ISSN:0975-9646

Raj Bharath.R et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1358-1362

Feature Extraction Based Dynamic Recommendation for Analogous Users

Raj Bharath.R [1], Thanigaivel.K[2], Alfahath.A[3], Prasanth.T[4]

[1]Asst.Prof, [2,3,4]B.Tech,
Department of Computer Science,
Manakula Vinayagar Institute of Technology.

Abstract- **In this work, a novel dynamic personalized recommendation is proposed based on feature extraction. Unlike previous works [1][4] this work provides an efficient technique to recommend product to the users with respect to the interest which changes dynamically. A method to classify users based on age groups is proposed to overcome the problem of cold start. Identifying the users of similar interest and providing recommendation based on the user contents of other users is proposed to enhance the recommendations techniques provided in existing work. To enhance e-marketing in the field of e-commerce a novel method of promoting new products to the users is proposed with periodically automated e-mail notifications which is purely based on recommendations provided in user's page. This method gives an efficient technique of providing recommendations by considering the multiple phases of interests of users and thereby providing recommendation dynamically. This enables to achieve efficient recommendation with accurate prediction of interests of users.**

Keywords- **feature extraction, time series analysis, collaborative filtering, personalized recommendation, adaptive weighting, ratings.**

I. INTRODUCTION

The internet has become an essential part of our lives, and it provides a manifesto for enterprises to distribute information about products and services to the customers conveniently. As this kind of information is increasing rapidly, one of the great challenges is to make sure that proper content must be delivered hastily to the suitable customers. Personalized recommendation is a substantial way to improve customer satisfaction.

Recommender systems apply knowledge discovery methods to the problem of making personalized recommendations for information about products or services during a live interaction. These systems, particularly the k-nearest neighbor collaborative filtering based methods, are achieving extensive success on the Web. The fabulous growth in the amount of available information and the number of guests to Web sites in recent years poses some key challenges for recommender systems. Consumer preferences for products are drifting over time. Product awareness and popularity are persistently changing as new selection emerges. Similarly, customer proclivity are evolving, leading them to ever redefine their interest. Thus, designing temporal dynamics is essential for modeling recommender systems or general customer penchant models. However, this raises unique challenges. Within the system, intersecting multiple products and customers and many different characteristics are shifting at regular time interval, while many of them influence each other and often those shifts are fragile and associated with a few data occurrences.

With the growth of e-commerce and the increase of easily accessible information, recommender systems have become a fashionable technique to prune large information spaces so that customers are directed toward those items that best meet their needs and preferences. Various techniques have been proposed for performing suggestions, including content-based and collaborative methods. Content-based filtering selects information based on semantic content, whereas collaborative filtering coalesce the opinions of other users to make a prediction for a target customer. There are essentially three approaches to recommendation engines based on different data analysis models, i.e., rule-based, content-based and collaborative filtering. Among them, collaborative filtering (CF) [1] needs only data about past user behavior like, ratings and its two major approaches are the neighborhood methods and latent factor methods.

The neighborhood methods can be customer-oriented or item-oriented. They attempt to find like-minded users or similar items on the basis of co-ratings, and forecast based on ratings of the nearest neighbors. While latent factor methods involves mostly on the ratings to capture the general curiosity of users, they still have intricacy in catching up with the drifting signal in dynamic recommendation because of sparse data, and it is hard to actually explain the reason of the involving. Generally, the interest cycle varies from user to user, and the pattern how user preferences vary cannot be precisely described by several simple functions. Moreover, CF approaches usually accounted the cold-start crisis which is amplified in the dynamic circumstances since the rate of new-fangled users and new items would be high.

The main contributions of the preceding works can be summarized as follows:

(a) More information can be used for recommender systems by examining the similar relation among correlated user profile and item content. Compared with the existing works, utilization of the similarity among content in each profile attribute is done so that more content information is used, particularly content in those attributes which are hard to be computed.

(b) A novel set of dynamic features is anticipated to illustrate users' preferences, which is more flexible and suitable to model the impacts of preferences in various phases of interest compared with dynamic methods used in existing works, since the features are designed according to intermittent characteristics of users' interest and a linear model of the features can catch up with changes in user preferences.

