IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 07, Issue 03 (March. 2017), ||V1|| PP 61-65

Social Recommender System Using Skyline Query and Reciprocal Scoring Technique

Poonam Rani[1], Ayudh Kumar Gupta[2]
[1](Assistant Professor, Computer Engineering Department, NetajiSubhas Institute of Technology, India)
[2](Undergraduate, Computer Engineering Department, NetajiSubhas Institute of Technology, India)

Abstract: - With the internet being flooded with various social networking sites, e-commerce websites, video streaming sites, news websites; a naive user would want to access relevant data which could be possibly of his/her interest and to achieve this, services such as a recommender system is required, which uses not only user's previous history of interests but also recommend social elements which are trending among similar users. Social recommender systems use a combination of social data and attributes including user's profile bio, likes, and dislikes, purchase history, reviews and other information. Since user information on various social networks might include information such as locations provided by the user, using this information and the data existing in the data base, it might actually be possible to suggest user with a much more apt recommendation. The proposed work uses skyline query algorithm which uses modified reciprocal scoring as an argument to recommend items to the users.

Keywords: - *Reciprocal Scoring, Recommender Systems,Similarity Measures, and Skyline Query*

I. INTRODUCTION

To avoid getting swamped by exponentially increasing irrelevant data, Recommender systems have played a major role for personalizing user experience over a variety of platforms by providing users optimized suggestion for news, other people, media content etc. Recommender Systems are a special class of filtering tools which filter out and recommend data by predicting the preference of an item that a user would give to the item which they had not yet considered. These systems can also be classified broadly in two categories viz. Content based and Collaborative filtering based. Content-based recommender systems are mainly used to recommend items based on the interrelationship and the user's interests. On the other hand, Collaborative filtering based systems predict and recommend items on the basis of similarity of interests between users and social data items.
1.1 Major applications of recommender systems: 1.1.1 Recommender systems find their application in social networking sites for suggesting other users to a target user. For Example, Facebook's friend suggestions are nothing but an implementation of social recommender systems.1.1.2. Recommender systems have also been used in various popular websites such as IMDB, Netflix, Amazon, Google News etc. i.e. where item recommendations are required. Item recommendations can be done by two different ways either on the basis of user's previous history or on the basis of what is trending among other users.

This paper focuses on the recommendation systems used in the field of social networking which is basically friend recommendations. Social recommendation takes social relations such as trust relations or friendships as input and provides recommendations as output. Friend recommendation can be done by a number of ways. These include methods such as FOF (Friend of Friend), SONAR, skyline query method [1] etc. The most common method is a Friend of Friend method, which gives friend suggestions on the basis of the number of mutual friends. Some other methods also use location of a user to suggest other users. These methods treat the social network as a graph where each user is represented as a node and the links between these various nodes represent that they are friends with each other. Each link has some attributes attached with it such as the number of mutual friends, distance between them etc. With the recent development in smart phone industry, popular shopping and social networking sites are just a click away from millions of users. These users provide all sorts of useful information which could be used, if analyzed properly, to provide suitable recommendations to them. Recommender systems can have a variety of use cases on various platforms including social networking sites, e-commerce websites, dating sites etc. The output of the recommendation function used by these systems could vary from suggesting a friend on a social network to suggesting products on a shopping website. The major application of these systems is however in the field of social networks and dating websites. Social networks not only help users to maintain their social relationships, but to find new people and explore new experiences. Therefore, potential friends recommendation function is necessary for each social networking site, and the output of this function should vary from user to user. With enhancements in global positioning systems (GPS), a new trend of location-based recommendations has also come into light. Popular social networks such as

Facebook and Foursquare, provide users to share their current location on their social network by using a feature called check-in, which uses these services to access the users's location after the user has granted these services the permission to access their location. The main idea behind the following proposed work is the amalgamation of the users's location based information (places user likes to visit frequently), number of mutual friends, and the reciprocal scoring method which uses similarity measures between two users based on their interests to provide recommendations.

