

THE METHOD OF CONSTRUCTING RECOMMENDATIONS ONLINE ON THE TEMPORAL DYNAMICS OF USER INTERESTS USING MULTILAYER GRAPH

Serhii Chalyi

*Department of Information Control Systems
Kharkiv National University of Radio Electronics
16 Nauka ave., Kharkiv, Ukraine, 61166
serhii.chalyi@nure.ua*

Inna Pribylnova

*Department of Information Control Systems
Kharkiv National University of Radio Electronics
16 Nauka ave., Kharkiv, Ukraine, 61166
inna.butsukina@nure.ua*

Abstract

The problem of the online construction of a rating list of objects in the recommender system is considered. A method for constructing recommendations online using the presentation of input data in the form of a multi-layer graph based on changes in user interests over time is proposed. The method is used for constructing recommendations in a situation with implicit feedback from the user. Input data are represented by a sequence of user choice records with a time stamp for each choice. The method includes the phases of pre-filtering of data and building recommendations by collaborative filtering of selected data. At pre-filtering of the input data, the subset of data is split into a sequence of fixed-length non-overlapping time intervals. Users with similar interests and records with objects of interest to these users are selected on a finite continuous subset of time intervals. In the second phase, the pre-filtered subset of data is used, which allows reducing the computational costs of generating recommendations. The method allows increasing the efficiency of building a rating list offered to the target user by taking into account changes in the interests of the user over time.

Keywords: recommender system, collaborative filtering, multi-layer graph, online recommendations, content personalization, area under the error curve.

DOI: 10.21303/2461-4262.2019.00894

© Serhii Chalyi, Inna Pribylnova

1. Introduction

Recommender systems personalize content provided to users. They take into account the limited knowledge of users and simplify their choice in a complex information environment [1]. Such systems are mostly used in the field of e-commerce. In this field, they form a rating list of goods and services, taking into account the projected interests of users. This list contains a limited set of items, ordered by decreasing of expected utility to the consumer [2].

Such systems make a forecast and construction recommendations using the available information about the interaction of two categories of entities: users and objects [3]. Users provide input information for making recommendations. Objects are products, services, sites, etc., that are interesting for the user.

Construction of recommendations for users is carried out both offline and online [4]. In the first case, the stored data about the selection, purchases, user ratings are used. Offline calculations are performed in batch mode, which allows the use of large amounts of data and resource-intensive models and algorithms. However, the results of these calculations may be outdated by the time recommendations are made. Therefore, in addition, calculations are performed online. Such calculations make it possible to take into account recent changes in the interests of users and, thus, it is better to satisfy their requests. Online calculations are performed in real time. Therefore, in this mode, it is necessary to take into account limitations on the amount of data and the computational complexity of the used methods.

Thus, the problem of construction online recommendations in recommendation systems with regard to changes in user priorities over time is relevant.

Existing approaches to the construction of recommendations use information about the interaction of users and objects. Such interaction occurs, for example, when buying goods online, grading these goods, when surfing pages with a description of goods and services. These approaches are divided into two groups: building recommendations based on the similarity of objects, as well as taking into account the similarity of consumers [5].

In the first case, the selection of objects is carried out on the basis of a comparison of their characteristics with those of the objects selected by the current user of the recommender system [6].

In the second case, a collaborative filtering algorithm is used. The main idea of collaborative filtering is to identify hidden factors that reflect the relationship between users and objects. These factors are used to find consumers with similar requests. Recommendations are based on information about the known choice of such consumers. As input data, collaborative filtering uses the information on the selection of users with similar characteristics. To reduce the amount of processed data, it is advisable to take into account the temporal dynamics of users, changes in their interests and behavior over time.