(c) An adaptive weighting algorithm is modeled to combine the dynamic features used for personalized recommendation, in which time and information density factors are considered to get used with dynamic recommendation on sparse data.

In most cases, the drifting of users' preferences or items' reputations is not too quick, which makes it possible to describe temporal state of them by using some features. In this segment, firstly we introduce a way to make use of profiles to extend the co-rating relation, and then we put forward a set of dynamic features to reflect users' preferences or items' reputations in multiple phases of interest, and after that we suggest an adaptive algorithm for dynamic personalized recommendation.

II. RELATED WORK

In case of recommendation data being sparse, the main difficulty of confining users' dynamic preferences is the lack of helpful information, which may come from three sources - user silhouette, item silhouette and historical rating proceedings. Traditional algorithms heavily rely on the co-rate relation (to the same item by different users or to different items by the same user), which is rare when the information is sparse. Useful ratings are discovered using the co-rate relation, which is simple, intuitional and actually significant when we go one or two steps along, but it sturdily limits the amount of data used in each prediction.

Instead of searching neighboring nodes along co-rate edges in the U×I plane, we try to find a different way to find useful ratings. We notice that when taking into consideration the factors which affect a rating r (u, i), we may focus more on some attributes of u and i in their profiles, as an alternative of the user himself or the item itself. For example, if the movie "Gone with the Wind" is given high ratings by middle-aged people and lower ratings by teenagers with no doubt, we would mainly check on the age attribute in a user's profile when predicting probable rating the user would give to the movie, instead of other descriptions of the customer or how the customer has rated further movies. As is evident, it may not be compulsory to stick only to the co-rate relation, and we bring in the semi-co-rate relation between ratings whose corresponding user profiles or item contents have similar or identical content in one or more attributes. Since semi-co-rate is much less limited, we extend the co-rate relation to it by means of user profile and item content, and suggest a new way of finding useful ratings for dynamic personalized recommendation.

Our method is based on personalized transition graphs over existing Markov chains. That means for every user an own transition matrix is learned - thus in whole the method uses a transition cube. As the observations for the estimation of the transitions are usually very less, our method factorizes the transition cube with a pair wise interaction model which is a special case of the Tucker Decomposition. We prove that our factorized personalized MC (FPMC) model subsumes both a common Markov chain and the normal matrix factorization model. To learn the model parameters, we bring in an adaption of the Bayesian Personalized Ranking (BPR) framework for sequential basket data. Empirically, we prove that our FPMC model outperforms both the common matrix factorization and the impersonalized MC model both learned with and without factorization.

Compared with the filtering algorithm, which is among the existing best algorithms of k-means clustering, our algorithm can efficiently reduce the computing time. It is observed that our proposed algorithm can generate the same clusters as that produced by hard k-means clustering. The advantage of our method is more remarkable when a larger data set with higher dimension is used.

With current projections regarding the growth of Internet sales, online retailing raises many questions about how to market on the Net. While convenience impels consumers to purchase items on the web, quality remains a significant factor in deciding where to shop online. The competition is increasing and personalization is considered to be the competitive advantage that will determine the winners in the market of online shopping in the following years. Recommender systems are a means of personalizing a site and a solution to the customer's information overload problem. As such, many e-commerce sites already use them to facilitate the buying process. In this paper we study the application of recommender systems for electronic retail sites, focusing on the peculiar characteristics and requirements of this environment. We also introduce a hybrid model supporting dynamic recommendations, which eliminates the problems the underlying techniques have when applied solely. We then discuss the application of the proposed solution in the case of an innovative research project on electronic retailing funded by the European Commission and conclude with some ideas for further development and research in this area.

III. FEATURE EXTRACTION FOR ANALOGOUS USERS

We can use only historical data but not future data for current prediction in real applications. In conventional RMSE evaluations training and testing data are randomly sampled and the train and test split is not based on time. This would produce current prophecy based on future data. It is clear that the proposed algorithm is quite robust in the phases, and we found it is true that the more recent ratings should have heavier weights across the whole time, which explains the advantages of the features such as light computation, flexibility and high accuracy.

A method to find like-minded users is proposed for recommending products using profile contents and ratings. In this work a novel dynamic recommendation is provided based on age, user interest and based on like-minded users. This method is proposed for providing recommendations for both new and existing users in case of arrival of a new product. When considering about the dynamic personalized recommendation, the recommendation is provided based on multiple phases of interest.

