Effectivecollaborative movie recommender system using asymmetric user similarity and matrix factorization

Rahul Katarya, Om Prakash Verma
Department of Computer Science & Engineering
Delhi Technological University, Delhi, India
rahulkatarya@dtu.ac.in

Abstract-- **Recommender systems are becoming ubiquitous these days to adviseimportant products to users. Conventional collaborative filtering methods suffer from sparsity, scalability, and cold start problem. In this work, we have implemented a novel and improved method of recommending movies by combining the asymmetric method of calculating similarity with matrix factorization and Tyco (typicality-based collaborative filtering). The asymmetric method describes that similarity of user A with B is not the similar as thesimilarity of B with A. Matrix factorization showsitems (movies) as well as users by vectors of factors derived from rating pattern of items (movies). In Tyco clusters of movies of the same genre are created, and typicality degree (a measure of how much a movie belongs to that genre) of each movie in that cluster was considered and subsequently of each user in a genre was calculated. The similarity between users was calculated by using their typicality in genres rather than co-rated items. We had combined these methods and employedPearson correlation coefficient method to calculate similarity to optimize results when compared to cosine similarity, Linear Regression to make predictions that gave better results. In this research work stochastic gradient descent is also used for optimization and regularization to avoid the problem of overfitting. All these approaches together provide better prediction and handle problems of sparsity, cold start, and scalability well as compared to conventional methods. Experimental results confirm that our HYBRTyco gives improved results than Tyco regardingmean absolute error (MAE)and mean absolute percentage error (MAPE),especially on the sparse dataset.**

Keywords--Recommender system; collaborative filtering; typicality; matrix factorization; asymmetric model

I. INTRODUCTION

Recommender system(RS)is automated systems that generate suggested items for users based on their likes and dislikes and save time searching for things online[1–6]. Different techniques are used in building RSs, and collaborative approach is the one most widely used[7]. However, most traditional techniques suffer from problems, which require them to be supplemented or made ahybrid with other techniques such as clustering or matrix factorization.The most significant use of recommendation systems is at e-commerce applications. Recommendations made by e-commerce applications are not random, but that are based on similarity of rated items with other items or the purchasing decisions made by same customers.Netflix provides recommendations for movies by comparing the watching as well as searching patterns of same users and by offering movies that have similar characteristics as films that a user has rated highly[8]. A framework was proposed that works in two-steps generate user ratings[9]. The first step predicts the probability of a user rating an item, and the second step predicts rating values. This framework forms the basis for a hybrid approach that used thetopic model in the first instance and matrix factorization in the second to solve the problem of recommendation and dataset used MovieLens and EachMovie. Recommender systems are also influenced with evolutionary computation such as, a recommendation system was proposed in which Particle swarm optimization algorithm (PSO) was used to study user weights on various features, and fuzzy sets for representing user features efficiently by MovieLens dataset[10]. The big challenging task of arecommender system is to handle big data, so theconcept of category experts was used to produce a novel method called CE method, that provide both performance and accuracy[11]. The CE method chooses a few users from each category as experts in each category and uses their ratings rather than ordinary neighbors'. Two variants of the CE methodare proposed, CES and CEP methods, which achieve higher accuracy. MovieLens and

ISBN:978-1-5090-1666-2/16/$31.00 ©2016 IEEE

Ciao dataset were used to show the efficiency of the method. In our work, we combined the asymmetric method of calculating similarity with matrix factorization and Tyco (typicality-based collaborative filtering)[12]. The similarity between users was computed by using their typicality in genres rather than co-rated items. Experimental results authorize that our HYBRTyco gives better-quality results than Tyco regarding MAE and MAPE, especially on the sparse dataset.The structure of this paper is as follows: related work and thebackgroundis introduced in section 2. In section 3 the improved typicality-based CF methods and asymmetric matrix factorization are discussed in details. We then explore the results of the proposed method with the MovieLens in section 4. Section 5 includes the conclusion and future work of this paper.

