

УДК 004.9+004.42+339.1

## МЕТОД І АЛГОРИТМ ПРОГНОЗУВАННЯ РЕКОМЕНДАЦІЙ ДЛЯ СПІЛЬНОТ КОРИСТУВАЧІВ

М. В. Лобур, Ю. В. Стех, М. Є. Шварц

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

*Розроблено метод і алгоритм прогнозування рекомендацій для спільнот користувачів. Метод використовує вектори профілів користувачів для знаходження спільнот користувачів і метод розрахунку адитивної корисності для прогнозування предметів для кожної спільноти користувачів.*

***Ключові слова:** рекомендаційна система, профіль користувача, метод к-середніх, матриця користувач-предмет, прогнозування рекомендацій, профіль предмета.*

**Постановка проблеми.** Системи рекомендацій або рекомендаційні системи є підкласом системи фільтрації інформації, які прагнуть передбачити «рейтинг» або «перевагу», які користувач дав би певному предмету [1, 2]. Рекомендаційні системи набули широкого застосування в останні роки. Вони широко застосовуються під час пошуку найрізноманітнішого контенту в Інтернеті (книги, музика, фільми, послуги). Сьогодні такі відомі пошукові системи, як Google, Yahoo, Yandex використовують рекомендаційні системи і ведуть інтенсивні дослідження в галузі розроблення і дослідження нових методів і алгоритмів роботи рекомендаційних систем [3, 4, 5]. Більшість існуючих сьогодні рекомендаційних систем прогнозують рекомендації для індивідуальних користувачів. Перспективним напрямом подальших досліджень в галузі рекомендаційних систем є розроблення рекомендаційних систем, які надають рекомендації для спільнот користувачів. Такі рекомендаційні системи мають узагальнювати інтереси всіх користувачів в спільноті та надавати найбільш корисні рекомендації.

**Аналіз останніх досліджень та публікацій.** Аналіз останніх публікацій [4, 5, 6, 7, 8, 9] та досліджень показує, що перспективними напрямами дослідження методів і алгоритмів побудови рекомендаційних систем для спільнот користувачів є: 1) дослідження методів і алгоритмів виділення спільнот користувачів; 2) дослідження методів і алгоритмів прогнозування рекомендацій для спільноти користувачів; 3) дослідження введення нових користувачів в спільноту; 3) дослідження методів і алгоритмів врахування інтересів індивідуальних користувачів у спільноті.

**Мета статті** — побудувати і дослідити метод і алгоритм прогнозування рекомендацій для спільнот користувачів. Основою виділення спільнот користувачів є множина векторів профілів користувачів. Досліджено можливість застосування методу неієрархічної кластеризації для пошуку спільнот користувачів. Висвітлено можливість прогнозування рекомендацій для кожної спільноти.

**Виклад основного матеріалу дослідження.** Нехай  $U$  — множина користувачів,  $|U| = m$ ,  $I$  — множина предметів,  $|I| = n$ ,  $D$  — множина дій надання персоналізованих рекомендацій. Потрібно здійснити прогноз персоналізованих рекомендацій  $U \times I \rightarrow D$ .

Нехай  $R$  — множина оцінок користувачів (рейтингів предметів),  $U_p$  — множина профілів користувачів,  $I_p$  — множина профілів предметів,  $U_{p_j} \in U_p$  — вектор профілю  $j$ -того користувача (1)

$$U_{p_j} = (r(u_{p_j}, i_1), r(u_{p_j}, i_2), \dots, r(u_{p_j}, i_n)), \tag{1}$$

де  $r(u_{p_j}, i_k)$  — оцінка  $j$ -того користувача для  $i_k$ -того предмета.

$I_{p_j} \in I_p$  — вектор профілю  $j$ -того предмета (2)

$$I_{p_j} = (r(u_{p_1}, i_j), r(u_{p_2}, i_j), \dots, r(u_{p_m}, i_j)), \tag{2}$$

де  $r(u_{p_k}, i_j)$  — оцінка  $j$ -того предмета для  $k$ -того користувача,

Профіль користувача містить оцінку користувачем предмета за 5-ти або 10-ти бальною шкалою. Приклад матриці користувач-предмет наведений на рис. 1.