A number of works on collaborative filtering take into account the temporal characteristics of users. In [7], the effect of seasonal changes on the choice of users was shown and a model of the time-varying factor was proposed. The model uses training based on the gradient descent algorithm. Session temporal graph (STG) as a model of long-term and short-term user preferences in the time proposed in the work [8]. To search for recommendations in this work, a personalized random search is used. The work [9] is devoted to generating recommendations based on the use of partial data filtering to reflect the dynamics of user preferences. This filtering is designed to build short-term forecasts and is based on the principles of self-study. The work [10] considers the situation of multiple user behavior when purchasing goods and services (purchases, clicks, browsing). This behavior is modeled using a neural network. The idea of evaluating the similarity of objects with regard to the time of formation of the ratings of these objects was implemented in [11]. This approach is based on the calculation of the proximity of objects and can be implemented only if there is explicit feedback from the user. In general, the models considered are not oriented towards online use, since they require resource-intensive computing. To overcome this drawback, it is advisable to reduce the amount of input data for building recommendations. To do this, it is possible to take into account the temporal aspect and select a subset of data taking into account the current interests of a group of similar users. Then, recommendations are formed on the basis of the obtained data subset using existing approaches, in particular, collaborative filtering.

The aim of research is development of a method for constructing online recommendations based on the typical behavior of users in order to use both popular objects and items of personal interest in the resulting list.

2. Method of construction recommendations online taking into account the dynamics of user interests

The developed method is based on the idea that from the sequence of records about the behavior of a certain system can be selected subsets of data corresponding to the processes in its subsystems [12], as well as reflecting the typical behavior of users of this system. These subsets contain behavioral patterns in the form of, for example, temporal dependencies [13], which reflect the user's implicit knowledge of the system and its elements [14].

This suggests the feasibility of selecting data corresponding to implicit temporal dependencies. Thus, to build recommendations in the online mode, it is advisable to perform a preliminary selection of a data subset, and then collaborative filtering of the resulting subset.

In the first phase of the pre-selection, it is necessary to find a subset of the source records about the user's choice (in the form of, for example, sales information or user ratings) that best fits its preferences given the changes in these preferences over time.

In the second phase, traditional collaborative filtering is performed for a selected subset of the relevant source records.

As the input data of the method the sales statistics of access to the site's pages as well as information about the interests (already known choice) of the target user are used.

Formally, the set of input data can be represented as a multi-layered bipartite graph, reflecting a change in interests (purchases, clicks) of users on a sequence of intervals

$$T = \langle T^1, T^2, \dots, T^k, \dots, T^K \rangle.$$

Each of these intervals is given by the initial τ_{start}^k and final τ_{fin}^k time stamps:

$$T^k = [\tau_{\text{start}}^k, \tau_{\text{fin}}^k], \quad k = \overline{1, K}.$$

Each k – graph layer describes the selection of objects by users on a priori specified interval T^k . For example, in each layer purchases can be represented on the site in one hour, day, week.

Bipartite graph k – layer contains a subset of vertices

$$U^k = \{U_i^k\}, \quad \bigcup_k U^k = U,$$

which correspond to users and a subset of vertices

$$I^k = \{I_j^k\}, \quad \bigcup_k I^k = I,$$

corresponding to objects. The set of the edge of the graph E^k , in each layer sets the connection between the vertices of users and objects in the time interval T^k .

$$E^k = \{E_{ij}^k : \forall k \exists T^k\}, E^k \subseteq U^k \times I^k, \bigcup_k E^k = E. \quad (1)$$

The edges of the graph E_{ij}^k are determined only for the case of selecting or evaluating the subject I_j^k by user U_i^k . The semantics of this relationship is different for explicit and implicit feedback from the user.

In the case of explicit feedback from the user, each edge E_{ij}^k corresponds to the user's rating and have a weight equal to the rating of the subject R_{ij}^k .

In the case of implicit feedback, the weight of the edge can correspond to the number Q_{ij}^k of items purchased, the number of clicks on the page with the description of the subject, the number of transitions to the page, etc.

Later, for the case of implicit feedback, we will use the number of items purchased by the user as weight, since this information is always available in these e-commerce systems. In the case, if there is additional information about clicks and transitions to the page is exist, then a weighted sum of the values of these indicators can be used as a weight.

