Item-Based Collaborative Filtering in Movie Recommendation in Real time

Mukesh Kumar Kharita
Computer Science & Engineering
National Institute of Technology
Hamirpur
Hamirpur, India
mukeshnithcse@gmail.com

Atul Kumar
Computer Science & Engineering
National Institute of Technology
Hamirpur
Hamirpur, India
atul.kumar.prajapati02@gmail.com

Pardeep Singh
Computer Science & Engineering
National Institute of Technology
Hamirpur
Hamirpur, India
avagaman@gmail.com

***Abstract*— Collaborative filtering is one of the most effective and adequate technique used in recommendation. The fundamental aim of the recommendation is to provide prediction of the different items in which a user would be interested in based on their preferences. Recommendation systems based on collaborative filtering techniques are able to provide approximately accurate prediction when there is enough data. User based collaborative filtering taechniques have been very powerful and success in the past to recommend the items based on user's preferences. But, there are also some certain challenges such as scalability and sparsity of data which increases as the number of users and items increases. In a large website, it is difficult to find the interested information in a certain time. But the recommendation system filter out information and items that are best suitable for us. Although there are different recommendation approaches, yet collaborative filtering technique is very popular because of the effectiveness. In this work, movie recommender system has been described, which basically uses item-based technique of collaborative filtering to provide the recommendations of items, which is dynamic and will learn from the positive feedback.**

Keywords—Movies Recommendation, Collaborative Filtering, Item Based Collaborative Filtering

I. INTRODUCTION

The World Wide Web is increasing at a tremendous rate and hence, the complexity and the size of different websites are also increasing. So finding information/data/items in these websites are very difficult, challenging and also time consuming. So any system or website can be made in such a way that it can recommend the items that the users have interests and the items that users are searching for. These recommender systems are able to improve any website for individual users and they do it by adding hyperlinks dynamically. The main purpose of these hyperlinks is to give suggestion for the interesting items to the user i.e. the items that he like and similar to the items he is looking for and thus boosting the communication between the user and the system. A user can browse for interested items through a website. These Interested items may be any kind of textual information, it can also be an index on a particular subject that a user is looking for. A website can be recognized as a compilation of these interested items [16].

Information filtering is the process of selecting those information from a large existing data set that a user is searching or a user may find interesting. Based on the given data (training data or training set) a user model is able to classify the input data that have unseen output, into two classes, a positive class P (significant to the user) or a negative class -P (insignificant to the user). These training set/training data consists those movies that users have already given some rating. These ratings from user to user vary for a given movie. In the given dataset, each item can be expressed as combination of different n components. These component represents items as a vector Y=(y1, y2, y3,…..yN). The components can be a nominal, numerical or binary attributes. These attributes are derived either from the properties of the items or from the knowledge about the user's preferences. In Item-based collaborative filtering in movie recommendation cosine similarity matrix are generated which contains the predicted ratings of the movies. From these predicted ratings top rated movies are selected and be recommend to the particular user. The recommendation task has to decide which movie should be recommended and which should not by classifying these items into classes.

II. EXISTING WORK

Jinbo Zhang et al;[1] presented a model where he first solved Cosine based Similarity, Adjusted Cosine Similarity, and Correlation based Similarity. Then they presented the algorithm that can find the neighbor or similar item more accurate. This optimized algorithm can solve the pinholes or drawbacks of the fundamental item based collaborative filtering algorithm.

Jun Wang et al;[2] proposed the algorithm that unify the item based and content based collaborative filtering. They represented a fusion framework that is able to solve the prediction problem more accurate and also solve the problem related to the data sparsity and new user problem.

The research on these filtering techniques and algorithms are going on from so many years. The different applications uses these systems for filtering different information such as movies, documents, articles and items from large data sets.

Tapestry was the first person who built the recommendation system that uses filtering technique based on collaboration. In this system, he first gathered the user's rating given by the users explicitly or implicitly. After this using the collaborative approach he provided the rating to an intended user.