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$u_1$	0	1	0	5	3
$u_2$	2	4	3	0	0
$u_3$	1	0	5	0	4
$u_4$	2	5	4	4	1
$u_5$	1	0	2	3	0

Рис.1. Приклад матриці користувач-предмет

Необхідно здійснити прогноз персоналізованих рекомендацій (3)

$$\hat{r}_{ij} = Predict(U_p, I_p), \tag{3}$$

де  $\hat{r}_{ij}$  — прогнозоване значення оцінки,  $r_{ij}$  — оцінка, яку виставив користувач. Завдання системи полягає в тому, щоб досягнути мінімальної різниці між  $\hat{r}_{ij}$  і  $r_{ij}$  (4)

$$|\hat{r}_{ij} - r_{ij}| \Rightarrow \min. \tag{4}$$

Прогнозування рекомендацій для спільнот користувачів потребує попереднього розбиття всієї множини користувачів на спільноти, в межах котрих користувачі мають подібні інтереси. Інтерес користувача до групи предметів визначається його вектором профілю. Вектор профілю користувача — це вектор в  $n$ -вимірному просторі. Тому подібність двох або більше користувачів відповідає подібності (близькості) їхніх векторів профілів в  $n$ -вимірному просторі. Найбільш вживаними мірами подібності в рекомендаційних системах є [4, 5, 10]: 1. Косинусна відстань; 2. Коефіцієнт кореляції; Пірсона; 3. Евклідова відстань.

У праці [11] зображено, що найкраще відповідає дійсній мірі подібності величина евклідової відстані між двома векторами профілів користувачів.

Пошук множини спільнот подібних користувачів здійснюється за допомогою кластеризації множини векторів профілів користувачів. Для кластеризації використовується метод  $k$ -середніх. Метод  $k$ -середніх має низку недоліків: потрібно до початку роботи алгоритму задати кількість кластерів і їх початкове розміщення. Окрім того, метод  $k$ -середніх локально збіжний. Тому в розробленому алгоритмі використовується модифікація методу  $k$ -середніх, яка забезпечує глобальну збіжність і наведена в [12].

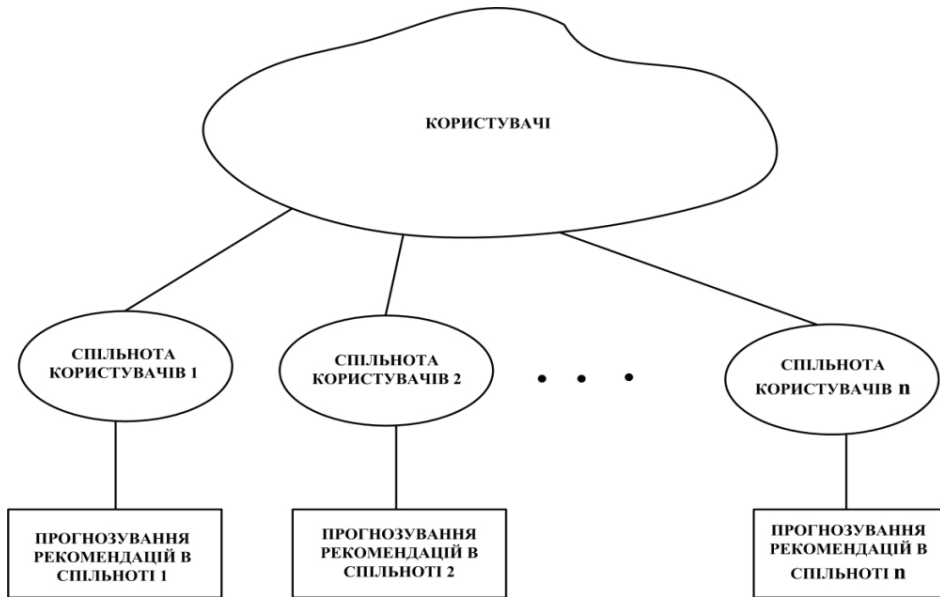


Рис. 2. Узагальнена структурна схема прогнозування рекомендацій для спільнот користувачів

