

Enhanced Cross Domain Recommender System using Contextual parameters in Temporal Domain

Swapna Joshi

Department of Computer Engineering, Smt.
Kashibai Navale College of Engineering, Pune,
Maharashtra, India

Prof. Manisha Patil

Department of Computer Engineering, Smt.
Kashibai Navale College of Engineering, Pune,
Maharashtra, India

Abstract: Cross-domain collaborative filtering (CDCF) is an evolving research topic in the modern recommender systems. Its main objective is to alleviate the data sparsity problem in individual domains by mixing and transferring the knowledge among the related domains. But there is also an issue of user interest drift over time because user's taste keeps on changing over time. We should consider various temporal domains to overcome user interest drift over time problem to predict more accurately as per the user's current interest. This paper discusses how to achieve effective ratings recommendations using contextual parameters in temporal domain in this research line. It calculates the contextual parameters as per user's current timestamp. This will enhance the recommendations more in line with the current temporal domain. It also deals with cross domain recommendations for both movies and novels based on their categories and similarities.

Keywords: Collaborative filtering; temporal domain; cross-domain; recommender systems

1. INTRODUCTION

The main problem of existing Collaborative Filtering (CF) methods is to find similar users or items and to measure similarities between them. Unit now, most of the existing CF methods are single-domain based, which generates predictions based on single rating matrix. These methods can only find similar users or items in a single domain. However, in practical recommendation scenarios, multiple related CF domains might be presented at the same time and then finding similar users or items across domains becomes possible, such that common rating knowledge might be shared among various related domains.

Cross-domain collaborative filtering (CDCF) intends to share the common rating knowledge across multiple related CF domains to enhance the prediction performance. CDCF methods extract knowledge from various related input domains containing additional user's preference data to improve recommendations in the target domain. CDCF can also benefit multiple data owners by improving quality of service in different related domains.

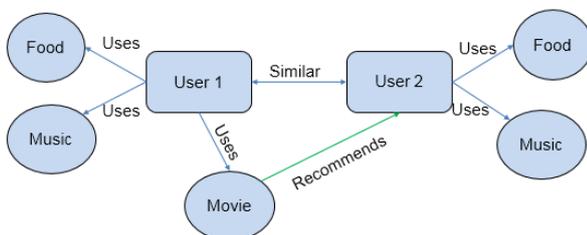


Fig 1: Generic Cross Domain CF Architecture

Major Issues in current CDCF system:

a. User Interest Drift Over Time:

User's taste keeps on changing over the changing time because it continuously gets affected by multiple factors such as moods, contexts, cultures, festivals, seasons etc. For

example, a person who does not like animation based movies, might like to watch them in future due to superior 3D animation technologies with very good visual and sound effects or if the person having baby might like to watch it now which he has never watched before. Another example would be 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.

b. Data sparsity:

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 and recommended for under 1 percent of the items (very negligible quantity out of total 2 million books). Accordingly, a recommender system based on nearest neighbor algorithms might not be able to suggest any item recommendations for a user. Thus, the accuracy of recommendations would be poor.

2. MOTIVATION

In most of the CDCF approaches, more stress is given on referring the historical ratings from multiple sites of the related domains. But there is a problem in recommender systems that the user interests keep drifting over a period. 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 point of time, we cannot simply refer the previous rating patterns of the same user in the historical temporal domains. We should also apply the current time's context to enhance the recommendations which will be more suitable with current time. For example, a person who has watched the movies of a particular category, might not like the recommendations on the movies of similar category again and again. His interest might change over the time context. Hence the recommender system should not just calculate the historical temporal ratings but should also apply

the contextual parameters for the current time domain's context variance.

3. LITERATURE REVIEW

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

[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 which deal with data sparsity reduction problem. In [2], Latent factor model is used based on the ONMTF framework to cluster the users and items in Rth domain simultaneously and latent space model is used to construct the cluster-level rating pattern of user-item co-clusters. But It considered only one dimension for rating the cross domains.

In [3], principled matrix-based transfer learning framework is used that considers the data heterogeneity. The principle coordinates of both users and items in the auxiliary data matrices are extracted and transferred to the target domain to reduce the effect of data sparsity.

[4], [5] and [6] deal with 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. [6] uses derived method based on a Bayesian latent factor model which can be inferred using Gibbs sampling. This deals with user interest drift over time in single domain.

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.

[8] uses Principled matrix-based transfer learning framework that considers the data heterogeneity. The principle coordinates of both users and items in the auxiliary data

matrices are extracted and transferred to the target domain to reduce the effect of data sparsity.

