого в матриці \mathbf{A} рекомендаційної системи порожній.
2. Рекомендація предметів користувачу, який вже звертався до рекомендаційної системи і профіль якого в матриці \mathbf{A} рекомендаційної системи містить інформацію про оцінку частини предметів.

Асоціативні правила — це імплікація

$$\mathbf{X} \Rightarrow \mathbf{Y}, \quad (3)$$

де $\mathbf{X} \in \mathbf{G}$, $\mathbf{Y} \in \mathbf{G}$, $\mathbf{X} \cap \mathbf{Y} = \emptyset$, \mathbf{G} — множина досліджуваних предметів (книги, фільми, наукові статті тощо).

Ця імплікація означає, що вибираючи предмети, які належать множині \mathbf{X} , вибираються предмети, які належать множині \mathbf{Y} .

Прогнозування рекомендацій для активного користувача в методі колаборативної фільтрації здійснюється на основі всіх векторів профілів користувачів, які

містяться в матриці $\tilde{\mathbf{A}}$ з урахуванням подібності між вектором профілю активного користувача і вектором профілю кожного користувача в матриці \mathbf{A} .

Позначимо через \mathbf{S}_v множини предметів, які вибрав користувач v , \mathbf{S}_u — множини предметів, які вибрав користувач u . Тоді \mathbf{S}_{uv} — множина предметів, які вибрали користувачі u і v

$$\mathbf{S}_{uv} = \{r_{ui} \neq 0 \wedge r_{vi} \neq 0\}, i = 1(1)n. \quad (4)$$

Нехай \bar{r}_v, \bar{r}_u — середні значення рейтингів предметів векторів профілів користувачів u і v відповідно.

Позначимо через \mathbf{T}_a множини користувачів, які мають спільно вибрані предмети з активним користувачем.

Прогноз рейтингу у методі колаборативної фільтрації, який орієнтований на використання векторів профілів користувачів, здійснюється за такою формулою:

$$r_{ai} = \bar{r}_a + \frac{\sum_{l \in \mathbf{T}_a} (r_{li} + \bar{r}_l) w_{al}}{\sum_{l \in \mathbf{T}_a} |w_{al}|}, \quad (5)$$

де w_{al} — міра подібності між вектором профілю активного користувача і вектором профілю користувача l .

Точність прогнозу рейтингу значною мірою залежить від точності розрахунку коефіцієнтів подібності w_{ij} . Зазвичай розрахунок коефіцієнтів подібності здійснюється за формулою косинуса кута між векторами профілів користувачів (6) або за формулою кореляції Пірсона (7):

$$w_{uv} = \frac{\sum_{i \in \mathbf{S}_{uv}} r_{ui} \times r_{vi}}{\sqrt{\sum_{i \in \mathbf{S}_{uv}} r_{ui}^2 \sum_{i \in \mathbf{S}_{uv}} r_{vi}^2}}, \quad (6)$$

$$w_{uv} = \frac{\sum_{i \in \mathbf{S}_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in \mathbf{S}_{uv}} (r_{ui} - \bar{r}_u)^2 \sum_{i \in \mathbf{S}_{uv}} (r_{vi} - \bar{r}_v)^2}}. \quad (7)$$

Особливість матриці $\tilde{\mathbf{A}}$ векторів профілів користувачів і предметів полягає в тому, що вона має велику кількість нульових елементів. Кожен користувач не вибирає усі предмети, і кожен предмет не вибирається усіма користувачами. Переважно відсоток ненульових елементів не перевищує 10 %.

Для нового користувача, який вперше звернувся до рекомендаційної системи, вектор профілю містить лише нульові елементи. Тому застосування виразів (5)–(7) методу колаборативної фільтрації неможливе. Точність прогнозування рекомендацій для вже зареєстрованого користувача також низька через високу розрідженість матриці $\tilde{\mathbf{A}}$.

Метод побудови асоціативних правил призначений для вирішення цих задач.

