

Ye. Meleshko

Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

## METHOD OF GENERATING RECOMMENDATIONS LISTS WITH CONSIDERING ACTIVITY INDEXES OF USERS IN A RECOMMENDATION SYSTEM

The **subject matter** of the article is the process of generating recommendations lists for users of a recommendation system. The **goal** is to develop a new method of building recommender systems to improve the quality of recommendations lists, increase user space coverage and item space coverage with considering information about user activity indexes; development of a hybrid of this method with the method of collaborative filtering. The **tasks** to be solved are: to develop the method of building recommendation systems based on considering activity indexes of users, develop software to test this method, conduct experiments with the developed software to test the effectiveness of the developed method, determine the quality of its work and compare this method with the standard method of collaborative filtering. The **methods** used are: graph theory, mathematical statistics, the theory of algorithms, object-oriented programming. The following **results** were obtained: the expert-oriented method for building recommender systems based on considering indexes of user activity and calculating expert coefficients has been developed, the hybrid of this method with the collaborative filtering method has been developed, software for implementing this method and this hybrid has been developed, experiments with developed software to test the developed method and the hybrid has been conducted. **Conclusions.** The possibility of using information about user activity in recommender systems to improve the quality of recommendations lists has been investigated. The calculation of expert coefficients is proposed to supplement the similarity coefficients in recommender systems. An expert-oriented method for constructing recommender systems based on considering activity indexes of users and its hybrid with collaborative filtering has been developed. Experiments has been conducted with the developed software have shown that the developed method significantly improves such indicators of the recommender system as user space coverage and item space coverage and allows to create higher quality of recommendation lists without significant fluctuations Precision and Recall of the recommender system, and in some cases even improve these indicators, it depends on the features of the input data.

**Keywords:** recommendation systems; collaborative filtering; similarity coefficients; expert-oriented filtering; expert coefficients.

### Introduction

Recommendation systems are important tools in the areas of digital marketing and search data [1–3].

Modern recommendation systems have a number of standard problems [4] that different scientists and developers try to solve in different ways. Most of these problems are solved only partially and not for all occasions.

In order to create high-quality recommendations lists for different users today, recommendation systems are developed as complex hybrids of various methods of data filtering. These may be the following methods [1–4]:

- *Collaborative filtering* (based on users behavior history, numerical ratings given to items and/or items previously purchased or selected);

- *Content-based filtering* (based on using descriptions and keywords for items);

- *Knowledge-based filtering* (based on explicit knowledge about the item assortment, user preferences, and recommendation criteria, i.e., which item should be recommended in which context);

- *Community-based filtering* (based on knowledge of social connections);

- *Context-based filtering* (based on using contextual information, such as demographic).

Modern recommender systems are often hybrids of methods of content-based and collaborative filtering.

Knowledge-based filtration is complicated in the implementation and collection of data and rarely used compared to others.

Context-based and community-based filtering can only be used as additional to others methods, and data for them is not available in all systems.

One of the important problems of collaborative filtering is that not all users can create recommendation lists due to cold start problem, continuous cold start problem, or an insufficient number of users that are similar to a particular user.

This problem can be partially solved by using a hybrid of collaborative filtration with the above methods. Because it's not always possible to create a hybrid of collaborative filtration with knowledge-based, context-based, or community-based filtering, and qualitative content-based filtering requires a lot of information about items and, therefore, more time to compute, and also depends on the quality of an item descriptions, it was decided to research alternative ways of solving the problem of lack of recommendations for some users, in particular, to try to take into account the user activity indexes for generating recommendations, and to estimate how much information on user activity will improve quality formation of recommendations lists.

Therefore, in this work, it is suggested to use the data on user activity indexes to form recommendations list. That is, recommendations lists for users for which it was not possible to form recommendations by other methods will consist of recommendations from most popular items among the most active users. This approach is proposed to call expert-oriented since each user of a recommendation system is considered as an expert on the content of this system and has a weight (expert coefficient), which value depends on his index activity. The

greater the weight of the expert, the greater the contribution to the formation of recommendations lists for other users does data about his preferences. In this method, it is proposed to determine the weight of an expert based on his activity index. Although in the future it is planned to develop and other methods for determining the weight of the expert.