Table 1: Overall Evaluation of Related work

Reference Number	Paper Name	Data Sparsity	User Interest drift
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

[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. Series of time slices are meant as related domains and a user at the current time slice depends on the previous historical time slices. 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.

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

The above evaluation of related work shows that the current work is mainly dealing with the temporal domain ratings from the historical data. The proposed work will deal with the historical ratings along with the current time frame's contextual parameters like season, holiday, locations, category etc. This will enhance the ratings and will be more current in time.

4. SYSTEM OVERVIEW

The aim of proposed work is to implement the concept of context based parameters in the cross-domain recommender system for historical and current time slices of the user over cross domain techniques. The ratings calculation is done to obtain the average recommendations for the movie/book selected by the user. The project aims at developing an efficient system architecture that enables the calculation of ratings based on various contextual parameters for both historical and current time slices.

For a recommender system, various types of contexts can be used as per the applications' requirements. Physical context represents the time, position, and activity of the user, but also the weather, light, and temperature when the recommendation is supposed to be used. Social context represents the presence and role of other people (either using or not using the application) around the user, and whether the user is alone or in a group when using the application. Interaction media context describes the device used to access the system (for example, a mobile phone or a kiosk) as well as the type of media that are browsed and personalized. The latter can be ordinary text, music, images, movies, or queries made to the recommender system. Modal context represents the current state of mind of the user, the user's goals, mood, experience, and cognitive capabilities.

Proposed idea calculates the recommendation ratings for the movie/book in which user is interested in using contextual parameters. Contextual parameters such as season, holidays, location and category are applied to the historical ratings of the users and to the current user's time context. Then to calculate the final recommendations, both the historical and current contextual parameters will be mapped together. It will also apply mixing of cross domain rating recommendations to handle the data sparsity in movies and books domains.

Database containing movies and books ratings given by various users is the input to the recommendation system along with the contextual parameters applied for each time slice. The output is final recommendation for the movies and books for which user is interested in along with other related movies and books in similar categories/context.

The proposed work considers 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 etc. 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.

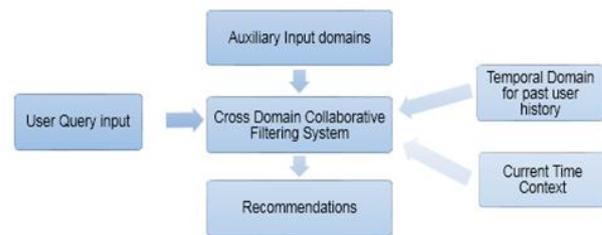


Fig 2: Architecture for Proposed System

4.1 Algorithm

Input: Movie and Novel ratings dataset

Output: The recommended movie and novel ratings as per user query

1. String event = {season, holiday, location, movie category}; // Initialize the contextual parameters
2. movie_mat=[user][movie]; // Initialize movies rating matrix.
3. book_mat=[user][book]; // Initialize books rating matrix.
4. current_user ← {season, holiday, location, movie category}; // Calculate user state
5. sliceM_mat ← movie_mat/12 ;
6. sliceB_mat ← book_mat/12 ;
7. For (int i=0;i<12;i++) {
8. //Compare user context with event context
9. sum_rating += rating; }
10. avg_rating ← sum_rating / total_ratings_cnt

4.1.1 Dataset pre-processing module

This is used to pre-process the historical movie and books ratings data. 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.

Step1: Collect ratings data from 3 different movie rating sites.

Step2: Apply data enrichments using various parameters such as of location, season, holidays and movie category.

Step3: Divide the user ratings in 12 slices with 1 slice per month so that we can get fine grained ratings

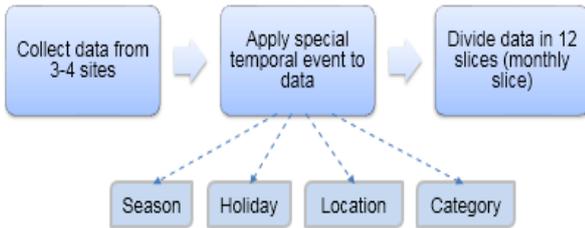


Fig 3: Flow diagram of the Data Preprocessing module

4.1.2 Main module

The pre-processed dataset will be used to calculate the ratings recommendations. The historical time slices along with contextual parameters will be stored in the database for past 3 years with monthly time frame per slice.

Datasets of two different domains for movies and books are used for cross-domain functionality. Bayesian network word analyzer is used to calculate the similarity index of the 2 domains. If the similarity index is 80% or above, then those two domains are similar and could provide better recommendation results in cross-domain framework.