II. PROBLEM DESCRIPTION

Social recommender system is a tool that can help users to extract high valued and personalized information among massive data sources. The main goal of using this information is to personalize user experience and provide unique recommendations. Recommender systems can be categorized into two classes viz. Content based systems and Collaborative filtering. Content-based methods compare the target item and target user with the help of similarity measures that measure the similarity of the recommended item to the similar types of items that a user likes or dislikes. Whereas, collaborative filtering, filters users with comparable interests as the intended users based on their past experiences. The system then makes suggestions to the intended user depending on their likes and dislikes.

2.1. Issues faced in the past in the field of Recommendation Systems:
Even after all the advancements in technology, these systems still have to go a long way to overcome various challenging issues. Some of them are as follows:
2.1.1. There is scope for further optimization on the accuracy of predictions.
2.1.2. Algorithms for recommender systems may need to overcome various bottlenecks. For example, to calculate item similarity, content-based systems use only discrete material item descriptions and not on abstract items such as ideas, thoughts or sentiments.
2.1.3. The time complexity of the algorithm needs to be improved for a much smoother experience.
2.1.4. A lot of data might be required to provide better recommendations.
2.1.5. As the size of the data which needs to be processed increases, processing time complexity may also worsens.
2.1.6. User's preferences keep on changing so, it's difficult to monitor these changes.

The idea is to propose an algorithm that could overcome the above mentioned problems and provide comparatively more accurate recommendations. To tackle these problems, few papers have been consulted and a hybridization of some algorithms used in them is suggested. The following work proposes the amalgamation of using Skyline query with a modified reciprocal scoring algorithm.

III. RELATED WORK

3.1. Recommendation Systems: A lot of research has been done in the field of recommender systems, which has shaped the level of personalization in user experience found in present services. Various approaches have been implemented and worked upon in the past including Friend of Friend, Content Matching, MinHash etc. The basis on which these systems recommend items to users vary from users behavior, preferences, opinions, location history, interests, profile etc. Since the following proposed work is based on Collaborative Filtering, we focus only on this field. In this type of recommender systems, instead of considering entire set of data items, only those items are considered which the target user might relate to. The following includes the introduction of filtering based works which rely on user locations and similarity scores.3.2. Local Social Network: With advances in GPS and internet, naive users can share their location information to social networks, which has led to the development of social networking location sharing services e.g Foursquare. Users can easily share their experiences; write reviews about locations and places with their social network through various applications of popular websites like Facebook, Zomato, and Foursquare etc. on their smart phones. User check-in information can be used to provide suitable recommendations if analyzed properly.3.3. Skyline Query: In the consulted paper [1], a Location Based Social Network friend recommendation has been proposed which uses skyline query. Skyline query is used to find optimal results when there are multiple constraints. For example, One needs to find a cheap hotel near the sea, but as the distance between a hotel and the sea decreases, the prices also increases, to get the most optimal hotel, Skyline query can be used. In the described method, not just common friend number is considered to recommend new friends, but also the user's check-in information to get the distance between two users is also considered, finally it uses the Location Based Social Network friend recommendation skyline query to search optimal friend for the user by keeping the number of mutual friends to be high and distance between two users to be low. Skyline query used in this method ensures that number of mutual friends between the target and candidate user is as high as possible while simultaneously keeping the distance factor as low as possible.

3.3.1. Advantages of using Skyline Query : Location based social network system can help not only in recommendations based on friends of friends approach but also on other parameters like distance between two people and friend influence among them. Users can easily share their experiences about locations on their social networks just through a click on their smart phones. User's check-in information is an important piece of information that can be used to reflect behavior and make recommendations accordingly. The paper consulted, uses the concepts of collaborative filtering that reflect common user interests and contribute in making recommendation more accurate. Skyline query is an approach to select a limited number of good items from a large database, and in some recommendation systems, skyline query is chosen to select the candidate friend for user.3.3.2. Disadvantages if using Skyline Query : Information from various sources are considered in recommendation systems for Location Based Social Networks, including user's profile, user's likes & dislikes, and past check in histories. This involves processing of huge volumes of data in real time. The major challenges to be addressed in Location Based Social Network recommendation are location-context awareness i.e. heterogeneous domain and rate of change in preferences. A User may not visit his place of interest as his likings may change. User preferences keep changing over time so it is difficult to monitor and analyze the changes to make new recommendations.3.4. Research in the Online Dating Field: A lot of research in the field of machine learning and reciprocity patterns has been done in the past few years. A machine learning based approach is proposed in the consulted paper [2] to find more suitable romantic partners via recommendations. When a user signs up for a dating profile, they are required to enter details regarding interests, age, height, income, education etc. Reciprocal Similarity scores are then calculated between the user and the candidate users. The candidate with a higher reciprocal score would have an edge over the candidate with lower reciprocal score in terms of recommendation. Recommendations were given on the basis of discrete explicitly defined parameters.

