International Journal of Engineering & Technology, 7 (3) (2018) 1504-1507

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i3.11630
Research paper

Collaborative filtering-based recommendation of online social voting

Dr. Mohammed Ismail [1] *, Dr. K. Bhanu Prakash [1], Dr. M. Nagabhushana Rao [1]

[1] Professor, Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation, AP, India
**Corresponding author E-mail: mdismail@kluniversity.in*

Abstract

Social voting is becoming the new reason behind social recommendation these days. It helps in providing accurate recommendations with the help of factors like social trust etc. Here we propose Matrix factorization (MF) and nearest neighbor-based recommender systems accommodating the factors of user activities and also compared them with the peer reviewers, to provide a accurate recommendation. Through experiments we realized that the affiliation factors are very much needed for improving the accuracy of the recommender systems. This information helps us to overcome the cold start problem of the recommendation system and also y the analysis this information was much useful to cold users than to heavy users. In our experiments simple neighborhood model outperform the computerized matrix factorization models in the hot voting and non hot voting recommendation. We also proposed a hybrid recommender system producing a top-k recommendation inculcating different single approaches.

Keywords: *Matrix Factorization; Nearest Neighbors; Recommendations; Recommender Systems; Social Voting.*

1. Introduction

The Internet rein has arrived and is ruling the needs of the human. From booking a flight ticket to ordering vegetables everything has a website or an application to make use with. In such vast and soon developing period, customers decision making ability is also getting effected. So, the main aim of any business website let it be with e-commerce, or a movie website is to provide the users with accurate information by using an efficient recommender system

Recommender system is basically an algorithm used to provide the user or the customer an accurate suggestion of the product or a movie review/rating they have been looking for [1]. This is basically done in two different ways, one of which includes suggesting a relevant item based on the user's history which is his/her previous activity related to it (Personalized method). Another one is the non-personalized method which can be described as the seasonal sale that is prediction based on stock availability.

Many Recommender systems basically work on collaborative filtering. Collaborative filtering is one of the best method and a backbone method of today's social recommender systems [2]. In this fast-growing era, there comes the problem of big data because of the growing users of the internet [3]. So, to give an accurate prediction to the user, any recommendation technique used should search the whole database and find an accurate prediction, which is a typical task. Collaborative filtering works in a similar manner but in a different way, where the database search is done with respect to user's previous activities, find similarities with the other database by using suitable algorithms and then find a top prediction for the user. In this way, it also paves a way to a vast research area with many arising complexities. Improvising these methods will be a great advantage as this is used by many leading business websites [14] like Amazon, Book my Show etc.

When coming to the websites like book my show, it not only enables us to book movie tickets but also lets us know what rating it has gained. These ratings are given based on a individual's voting or opinion. But if we consider a case as an example where only a single user or a less number of users have given the review as good or above average, the overall review delivered would be a good or above average because only some users have rated it. In this way an accurate prediction is not delivered to the end user. So, this paper proposes a hybrid collaborative system, which calculates the movie overall review by comparing the individual's review based on previous activities, based on the comparison of the users with the others who gave similar ratings and compares the individual ratings to all other people's rating for the very same movie and ranks it based on Top 10, Top 20 and Top 50. The review is based on hot voting and cold voting where the hot voting is based on the user's participation in giving the rating and a cold user who rarely gives the review. We show that simple meta path-based NN models outperform computation-intensive MF models in hot-voting recommendation, while users' interests for nonhot voting can be better mined by MF models. Also, this paper proposes a method to know whether a user is likely to watch a movie or not based on the k-nearest neighbor algorithm.

2. Background

Recommender systems makes use of many filtering algorithms for giving a accurate prediction. Some of the frequently used algorithms are Random Prediction algorithm, Frequent Sequence Algorithm, Collaborative Filtering, Content Based Filtering, Demographic Filtering etc. [4].

Random Prediction algorithm suggests the top product by picking up the products without any criteria. Therefore, this algorithm accuracy depends only on luck, as they are more chances of failure.

