2014 X International Symposium on Telecommunications (BIHTEL)
October 27-29, 2014, Sarajevo, Bosnia and Herzegovina

An approach to design of time-aware recommender system based on changes in group user's preferences

Bakir Karahodza, Haris Supic, Dzenana Donko
University in Sarajevo, Faculty of Electrical Engineering
Bosnia and Herzegovina
bakir.k@gmail.com, haris.supic@etf.unsa.ba, dzenana.donko@etf.unsa.ba

Abstract—**Traditional recommender systems use collaborative filtering or content-based methods to recommend new items for users. New users and items are continuously updated to the system bringing changes in user's preferences, as well as the additional context in form of temporal information. The continuous system updates change not just individual user's preferences, but also group user's preferences affecting prediction of ratings for individual users. In this work is presented improved user-based collaborative filtering algorithm using temporal contextual information. With difference to other approaches, we propose using weight function based on changes in the group user's preferences over time that increases prediction accuracy of collaborative filtering prediction algorithm.**

Index Terms—**Recommender systems, collaborative filtering, algorithm, temporal information, user preferences.**

I. INTRODUCTION

The general task of recommender system is to identify items for each user that follows some of the goals such as maximizing user's utility or optimizing the lifetime of user's utility. The recommendation problem is formalized in [1] assuming utility function r that measures usefulness of item i ε I for user u ε U as $U x I \rightarrow R$, where R is a totally ordered set of numbers within a certain range. Consequently, the task of recommendation is defined as finding item i_j that maximizes a given user's utility. In general terms, the recommendation problem relies on the notion of rating as a mechanism to capture user preferences for different items.

The recommendation systems differentiate themselves by the type of knowledge they exploit, e.g. product knowledge or some domain expert's heuristics. The usual strategies in recommender systems are [1]:

- *Content-based (CB) recommendations*, which recommends items similar to those preferred by the user in the past,
- *Collaborative filtering (CF)*, which recommends items preferred in the past by similar users, using approaches with item-based or user-based filtering, and
- *Hybrid recommenders*, that combine both CB and CF in order to overcome particular limitations of each individual strategy.

All traditional recommender systems exploit only user and item profiles associated to past ratings in order to predict ratings of unseen items [1], and they do not take any contextual information into considerations. In extension of the rating notion, Context-Aware Recommender Systems (CARS) [2] additionally take into consideration contextual information (e.g. time, location, and weather) associated to previously collected user preferences. In this way, CARS can differentiate the interest a user may have in a particular item within different contexts.

Among the existing contextual dimensions, the time context can be considered as the most variable one. With difference to other contextual information that need to be collected, identified and utilized using a certain efforts, collecting of time information does not need any additional user or device resources, neither impose additional system requirements. Moreover, the time context information has been used as a key input for achieving significant improvement on recommendation accuracy [3]. These facts has encourage the research and development of Time-Aware Recommender Systems, that are actually CARS that exploit the time information in recommendation strategies.

The main motivation for new recommendation strategies that take into consideration temporal rating dimension is that user preferences are naturally changing over time. However, in many recommender systems with collaborative filtering methods, the data collection is regarded as static. Although a great effort has been conducted in development of new algorithms and improvements of CF methods, many challenged are still present, mostly related with the expected accuracy of CF recommender systems. Apart from this issue, the problem with the changing data in collaborative filtering over time is not just about the change of individual user's preferences. There is also presented a change of preferences within a group of users, or at least change of preferences within similar users that influences the predicted ratings in CF algorithms.

In this paper we propose an improved approach to design of user-based collaborative filtering method, using temporal information for weighting neighbor's ratings.

The rest of the paper is organized as follows: Section 2 shortly describes some of the research work related to our paper. In Section 3, we propose improved approach for user-based time-aware collaborative filtering recommender system

978-1-4799-4136-0/14/$31.00 ©2014 IEEE

and describe its steps in details. Section 4 presents results of our experimental work. It describes details of our approach, data sets used, evaluation metrics, and results of different experiments with discussion. The last section presents conclusion remarks and direction for future work.

II. RELATED WORK

Previous work that exploit the advantage of inclusion of temporal context information in recommendation systems are presented in [4], which takes continual variable to represent fluctuations in user preferences over time, and in [5] that identifies repetitive patterns through time using categorical variable (periodic, discrete information), or using both representations [3]. Beside these approaches of including temporal information, there are recommender systems that include time-aware contextual information without determining rating prediction based on the future recommending time. They rather dynamically set or adopt some of the parameters according to changes of characteristic data over time.