3.4.1. Advantages of Reciprocal Similarity: The consulted paper elaborates the concept of social recommendation systems to online dating sites to look for potential romantic partners. It enables greater communication among recommended users which have similar interests. Unlike the traditional user-item recommendation systems, which predict a user's opinion towards material items such as books or movies, the paper considers a variety of set of attributes to provide most suitable recommendations. The algorithm uses an array of user information available in the database that can be utilized for better predictions. These characteristics include user's interests, age, sex, demographic details (height, weight, body build), education background, hometown, occupation etc. In addition to this, algorithms consider a number of user's preference specific information like a user's preference for potential partners such as age, body build etc. making the process of recommendation more efficient.3.4.2. Disadvantages of Reciprocal Similarity: Algorithm used in this paper is able to provide new and reactive users with recommendations well enough using only the one-way preferences of the candidates. It does not fair so well in recommending cold start users to old users. Secondly, it is not easy to explain to a user how their actions have led to the presented recommendations because it is collaborative filtering which is affected by opinions of other users as well and not only on a single user. It does not maintain a history record of users having dated another user. This makes it likely for recommendations to be redundant for active users who are looking for online dates frequently.

IV. PROPOSED SCHEME

The idea of a hybrid proposal consists of using the reciprocal score between two elements as an argument with other arguments, which can be passed in the skyline query. When a user signs up for a social networking site, they need to provide an array of personal details like user' s sex, age, city, demographic details (height, weight), occupation, education level etc., which would be displayed as the bio of the user. Most of this information are categorical features that can fed into an algorithm and then can be used as similarity measures which can further be fed into a recommender system to generate recommendations.4.1. Modified Skyline Query: The Skyline operator is used in a query and performs a filtering of results from a database so that it keeps only those objects that are not worse than any other. Reciprocal Scores can be passed in the skyline query as an argument so as to produce optimal results keeping the priority of recommending based on the distance factor between two users maximized, number of mutual friends and the reciprocal scoring. The distance between two users used here is the geodetic distance between two points and not the Euclidean distance. The locations used here are the three most frequently checked in locations. So the distance can be calculated by just entering the coordinates of two locations in the geodetic distance formula. The friend list of a user is implemented using a set, so the mutual friends between two users can be found out by taking the intersection of the friend list set of these users. Finally, distance, number of mutual friends and the reciprocal score are passed as the parameters in the skyline query. The aim of using the skyline query is to filter out recommendations, keeping the distance between the users low, number of mutual friends high and reciprocal similarity score also high at the same time.4.2. Modified Reciprocal Score: Each candidate user can be considered as a vector of various discrete attributes whose value depends on the users profile. This information can have a wide range of domain ranging

from integers to some other discrete values representing characteristics of the user including demographic features (age, height etc.), income, interests etc.

Algorithm for calculating modified reciprocal scoring by modifying consulted algorithm [2]:

//initialise values

TempSim(x,y)=TempSim(y, x) = 0.0 (1)

For each User u in Friends(y):

TempSim(x,y)=TempSim(x,y)+ Sim (x, u) (2)

For each User v in Friends(x):

TempSim(y,x)=TempSim(y,x)+ Sim (y, v) (3)

// normalize both temporary scores

If |Friends (y)| > 0 then, //represents number of friends of User y

TempSim(x,y)=TempSim(x,y)/ |Friends (y)| (4)

If |Friends (x)| > 0then, //represents number of friends of User x

TempSim(y,x)=TempSim(y,x)/ |Friends (x)| (5)

// calculate reciprocal score

If TempSim (x, y) > 0 &TempSim (y, x) > 0

Return2/(TempSim(x,y)^{-1}+ TempSim (y, x)^{-1}) (6)

Else return 0.0

Where, Friends (a): Set of friends in User a's friend list.

