International Journal for Modern Trends in Science and Technology
Volume: 02, Issue No: 10, October 2016
http://www.ijmtst.com
ISSN: 2455-3778

Effective Cross-Domain Collaborative Filtering using Temporal Domain – A Brief Survey

Swapna Joshi[1] | Prof. Manisha Patil[2]

[1,2] Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

To Cite this Article
Swapna Joshi and Prof Manisha Patil, Effective Cross-Domain Collaborative Filtering using Temporal Domain – A Brief Survey , International Journal for Modern Trends in Science and Technology, Vol. 02, Issue 10, 2016, pp. 88-92.

ABSTRACT

Cross-domain collaborative filtering (CDCF) is an evolving research topic in recommender systems. It aims to alleviate the data sparsity problem in individual domains by transferring knowledge among related domains. But it has an issue of user interest drift over time. Along with data sparsity, we should also consider the temporal domains to overcome user interest drift over time problem to predict more accurately as per the current user's interest. This paper surveys few of the pilot studies in this research line and the methods of how to add the temporal domains in the recommender systems. The paper also proposes possible extensions of using temporal domains with different contexts in current timestamp.

KEYWORDS: *Survey, collaborative filtering, temporal domain, cross-domain, recommender syste*

I. Introduction

The essential problem of Collaborative Filtering (CF) methods is how to find similar users/items and how to measure similarities between them. Thus far, most existing CF methods are single-domain based, which make predictions based on one rating matrix. In other words, these methods can only find similar users/items in a single domain. However, in many recommendation scenarios, multiple related CF domains may be presented at the same time and finding similar users/items across domains becomes possible, such that common rating knowledge can be shared among related domains.

Cross-domain collaborative filtering (CDCF) aims to share common rating knowledge across multiple related CF domains to boost the CF performance. CDCF methods exploit knowledge from auxiliary domains (e.g., movies) containing additional user preference data to improve recommendation on a target domain (e.g. books). While relying on a broad scope of existing data in many cases is a key to relieving the problems of sparse user-item data in the target domain, CDCF can also simultaneously benefits different data owners by improving quality of service in different domains. Fig. 1 shows the general architecture of the cross domain collaborative filtering system.

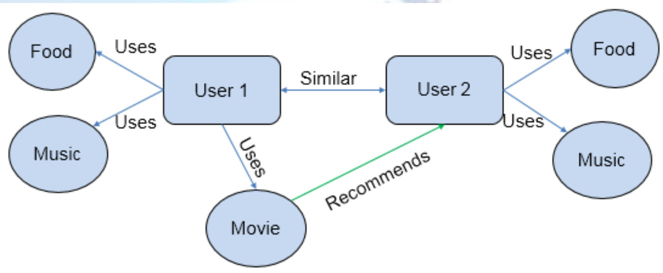

Fig. 1: Cross Domain CF architecture

Major Issues in current CDCF system –

A. User Interest Drift Over Time -

User's interests keep changing over time since it continuously getting affected by multiple factors such as moods, contexts, cultures, festivals and different seasons. An example is a person who does

not like animated movies, may like to watch them in future due to evolving superior 3D animation technologies with very good visual effects or else if the person having baby may like to watch it now which he has never watched before. Another example is a person buying outfit items of a specific type may like to buy the outfits of different style if he/she is relocating to an altogether different geographical area.
Reviewers express their opinions about a particular product or a service (or restaurants, hotels, movies, etc.) through some review sites, social networking sites like facebook, twitter, etc. or even over blogs. All these act as sources for opinions.

B. *Data sparsity –*

In practice, many commercial recommender systems are used to evaluate large item sets (e.g., Amazon.com recommends books and CDnow.com recommends music albums). In these systems, even active users may have purchased well under 1% of the items (1% of 2 million books is 20,000 books). Accordingly, a recommender system based on nearest neighbor algorithms may be unable to make any item recommendations for a particular user. As a result the accuracy of recommendations may be poor.

C. *Scalability –*

Nearest neighbor algorithms require computation that grows with both the number of users and the number of items. With millions of users and items, a typical web-based recommender system running existing algorithms will suffer serious scalability problems.

D. *Synonyms –*

Synonyms refers to the tendency of a number of the same or very similar items to have different names or entries. Most recommender systems are unable to discover this latent association and thus treat these products differently. Topic Modelling (like the Latent Dirichlet Allocation technique) could solve this by grouping different words belonging to the same topic.

E. *Diversity and the Long Tail –*

Collaborative filters are expected to increase diversity because they help us discover new products. Some algorithms, however, may unintentionally do the opposite. Because collaborative filters recommend products based on past sales or ratings, they cannot usually recommend products with limited historical data. This can create a rich-get-richer effect for popular products, akin to positive feedback. This bias toward popularity can prevent what are otherwise better consumer-product matches.

II. MOTIVATION

Many In most CDCF approaches more stress is given on referring the ratings from multiple sites of the related domains. But there is one more problem in recommender systems that the user interests keep drifting over a period of time. Hence the temporal domains should also be considered in the cross domain recommender systems. As users' interests keep changing over time and a user is more likely to be interested in different item categories at different time, we cannot simply view the multiple historical buying patterns of the same user in different historical temporal domains.