The usual approaches for adoption of temporal effect in CF algorithms generally boost recent ratings and penalize older ratings that are presumed less relevance at recommendation time. This can be accomplished using a series of different techniques, mostly based on discrete time windows [6] or continuous decay function [4], [8]. An example of recommender system that penalize older preferences considered as less valuable at the recommendation time can be found in [7], where implicit purchase information is transformed into ratings by assigning increased weight to more recent purchases. This is presented as a time-dependent function w_t' that assigns a weight to each rating according to available rating timestamp:

$$F(u,i) = \underset{v \in N(u)}{aggr}\, r_{v,i} \times w_t'\big(\tau(r_{v,i})\big) \quad (1)$$

Another formulation for time weight function for item-based CF algorithm is presented in [8] as a monotonic decreasing time function $f(t) = e^{\alpha\tau}$ in the recommendation step, giving more weight to recently rated items. In function f, τ is the time at ratings was given and α is parameter that controls the decay date relative to time.

Other authors propose a similar approach that derives time series of user ratings, aiming to establish current user interests for individual item or group of items. In order to build time series, items are grouped according to item category in [9], leading to several time series for each user's category, or grouping all ratings using an interest measure that takes into account item similarity [10], leading to a unique time series for each user. In [11], users' rating history has been divided into several periods and then quantized every user's interest, with setting a time-window to find user's recent interest. In [12] is presented a time-driven filtering strategy that link items and their features to time functions corrected by temporal curves built from individual user consumption histories. However, all related work is based on item-based CF algorithm and follows changes in individual user preferences without taking into account changes in the group to which the user belongs.

III. PROPOSED RECOMMENDER SYSTEM

Our new approach is based on a user-based collaborative filtering method and gives an improvement of the prediction accuracy.

The basic idea includes assumption that similar users follow the same shift in item ratings trend, and also that recent ratings are more relevant. For example, Fig. 1 is used to represent all ratings one the item i over time with circles as rating data points. From the entire set of the ratings, standard CF approach detects neighbor's ratings for each individual user, presented with the bolded circles. From the Figure 1, we can see that neighbor's ratings follow a certain trend over time, until the most recent rating at the *max(t)* time.

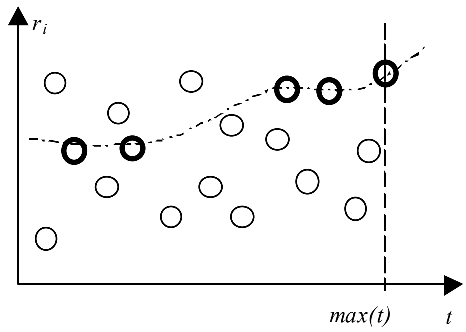

Fig. 1. Trend in neighbor's rating over time.

Based on these assumptions, in this work we propose design of improved user-based collaborative filtering recommender system with high accuracy:

We find set of users as closest neighbors. To exceed the problem of shift in user preferences, we assign a certain weight function to each neighbor's ratings dependant on the time of the ratings. We assume that the most recent rating of all the n similar neighbors is the closest to a future preference of active user a. Consequently, each neighbor's rating is weighted by computing difference between the item rating and the most recent neighbor's rating.

In the following sections, we first describe standard user-based collaborative filtering algorithm, and then we define the proposed improvement.

The prediction function in the traditional user-based collaborative filtering is defined as [13]:

$$p_{a,i} = r_a + \frac{\sum_{u=1}^{n} (r_{u,i} - r_u) \times P_{a,u}}{\sum_{u=1}^{n} P_{a,u}}$$

where $p_{a,i}$ represents prediction for active user a for item i; r_a is the mean rating given by user a; $r_{u,i}$ is the rating given to an item i by neighbor's user u; r_u is the neighbor's mean rating;

$P_{a,u}$ is the similarity between users a and u; and n is the number of neighbors.

The similarity between users is computed using the Pearson correlation coefficient that is defined as [13]:

$$P_{a,u} = \frac{\sum_{i=1}^{m}(r_{a,i} - r_a)\times(r_{u,i} - r_u)}{\sqrt{\sum_{i=1}^{m}(r_{a,i} - r_a)^2 \times \sum_{i=1}^{m}(r_{u,i} - r_u)^2}} \quad (3)$$

where $r_{a,i}$ is the rating given to an item i by user a; r_a is the mean rating given by user a; m is the total number of items.