Узагальнена структурна схема прогнозування рекомендацій для спільнот користувачів зображена на рис. 2. Формальна постановка задачі пошуку спільнот користувачів така. Нехай задано множину векторів профілів користувачів (5)

$$\mathbf{U}_p = (\mathbf{U}_{p_1}, \mathbf{U}_{p_2}, \dots, \mathbf{U}_{p_m}). \quad (5)$$

Потрібно знайти множину спільнот користувачів (6)

$$\mathbf{G}_p = (\mathbf{G}_{p_1}, \mathbf{G}_{p_2}, \dots, \mathbf{G}_{p_k}). \quad (6)$$

Множина спільнот користувачів має такі властивості (7), (8), (9):

$$\mathbf{G}_p = \mathbf{G}_{p_1} \cup \mathbf{G}_{p_2} \cup \dots \cup \mathbf{G}_{p_k}, \quad (7)$$

$$\mathbf{G}_{p_i} \cap \mathbf{G}_{p_j} = \emptyset \forall i, j = 1(1)k, i \neq j. \quad (8)$$

$$\mathbf{G}_{p_j} = (\mathbf{U}_{p_{j1}}, \mathbf{U}_{p_{j2}}, \dots, \mathbf{U}_{p_{jm}}). \quad (9)$$

Для кожної спільноти користувачів  $\mathbf{G}_{p_i}$  евклідова відстань між будь-якими двома векторами профілів користувачів, які належать до цієї множини, є меншою ніж відстань між будь-яким вектором профілю множини  $\mathbf{G}_{p_i}$  і довільної множини  $\mathbf{G}_{p_j}$ ,  $\mathbf{G}_{p_i}, \mathbf{G}_{p_j} \in \mathbf{G}_p$ . Кількість ненульових елементів у матриці користувач-предмет не перевищує 6–7% від загальної кількості елементів [4, 5]. Тому перед виконанням кластеризації матриця користувач-предмет заповнюється за таким алгоритмом: кожен нульовий елемент матриці заповнюється середнім арифметичним значенням від середнього арифметичного усіх елементів рядка і усіх елементів стовпця в якому є нульовий елемент. Приклад заповнення матриці, яка наведена на рис. 1, приведений на рис. 3.

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$u_1$	3	1	4.6	5	3
$u_2$	2	4	3	4.2	3.4
$u_3$	1	4	5	4.4	4
$u_4$	2	5	4	4	1
$u_5$	1	3.2	2	3	2.8

Рис. 3. Приклад заповнення нульових елементів матриці користувач-предмет

Для кожної спільноти користувачів обчислюється адитивна корисність кожного предмета.

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$u_1$	0	1	0	5	3
$u_2$	2	4	3	0	0
$u_3$	1	0	5	0	4
$u_4$	2	5	4	4	1
$u_5$	1	0	2	3	0
Адитивна корисність	6	10	14	12	8

Рис. 4. Приклад обчислення адитивної корисності для спільноти користувачів

Адитивна корисність — це сума рейтингових оцінок, які виставлені для кожного предмета кожним користувачем. Надалі для усіх членів спільноти рекомендується топ N предметів з найвищою адитивною корисністю.

Приклад розрахунку адитивної корисності для спільноти користувачів (рис.1) зображений на рис. 4. На рис. 5 наведено блок-схема алгоритму прогнозування рекомендацій для спільнот користувачів.