III. RELATED WORK

The Cross domain collaborative filtering is an evolving research topic in recommender systems. The survey was cried to study two major issues in current CDCF systems, data sparsity and user interest drift over time.

[1] was studied to understand the basic Cross-domain recommender systems. In cross-domain, we refer multiple auxiliary domains for user ratings/inputs from related domains and then use these ratings to transfer the collective rating matrix to the target domain for recommendations. For example, to recommend a movie of 3D animation genre to the end user, the target domain is 3D animation genre movie. So we consider multiple input auxiliary domains as books, music, and other movie genre similar to the movie subject etc. It used a derived method based on a Bayesian latent factor model which can be inferred using Gibbs sampling. This addresses the challenge of modeling user-interest drift over time by considering the historical time slice patterns of the user.

[2] and [3] were studied for dealing with data sparsity reduction problem. In [3], principled matrix-based transfer learning framework is used that takes into account the data heterogeneity. The principle coordinates of both users and items in the auxiliary data matrices are extracted and transferred to the target domain in order to reduce the effect of data sparsity.

The survey was carried out on [4], [5] and [6] to study different methods of cross domain collaborative filtering with temporal domain.

In [4], the temporal domains are considered. The ratings provided by the same user at different time may reflect different interests as those ratings are provided by different users. The proposed algorithms are a variant of standard neighborhood-based (either user based or item based) CF. The main idea behind the proposed approaches is to enhance inter-domains edges by both discovering new edges and strengthening existing ones. In [5], Factorization model and item-item neighborhood model are used. In both factorization and neighborhood models, the inclusion of temporal dynamics proved very useful in improving quality of predictions, more than various algorithmic enhancements

While working in CDCF recommender system, one critical aspect is how to collect and transfer the knowledge from different input auxiliary domain to the target domain for giving the recommendations. The survey was done on [7], [8] and [9] to understand the methods of transferring the knowledge across the domain. These Transfer learning methods are used to first individually collect the ratings matrix for each auxiliary input domain and then transferring the collective ratings to the target domain.

[10] was studied rigorously which deals with the cross domains over different sites (domains), transferring the rating knowledge from these sites to recommend in target domain by using knowledge transfer method. Along with using multiple sites/domain to collect the user ratings, it also extends the model of [4] for using temporal domain to deal with the issue of user interest drift over time. It uses the principle that user has multiple counterparts across temporal domains and the counterparts in successive temporal domains are different but closely related. A series of time-slices are viewed as related CF domains and a user at current time-slice depends on herself in the previous time-slice. Model is built on a cross-domain CF framework by viewing the counterparts of the same user in successive temporal domains are different but related users. If we can find the unchanged rating patterns (static components) shared across temporal domains, the drifting factors of users (changing components) in each temporal domain can be easily captured. Bayesian generative model to generate and predict ratings for multiple related CF domains on the site-time coordinate system, is used as the basic model for the cross-domain CF framework which is extended for modeling user-interest drifting over time, where a series of time-slices are viewed as related CF domains and a user at current time-slice depends on herself in the previous time-slice.

Table 1 provides an overall evaluation of related work.

Table 1: Overall Evaluation of Related work

Reference Number	Paper Name	Data Sparsity	User Interest drift over time
1	Cross-domain recommender systems	No	Yes
2	Can movies and books collaborate? Cross domain collaborative filtering for sparsity reduction	Yes	No
3	Transfer learning in collaborative filtering for sparsity reduction	Yes	No
4	Cross-domain collaborative filtering over time	No	Yes
5	Collaborative filtering with temporal dynamics	No	Yes
6	A spatio-temporal approach to collaborative filtering	Yes	Yes
8	Transfer learning for collaborative filtering via a rating-matrix generative model	Yes	No
10	Rating Knowledge Sharing in Cross-Domain Collaborative Filtering	Yes	Yes

Table 2: Specific Evaluation of Related work (for user interest drift over time issue)

Paper Name	Using historical time slice data	Using Current time context
Cross-domain recommender systems [1]	Yes	No
Cross-domain collaborative filtering over time [4]	Yes	No
Collaborative filtering with temporal dynamics [5]	Yes	No
A spatio-temporal approach to collaborative filtering [6]	Yes	No
Rating Knowledge Sharing in Cross-Domain Collaborative Filtering [10]	Yes	No
Proposed System	Yes	Yes

IV. PROPOSED WORK

After the rigorous survey, the issue specific to user interest drift over time needs to be handled effectively using current time context. As per the survey of existing work [10], the user interest drift over time problem in Cross domain collaborative filtering techniques is handled by introducing the concept of temporal domain. The historical rating data is time stamped. With the input historic data from the same or different user having similar interests of the past buying patterns, the data set is divided in multiple time slices. Each time slice is considered as one time domain and then the principles of cross domain filtering are applied on these multiple time domains. The knowledge is transferred along successive temporal domains to benefit the user-counterparts in each domain. A user in different temporal domains can be treated as a set of user-counterparts with similar but different interests. Thus, the ratings provided by the same user at different time may reflect different interests as those ratings are provided by different users. In this current work, only the historical time domain is being considered by gathering the historical ratings provided by same or different user having same taste.