When the current user enters a movie/book name to see the ratings, the current user's state will be assigned with the same contextual parameters of location, season, holidays and movie category. Then the mapping of current user's state will be done with the historical ratings to calculate the actual rating recommendations.

From the mapping calculations for all the parameters, the average movie ratings will be calculated and the average from all the time slices will be finally calculated for presenting the recommendations result to the end user.

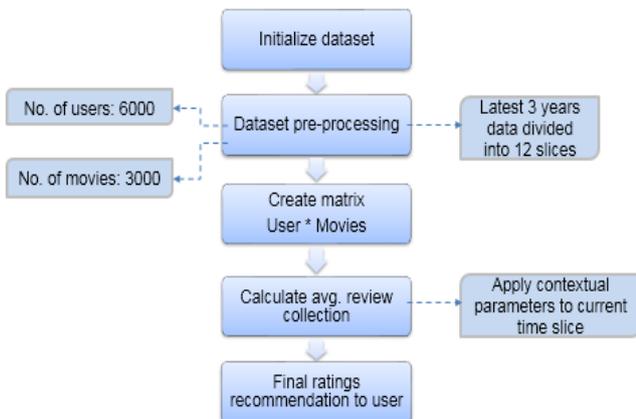


Fig 4: Flow diagram of the main system

Also, the rating recommendations for some other related movies and books in the similar movie category will be presented to the end user. It can be customized as per user's interest and then filtering as per the contextual parameters by

applying post-filtering contextual techniques. This will help the user to choose from other similar movies and similar books which are available in the database.

Contextual postfiltering (or contextualization of recommendation output) is used in the proposed system. In this paradigm, contextual information is initially ignored, and the ratings are predicted using any traditional 2D recommender system on the entire data. Then, the resulting set of recommendations is adjusted (contextualized) for each user using the contextual information.

System Components:

- Data Pre-processor component:**
 - Load ratings data from multiple sites of movies and books
 - Apply special contextual parameters of different events such as location, season and category to the data
 - Divide data in 12 monthly slices
- Rating Analyzer:**
 - Get user query for the movie or novel search and customization details if applied by the user
 - Create user's contextual event parameters
 - Compare user's query contexts and customizations with existing data set by mapping with the historical ratings to calculate the rating recommendations.
 - Provide ratings for the queried movie or novel.
 - Also, provide the ratings and recommendations for other related movies and novels in the similar context. It can be customized as per user's interest and then filtered as per the contextual parameters.

4.2 Mathematical Model

Step 1:

Select the highly available movies from the Netflix.com dataset for past 3 years. Initialize User * Movie matrix for the users and movies.

Let c be the number of movies and r be the number of users.

The ratings will be denoted as $U_i M_j = R_{ij}$

$$\begin{matrix}
 & M_1 & M_2 & M_3 & \dots & M_c \\
 U_1 & R_{11} & R_{12} & R_{13} & \dots & R_{1c} \\
 U_2 & R_{21} & R_{22} & R_{23} & \dots & R_{2c} \\
 U_3 & R_{31} & R_{32} & R_{33} & \dots & R_{3c} \\
 \vdots & & & & & \\
 U_r & \dots & \dots & \dots & \dots & R_{rc}
 \end{matrix}$$

Similarly prepare the rating matrix for books ratings.

The ratings will be denoted as: $U_i B_j = R_{bij}$

Step 2:

Divide the ratings matrix data in 12 different slices with one slice per month and save in the Slice Matrix set table.

$$\text{Slice_matrix} = \text{User_matrix} / 12$$

Assign season, holiday, location and category as the contextual parameter values for each movie/book rating in the Slice matrix set.

$U_m = [U_{m_season}, U_{m_holiday}, U_{m_location}, U_{m_category}]$

$U_b = [U_{b_season}, U_{b_holiday}, U_{b_location}, U_{b_category}]$

Step 3:

Calculate the similarities between movies and the books in the matrix based on categories using Bayesian theory. Similarity index should be 80% or more to have the cross-domain functionality.

Step 4:

Get user input query for the interested movie/book as U_q . Assign season, holiday, location, category contextual parameter values for U_q as

$U_q = [U_{q_season}, U_{q_holiday}, U_{q_location}, U_{q_category}]$

Step 5:

Compare event contexts array in different historical time slices with current user's context array to calculate the review ratings:

$sum_review = \sum Rum \mid Rum \Rightarrow Ruq$

Calculate average review ratings from all time slices:

$avg_review = sum_review / review_count$

Step 6:

Provide the cross-domain recommendations of both movies and books for context based filter criteria if applied.