As previously stated, traditional approaches for user-based collaborative filtering use ratings given by the most similar neighbors to predict user preferences for specific item. However, in a real situation, user preferences are sensitive on rating context in terms of changes over time. To determine the current trend and get more accurate predictions, we propose use of time-dependant weight function to each rating in the preference prediction step as the function of neighbor's rating times. In that way, Equation 3 is modified as:

$$p_{a,i} = r_a + \frac{\sum_{u=1}^{n}(r_{u,i} - r_u)\times P_{a,u}\times W(t(r_{u,i}))}{\sum_{u=1}^{n} P_{a,u}} \quad (4)$$

where $W(t(r_{u,i}))$ is a time-dependant weight function.

According to previous assumptions, as the base for time-dependant weight function is used difference between time of neighbor's rating on item i and the newest time in all neighbor's ratings:

$$W(t(r_{u,i})) = e^{-\frac{1}{\lambda}\times(\max(t(r_{u,i})) - t(r_{u,i}))} \quad (5)$$

where λ is a parameter for precise accuracy tuning: $\max(t(r_{u,i}))$ represents the time of the newest time in all neighbor's ratings; $t(r_{u,i})$ represents the time of u-th neighbor's rating on the item i.

As in [8], where weight function is a monotonic time decreasing function for item-based CF, we can observe from Equation 5 that the value of weight function is in the range of [0,1]. Since the weight is exposed as the decreasing time function, the more the rating's time of an item deviates from the most recent time, the lower is the weight of the neighbor's rating.

IV. Experiments and results

We've performed a set of experiments in order to examine the performance of our proposed approach. In particular, we refer to the issue of how the change in preferences of similar users in time-dependant context influences the prediction accuracy for specific user, and how is our proposed approach compared to usual collaborative filtering method.

As previously discussed, recommender systems are characterized by changes of concept in real situations and this fact degrades the prediction accuracy. To overcome the problem, we assume that similar users follow the same rating trend over time, and that older neighbor's ratings are less relevant.

In our experiment, we compare accuracy of the proposed method with user-based method without weight function.

A. Experiment Design

In our experiments, we use two data sets: MovieLens 100k[1] and MovieLens + IMDb/Rotten Tomatoes (HetRec2011[2]) which broadly used data sets in collaborative filtering research projects. MovieLens 100k data consists of 100,000 ratings from 1,000 users for 1,700 movies. HetRec2011 consists of 855,598 ratings from 2,113 users for 10,197 movies.

We've used the usual approach for experiments with a number of the ratings as training data and built correlations between users using these training transactions. The remaining of transactions were used as test data in the prediction step to demonstrate effectiveness of our proposed system.

We've also rearranged data sets and created secondary versions where the newest ratings are used for test data. In each experiment, the number of the maximum neighbors used has been set to 30 according to recommendations in [14]. The minimum number of the neighbors has been set to 3 to ensure acquisition of accurate rating trend over time within neighbor's ratings.

The evaluation metrics used in experiments is Mean Absolute Error (MAE). MAE is defined as the average absolute difference between predicted ratings and their true user-specified values [12]:

$$MAE = \frac{\sum_{i=1}^{N}|p_i - q_i|}{N} \quad (6)$$

where p_i represents the predicted rating for i-th item; q_i represents the user-specified rating value; N represents the number of user's ratings.

B. Results of experiments

The experiments have been conducted on both MovieLens and HetRec2011 data sets with their corresponding generated subsets. The HetRec2011 data set has been splitted into several smaller parts to ensure similarity in size with MovieLens. The final data sets have been adjusted to time-order ratings in order to use the newest ratings for test data.

We've tested different values for parameter α for each data set in order to get the most accurate prediction. We've found that the value of $\alpha = 2/\Delta T$ provides the best results, where ΔT presents the total time rating period in the training set. For example, if the total rating period is 300 days, then $\alpha = 1/150$.