Він складається з таких кроків:

1. Матриця користувач-предмет перетворюється в матрицю користувач-інтерес. Не всі предмети, які вибрав користувач, мають однаковий інтерес для нього. Рейтинги предметів, які виставляють користувачі, належать інтервалу $[0,5]$ або $[0,10]$. Перетворення матриці користувач-предмет відбувається за таким алгоритмом: 1) якщо $r_{ij} < 3$ для інтервалу $[0,5]$ або $r_{ij} < 6$ для інтервалу $[0,10]$, тоді $r_{ij} = 0$;

2) якщо $r_{ij} \geq 3$ для інтервалу $[0,5]$ або $r_{ij} \geq 6$ для інтервалу $[0,10]$, тоді $r_{ij} = 1$. Матриця користувач-предмет перетворюється в матрицю користувач-інтерес, яка містить лише двійкові елементи 0 або 1.

2. Здійснюється пошук асоціативних правил згідно з елементами матриці користувач-інтерес

$$\mathbf{X} \Rightarrow \mathbf{Y}, \quad (8)$$

де множини \mathbf{X} і \mathbf{Y} мають такі властивості:

$$\mathbf{X} \subset \mathbf{I}_p, \mathbf{Y} \subset \mathbf{I}_p, \mathbf{X} \cap \mathbf{Y} = \emptyset. \quad (9)$$

Кожен рядок матриці користувач-інтерес вважається окремою транзакцією.

3. Отримані асоціативні правила використовуються для прогнозування рекомендацій.

Задача прогнозування рекомендацій для нового користувача вирішується за допомогою таких кроків:

1. Рекомендаційна система пропонує новому користувачу top 10 предметів, які мають найбільшу кількість запитів з боку інших користувачів системи. Новий користувач вибирає один із запропонованих предметів.

2. Серед асоціативних правил, в яких $|\mathbf{X}| = 1$, вибирають ті правила, в котрих $i_k \equiv \mathbf{X}$.

Нехай це будуть правила:

$$i_k \Rightarrow \mathbf{Y}_1, i_k \Rightarrow \mathbf{Y}_2, \dots, i_k \Rightarrow \mathbf{Y}_l, \quad (10)$$

$$\mathbf{Y}_1 \subset \mathbf{Y}, \mathbf{Y}_2 \subset \mathbf{Y}, \dots, \mathbf{Y}_l \subset \mathbf{Y}. \quad (11)$$

3. Новому користувачу пропонуються предмети з множини

$$\mathbf{N} = \mathbf{Y}_1 \cap \mathbf{Y}_2 \cap \dots \cap \mathbf{Y}_l. \quad (12)$$

Задача прогнозування рекомендацій для активного користувача, для якого вже існує вектор профілю в матриці $\tilde{\mathbf{A}}$, вирішується за допомогою таких кроків:

1. Будується вектор профілю активного користувача $\tilde{\mathbf{U}}_p$ по даних матриці користувач-інтерес.

2. Серед побудованих асоціативних правил шукають такі, для яких виконуються умови $\tilde{\mathbf{U}}_p \equiv \mathbf{X}$.

Нехай це будуть правила:

$$\tilde{\mathbf{U}}_p \Rightarrow \mathbf{Y}_1, \tilde{\mathbf{U}}_p \Rightarrow \mathbf{Y}_2, \dots, \tilde{\mathbf{U}}_p \Rightarrow \mathbf{Y}_l, \quad (13)$$

$$\mathbf{Y}_1 \subset \mathbf{Y}, \mathbf{Y}_2 \subset \mathbf{Y}, \dots, \mathbf{Y}_l \subset \mathbf{Y}. \quad (14)$$

3. Активному користувачу пропонуються предмети із множини

$$\mathbf{N} = \mathbf{Y}_1 \cap \mathbf{Y}_2 \cap \dots \cap \mathbf{Y}_l. \quad (15)$$

Для побудови асоціативних правил для нового користувача і для активного користувача використовується алгоритм Apriori [10].