On the basis of this approach, the expert-oriented filtering method based on activity indexes of users and the hybrid of this method with the method of collaborative filtration, was developed.

The software was developed and the series of experiments to test the effectiveness of the developed method and developed hybrid and compare their with the standard collaborative filtration method was conducted.

### The main material

For recommendation systems cold start problem and continuous cold start problem for users and for items are relevant, they mean it is impossible to create recommendations for new elements of a system or elements that have changed their properties [4].

It should also be noted that other problems and obstacles are possible for building recommendations lists.

So the work of modern recommendation systems is based on the definition of similarity coefficients between users and/or items of a system [1–3, 5, 6].

The similarity of users each other is determined by their actions with a same items if these actions are comparable, for example, can compare ratings which different users were put the same films.

If users do not have the same actions for comparison, then their similarity coefficients cannot be determined by standard methods. If for a particular user it was not possible to identify similar users, then it is not possible to form recommendations list based on collaborative filtering for him. In such cases, additional information is used, for ex-ample, contextual information, or information about social connections, or content filtering [1, 4], and it was also suggested to determine a missing similarity coefficients between users through using associative networks [7].

The purpose of this work is to develop the method of filtering for recommendation systems that would solve the above problems by using information about a degree of users activity in a system.

It is suggested to consider most active users of the system, which put the most ratings of items, experts in the content of the system. It is also suggested to calculate the value of the expert coefficient for each user of the system and use such coefficients in the algorithm of formation recommendations list. This approach is proposed to be called expert-oriented filtering.

It was decided to calculate the expert coefficient for each user as follows:

$$k_{exp}(u) = \frac{m_u}{\max(m_i)}, \quad (1)$$

where  $m_u$  – the number of items that the user  $u$  has rated;  $\max(m_i)$  – the number of items that were rated by the user who rated the more items in the system among all users.

The more items a user has rated, the higher he will get the expert coefficient. The value of this coefficient can vary from 0 to 1. It will be 0 if the user has not evaluated any item, and is equal to 1 if the user has evaluated the more items in comparison with others.

After calculating expert coefficients, recommendations list for a specific user  $u_i$  can be formed as follows:

1. Choose all users except the user  $u_i$ .
2. Select all the items, which were rated by choose users, and calculate for each item his coefficient of interest for the user  $u_i$ :

$$k_{int}(o, u_i) = \sum_{j \geq 1, j \neq i}^n r_j \cdot k_{exp}(u_j), \quad (2)$$

where  $o$  – the item for which calculate the coefficient of interest for the user  $u_i$ ;  $n$  – the number of users in the system;  $k_{exp}$  – the expert coefficient for  $j$ -th user;  $r_j$  – the rating given by  $j$ -th user of the item  $o$ .

3. Create the list of all selected items, if necessary, remove items that have already been evaluated by the user  $u_i$ .

4. Sort the list of all choose items in descending order of the coefficient of interest for the user  $u_i$ .

5. If necessary, select TopN of items from the resulting list.

The need to perform clauses 3 and 5 depends on for which application or web resource the recommendation system is created. There may be cases when can recommend items for users that they have already viewed (for example, in a recommendation system for Internet-radio). There may be cases when need to show all recommendations list, not just TopN (for example, in search results of search engine).

Also, in this work, the hybrid of the developed method with the method of collaborative filtration has been developed.

Since all the data in a system is conveniently written in the form of a graph [8], where users are vertices of the graph, and similarity coefficients and expert coefficients are written as values of the properties of the corresponding edges of the graph, in order to explain the principle of the work of the developed hybrid, introduce the following concepts:

- bidirectional relationship "Similarity" between users (Fig. 1);
- unidirectional relationships "Expert" between users (Fig. 2).