A recommender system named Ringo video is also a web application that recommend the movies, music, videos to the intended users.

Group lens has also developed a system that recommend movies and news etc. which uses item based collaborative approach.

There are so many other different techniques that have been developed to implement a recommendation system. These techniques include various diverse fields of clustering,

data mining and Bayesian Network Methodology. These methodologies work effectively that construct a model using large amount of data that is training data and then use this model for the prediction of the output for new input. Model, constructed using these methodologies, works fine and effectively.

III. DATASETS

The dataset that we are using in recommendation is MovieLens dataset. Basically, it is gathered and handled by GroupLens Organization. For our engine, we used the data set containing more than a million ratings. These ratings have been given by 943 users for more than 1682 movies[20].

Rating.csv file contains the ratings given to different movies by different users in the format as UserID: MovieID: Rating: timestamp in which the range of MovieID is in between 1 and 1682 IDs. Ratings of the movies vary from 0 to 5. Here in our project we are not using the timestamp which is in the Rating.csv file. The user has to give ratings to at least 20 movies to run algorithm. Properly for the recommendation.

IV. IMPLEMENTATION OF ITEM BASED COLLABORATIVE FILTERING IN MOVIE RECOMMENDATION

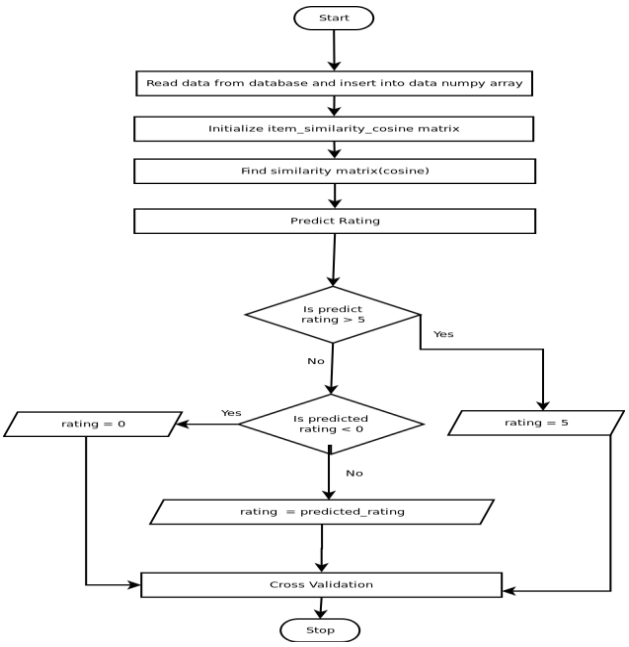

Fig.1. Flow Chart of Item-Based Collaborative Filtering

These systems are used in plentiful number of problems and application such as Documents, books, movies, songs, articles, and items. We can develop a recommendation engine to the user based on his/her preferences and past history.

Here, we have provided the theoretical and implemented method of the item based collaborative technique in movie recommendation.

The rating given to the similar items is looked in the collaborative process. In case of user-item based collaborative filtering technique, we find the similar users for a particular user to whom the system is going to recommend. While in process of item-item based collaboration we have used the ratings of those movies that are so much similar to the rating of the movie, given by intended user, by using the item similarity weights in the item similarity weight matrix. And then selecting the K most similar items with higher ratings and recommend these movies to the intended user.

TABLE I. MOVIE-USER RATING MATRIX

Movies/Users	UserA	UserB	UserC	UserD
MV1	?	5	?	5
MV2	?	0	5	?
MV3	?	?	5	1
MV4	4	5	?	1
MV5	5	5	?	1

Above table shows the relation between the movies and the user. The '?' field represent that particular user in that column has not given any rating for the movie in that row. However the rating lies from 0 rating represent that user has not liked that movie and the rating 5 represent that user has liked that movie so much. Here each movie and each user will be represented by particular unique ids.