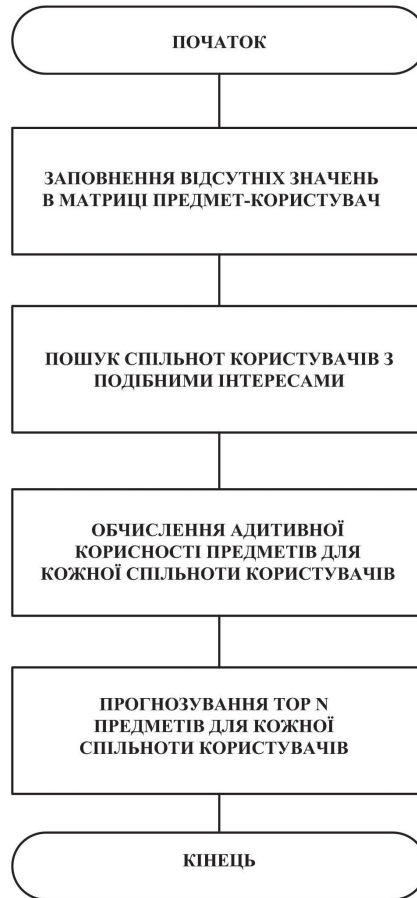


Рис. 5. Блок-схема алгоритму прогнозування рекомендацій для спільнот користувачів

**Висновки.** Розроблено метод і алгоритм прогнозування рекомендацій для спільнот користувачів. Пошук спільнот користувачів базується на векторах профілів користувачів, які містять рейтингові числові оцінки корисності використаних предметів. Для пошуку спільнот користувачів використовується глобально збіжний метод кластеризації  $k$ -середніх. Для кожної спільноти обчислюється адитивна корисність кожного предмета. Кожному користувачу спільноти рекомендується топ- $N$  предметів з найвищою адитивною корисністю.

#### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ricci F., Rokach L. Shapira B. Introduction to Recommender Systems Handbook. Springer, 2011. 35 p.
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. Affan Y., Liu L., Rubia F., Wang J. «Designing the Next Mobile App Recommender System for the Globe» In 2017 14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC), 21–23 June 2017. Pp. 491–500.
7. Minh N. H. N., Chuan P., Jaehyeok S., Choong S. H. «Online learning-based clustering approach for news recommendation systems», In 2016: 18th Asia-Pacific Network Operations and Management Symposium (APNOMS), 5–7 Oct. 2016. Pp. 1–6.
8. Quintarelli E., Rabosio E., Tanca L. «Recommending New Items to Ephemeral Groups Using Contextual User Influence», In Proceedings of the 10th ACM Conference on Recommender Systems. 15–19 September, 2016. Pp. 285–292.
9. Irfan A., Sang-Wong K., An effective approach to group recommendation based on belief propagation, In Proceedings of the 30th Annual ACM Symposium on Applied Computing. Spain. 13–17 April, 2015. Pp. 1148–1153.
10. A new user similarity model to improve the accuracy of collaborative filtering / H. Liu, Z. Hu, A. Mian, H. Tian, X. Zhu. *Knowledge-Based Systems*. 2014. Vol. 56. Pp. 156–166.
11. Stekh Y., Artsibasov V. Some methods in software development recommendation systems. *Вісник НУ «Львівська політехніка»: Комп'ютерні системи проектування. Теорія і практика*. 2013. № 777. С. 74–78.
12. Stekh Y., Artsibasov V. Adaptive clustering algorithm for recommender systems. *Вісник НУ «Львівська політехніка»: Комп'ютерні системи проектування. Теорія і практика*. 2012. № 747. С. 75–78.

## METHOD AND ALGORITHM FOR PREDICTING RECOMMENDATIONS FOR THE USER COMMUNITY

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 predicting recommendations for user communities have been developed. The method uses the user profile vectors to locate the user communities and a method of calculating additive utility for predicting items for each user community.*

**Keywords:** *recommender system, user profile, k-means method, user-item matrix, predicting of recommendations, subject profile.*

*Стаття надійшла до редакції 00.00.2018.*