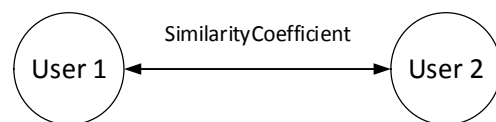


Fig. 1. Relationship "Similarity" between users

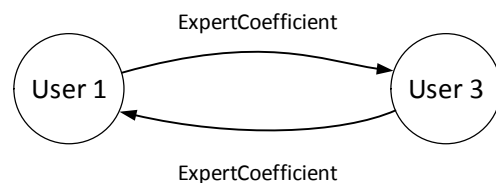


Fig. 2. Relationships "Expert" between users

The developed hybrid works as follows:

1. Calculate similarity coefficients between system users.
2. For those users who do not have same actions with other users (Figure 3, User3), therefore, it is impossible to calculate similarity coefficients for them with other users and to create relationships "Similarity", the expert-oriented filtering method based on considering activity indexes of users was developed. In this case, relationships "Expert" will be directed from all users to the user without relationships "Similarity", and from this user to all other users (Fig. 3).

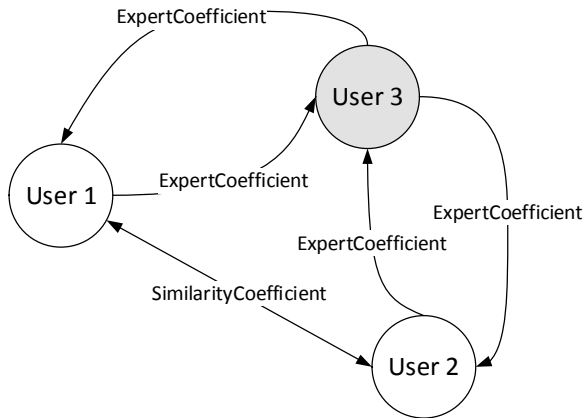


Fig. 3. The example of constructing relationships "Similarity" and "Expert" in the hybrid method

This solves the problem with users who do not have any relationships "Similarity", recommendations for them are built on the basis of the preferences of the most active users in the system. Also, there is a problem with items that were rated only by those users who have no relationships "Similarity" with anyone, and therefore it is unknown to whom can recommend them; relationships "Expert" from these users will be directed to all other users of the system, and the more items were rated by them, the greater the contribution to recommendations lists of other users will make their preferences. It possible to say that user who has no relationships "Similarity" is an expert of rare content of the system.

For conducting experiments, a recommendation system was constructed using the programming language Python and the graph database Neo4j, since the database Neo4j provides extensive opportunities for working with data [9]. In the developed recommendation system, the collaborative filtration was implemented, in which the similarity coefficients between users were determined on the basis of the cosine distance:

$$d(x_1, x_2) = \frac{\sum_{i=1}^m (x_{1i} - X_1) \cdot (x_{2i} - X_2)}{\sqrt{\sum_{i=1}^m (x_{1i} - X_1)^2 \cdot \sum_{i=1}^m (x_{2i} - X_2)^2}} \quad (3)$$

where  $d(x_1, x_2)$  – the distance between users  $x_1$  and  $x_2$ ;  $x_{1i}, x_{2i}$  – the value of the  $i$ -th property in the 1st and 2nd users, respectively;  $X_1, X_2$  – the set of values of properties in the 1st and 2nd users.

The developed recommendation system was tested on the open data sets MovieLens, that was created in the research laboratory at the Department of Computer Science and Engineering at the University of Minnesota [10]. This recommendation system is developed for the movie trailers web site. Users set ratings and hashtags for movies and receive recommendations lists of movies for future viewing

For each experiment, the numbers of users  $N_u$  from MovieLens Datasets were selected. Ratings that they put movies were split based on timestamps into two parts for the calculation of recommendations ("current data") and for testing the system ("future data").

For each data set the system was launched in three modes:

- the standard collaborative filtering;
- the expert-oriented method based on considering activity indexes of users;
- the hybrid of the first two methods.

In this data set, ratings may take the following values: 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0. It was decided to divide them into positive (from 3.5 to 5.0) and negative (from 0.5 to 3.0) ratings.

Forecasts of users' preferences are divided into positive ones when a positive rating is predicted for a user-item pair, and negative ones when a negative rating is forecasted.