TABLE II. MOVIE-USER BINARY MATRIX

Movies/Users	UserA	UserB	UserC	UserD
MV1	0	1	0	1
MV2	0	0	1	0
MV3	0	0	1	0
MV4	1	1	0	0
MV5	1	1	0	0

Above matrix is the representation between the movies and the users, called the binary matrix, where the field with entry 0 represents that the user in that column has not yet rated the movie in that row.

V. COMPUTATION OF SIMILARITY MATRIX

The similarity matrix is formed that consists of the weights, which plays an essential role in the technique of item based approach in collaborative filtering. The similarity matrix consisting the values that are the weights that represents the relation between two items. Higher the value, strong the relation between the items. In movie recommendation engine, modified cosine similarity matrix is generated. This matrix finds the items with similar weights.

$$S(\mathrm{j},\mathrm{k}) = \frac{\sum_{u\in U}\left(R_{(\mathrm{u,j})}-\bar{R}_u\right)\left(R_{(\mathrm{u,k})}-\bar{R}_u\right)}{\sqrt{\sum_{u\in U}\left(R_{(\mathrm{u,j})}-\bar{R}_u\right)^2}\sqrt{\sum_{u\in U}\left(R_{(\mathrm{u,k})}-\bar{R}_u\right)^2}} \qquad (1)$$

Where S(j, k) is the similarity weight value with movie j and movie k which is calculated by using above given equation no.(1). R(u, j) represents the rating provided by an intended user u to the particular movie j and value R(u, k) is

the rating provided to movie k by the same user. Here R_u vector represents the average of the rating provided by the user u to different movies in the matrix of user-movie.

VI. PREDICTION OF UNKNOWN RATINGS

Here we intend to give the prediction of those movies to the current users for which he has not yet given ratings. Here we use weight matrix for the prediction of items and then by selecting k similar weights with highest values. This calculation is done using the formula given below .

$$r_{(x,i)} = \frac{\sum_{j \in N(i;x)} S_{(i,j)} \cdot r_{(x,j)}}{\sum_{j \in N(i;x)} S_{(i,j)}} \quad (2)$$

Here, In equation (2) $S_{(i,j)}$ represents similarity weight value between movie i and movie j in the similarity weight matrix. $r_{(x,j)}$ is the rating provided to the particular item j by the current user x.

VII. RECOMMENDATION OF TOP K ITEMS

As we have find those movies which are similar to movie rated by the current user by using similarity weight matrix. Among these items we select the top k most similar items and hence these are the items which our model will recommend.

VIII. COMPARISION

In our study, we propose a novel approach for recommendation systems in a very optimal & speedy manner. Since, we are using only movie rating as the principle component of the feature set, the derived system is quite optimized and real-time analytics is also possible. The recommender model is quite dynamic and is able to regenerate fresh recommendations based on the updates done by the subscribers in real-time. Even though the model accuracy is somewhat low as compared to contemporary recommendation models, yet the real-time recommendations is what make our research standout from the rest.

IX. CONCLUSION

In this study, we propose a novel approach for recommendation systems in a very optimal & speedy manner. Since, we are using only movie rating as the principle component of the feature set, the derived system is quite optimized and real-time analytics is also possible. The recommender model is quite dynamic and is able to regenerate fresh recommendations based on the updates done by the subscribers in real-time. Even though the model accuracy is somewhat low as compared to contemporary recommendation models, yet the real-time recommendations is what make our research standout from the rest.

The work, item-based movie recommendation system has been described which uses the rating of movies as its feature. It uses the modified cosine similarity matrix to find movies needed to recommend. The root mean square error (RMSE) that we get in the experiment is 1.01. We give the rating out of 5. So the accuracy is 79.72% in the algorithm of item based collaborative filtering.