II. BACKGROUND AND RELATED WORK

In this section, we will discuss some contribution provided by authors. To addresses the problem of cold start author proposed a method that includes tags and keywords that provide information about user and item[13]. The direct prediction on the user-item rating matrixwastransformed into the indirect prediction on the tagkeyword relation matrix that adapts to the emergence of new data. The problems of scalability, sparsity, and cold start problems were addressed in the framework for movies recommendation that couldbe easily stretched to other domains[14]. Depending on the situation, it ran different predictive models for making recommendations. A collaborative user-centered recommendation approach was proposed that took into account the preferences, opinions, behavior and feedback of the user that were considered and integrated together with items' features and context information to support different applications using proper customizations[15]. The model was proposed to handle the issue of overspecialization or serendipity problem, by fostering the suggestion of items the user might be unlikely to discover[16]. The strategy uses background knowledge with a graph-based recommendation algorithm to allow the system to understand the items fully. The infused knowledge can help discover hidden correlations among items and, therefore, promote non-obvious suggestions. The novelty can be summarized in it is the ability to show how to add different weights on the social trust relationships among users based on the trustee's competence and trustworthiness.Incorporation of the similarity relationships among users as a complement into the social trust relationships to enhance the computation of user's neighborhood, and balance the influence of these two kinds of relationships based on user's individuality adaptively. Experiments were accomplished onEpinions and Ciao[17].In another approach author also proposed a new model that helps recommendation systems select the most appropriate content by using contextual parameters. The proposed model provides with thecontextthat helps tailor the recommendation results to individual users. By identifying the similarities in the user's contexts and the contexts and items, it is possible to determine the suitable items given in a particular context. The context was then representedby the items depending on an individual user, so as to recommend the most appropriate item suited to the needs of the user[18]. Collaborative filtering builds a model based on user's past preferences of items and his ratings. Then it attempts to find other users (neighbors) similar to the one for whom recommendations is being made (active user)[19]. Traditional methods of calculating thesimilarity between to users are based on the assumption that the effect user A receives from predicting arating for another user B is same as that B receives by aprediction for A[20]. Matrix factorization relies on input data in theform of amatrix where one dimension represents the user and another item.It usually takes into account to types of data type-explicit feedback and implicit feedback[21]. Typicality based approach (TYCO)utilize cognitive typicality to reduce the common problems of collaborative filtering like data sparsity, recommendation inaccuracy and errors in predictions[12]. Tyco finds aneighbor of users by forming user and item groups based on their genres and rated movies. It further calculates the typicality of a user in the user group and an item in the item group. This information is used to calculate then the neighbors of a user and predict his ratings for other movies.

III. PROPOSED HYBRTYCO SYSTEM

a)Asymmetric model for similarity calculation

We first calculate the asymmetric similarity, measure as

$$\text{Sim}(a,b) = \frac{|Ia \cap Ib|}{Ia} \quad (1)$$

Asymmetric similarity measure computed as theratiobetween common rated items to items rated by one user[22]. Where I_a and $I_b$are movies rated byuser a and user b respectively. So to take that into account, a new parameter called Sorenson index is multiplied by equation 1.

$$\text{Sim}(a,b) = \frac{|Ia \cap Ib|}{Ia} \cdot 2*\frac{|Ia \cap Ib|}{|Ia|+|Ib|} \quad (2)$$

The asymmetric similarity measures are taking total ratings into account.This similarity measure is combined with the actual formula of cosine similarity to give rise to formula

$$\text{ACOS}(a,b) = \frac{\vec{ra}.\vec{rb}}{||\vec{ra}||.||\vec{rb}||} \cdot \frac{|Ia \cap Ib|}{Ia} \cdot 2*\frac{|Ia \cap Ib|}{|Ia|+|Ib|} \quad (3)$$

Asymmetric similarity measure combined with cosine similarity to yield ACOS (asymmetric cosine. This model not only differentiates sim (a,b) and sim(b,a) but also provides better ratings for multiple users.

b) *Matrix factorization*

Matrix factorization(MF) approach depicts the user vs. user similarity matrix R as two low-rank matrices, P, and Q where Pi and Qi are the column vectors and indicate K-dimensional latent feature vectors of user i and j, respectively[21,23].

c) *Flow of work*

In this effort, acombination of theasymmetric model with Pearson correlation coefficient is used instead of COS similarity and also used regularization with gradient descent optimization in matrix factorization for better predictions and avoid theproblem of overfitting. Research work combined the above with Tyco method for clustering and handling sparsity and scalability problems. We have used linear regression instead of weighted average to predict rating that a user would give if he were to rate that item based on his past preferences. The HYBRTyco model pseudo code is givenbelow, and its components are showninfigure 1.