To test the quality of the recommendation system, the following indicators of the quality of a recommendation systems work was selected [11]:

1. The absolute number of users for whom the system was unable to create a recommendations list.
2. **Precision** – the measure that characterizes how many positive predictions of a recommendation system are correct. It was calculated by the formula:

$$Precision = tp / (tp + fp), \quad (4)$$

where  $tp$  – positive predictions of the recommendation system, which proved to be true;  $fp$  – positive predictions of the recommendation system, which turned out to be false.

3. **Recall** (also known as Sensitivity) – the measure characterizing the possibility of a recommendation system to create correct positive predictions, while failing to take into account wrong positive predictions. It was determined by the formula:

$$recall = tp / (tp + fn), \quad (5)$$

where  $tp$  – positive predictions of the recommendation system, which proved to be true;  $fn$  – negative predictions of the recommendation system, which turned out to be false.

4. **User space coverage** – percentage of all users for whom the system can provide recommendations. It was determined by the formula:

$$User\ Space\ Coverage = \frac{N_{ur}}{N_u}, \quad (6)$$

where  $N_{ur}$  – the number of users for which it turned out to create recommendations;  $N_u$  – the number of users that was taken in the experiment.

5. **Item space coverage** – percentage of all items that can be recommended to users. It was determined by the formula:

$$\text{Item Space Coverage} = N_{mr} / N_m, \quad (7)$$

where  $N_{mr}$  – the number of items that have fallen into recommendations lists;  $N_m$  – the total number of items in the experiment.

The series of experiments were conducted for users randomly selected from the MovieLens Datasets. For each experiment, another set of users has been selected. Results of experiments are given in Table 1. It used the following abbreviations: CF – the collaborative filtration method; AU – the developed method based on considering activity indexes of users in a recommendation system; H – the hybrid of these two methods.

Table 1 – Test results of the developed methods of building recommendation systems

Experiment number	total number of users	Number of users without recommendations			Precision, %			Recall, %			User space coverage, %			Item space coverage, %		
		CF	AU	H	CF	AU	H	CF	AU	H	CF	AU	H	CF	AU	H
1	30	13	0	0	82.75	72.81	81.68	85.71	68.81	67.92	56.66	100	100	33.79	72.81	72.81
2	30	16	0	0	92.30	81.55	81.73	77.41	79.66	79.85	46.66	100	100	22.19	63.08	63.08
3	30	10	0	0	58.18	67.45	67.45	58.18	75.44	77.22	66.66	100	100	27.29	67.45	67.45
4	50	18	0	0	72.13	83.08	84.58	86.27	77.29	77.23	64.00	100	100	33.14	85.57	85.57
5	50	21	0	0	41.66	66.55	66.27	83.33	71.31	84.15	57.99	100	100	28.88	71.31	71.31
6	50	22	0	0	65.71	74.24	74.13	69.69	72.22	71.99	56.00	100	100	21.85	52.65	52.65
7	100	43	0	0	72.58	65.24	65.10	77.58	72.92	72.92	56.99	100	100	26.74	68.43	68.43
8	100	54	0	0	73.43	81.72	81.75	71.21	68.53	68.60	46.00	100	100	20.21	47.80	47.80
9	100	46	0	0	83.33	75.52	75.48	62.50	77.71	77.27	54.00	100	100	37.77	77.71	77.71
<b>Mean values:</b>					71.34	74.24	75.35	74.65	73.76	75.23	56.10	100	100	27.98	67.42	67.42

As the results of the experiment showed, the developed method, unlike the collaborative filtering method, allows to create recommendations for all users of the system. Precision and Recall of the developed method can be both worse and better than in the collaborative filtering method, that completely depends on the features of a set of input data. But on average, the developed method and the developed hybrid have greater Precision and almost the same Recall as compared with the method of collaborative filtering. User space coverage in the developed method and in the developed hybrid is always 100%, and in the method of collaborative filtration, on average, this figure is 56.1%. Also, the developed method and the developed hybrid always give better item space coverage, on average more by 39,44% than the method of collaborative filtration.