Frequent Sequence algorithm predicts using the user's purchase history where in which it calculates the frequently bought items and calculates the similarity and gives the suggestion as the top product

Copyright © 2018 Dr. Mohammed Ismail et. al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Demographic Filtering is the filtering method, where prediction is made based on the personal information like name, age, sex, etc. This is best used to solve cold start problems where the recommender system has no user history to start with.
Content Based Filtering is one such method where the recommendations are based on the information about the product and predict it to the user who has a similar need. [5]
Collaborative Filtering is the method which filters not only based upon the user's past activity but also compares it with the several other users who gave similar type of rating and then give an accurate prediction. [6]
They are two types of collaborative filtering techniques where the prediction is done as:

1) User based collaborative filtering
2) Item-based collaborative filtering

User based collaborative filtering which is also called as memory-based filtering recommends the products based on the user's likes and interests. This is done by analyzing two or more users who are having the similar tastes. For example, let us consider an example of three users User 1, User 2, User 3, two (User 1 and User 3) of which looks for the same product to purchase and user 1 likes a product which he is looking for. So, this product is recommended for user 3, as two of them are searching for the same product. In this way multiple users who are having similar interests the top products are suggested. [7]
Item Based Collaborative filtering also called as model based collaborative filtering where the similar between products/items is analyzed and recommended. For example, let us consider four products Product A, Product B, Product C, Product D. When analyzed, it showed that product A and product C are similar, and Product B and Product D are similar. So, when a user likes a product A, he is recommended with product C and vice-versa. Similarly, when a user likes product B, he is recommended with product D and vice versa. In this way the products are analyzed and recommended to the users. [7]
Although collaborative filtering is one such method with some advantages, they are some challenges and issues with it.
The main challenge faced is the cold start problem where if the user has no activity before that is a new user, then the prediction might not be too accurate [8]. To eliminate this problem the recommendations are given based on some user's personal information like age, sex etc (demographic filtering). Another major issue faced for the filtering techniques is the scalability problem. As the users are increasing day by day, data gets increasing due to which efficiency of the techniques gets decreased. For this issue to be encountered several filtering techniques are applied together. [9]Sparsity is the major problem that can occur in tremendous amount of data as when two items are similar only for some features, the algorithm doesn't look the other features, which can be missing for some[10]. Collaborative filtering avoids this issue by forming a profile for the similar neighborhood users.

3. Related work

Ofer Arazy et al [11] have proposed a personalized recommender system of hybrid collaborative filtering architecture. The research also contains some information about what key recommendation features have been implemented. For example, with the movie lens data set, shared preferences and trust relationship-based movie recommendations have been established. They proposed a main framework where to calculate the profile similarity they have taken the consumption history of the source and the receivers. Through social network they have derived the trust propagations and the social network analysis. By observing the ratings of recommendations, they have calculated the reputation mechanisms and by the online communications they have calculated the interaction frequency which then gives the total tie strength. Through profile similarity calculation they have obtained the shared preferences data and through the trust propagation trust percentage has been calculated. The source reputation is calculated by using data obtained through the social network analysis and reputation mechanisms. All the obtained assets (Tie strength, Trust, Source Reputation and shared preferences) gives them a source's qualification component through which an accurate prediction is generated. The model has maximum accuracy but slightly effects the privacy of the user because they alleviated the cold start recommendation problem by considering the user personal data where maximum users refuse to give it.
Chacho Chen et al [12] showed today's social recommender system doesn't fully use the social trust relationships, which could lead to an inaccurate recommendation. Therefore, they have used the social trust relationships in two ways, Implicit and Explicit ways. The metrics of the social recommender system has been combined with trust relationships through a probabilistic matrix factorization method. They have used the trust relationship in an implicit way by combining the user rating item matrix with social trust network by shared user latent space. Then the preference of the user is combined with his/her trusted friend for a final rating. A gradient is then performed on both the ratings and then give the final accurate recommendation based on the social trust. This research encourages the use of social trust relationships by using them implicitly and explicitly simultaneously. The accuracy has been improved because of the usage of the gradient method which reduces the number of iterations. The system is trained to recommend by learning the social trust ensemble. Though the social trust relationships are a key element many other elements such as trust propagation and comments because problems like data sparsity can be surmounted by using the trust propagation and through contents in the comments, analysis of the users can be estimated.
Anahita Davoudi in [13] proposed a unique way to improve the accuracy of the recommender system by considering the elements like similarity, importance of the user and social trust relationships. They have used the Epinions dataset to prove their proposed model. They considered user-item matrix and item-item matrix and performed a gradient descent on the objective function to find the difference between the actual ratings and the predicted ratings. The elements of similarity, user importance was used to model the parameter of trust. This model also minimized the errors of the squared sum between the actual and the predicted rating. The results where compared by the absolute error technique where through the results has been proved that the model is producing more accurate results. As the data inputs increase for a model, the potential risk for attacks will also increase. This is done so by including the fake profiles, which is the profile injection attack. The model has also proposed a way, which could detect the fake profiles by observing the ratings given by the user. Either a user gives a full rating or a very low rating, which is termed as extreme rating behavior. Through this the rating is also used to detect the anomalies
[14] Many of the social recommender systems group the neighbors who are similar in terms of ratings and leave the other profiles. Due to this some important information Is lost. So, this research tackled the above problem, by considering the similar profiles even though they don't have rated the target item. They considered the similar user with respect the active user and predicted the rating prediction of the neighbor by the proposed rating algorithm. A similarity metric is proposed and calculated for the similar user profile and the rating is predicted. A memory based collaborative filtering approach is used. A weighted average of the rating is calculated to predict the rating of the active user. Two parameters rating similarity, social similarity has been calculated where the similarity degree is predicted by their ratings given in the past. The social similarity is calculated through the social trust relationships by social network trust indicators and metrics such as closeness etc. The neighbor selection is a two-fold as proposed. Firstly, the similar users of the active user are considered as the neighborhood set and Next a modularity-based community algorithm is applied to identify the nodes. A higher accuracy compared to the other models has been achieved.
[15] Social information present in the social information network is not fully useful in the social recommender systems.
Due to this scalability will be an issue. So, to overcome this problem this research has taken only informational friends into account.