A method of feature extraction is used to find the user profile contents and ratings. For providing recommendations through e-mail, an automated notification mail regarding the personalized recommendation based on age, user interest and like-minded users will be sent to all the users. The drifting of user preferences focuses mainly on feature extraction of user profile contents which enable to find like-minded users and users of similar age group. This novel method overcomes the problem of cold-start.

The proposed algorithm works on different phases of historical data. We have applied the proposed algorithm on the data in each single phase which are defined before, and the RMSEs are calculated separately according to the definition of users' multiple phases of interest. More information and may achieve better recommendation accuracies if the information mined is sufficient and the dynamic nature of data is well handled.

The data in different phases of interest at different training ratios is quantified. It is clear that the proposed algorithm is quite robust in the phases, and we found it is true that the more recent ratings should have heavier weights across the total time, which illustrates the advantages of the features – light computation, flexibility and high accuracy. More information can be used for recommender systems by investigating the similar relation among related user profile and item content.

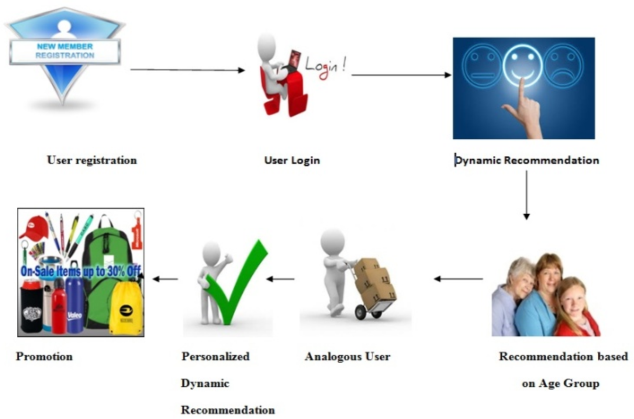


Fig 1. Personalized recommender system for analogous users

IV. DYNAMIC FEATURE EXTRACTION

Users' preferences or items' reputations changes with regular time interval, thus we have to deal with the dynamic nature of data to enhance the precision of recommendation algorithms, and latest ratings and remote ratings should have different weights in the prophecy. Three kinds of methods were proposed in concept drift to deal with the drifting problem as instance selection, time-window (usually time decay function) and ensemble learning. These methods help to make progress in precision of dynamic recommendation, but they also have their drawbacks: decay functions cannot precisely describe the evolution of user preferences and only isolating transient noise cannot catch up with the change in data.

To enable the features to describe users' preferences in multiple phases of interest, we divide each rating subset *Rs* into several disjoint secondary subsets where each secondary subset is manually assigned with a range of time-distance (corresponding to multiple phases of interest), and then we analyze the features on each secondary subsets using some basic algorithms such as time series analysis (TSA). TSA is the fundamental feature extraction method. In fact, methods for concept are also variants of TSA algorithms in the angle of prediction. Most importantly, since the results of TSA are generally representative and predictive of the utilized data in pertinent time ranges, we could conveniently use and update the results as features and "expectations" of certain phases of interest for further analysis. In the hypothesis of time series analysis, past ratings should impact the predictive features less, and thus they should have lesser weights. So if we perform TSA algorithm on a secondary subset of R (i.e. R_s^d) to get a feature $fea_{s,d}$, there would be a uniform formulation as:

$$fea_{s,d} = \sum_{l=1}^{o} \frac{w_l}{w} R_{s,l}^d,$$

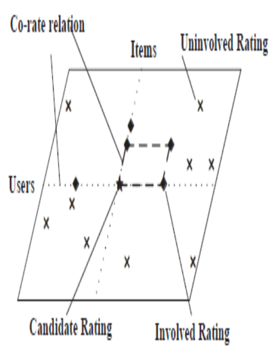
Where $\#R_s^d = o$, $R_{s,l}^d (l = 1, 2, ..., o)$ are the rating values which are from the subset R_s^d and listed in reversed time order. And positive weight parameters w_l, $(l = 1, 2, ..., o)$ and normalization factor w should satisfy

$$\begin{cases} w = \sum_{l=1}^{o} w_l, \\ w_{l_1} \geq w_{l_2} \ if \ l_1 < l_2. \end{cases}$$