Step 7:

The evaluation matrix of the system is calculated by Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) parameters.

$$RMSE = \sqrt{\sum_{i \in S} (r_i - \hat{r}_i)^2 / |S|}$$

$$MAE = (\sum_{i \in S} |r_i - \hat{r}_i|) / |S|$$

4.3 User Interface of the Application

The graphical user interface of the application is simple and self-explanatory. The user should provide the user id and password to the application as the system allows only authorized users to enter. Admin and user are the two roles provided. After authentication, the Admin is asked to select the datasets of different domains and then click to calculate the similarity index. Also, the contextual events should be applied and results are to be saved. With the user login, any user can provide his query for interested movies or novels. User can select the movie/novel name from the provided drop down box if available and click on submit. The results are displayed in 3 different frames. First frame shows the rating

recommendations of the selected movie/novel. Second frame shows the other related movies and their ratings. Third frame shows the other related novels and their ratings.

The user can also search as per the context as in category, writer name, season and location. Then the recommendation results are displayed based on those selected contextual parameters.

5. RESULTS AND DISCUSSION

The contextual parameters based temporal domain is used for recommendation system. We have used dataset from MovieLens.com. For the historical movies and books ratings dataset, the contextual parameters of season, holiday, location, and category are applied in the database. Consider a user submits the query for the recommendations of a particular movie or book. The system calculates the contextual parameters for the user's current state before calculating the ratings recommendations. These contextual parameters are compared with the parameters of historical movies and books ratings and those with matching parameters will be picked up to calculate the rating averages. This fetches the ratings recommendations which are in close match with the current user's likings and the current temporal domain scenario.

Also, the other related movies and books with similar category and closely matching contextual parameters will be presented to the user. These recommendations will be more in line with the user's current time domain and will be based on the customizations as per his own taste if he has rated any movies or books in the past. By this, the system takes care of the same user's historical selection patterns and the similar likings of other categories which he might not have seen/rated before but those movies or books which might closely match with the user's current temporal domain context such as season, holiday and location.

For example, the user is searching for a movie in comedy category then the system will provide the ratings of that movie by matching the user's contextual parameters with historical ratings' contextual parameters. System will also provide the recommendations of some other movies in let's say children category movies if the current season is of summer vacations and user has not customized input for movie category. Also by applying cross domain parameters, the books based on the queried movie and its category, will also be recommended to the user.

Below graphs show the expected results of the qualitative parameters of the base system with the proposed system. By using the contextual parameter based temporal domain system, the quality of the recommendations is more granular and closer with the user's interest in current time context scenario.

The performance of the proposed system is calculated based on the qualitative parameters of the rating accuracy. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) parameters of the proposed system are lower than the base system due to mixing cross domains. They are calculated for 300 samples of the dataset. The performance of the proposed system on the qualitative parameters of the rating accuracy is better with more granular recommendations and mixing of sites. Hence the recommendations are closer with user's interest in current time contextual scenario.

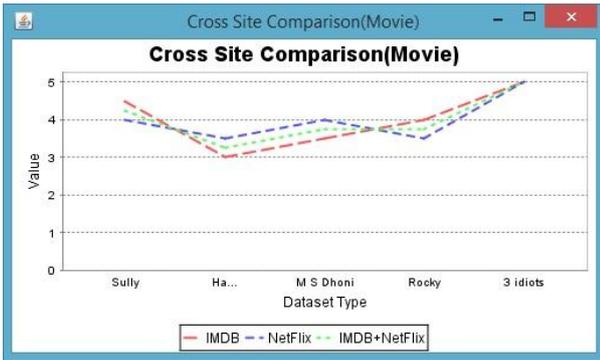


Fig 5: Cross Site Comparison for Movies

Below table shows the comparative values of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) parameters of the base and proposed system

Table 9.1: Comparison Results of RMSE and MAE

Parameter name	Base System	Proposed System
RMSE Movie	0.860	0.846
RMSE Novel	0.916	0.887
MAE Movie	0.671	0.632
MAE Novel	0.806	0.796