REFERENCES

[1] Jinbo Zhang, Zhiqing Lin, Bo Xiao and Chuang Zhang, "An optimized item-based collaborative filtering recommendation algorithm," 2009 IEEE International Conference on Network Infrastructure and Digital Content, Beijing, 2009.

[2] Jun Wang , Stephen Robertson , Arjen P. Vries , Marcel J. Reinders, "Probabilistic relevance ranking for collaborative filtering", Information Retrieval, Volume 11, Issue 6, pp. 477-497, Springer, December 2008.

[3] Schafer J. B., Frankowski D., Herlocker J., and Sen S., "Collaborative filtering recommender systems," Lecture Notes In Computer Science, vol. 4321, pp. 291-324, 2007.

[4] Mohammad Yahya H. Al-Shamri , Kamal K. Bharadwaj; "A Compact User Model for Hybrid Movie Recommender System " International Conference on Computational Intelligence and Multimedia Applications, 2007.

[5] Christina Christakou, Leonidas Lefakis, Spyros Vrettos and Andreas Stafylopatis; "A Movie Recommender System Based on Semi-supervised Clustering ", IEEE Computer Society Washington, DC, USA, 2015.

[6] Mukherjee, Rajatish, Neelima Sajja, and Sandip Sen. "A movie recommendation system–an application of voting theory in user modeling." User Modeling and User-Adapted Interaction 13.1, pp. 5-33, 2003.

[7] Ahn, Shinhyun, and Chung-Kon Shi. "Exploring movie recommendation system using cultural metadata." Transactions on Edutainment II, Springer, Berlin, Heidelberg, 2009. 119-134

[8] Miller, Bradley., et al. "MovieLens unplugged: experiences with an occasionally connected recommender system", Proceedings of the 8th international conference on Intelligent user interfaces, ACM, 2003

[9] Lekakos, George, and Petros Caravelas. "A hybrid approach for movie recommendation", Multimedia tools and applications36.1, pp. 55-70, 2008.

[10] Deng Ai-lin, Zhu Yang-yong and Shi Bo-le, "A collaborative filtering recommendation algorithm based on item rating prediction", Journal of Software, 2003

[11] John S. Breese, David Heckerman, Carl Kadie.Empirical Analysis of Predictive Algorithms for Collaborative Filtering. Proc of the 14 th Conf on Uncertainty in Artificial Intelligence. Madison. WI: Morgan Kaufmann Publisher, 1998

[12] Sarwar B, Karypis G and Konstan J., "Item-based collaborative filtering recommendation algorithms", Proceeding of the 10th International World Wide Web Conference, 2001.

[13] Wu Yan, Shen Jie, Gu Tianzhu, CHEN Xiaohong, LI Hui, ZHANG Shu. Algorithm for Sparse Problem in Collaborative Filtering. In: Application Research of Computers. Vol.24, Issue 6, June, 2007.

[14] Gong Song-jie and Ye Hong-wu,"An Item Based Collaborative Filtering Using BP Neural Networks Prediction", International Conference on Industrial and Information Systems, 2009.

[15] Kwok-Wai Cheung and Lily F. Tian, "Learning User Similarity and Rating Style for Collaborative Recommendation", Information Retrieval, 2004

[16] Xiaoyuan Su and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", Advances in Artificial Intelligence, Hindawi 2009.

[17] Intro to Recommender Systems: Collaborative Filtering, 2015. Available: https://www.ethanrosenthal.com/2015/11/02/intro-to-collaborative-filtering/

[18] Rahul Katarya and Om Prakash Verma, "An effective collaborative movie recommender with cuckoo search", Egyptian Informatics Journal, pp. 105-112, Volume 18, Issue 2, July 2017.

[19] Machine Learning. Available: https://www.coursera.org/learn/machine-learning; Retrieved November 10, 2017

[20] Movie Lens | Group Lens Dataset Available: https://grouplens.org/; Retrieved October 20, 2017.