Consequently, without significant fluctuations of Precision, the developed method and the developed hybrid increase item space coverage by 2.5 times, and user space coverage is 100%, solving the problem with users for whom there was no recommendation in the method of collaborative filtration.

## Conclusions

The method of building recommendations lists based on considering activity indexes of users is proposed, which is proposed to be called expert-oriented filtering based on data about user activity indexes.

The hybrid of this method with the method of collaborative filtration has been developed.

Software has been developed, in which a series of experiments was conducted to test the effectiveness of the developed method and the developed hybrid.

The conducted experiments showed that the developed method and the developed hybrid in general increases the quality of the recommendation system without significant fluctuations of Precision and Recall of the system. Precision and Recall can decrease slightly or increase, depending on the characteristics of the incoming data set.

The developed method and the developed hybrid 2.5 times increases item space coverage and solves the problem with users for which it was impossible to create recommendations by other methods, increasing user space coverage to 100%.

## REFERENCES

1. Francesco, Ricci, Lior, Rokach, Bracha, Shapira and Paul B., Kantor (2010), *Recommender Systems Handbook*, Springer-Verlag, New York, USA, 842 p.
2. Meleshko, E.V., Semenov, S.G. and Khokh, V.D. (2018), "Research of methods of building advisory systems on the internet", *Control, Navigation and Communication Systems*, Issue 1(47), PNTU, Poltava, pp. 131–136.
3. Segaran, T. (2008), *Programming Collective Intelligence. Building Smart Web 2.0 Applications*, O'Reilly Media, 368 p.
4. Meleshko, E.V. (2018), "Problems of modern recommendation systems and methods of their solution", *Control, Navigation and Communication Systems*, Issue 4 (50), PNTU, Poltava, pp. 120–124.
5. Saravana, Balaji B., Karthikeyan, N.K. and Raj Kumar, R.S., (2018), "Fuzzy service conceptual ontology system for cloud service recommendation", *Computers & Electrical Engineering*, Vol. 69, pp. 435–446, DOI: <https://doi.org/10.1016/j.compeleceng.2016.09.013>.
6. Ruban, I., Kuchuk, H., Kovalenko, A. (2017), "Redistribution of base stations load in mobile communication networks", *Innovative Technologies and Scientific Solutions for Industries*, No. 1 (1), pp. 75–81, DOI: <https://doi.org/10.30837/2522-9818.2017.1.075>

7. Meleshko, Ye. (2018), "Method of collaborative filtration based on associative networks of users similarity", *Advanced Information Systems*, Vol. 2, No. 4, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, pp. 55–59.
8. Meleshko, E.V., Bosko, V.V. and Reznichenko, V.A. (2018), "Development of the recommendation system based on DBMS neo4j", *VInt. Scientific and Practical Conf. "Information Technologies and Interactions"*, Kyiv, pp. 351–352.
9. *Neo4j Documentation* (2018), available at: <https://neo4j.com/docs/>.
10. Harper, F.M. and Konstan, J.A. (2016), "The MovieLens Datasets: History and Context", *ACM Transactions on Interactive Intelligent Systems (TiiS)*, DOI: <https://doi.org/10.1145/2827872>.
11. Meleshko, E.V. (2018), "Quality assessment methods of work of recommendation systems", *Control, Navigation and Communication Systems*, Issue 5 (51), PNTU, Poltava, pp. 92–97.