Висновки. Розроблено метод і алгоритм побудови асоціативних правил для прогнозування рекомендацій в колаборативних рекомендаційних системах. Побудова асоціативних правил базується на векторах профілів користувачів, які містять рейтингові числові оцінки корисності використаних предметів. Розроблено алгоритм побудови матриці користувач-інтерес на основі даних матриці користувач-предмет. Для побудови асоціативних правил використовується алгоритм Apriori. Розроблено методи і алгоритми прогнозування рекомендацій для нового

користувача рекомендаційної системи і для активного користувача, профіль котрого вже існує в рекомендаційній системі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ricci F., Rokach L. Shapira B. Introduction to Recommender Systems Handbook, Springer, 2011, 35pp.
2. Resnick P., Varian H. R. «Recommender system». Communications of the ACM. 1997. № 3. 40. Pp. 56–58.
3. Linden G., Smith B.J. «Amazon.com recommendations: Item-to-Item collaborative filtering». IEEE Internet Computing. 2003. Vol. 7, Pp. 76–80.
4. Adomavicius B., Tuzhilin A. «Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions». IEEE Transactions on Knowledge and Data Engineering. 2005. Vol.17. Pp. 734-748.
5. Bobadilla J., Ortega A., Gutierrez H. A. «Recommender systems survey». Knowledge Based System. 2013. Vol. 46. Pp.109–132.
6. Паклин Н., Орешков В. Бизнес-аналитика: от данных к знаниям Москва, Санкт-Петербург, 2013. 693 с.
7. Herlocker J. L., Konstan J. A., Terveen L. G., Riedl J. T. «Evaluating Collaborative filtering recommender systems». ACM Transactions on Information Systems .Vol 22. Pp.5–53.
8. Wang J., de Vries A. P., Reinders M. J. T. «Unifying userbased and itembased collaborative filtering approaches by similarity fusion», In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. 2006. Pp. 501–508.
9. Agrawal, R., Imielinski, T., Swami, A. N. «Mining association rules between sets of items in large databases» In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. 1993. Pp. 207–216.
10. Agrawal R. Srikant R. «Fast algorithms for mining association rules» In Proc. 20th Int. Conf. Very Large Data Bases. 1994. Pp.487–499.
11. Han J., Pei J. 2000. Mining frequent patterns by pattern-growth: methodology and implications. ACM SIGKDD Explorations Newsletter . 2000. Vol. 2. Pp.14–20.
12. Zaki M.J. «Mining Non-Redundant Association Rules». Data Mining and Knowledge Discovery. 2004. Vol. 9. Pp. 223–248.
13. Kotsiantis S., Kanellopoulos D. «Association Rules Mining: A Recent Overview». International Transactions on Computer Science and Engineering. 2006. Vol. 32. Pp. 71–82.
14. Tyag S., Bharadwaj K. K. «Enhanced New User Recommendations based on Quantitative Association Rule Mining» In 3rd International Conference on Ambient Systems, Networks and Technologies (ANT). 2012. Pp. 102–109.

FORMATION OF ASSOCIATE RULES FOR PREDICTING RECOMMENDATIONS IN COLLABORATIVE RECOMMENDER SYSTEMS

M. V. Lobur, Yu. V. Stekh, M. E. Shvarts

*Lviv Polytechnic National University,
12, Bandera St., Lviv, 79013, Ukraine*

A method and an algorithm for the formation of associative rules for predicting recommendations in collaborative recommender systems have been developed. The method uses the user profile vectors that are stored in the user-item matrix. The method uses the Apriori algorithm to form associative rules. The method allows you to predict recommendations for a new user of the recommender system and for the active user with a filled profile vector.

Keywords: *recommender system, collaborative filtering, associative rules, predicting of recommendations.*

Стаття надійшла до редакції 00.00.2018.