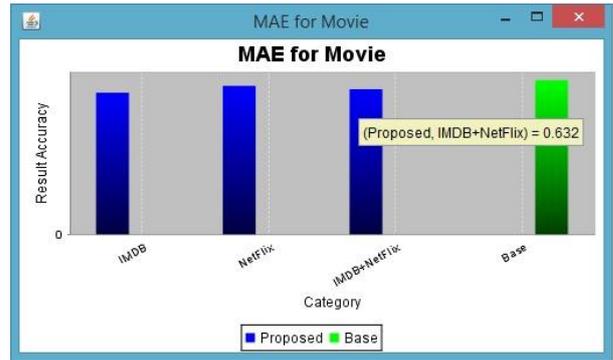


Fig. 8 MAE for Movies

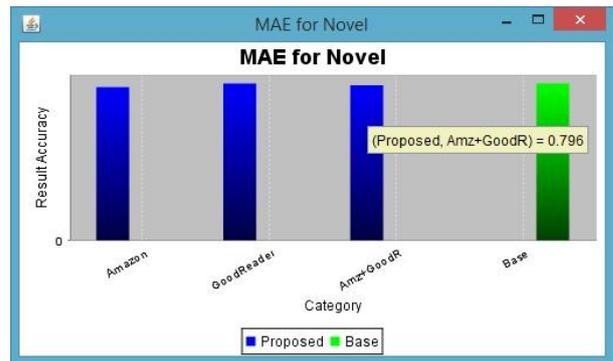


Fig. 9 MAE for Novels

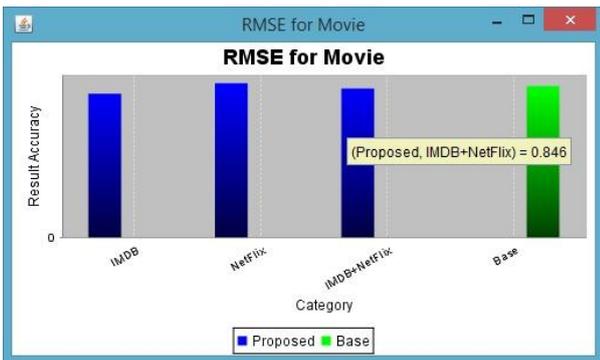


Fig 6: RMSE for Movies

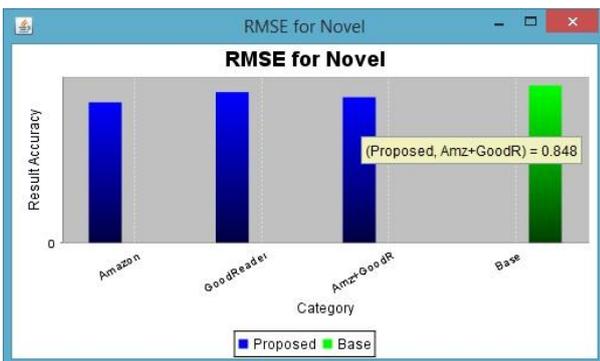


Fig 7: RMSE for Novels

6. CONCLUSION AND FUTURE WORK

The proposed work deals with the addition of temporal domain and contextual parameters to the historical movie and book ratings and slicing them on more granular level. The current user's state is calculated based on these contextual parameters. As per user's input query, the enhanced ratings recommendations for similar movies and books are presented by matching the user's contextual parameters in cross-domain matrix.

The future work may involve addition of more related domains to deal with data sparsity. It can also be integrated with social networking sites to provide rating recommendations on related domains. We can also implement machine learning system for the ratings datasets for analytics purposes.

7. ACKNOWLEDGEMENT

This work would not have been possible without the support and help of many individuals. I would like to extend 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 work.

8. 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.
- [12] 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 Volume: 02, Issue No: 10, October 2016 ISSN: 2455-3778.Pg.88-92 Oct 2016.
- [13] Swapna Joshi and Prof. Manisha Patil "Enhanced Rating Recommendations using CDCF in Temporal Domain" at, Smt. Kashibai Navale College of Engineering, in International conference on Internet of Things, Next Generation Networks and Cloud Computing 2017 (ICINC-2017).
- [14] <https://www.amazon.com>
- [15] Ms. Sanjeevani Dhaneshwar , Mrs. Manisha Patil "Context Awareness in Data Mining Applications – A Survey" International Journal of Science and Research (IJSR) Volume 5 Issue 1 ISSN (Online): 2319-7064 Pg no 253-255 January 2016.
- [16] Ms Manisha Kumbhar, Dr. Rajesh S. Prasad "Genres based Collaborative filtering: An approach to improved Quality of Recommendation". Proceeding of International Conference on Internet of Things, Next Generation Network and Cloud Computing 2016 ISSN: 0975 – 8887.Pg. no 423-427 March 2016.
- [17] Ms. Jyoti Pandey , Mrs. Manisha Patil "Recommender System Using Clustering Based On Collaborative Filtering Approach" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (4) , , 3419-3423 ISSN-0975-9646 Pg.3419-3423 2015.