They proposed six methods which improve the impact of social media information in the social recommender systems. The most informational friends have been taken y considering the friends with similar preferences or by filtering the user which have less preferences in common. The first method was clustering and filtering the users with same similarity, where only similar users in the same cluster are treated as informational friends. The second method is about considering a static threshold where a fixed limit of top friends ae considered for similarity. The third method includes having the informational friends with dynamic threshold where the variable limit of users is taken according to the friends of the active user but with maximum similarities than the targeted user. The fourth method included informational friends through the entropy partioning technique where the average error value of the informational friends is calculated. Fifth method included having the weighted friend impact is calculated by taking the immediate friends and performing the Pearson correlation and the weighted average technique. The last method includes calculating the informational friends through social influence is calculated in this method other the immediate friends the friends with similar tastes have been considered. The grouping is done based on the same information along with item acceptance. By performing all the methods, the results have been compared and then proved that the method of calculating the social information impact on the social recommender system can be accurately calculated by the method of dynamic threshold by with Pearson correlation.

4. Used database description

The dataset used for the experiment is MovieLens Dataset [16]. Movie lens dataset, which is collected by the GroupLens, which is a research lab in the department of computer science and engineering belonging to University of Minnesota. The dataset is divided into three files movies.dat, ratings.dat and users.dat. The users.dat file consists of 6050 users, which consists of five columns namely User-id, Gender, Age, Occupation, and Zip Code. The movies.dat file consists of columns 'Movie ID', 'Movie Title', 'Genre'. The ratings.dat file consists of 'User ID','Movie ID', 'Rating', 'Timestamp'.

To visualize the results MATLAB2013a has been used. The data was separately loaded as movies, ratings and users through the MYSQL software.

5. Proposed model

The proposed architecture is as shown in the figure below.

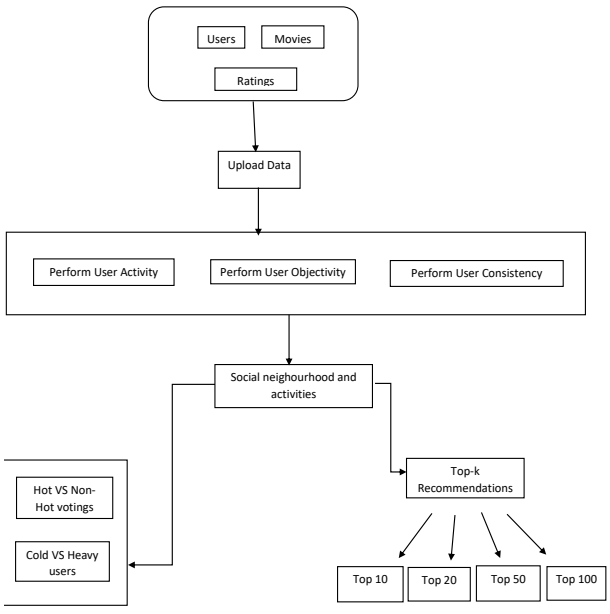


Fig. 1: Proposed Architecture.

As shown in the above figure the proposed system first includes uploading the movies, ratings and users into the database. We could ensure whether the data really got uploaded or not by logging into the database and giving the necessary queries. The uploaded dataset is trained accordingly. A graphical User interface (GUI) which enables us to perform User Activity: On clicking this button, the users are compared and resulted according to the ratings they give which is basically a one to one comparison.