1. Set up movielens dataset
2. Calculate User item typicality matrix t=Tyco()
3. Calculate User-user similarity matrix u=asymmetric()
4. Calculate U=gradient descent matrix factorization(u)
5. Calculate Ratings r=linear_regression(u)
6. Calculate MAE=calculate error(r)

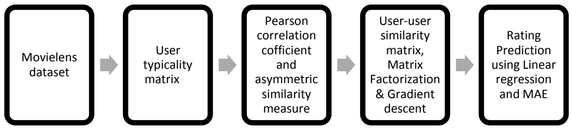

Figure 1.HYBRTyco recommender system

IV EXPERIMENT AND RESULTS

We employed MovieLens dataset(http://grouplens.org/datasets/movielens/) to evaluate our approach that contains 100,000 ratings given by 943 users on 1683 movies. Each user has rated at least 20 movies. The ratings have a measure between 1(very bad) to 5(excellent). Mean absolute error(MAE) metrics is used to measure statistical accuracy and for comparison. It is used to measure that how much our recommended ratings deviate from the actual ratings given by the user.

$$\text{MAE} = \frac{\sum_{i=1}^{n} |f_i - h_i|}{n} \quad (4)$$

Next, we calculate mean absolute percentage error (MAPE) it is used to measure that how much the user gives our recommended rating deviate relatively from the actual ratings.

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \quad (5)$$

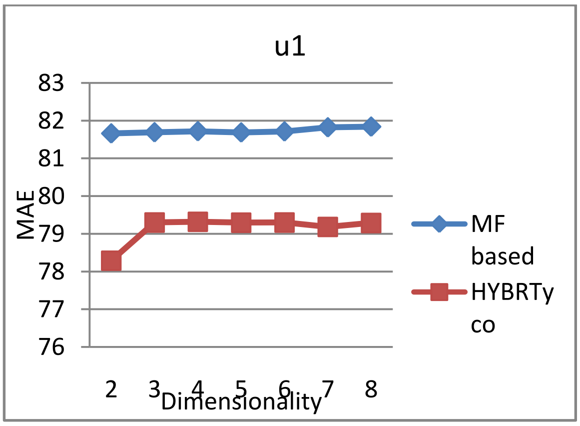

Figure 2. Comparison of MAE between MF based and HYBRTyco on U1 dataset by dimensionality.

Also, from the figure 2, we can easily interpret that HYBRTyco gives a better result and improves the accuracy by an average of 3% from the MF-based method.

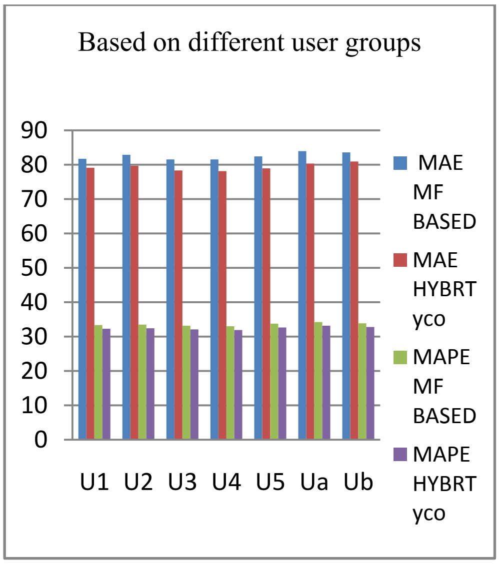

Figure 3.Comparison of results of HYBRTyco and MF-based approach on MAE and MAPE

From figure 2 and 3, we concluded that accuracy is improvedas compared to the previous collaborative approach used. The similarity between two users to be different when considered from both the sides. Reduction in the number of bigerror predictions.It has higher accuracy for large no of users.It has higher efficiency when compared to other methods

V. CONCLUSION

In this paper, we attempted to present HYBRTycosystemthat combines the classic matrix factorization technique with an asymmetric model for similarity and typicality-based collaborative filtering. HYBRTyco employs Pearson correlation coefficient, linear regression, and gradient descent optimization along with regularization and produces better resultsthatare evident from the facts that it generates a lesser MAE value and MAPE. Possible future work may be the application of our system on an even larger dataset without an enormous increase in execution time.

References

[1] D.H. Park, H.K. Kim, I.Y. Choi, J.K. Kim, A literature review and classification of recommender systems research, Expert Syst. Appl. 39 (2012) 10059–10072. http://linkinghub.elsevier.com/retrieve/pii/S0957417412002825 (accessed May 22, 2013).

[2] L. Lü, M. Medo, C.H. Yeung, Y.-C. Zhang, Z.-K. Zhang, T. Zhou, Recommender Systems, (2012) 97. http://arxiv.org/abs/1202.1112 (accessed August 22, 2014).

[3] Z. Sun, L. Han, W. Huang, X. Wang, X. Zeng, M. Wang, et al., Recommender systems based on social networks, J. Syst. Softw. 99 (2015) 109–119. doi:10.1016/j.jss.2014.09.019.