In the following, a weighted graph is considered for the case of implicit feedback

$$E^k = \{Q_{ij}^k\}.$$

The construction of recommendations in case of explicit feedback is performed using the same sequence of steps.

Elements P_j^{kl} of the set P relations between layers k and l set through connections between identical items I_j^k and I_j^l , belonging to these layers. This choice of relationships is due to the fact that users can select objects only at certain points in time. But objects are often used by users. For example, a consumer buys a computer mouse every few months or years. But these mice are sold to various users regularly.

A necessary condition for having relations P_j^{kl} is the existence of relations E_{ij}^k and E_{im}^l on the respective layers. In other words, objects I_j^k and I_j^l must be selected by i and m by users during time intervals T^k and T^l :

$$P = \{P^{kl}\}, \quad P^{kl} \subseteq I^k \times I^l, \quad P^{kl} = \{P_j^{kl} : \forall k, j \exists E_{ij}^k, \forall l, j \exists E_{im}^l\}. \quad (2)$$

Multi-layer graph G , describing the input data in accordance with (1) and (2), includes all the vertices and edge considered: $G = \{U, I, E, P\}$.

Recommendations for the current user U_α^K are made on the last known interval

$$T^K = [\tau_{\text{start}}^K, \tau_{\text{fin}}^K]$$

or after its completion. Then on the current layer K it is advisable to take into account only those users who have chosen (bought, viewed) the same products as the target user U_α^K . In other words, the condition $Q_{\alpha j}^K \geq 1$ must be met. And the number of choices of each object should be close to the user's U_α^K purchases, i. e.

$$|Q_{ij}^K - Q_{\alpha j}^K| \leq 1.$$

Accordingly, at the last K – layer it is needs to form a subgraph G_α^K , which has the following form:

$$G_\alpha^K = \{U_i^K, I_j^K, E^K : \forall i \forall j \forall Q_{\alpha j}^K \geq 1 |Q_{ij}^K - Q_{\alpha j}^K| \leq 1\}. \quad (3)$$

Filtering by layer according to (3) allows to select users with similar interests.

The change in user interests over time is calculated based on the dynamics of the popularity of each object, which will be included in the set of data selected in the first phase.

$$G_\alpha = \left\{ G_\alpha^k : \forall k \forall j \sum_{\substack{i: \\ U_i^K \in G_\alpha^K}} (Q_{ij}^{k-1} - Q_{ij}^k) \geq -1 \left| \sum_{\substack{i: \\ U_i^K \in G_\alpha^K}} (Q_{ij}^k) > 1 \right. \right\}. \quad (4)$$

The method includes the following steps.

Stage 1. Splitting the source data set D on an ordered sequence of disjoint subsets D^k . Each item D_s^k contains a timestamp τ_s^k from the interval T^k :

$$D = \langle D^k = \{D_s^k\} : \forall k \tau_s^k \in T^k, k = \overline{1, K} \rangle. \quad (5)$$

The result of this phase may be, for example, a subset of sales records for each day (for an hour, for a week).

Stage 2. Selection of data from a subset D^K , for which condition (3) is satisfied. This stage is based on such assumptions. First, the request for recommendations from the target α – user comes at the last known time interval T^K . Secondly, the subset D^K contains the most up-to-date data appearing in the recommendation request interval, which is most important to online recommendations.

The result of this phase is a subset of records from D^K , which contains information about the choice of users, similar to the choice of user U_α^K . The similarity is that users U_i^K select the same or 1 more object instances I_j^K , as the target user U_α^K .

Stage 3. Selection of data subsets D^{k*} from D^k , for which the condition is fulfilled:

$$D^* = \langle D^{k*} = \{D_s^{k*}\} : \forall k \forall i \exists U_i^K \in G_\alpha^K \rangle. \quad (6)$$

The result of this step is a subset of the input data D^* , that contains records about the selection of users who are similar to the α – user.

Stage 4. The selection of such data from a subset D^* , that would satisfy the conditions of the popularity of objects in time (4).