User Objectivity: When this activity is performed, a single user's rating is compared to a bunch of users who are having a similar profile (similar preferences) and resulted, a one to many comparisons.

User Consistency: This activity lets us know how consistent a user is based on the previous activities he/she performed and compared to the other ratings which also helps us to know whether he is a fake person or not. This can be quoted as many to many relationship comparisons.

Social neighborhood and activities: The main part where the users are clustered where k-means clustering is used, and the user's activities are compared with the other users and resulted. Once the k-means clustering is done, we can predict whether a person is most likely to watch the movie or not.

On performing the related activities, we can use the data resulted from the above activities either for comparing the hot voting vs non-hot voting and cold vs heavy users or for getting the top -k recommendations [17].

Hot voting vs non-hot voting: Hot voting and Non-hot voting is the term used how much the movie is rated and compared with the other movie's rating, which helps us to know whether the movie is in trending and how many people rated it as Top movie list.

Cold vs Heavy users: Cold users are the people who do not have a profile, a random guest who rated the movie or a person who does not review often whereas the heavy user can be termed as the user who reviews more often and is a good user of the website. All such users are compared to obtain a top-rated movie.

Top-k recommendations: The top k-recommendations are given based on the result of all the activities performed above. The obtained result from the user objectivity, user consistency and user objectivity give us a brief information about the user behavior in every perspective and helps the model to obtain a valid input for the next step, which is social neighbor, and activities where from here the users are clustered based on similar preferences and similar ratings. After this based on a constant ($0<k<1$), the ratings are clustered for four clusters namely Top 10, Top 20, Top 50 and Top 100. After performing the Top-k Recommendations we can predict a single user whether he/she is likely to watch a movie or not by taking in the information such as his age, gender, occupation, movie name, and the zip code. When this information is obtained from the user, the result is based from the information which is already clustered. If the user occupation is student and likes a genre, based on such demographic information and previous activities, the proposed model predicts the likeliness of the user to watch the movie or not.

6. Results

The below table is the sample result of top 10 movie recommendations which are sorted and ranked based on the rating which is above 0.08 which is incorporated as recommended. The result includes user-id, gender, movie-id, rating given and the movie with its genre. These were based on a constant k and clustered and derived limited to 10. We also performed the likeliness of the individual user based on the overall recommendation and appended it to the result. The overall results of ranking the Top 10, Top 20, Top 50, and Top 100 can be visualized in a graph as shown below. The graph is plotted by taking the rating on X-axis and the no of users on y-axis. As the graph, shows based on the users taken the movies are ranked and plotted on a scale of [5] rating.

Table 1: Top 10 Recommendations

USER ID	GENDER	MOVIE ID	RATING	MOVIE NAME	GENRE	Likeliness
14	M	3354	[1.3775]	Mission to Mars (2000)	Sci-Fi	Most Likely
11	F	1753	[1.3340]	Half baked(1998)	Comedy	Not Most Likely
18	F	2987	[1.3217]	Who framed rogger rabbit(1988)	Adventure, Animation	Most Likely
16	F	2987	[1.2302]	Who framed rogger rabbit(1988)	Adventure, Animation	Most Likely
5	M	2987	[1.1298]	Who framed rogger rabbit(1988)	Adventure, Animation	Not Most Likely
12	M	1252	[1.0896]	Chinatown(1974)	Mystery,Thriller	Not Most Likely
4	M	3468	[1.0519]	Hustler(1961)	Drama	Not Most Likely
2	M	1357	[0.9976]	Shine(1996)	Drama,Romance	Not Most Likely
3	M	3421	[0.9753]	Animal House(1978)	Comedy	Most Likely

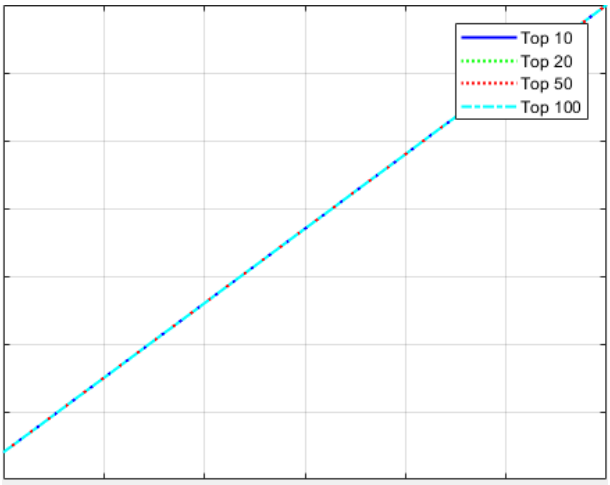


Fig. 2: Graph Plot Showing Respective Recommendations

7. Conclusion and future work

We can conclude that the recommendations are made not only based on a single user for a new movie, but the overall review for that movie not only based on the single user rating but provide a rating based on the similar users to that user and the user previous activities. The likeliness of the user is also tested and obtained. Through this, we are able to build a personalized hybrid recommender system. Our future work includes validating the user profiles, which helps to develop a more secure recommender system.