The results are presented in Table 1. Our new proposed method is able to outperform pure user-based CF algorithm in prediction accuracy with different success on used data sets. The most accurate prediction values, where the lowest MAE

[1] http://grouplens.org/datasets/movielens/
[2] http://grouplens.org/datasets/hetrec-2011/

value presents the lowest error, are recorded using MovieLens 100k data set. On the other side, experiments with HetRec2011 data sets have shown different success. We've found that data subsets with a long rating period are not good candidates for proposed algorithm. We believe that the results are affected by too many frequent changes in the neighbor's rating trend over time, which depreciates the impact of the weighting function in maximizing user's utility. The subsets *s1* and *s2* were also affected by a lack of ratings needed to establish similar neighbors, and the ratings were predicted for only a few users.

TABLE I. RESULTS

Data set	α	MAE	
		Pure CF	***Proposed***
MovieLens 100k (entire)	83	1.0428	0.9866
HetRec2011, subset *s1*	50	0.8866	0.8679
HetRec2011, subset *s2*	110	0.9058	0.8944
HetRec2011, subset *s3*	1800	0.8321	0.8261

On the other side, data subsets with a shorter rating period (lowest α) shows better results since the rating trend within a group of similar neighbors is more stable. This is expected since the user's opinion has tendency to change less over a short period of time

V. CONCLUSIONS AND FUTURE WORK

In this paper we've presented an improved user-based CF algorithm using time-aware context information. Unlike the traditional user-based CF algorithms, we propose using weight function based on the changes in time contextual information. The main contributions of our research are in design and testing a more appropriate prediction function including time context information that overcome the problem of changes in group preferences.

The results have shown that our new approach can improve prediction accuracy of user-based CF algorithm. The future challenge for our work is to find rating period where the time-dependant weighting function shows best results based on the rating domain. There is also the challenge to deal with changes of group user preferences within an item category over time and prediction of their rating trend.

REFERENCES

[1] Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A. Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach. ACM Transactions on Information Systems 23(1):103-145. 2005.

[2] Adomavicius, G., Tuzhilin, A. Context-Aware Recommender Systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., Recommender Systems Handbook, pp. 217-253, Springer US. 2011.

[3] Koren, Y., Bell, R., Volinsky, C. Matrix Factorization Techniques for Recommender Systems. Computer 42(8):30-37. 2009.

[4] Ding Y. and Li X. Time weight collaborative Filtering.In CIKM 2005, 485-492, Bremen, Germany, 2005.

[5] Baltrunas L. and Amatriain X. Towards time-dependant recommendation based on implicit feedback. In CARS 2009, New York, NY, USA, 2009.

[6] Nasraoui, O., Cerwinske, J., Rojas, C., Gonzalez, F.A.: Performance of recommendation systems in dynamic streaming environments. In: Proceedings of the SeventhSIAM International Conference on Data Mining, Minneapolis, Minnesota, USA. SIAM, 2007.

[7] Lee, T.Q., Park, Y., Park, Y. A, Time-based Approach to Effective Recommender Systems Using Implicit Feedback. Expert Systems with Applications 34(4):3055-3062. 2008.

[8] Ding Y., Li X., Orlowska, M.E. Recency-based Collaborative Filtering. In: Proceedings of the 17th Australasian Database Conference - Volume 49 (ADC'06), pp. 99-107, Australian Computer Society, Inc., Darlinghurst, Australia, Australia. 2006.

[9] Min, S., Han, I. Detection of the Customer Time-variant Pattern for Improving Recommender Systems. Expert Systems with Applications 28(2):189-199. 2005.

[10] Cao, H., Chen, E., Yang, J., Xiong, H. Enhancing Recommender Systems Under Volatile User Interest Drifts. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM'09), pp. 1257-1266, ACM, New York, NY, USA. 2009.

[11] Yuchuan Zhang, Yuzhao Liu, A Collaborative Filtering Algorithm Based on Time Period Partition, Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on, 777 – 780, 2010.

[12] Y. Blanco, M. López-Nores, J. Pazos-Arias, J. García-Duque, An improvement for semantics-based recommender systems grounded on attaching temporal information to ontologies and user profiles, Engineering Applications of Artificial Intelligence. Volume 24, Issue 8, pages 1385–1397, 2011.

[13] Resnick, P.; Iacovou, N.; Sushak, M.; Bergstrom, P.; and Reidl, J. GroupLens: An open architecture for collaborative filtering of netnews. In Proceedings of the 1994 Computer Supported Cooperative Work Conference. New York: ACM, 1994.

[14] J. Herlocker, J. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative fltering. In SIGIR '99: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 230-237, 1999.