Received (Надійшла) 15.01.2019

Accepted for publication (Прийнята до друку) 06.03.2019

### Метод генерації списків рекомендацій з врахуванням показників активності користувачів рекомендаційної системи

Є. В. Мелешко

**Предметом** вивчення у статті є процес генерації списків рекомендацій для користувачів рекомендаційної системи. **Метою** є розробка нового методу побудови рекомендаційних систем для підвищення якості списків рекомендацій, підвищення покриття простору користувачів та покриття каталогу об'єктів за допомогою врахування інформації про показники активності користувачів; розробка гібриду даного методу з методом колаборативної фільтрації. **Завдання:** розробити метод побудови рекомендаційних систем на основі врахування показників активності користувачів системи, розробити програмне забезпечення для тестування даного методу, провести експерименти на розробленому програмному забезпеченні для перевірки ефективності застосування розробленого методу, визначення якості його роботи та порівняння даного методу зі стандартним методом колаборативної фільтрації. Використовуваними **методами** є: теорія графів, математична статистика, теорія алгоритмів, об'єктно-орієнтоване програмування. Отримані такі **результати:** розроблено експертно-орієнтований метод побудови рекомендаційних систем на основі врахування показників активності користувачів та розрахунку коефіцієнтів експертів, розроблено гібрид даного методу з методом колаборативної фільтрації, розроблено програмне забезпечення для реалізації даного методу та гібриду, проведено експерименти на розробленому програмному забезпеченні для тестування розробленого методу та гібриду. **Висновки.** Досліджено можливість використання інформації про активність користувачів у рекомендаційних системах для покращення якості списків рекомендацій. Запропоновано розрахунок коефіцієнтів експертів для доповнення до коефіцієнтів подоби у рекомендаційних системах. Розроблено експертно-орієнтований метод побудови рекомендаційних систем на основі врахування показників активності користувачів та його гібрид з колаборативною фільтрацією. Проведені на розробленому програмному забезпеченні експерименти показали, що розроблений метод суттєво підвищує такі показники роботи рекомендаційної системи, як покриття простору користувачів, покриття каталогу об'єктів, та дозволяє створювати більш якісні списки рекомендацій без суттєвих коливань точності та повноти роботи рекомендаційних систем, а в деяких випадках навіть підвищує ці показники, що залежить від особливостей вхідних даних.

**Ключові слова:** рекомендаційні системи; колаборативна фільтрація; коефіцієнти подоби; експертно-орієнтована фільтрація; коефіцієнти експертів.

### Метод генерации списков рекомендаций с учетом показателей активности пользователей рекомендательной системы

Е. В. Мелешко

**Предметом** изучения в статье является процесс генерации списков рекомендаций для пользователей рекомендательной системы. **Целью** является разработка нового метода построения рекомендательных систем для повышения качества списков рекомендаций, повышения покрытия пространства пользователей и покрытия каталога объектов с помощью учета информации о показателях активности пользователей; разработка гибрида данного метода с методом колаборативной фильтрации. **Задачи:** разработать метод построения рекомендательных систем, основанный на учете показателей активности пользователей, разработать программное обеспечение для тестирования данного метода, провести эксперименты на разработанном программном обеспечении для проверки эффективности применения разработанного метода, проверить качество его работы и сравнить данный метод со стандартным методом колаборативной фильтрации. Используемыми **методами** являются: теория графов, математическая статистика, теория алгоритмов, объектно-ориентированное программирование. Получены следующие **результаты:** разработан экспертно-ориентированный метод построения рекомендательных систем на основе учета показателей активности пользователей и расчета коэффициентов экспертов, разработан гибридный метод с методом колаборативной фильтрации, разработано программное обеспечение для реализации данного метода и гибрида, проведены эксперименты на разработанном программном обеспечении для тестирования разработанного метода и гибрида. **Выводы.** Исследована возможность использования информации об активности пользователей в рекомендательных системах для улучшения качества списков рекомендаций. Предложен расчет коэффициентов экспертов для дополнения к коэффициентам подоби в рекомендательных системах. Разработан экспертно-ориентированный метод построения рекомендательных систем на основе учета показателей активности пользователей и его гибрида с колаборативной фильтрацией. Проведенные на разработанном программном обеспечении эксперименты показали, что разработанный метод существенно повышает такие показатели работы рекомендательной системы, как покрытие пространства пользователей, покрытие каталога объектов, и позволяет создавать более качественные списки рекомендаций без существенных колебаний точности и полноты работы рекомендательных систем, а в некоторых случаях даже повышает эти показатели, что зависит от особенностей входных данных.

**Ключевые слова:** рекомендательные системы; колаборативная фильтрация; коэффициенты подоби; экспертно-ориентированная фильтрация; коэффициенты экспертов.