The result of this stage is a subset of the input data D^{**} , that contains records about the selection of objects by similar users. These objects are popular for several consecutive time intervals T^k , including the interval T^K .

Stage 5. Formation of recommendations using collaborative filtering based on a subset of the source data D^{**} .

The result of this stage is an ordered set of recommended objects, which takes into account the interests of the target user, as well as the most popular objects for similar users at present.

3. Experiment

The purpose of the experiment is testing the possibility of reducing the amount of input data for collaborative filtering using matrix factorization while maintaining a comparable accuracy of recommendations. The computational complexity of this method corresponds to a quadratic dependence, which indicates the importance of reducing the amount of input data for building online recommendations.

To assess the accuracy of the recommendations metric AUC (Area Under the Curve) was used. This metric determines the area under the error curve (Receiver Operating Characteristic, ROC). The higher the AUC value, the higher the accuracy of the recommendations.

The Online Retail data set from the UCI repository was used for the experiment. This set contains a list of transactions from the British chain of stores, which sells various gifts. The company works with wholesale customers, so one customer can buy a set of different goods. Input data has the following structure: invoice number; product code; product name; the number of units of goods per transaction; invoice date; unit price; client code; a country. Therefore, the input data are implicitly dependent data. The data contains temporal labels, which allows using the method. The complete data set contains 541910 records. Code of the target user and start interval T^k was chosen randomly. Duration of intervals T^k was 1 day. The test was performed on several samples of input data with duration of 2 to 6 days. This duration depended on the selected minimum quantity of the purchased item.

4. Results

The results of the evaluation recommendations are in the **Table 1**. The resultant sets that are obtained after filtering the original sample include several thousand lines. However, it is two orders of magnitude smaller than the size of the original sample.

Table 1

The results of the evaluation of recommendations taking into account the dynamics of user interests

Number of rows after filtering at stages 1–4, thousand records	AUC
3,7	0,637
8,2	0,712
11,6	0,733

In order to compare the accuracy of recommendations online, 3 samples of the same size were selected from the source data and the recommendations obtained as a result of collaborative filtering based on matrix factorization were evaluated. The results of this assessment are in the **Table 2**.

Table 2

The results of the comparative evaluation of recommendations

Number of lines, thousand records	AUC
4	0,61
8	0,68
12	0,705

From the presented data it is clear that in the case of taking into account the dynamics of the interests of users, the ratings of the received recommendations increase by approximately 5 %.

At the same time, to achieve the minimum acceptable accuracy of online recommendations, the size of the received sample should exceed 10 thousand records. To obtain higher AUC values, it is necessary to use additional information about objects.

The considered method allows reducing computational costs due to their quadratic dependence on the amount of input data.

5. The discussion of the results

The result of the publication is the method of construction recommendations in the online mode, taking into account changes in the interests of users over time.

The resulting sequence of recommended objects for the user makes it easier to make decisions about the choice of objects by the user if he does not have a full set of knowledge about the characteristics of these objects.

The difference of the proposed method is in the preliminary filtering of the input data taking into account the choice of similar users, as well as taking into account the time interval during which the interest of these users in choosing a subset of objects does not change significantly.

The advantage of the method is that it allows to flexibly reduce the amount of input data about a known choice of users and, thereby, reduce the computational cost of constructing recommendations, which is important for working online.

The disadvantage of the method is that the size of the pre-filtered sample and, as a result, the evaluation of the obtained solutions largely depends on the characteristics of the source data.

The developed method is for use online in the recommendation system and complements the offline approaches to building recommendations.

6. Conclusions

The problem of the formation of rating lists of objects online for users of recommendation systems is considered. The expediency of building online recommendations is justified by the fact that offline approaches do not always take into account the latest actual data on the choice of users. The construction of such recommendations is executed in real time, so before applying traditional approaches it is advisable to reduce the amount of input data by removing outdated information.