[4] J. Borràs, A. Moreno, A. Valls, Intelligent tourism recommender systems: A survey, Expert Syst. Appl. 41 (2014) 7370–7389. doi:10.1016/j.eswa.2014.06.007.

[5] Z. Dehghani, S. Reza, S. Salwah, B. Salim, A systematic review of scholar context-aware recommender systems, Expert Syst. Appl. 42 (2015) 1743–1758. doi:10.1016/j.eswa.2014.09.017.

[6] J. Beel, B. Gipp, S. Langer, C. Breitinger, Research-paper recommender systems: a literature survey, Int. J. Digit. Libr. (2015). doi:10.1007/s00799-015-0156-0.

[7] 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. 69 (2005) 734–749. doi:10.1007/978-3-540-72079-9_9.

[8] C. a Gomez-uribe, N. Hunt, The Netflix Recommender System : Algorithms, Business Value, 6 (2015). doi:10.1145/2843948.

[9] X. Zhao, Z. Niu, W. Chen, C. Shi, K. Niu, D. Liu, A hybrid approach oftopic model and matrix factorization based on two-step recommendation framework, J. Intell. Inf. Syst. (2014). doi:10.1007/s10844-014-0334-3.

[10] M. Wasid, V. Kant, A Particle Swarm Approach to Collaborative Filtering based Recommender Systems through Fuzzy Features, Procedia Comput. Sci. 54 (2015) 440–448. doi:10.1016/j.procs.2015.06.051.

[11] W.-S. Hwang, H.-J. Lee, S.-W. Kim, Y. Won, M. Lee, Efficient recommendation methods using category experts for a large dataset, Inf. Fusion. 28 (2015) 75–82. doi:10.1016/j.inffus.2015.07.005.

[12] Y. Cai, H. Leung, Q. Li, S. Member, H. Min, J. Tang, et al., Typicality-Based Collaborative Filtering Recommendation, 26 (2014) 766–779.

[13] K. Ji, H. Shen, Addressing cold-start: Scalable recommendation with tags and keywords, Knowledge-Based Syst. 83 (2015) 42–50. doi:http://dx.doi.org/10.1016/j.knosys.2015.03.008.

[14] M.N. Moreno, S. Segrera, V.F. López, M.D. Muñoz, Á.L. Sánchez, Web mining based framework for solving usual problems in recommender systems. A case study for movies' recommendation, Neurocomputing. (2015) 1–9. doi:10.1016/j.neucom.2014.10.097.

[15] F. Colace, M. De Santo, L. Greco, V. Moscato, A. Picariello, A collaborative user-centered framework for recommending items in Online Social Networks, Comput. Human Behav. 51 (2015) 694–704. doi:10.1016/j.chb.2014.12.011.

[16] M. de Gemmis, P. Lops, G. Semeraro, C. Musto, An investigation on the serendipity problem in recommender systems, Inf. Process. Manag. 51 (2015) 695–717. doi:10.1016/j.ipm.2015.06.008.

[17] M. Wang, J. Ma, A novel recommendation approach based on users' weighted trust relations and the rating similarities, Soft Comput. (2015). doi:10.1007/s00500-015-1734-1.

[18] M.F. Alhamid, M. Rawashdeh, M.A. Hossain, A. Alelaiwi, A. El Saddik, Towards context-aware media recommendation based on social tagging, J. Intell. Inf. Syst. (2015). doi:10.1007/s10844-015-0364-5.

[19] J.S. Breese, D. Heckerman, C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, UAI'98 Proc. Fourteenth Conf. Uncertain. Artif. Intell. (1998) 43–52. doi:10.1111/j.1553-2712.2011.01172.x.

[20] P. Pirasteh, D. Hwang, J.J. Jung, Exploiting matrix factorization to asymmetric user similarities in recommendation systems, Knowledge-Based Syst. 83 (2015) 51–57. doi:10.1016/j.knosys.2015.03.006.

[21] Y. Koren, R. Bell, C. Volinsky, Matrix Factorization Techniques for Recommender Systems, Computer (Long. Beach. Calif). 42 (2009) 30–37. http://www.computer.org/csdl/mags/co/2009/08/mco2009080030-abs.html (accessed March 11, 2012).

[22] P. Pirasteh, D. Hwang, J.J. Jung, Exploiting matrix factorization to asymmetric user similarities in recommendation systems, Knowledge-Based Syst. 83 (2015) 51–57. doi:10.1016/j.knosys.2015.03.006.

[23] U. Ocepek, J. Rugelj, Z. Bosnić, Improving matrix factorization recommendations for examples in cold start, Expert Syst. Appl. 42 (2015) 6784–6794. doi:10.1016/j.eswa.2015.04.071.