The proposed work take into account the current time domain context along with historical data. This can be viewed in multiple contexts/aspects. Contexts can be added based on the season changes, location changes, cultural changes, festival periods ect. By using these current contextual parameters along with the existing work of historical time domain, the user recommendations will be more effective and more current addressing the current context of the user.

The dataset pre-processing will divide the input dataset of each year into 12 different slices having one slice per month as opposed to the 4 slices per year in the base system. By this more granular level rating matrix can be generated which is more aligned with the ever changing user interests. This can provide better ratings recommendations. Also the contextual parameters will be applied on these slices on granular levels so that the season, holidays, locations will be applied more closely to the current matching context.

The block diagram in fig. 2 shows the proposed work. First the dataset will be initialized by acquiring the data from various sites. Dataset pre-processing will create the user * movies matrix in 12 different time slices.

Then the average rating will be calculated by applying the current contextual parameters and comparing the historical ratings as per the time slices. Ten the final ratings recommendations will be presented to the user.

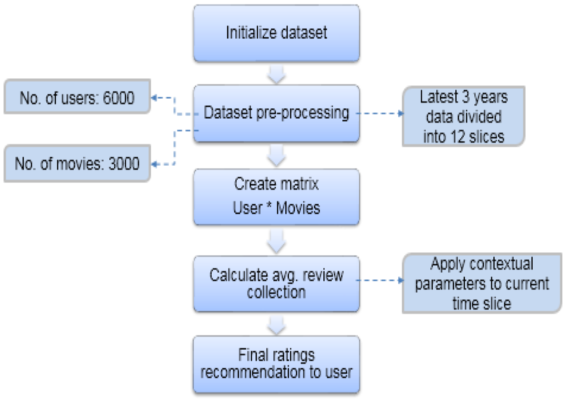

Fig. 2: Proposed work

V. CONCLUSION

We have given a brief survey of cross domain CF over Temporal domain and summarized the related works for different models dealing with the issues mentioned above. Studied papers are using only the previous history of either same user or user groups for recommendation of current time slice. As a part of future work, we can consider the current time slice parameters as well depending on different time aware contexts such as festivals, seasons, regional and cultural changes etc. as per current user's geographic locations. This will enhance the recommendations more in line with the current time domain.

ACKNOWLEDGEMENT

This survey would not have been possible without the support and help of many individuals. I would like to extent my sincere thanks to all of them. I am indebted to Prof. Manisha Patil, Department of Computer Engineering of Smt. Kashibai Navale College of Engineering affiliated to Savitribai Phule Pune University for her guidance and constant supervision as well as all the other staff members for providing important information regarding the survey.

REFERENCES

[1] P. Cremonesi, A. Tripodi, and R. Turrin, "Cross-domain recommender systems," in Proc. IEEE 11th Int. Conf. Data Mining Workshops (ICDMW), Vancouver, BC, Canada, 2011, pp. 496–503.

[2] B. Li, Q. Yang, and X. Xue, "Can movies and books collaborate? Cross domain collaborative filtering for sparsity reduction," in Proc. 21st Int. Joint Conf. Artif. Intell. (IJCAI), 2009, pp. 2052–2057.

[3] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang, "Transfer learning in collaborative filtering for sparsity reduction," in Proc. 24th Conf. Artif.Intell.

(AAAI), 2010, pp. 230–235.
[4] B. Li et al., "Cross-domain collaborative filtering over time," in Proc. 22nd Int. Joint Conf. Artif. Intell. (IJCAI), 2011, pp. 2293–2298.
[5] Y. Koren, "Collaborative filtering with temporal dynamics," in Proc. Int. Conf. Knowl. Discov. Data Mining (KDD), Paris, France, 2009, pp. 447–456.
[6] Z. Lu, D. Agarwal, and I. S. Dhillon, "A spatio-temporal approach to collaborative filtering," in Proc. 3rd ACM Conf. Recommender Syst., New York, NY, USA, Oct. 2009, pp. 13–20.
[7] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in Proc. 25th Int. Conf. Mach. Learn. (ICML), Helsinki, Finland, 2008, pp. 880–887.
[8] B. Li, Q. Yang, and X. Xue, "Transfer learning for collaborative filtering via a rating-matrix generative model," in Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML), 2009, pp. 617–624.
[9] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[10] Bin Li, Xingquan Zhu, Ruijiang Li, and Chengqi Zhang "Rating Knowledge Sharing in Cross-Domain Collaborative Filtering" in IEEE TRANSACTIONS ON CYBERNETICS, VOL. 45, NO. 5, MAY 2015
[11] B. Li, "Cross-domain collaborative filtering: A brief survey," in Proc. 23rd IEEE Int. Conf. Tools Artif. Intell. (ICTAI), Boca Raton, FL, USA, Nov. 2011, pp. 1085–1086.