A method for building recommendations online is proposed, taking into account the dynamics of changes in user interests. The method includes the steps of selecting users with similar interests, as well as filtering input data based on changes in priorities of these users and the subsequent constructing of recommendations based on collaborative filtering.

The proposed method has the following differences from existing hybrid methods for construction recommendations.

First, the pre-filtering of the input data is performed, which allows reducing the dimension of the task both by the characteristics of the user choice and by the time aspect of this choice, reflecting the change of users' interests in time.

Secondly, the method allows combining personalization for the target user with a choice of the most popular recommendations, which is essential in the case of a cold start, when the recommendation system contains a small amount of information about the target user.

In practical terms, the developed method allows increasing the efficiency of the formation of recommendations in the online mode, taking into account the latest relevant data on user choice.

With the further development of the method in order to improve the accuracy of recommendations, it is supposed to take into account the properties of objects and changes in the users' need for these properties over time.

References

- [1] Ricci, F., Rokach, L., Shapira, B. (Eds.) (2015). Recommender Systems. Springer, 1008. doi: <https://doi.org/10.1007/978-1-4899-7637-6>
- [2] Amatriain, X. (2013). Mining large streams of user data for personalized recommendations. ACM SIGKDD Explorations Newsletter, 14 (2), 37–48. doi: <https://doi.org/10.1145/2481244.2481250>

- [3] Aggarwal, C. (2017). Recommender Systems. Springer, 498. doi: <https://doi.org/10.1007/978-3-319-29659-3>
- [4] Amatriain, X. (2013). Big & personal. Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining Algorithms, Systems, Programming Models and Applications – BigMine '13. doi: <https://doi.org/10.1145/2501221.2501222>
- [5] Jones, M. T. (2013). Recommender systems, Part 1. Introduction to approaches and algorithms. Available at: https://www.ibm.com/developerworks/opensource/library/os-recommender1/index.html?S_TACT=105AGX99&S_CMP=CP
- [6] Schafer, J. B., Konstan, J., Riedi, J. (1999). Recommender systems in e-commerce. Proceedings of the 1st ACM Conference on Electronic Commerce – EC '99. doi: <https://doi.org/10.1145/336992.337035>
- [7] Koren, Y. (2009). Collaborative filtering with temporal dynamics. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '09. doi: <https://doi.org/10.1145/1557019.1557072>
- [8] Xiang, L., Yuan, Q., Zhao, S., Chen, L., Zhang, X., Yang, Q., Sun, J. (2010). Temporal recommendation on graphs via long- and short-term preference fusion. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '10. doi: <https://doi.org/10.1145/1835804.1835896>
- [9] Luo, C., Cai, X., Chowdhury, N. (2014). Self-training Temporal Dynamic Collaborative Filtering. Lecture Notes in Computer Science, 461–472. doi: https://doi.org/10.1007/978-3-319-06608-0_38
- [10] Gao, C., He, X., Gan, D., Chen, X., Feng, F., Li, Y., Chua, T., Jin, D. (2018). Learning Recommender Systems from Multi-Behavior Data. IEEE Transactions on knowledge and data engineering, 1–12.
- [11] Bakir, C. (2018). Collaborative Filtering with Temporal Dynamics with Using Singular Value Decomposition. Technical gazette, 25 (1), 130–135. doi: <https://doi.org/10.17559/tv-20160708140839>
- [12] Chalyi, S., Levykin, I., Petrychenko, A., Bogatov, I. (2018). Causality-based model checking in business process management tasks. 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). doi: <https://doi.org/10.1109/dessert.2018.8409176>
- [13] Kalynychenko, O., Chalyi, S., Bodyanskiy, Y., Golian, V., Golian, N. (2013). Implementation of search mechanism for implicit dependences in process mining. 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS). doi: <https://doi.org/10.1109/idaacs.2013.6662657>
- [14] Levykin, V., Chala, O. (2018). Method of automated construction and expansion of the knowledge base of the business process management system. EUREKA: Physics and Engineering, 4, 29–35. doi: <https://doi.org/10.21303/2461-4262.2018.00676>