Since the subsets are updated frequently, index smoothing, which is a classic TSA algorithm, is selected as the basic TSA algorithm:

$$\begin{cases} R_s^d = \{R_{j',k'} | R_{j',k'} \in R_s \ and \ T_{j,k} - T_{j',k'} \geq T_d\}, \\ fea_{s,d} = \sum_{l=1}^{o} \mu(1-\mu)^{l-1} R_{s,l}^d, \end{cases}$$

where R_s^d (d = 1, 2, ...) are the secondary subsets, T_d (d = 1, 2, ...) are a sequence of time differences manually set,$R_{s,l}^d (l = 1, 2, ..., o)$are the rating values listed in reversed order in the subset, μ is the forgetting element for index smoothing. We have tested diverse values for mu in the experiments and set μ = 0.95 experimentally.



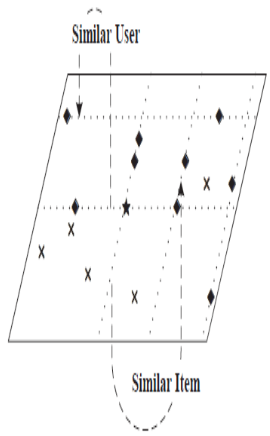


Fig 2. Common hybrid approach

Fig 3. Proposed system

V. ADAPTIVE WEIGHTING ALGORITHM

As features like feature subsets and division of time gained by applying Multiple Phase Division are all normalized rating values, or in other words, as content of user and item profiles have been quantified in the feature extraction, it is convenient for us to classify them for accurate rating estimation by adaptive weighting. Sizes of the relevant subsets are also analyzed in MPD and could reflect data density. We integrate these features for recommendation with a linear model since they are homogeneous and it is efficient to learn their weights. We incorporate these features for recommendation with a linear model since they are homogeneous and it is efficient to learn their weights. $R_{j,k}$ is used to note the estimated rating that user u_j could give to item i_k at time point $T_{j,k}$, and the adaptive linear model can be formulated as:

$$\hat{R}_{j,k} = \sum_{s}\sum_{d}(\alpha_{s,d} + \beta(\#R_s^d))b_{u_j}(s)b_{i_k}(s)fea_{s,d},$$
$$\text{with} : \alpha_{s,d} \geq 0,\ \beta \geq 0,$$

where sizes of relevant subsets are used as prior information in weighting the features to improve recommendation accuracy, $fea_{s,d}$(s = 1, 2, ..., d = 1, 2, ...)are the features calculated in Eq.(3), R_s^d (s = 1, 2, ..., d = 1, 2, ...) denote their relevant secondary rating subsets, b_{uj} and b_{ik} are binary functions denoting the relating state of candidate rating and relevant subset and $\alpha_{s,d}$ and β are weighting factors which should balance the weights of features and data density, or, maintain a balance of the affection of data consistency and quantity of information. In detail, b_{uj} (s) = 1 if $R_{j,\,k}$ is semi-co-rate related with all ratings in secondary subset R_s through attribute of the user u_j denoted by s, else b_{uj} (s) = 0, b_{ik} (s) = 1 if $R_{j,\,k}$ is semi-co-rate related with all ratings in secondary subset R_s through attribute of the item i_k also denoted by s, else b_{ik} (s) = 0.

This optimization problem has clear solution as:

$$\begin{cases} \beta = \dfrac{\sum_s \sum_d \delta_{s,d}}{2\sum_s \sum_d (\#R_s^d)}, \\ \alpha_{s,d} = \delta_{s,d} - \beta(\#R_s^d)\ for\ all\ s, d. \end{cases}$$

VI. EVALUATION OF DYNAMIC RECOMMENDATION

A. Age Classification

Recommendation is provided based on range of ages. This method mainly focuses on recommendation for new users. It is mainly based on user interest and ratings of existing users. Here grouping of ages is done and categories are divided based on age groups and gender. This method helps to overcome the problem of cold start where the new users gets the recommendation based on ratings and user interest of existing users. It will be easy for the new users to get recommendation based on likes of others users of similar age group. This method proves to be very efficient as it overcomes the limitations of existing works.