References

[1] Soghra Lazemi Hossein Ebrahimpour-komleh ,Improving Collaborative Recommender Systems via Emotional Features ,IEEE, Application of Information and Communication Technologies (AICT), (2016), IEEE 10th International Conference.

[2] Ashish Pal, Prateek Parhi and Manuj Aggarwal, An Improved Content Based Collaborative Filtering Algorithm For Movie Recommendations, IEEE, Proceedings of 2017 Tenth International Conference on Contemporary Computing (IC3),10-12 August2017,Noida,India.

[3] Manami Kawasaki, Takashi Hasuike, A Recommendation System by Collaborative Filtering Including Information and Characteristics on Users and Items , IEEE, Computational Intelligence (SSCI), 2017 IEEE Symposium Series.

[4] Samundeeswary K, Vallidevi Krishnamurthy, Comparative Study of Recommender Systems Built Using Various Methods of Collaborative Filtering Algorithm, IEEE, 2017 International Conference on Computational Intelligence in Data Science (ICCIDS).

[5] Mohammed Ismail. B, B. Eswara Reddy, T. Bhaskara Reddy, Cuckoo Inspired Fast Search Algorithm for Fractal Image Encoding, Journal of King Saud University Computer and Information Sciences November 2016.

[6] Surong Yan, Kwei-Jay Lin, Xiaolin Zheng, Wenyu Zhang, and Xiaoqing Feng ,An Approach for Building Efficient and Accurate Social Recommender Systems using Individual Relationship Networks ,IEEE, IEEE Transactions on Knowledge and Data Engineering (Volume: 29, Issue: 10, Oct. 1 2017).

[7] Kolla Bhanu Prakash "Mining issues in traditional Indian web documents", Indian Journal of Science and Technology, 2015.

[8] Luis G. Perez, Francisco Chiclana, Samad Ahmadi,A Social Network Representation For Collaborative Filtering Recommender Systems, IEEE, Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference, PP:438 – 443.

[9] Yingtong Dou, Hao Yang, Xiaolong Deng, A Survey of Collaborative Filtering Algorithms for Social Recommender Systems ,IEEE, Semantics, Knowledge and Grids (SKG), 2016 12th International Conference, PP:40 – 46.

[10] Mohammed Ismail, C.H.S.L. Sowmya, T. Sai Sudheera, P Ravi Teja ,A Comparative Study on Dealing with Sparsity in E-Commerce, International Journal of Pure and Applied Mathematics Volume 118 No. 5 Jan 2018, 185-194.

[11] Ofer Arazy, Nanda Kumar, *Ci* Bracha Shapira,Improving Social Recommender Systems, ,IEEE, IT Professional (Volume: 11, Issue:4,July-Aug.2009

[12] Chaochao Chen, Jing Zeng, Xiaolin Zheng, Deren Chen Recommender System Based on Social Trust Relationships IEEE, e-Business Engineering (ICEBE), 2013 IEEE 10th International Conference. PP: 32 – 37.

[13] Anahita Davoudi, Effects of User Interactions on Online Social Recommender Systems, IEEE, Data Engineering (ICDE), 2017 IEEE 33rd International Conference. PP: 1444 – 1448.

[14] Fahimeh Ebrahimi, S. Alireza Hashemi Golpayegani Personalized recommender system based on social relations ,IEEE, Electrical Engineering (ICEE), 2016 24th Iranian Conference,PP:218 – 223.

[15] Mohammed Ismail. B, Dr. T. Bhaskara Reddy, Dr. B. Eswara Reddy "Hybrid Fractal Image Compression Based On Range Block Size" Graphics, Vision and Image Processing Journal, ISSN 1687-3988, Volume 16, Issue 2, December. 2016 Pages: 23-32.

[16] Movie Lens Dataset https://grouplens.org/datasets/movielens/.

[17] Kolla Bhanu Prakash, Dorai Rangaswamy M.A. and Ananthan T.V. (2014), "Feature extraction studies in a heterogeneous web world", International Journal of Applied Engineering Research, Vol.9, No. 22, pp- 16571-79.2014.