Sim (a, b): Similarity score between User a and User b.

Int(a) : Set of discrete values characterising User a's Interests.

Mathematically,Sim(a,b) = (No.of fields in|Int(a)**cap**Int(b)|) / (Total No. of fields in Interests vector) (7)

Also, **cap** implies the intersection of sets on which it is being implied.

Assuming U signifies the candidate friend list from which candidates can be recommended to users, dist denote the distance between user and candidate friend,CF denote no. of common friends between user and candidate friend and RCS denote reciprocal similarity score between user and candidate friend. For two candidate friend and in the set U, candidate is said to have higher preference over the other candidate friend, if is better or equal to in all three dimensions than and at least better in one dimension. The lower the distance, the higher the number of mutual friends and the higher the reciprocal similarity score is considered better.

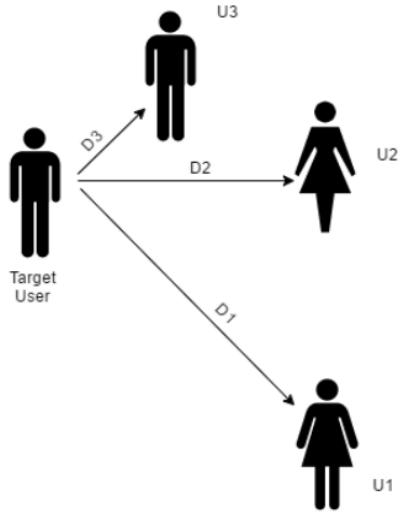

Figure 1. Example of application of proposed scheme

With reference to Fig.1, assume a target user, say T, and three candidate users (with equal number of mutual friends with the target user) which can be recommended to the target user. Assuming Reciprocal scores of U1, U2 and U3 be R1, R2 and R3 respectively. Also, assume that the distance between Target user and candidate users be D1, D2, D3. Let R1, R2 and R3 be calculated on the basis of similarity measures such that R2>R3>R1. Also, Let D1>>D2 and D2>D3. Since D1 is the farthest, the recommendation priority would be the least; also R1 is the lowest, which implies that U1 would be the least suitable recommendation out of the candidate users. Now, considering the remaining users, if we consider the users U2 and U3 assuming D2 and D3 are approximately equal but D2 slightly greater then D3, so the recommendation priority of U3 would be greater and U3 would be recommended according to previous works, but according to the proposed work, the skyline query would consider the reciprocal scores too and would recommend U2, since it needs to maximize the priority (by minimizing distance and maximizing number of mutual friends, which in this case are equal) while

simultaneously keeping the reciprocal scores maximized. The proposed scheme ensures that recommendation is not only made on the basis of location distances or number of mutual friends but also on the similarity of interests between two users. Since accessing user home locations can be a breach of security, check in information provided by the user can be used to infer priority. The reciprocal similarity score is calculated using a parameter for location similarity, friend influence, and content similarity, taking into consideration the influence of characteristics including interest, height, age, education level, income etc. of a user. The former parameters are taken from the algorithm suggested by skyline query method and the latter one from reciprocal score is passed into the skyline query.

V. CONCLUSION

By making use of a combination of modified reciprocal scoring along with skyline query, there shall be a synergy, wherein the results are much better had the algorithms been used separately. This proposed scheme can be extended to the analysis of data, linked across multiple social networking sites and an exhaustive evaluation of community and enterprise social networks. It can also be used to discover and predict trends of social media sites. We can address scalability concerns, improve accuracy and performance, and get results at a faster rate.

REFERENCES

[1] ShufengZheng, AsifZaman, Yasuhiko Morimoto, Friend Recommendation by Using Skyline Query and Location Information , Bulletin of Networking, Computing, Systems, and Software, Volume 5, Number 1, pages 68–72, January 2016

[2] Peng Xia, Benyuan Liu, Yizhou Sun, Cindy Chen, Reciprocal Recommendation System for Online Dating , *arXiv:1501.06247v2* [cs.SI], January 2015