B. Dynamic Feature Extraction

We propose a concept of finding like-minded users. This is mainly based on Feature Extraction of user contents and rating provided by the user. This enables to provide recommendation based on similarities of user interests. Finding of like-minded users based on Feature Extraction proves to be efficient in recommending the products of similar interest. The dynamic recommendation is also provided based on the change of interests of users which uses the concept of multiple phases of interest. Such that the like-minded user's recommendation changes periodically when there is a change of interest. This efficient method overcomes the limitation of previous works by providing recommendation dynamically, not only based on personalized usage but also based on other user's interests.

C. Heavy Weight Ratings

The concept of multiple phases of interest is used to find the interest of the users. This is highly recommended because of the non-static recommendation for users. User interests changes across time period. It is recommendation based on user's full history. A method to find the user's recent ratings is proposed which comes under the concept of Time Series Analysis and Multiple Phases of Interest such that the recent ratings have heavy weight and old ratings have light weights. The personalized recommendation is provided according to the user's heavy weight ratings which changes periodically and dynamically. This efficient method provides the way to recommend products to the users with their current interest and ratings.

D. Automated E-mail Notification

All the previous works of recommending products to the user's through email is based on organization's suggestion. This method overcomes that problem, by providing notification emails only based on users personalized content, classification and Heavy Weight ratings. The notification mails are sent periodically according to the user's choice. The arrival of new products is also recommended through email notification to enhance the business. The scope of the email notification is to provide highly recommended products to the users. Automated email notification is fully based on the concepts used in proposed work.

E. Promotions

The previous works of online shopping does not provide the way to the new arrivals. This method gets the information of the new products which comes into existence and classifies the products. The method of classification is based on the age and gender. Finding the age groups, classifying the new products and recommending the product to particular age provides an efficient way to publish and market the new products. This method helps in the field of commerce by making the product to be reached to the users of particular groups. This proves to be an efficient method of e-marketing the products.

VII. CONCLUSION

In this paper, accurate recommendation is provided based on feature extraction. The concept of recommending the products based on age and gender helps to overcome the cold start problem. The way of finding the personalized content using multiple phases of interest provides a dynamic personalized recommendation for existing users. The recommendation process to be more effective and efficient, like-minded users are identified by the way they use and rate the products. Also for e-shopping to be more effective and e-marketing to be more productive, promotions of new products is proposed. The notification mail received by each user is based on the recommendation the user gets in the customer recommendation page. This work proves to be more efficient in the field of recommending the products in e-marketing.

REFERENCES

[1] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl, Item-based collaborative Filtering recommendation algorithms, in: WWW, 2001, pp. 285–295.
[2] P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.),The Adaptive Web, Methodsand Strategies of Web Personalization, Lecture Notes in ComputerScience, Springer, 2007.
[3] G.Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of –the-art and possible extensions, IEEE Trans. Knowl. Data Eng. 17(6) (2005) 734-749.
[4] Y. Koren, Collaborative filtering with temporal dynamics, Communications of the ACM 53 (4) (2010) 89-97.
[5] L. Candillier, F. Meyer, M. Boull´e, Comparing state-of-the-art collaborativefiltering systems, in: P. Perner (Ed.), MLDM, Vol. 4571 of LectureNotes in Computer Science, Springer, 2007, pp. 548–562.
[6] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, H. Kriegel, Probabilisticmemory-based collaborative filtering, IEEE Transactions on Knowledgeand Data Engineering 16 (1) (2004) 56–69.
[7] F. Fouss, A. Pirotte, J. Renders, M. Saerens, Random-walk computationof similarities between nodes of a graph with application to collaborativerecommendation, IEEE TKDE 19 (3) (2007) 355–369.
[8] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques forrecommender systems, Computer 42 (8) (2009) 30–37.
[9] S. Boutemedjet, D. Ziou, Long-term relevance feedback and feature selectionfor adaptive content based image suggestion, Pattern Recognition 43 (12) (2010) 3925–3937.
[10] B. M. Kim, Q. Li, C. S. Park, S. G. Kim, J. Y. Kim, A new approachfor combining content-based and collaborative filters, J. Intell. Inf. Syst.27 (1) (2006) 79–91.