A Friend Recommendation System based on Similarity Metric and Social Graphs

Rashmi. J, Dr. Asha. T
Department of Computer Science
Bangalore Institute of Technology, Bangalore, Karnataka, India
rash003.j@gmail.com, asha.masthi@gmail.com

ABSTRACT-All the existing social networking services recommend friends to users based only on their social graphs, which is not very appropriate in reflecting user's preferences in selecting a friend in real life. In this paper, we present a friend recommending system for social networks, which recommends friends to users based on their life styles as well as social graphs, as the proposed friend recommending mechanism is being integrated into social network. By taking the advantage of sensor-rich smart phones, the proposed system discovers the life styles of users , measures the similarity of life style existing between the users, calculates the friend recommendation score using the proposed similarity metric, and recommends friends to query user who are having high friend recommendation scores. Since the proposed system is integrated into the social network, the existing feature of social network i.e. the social graphs is used for recommending social friends to the user. Therefore on receiving the request, the proposed system returns a list of people with high friend recommendation scores as well as a list of social friends to the query user.

Keywords – *Friend Recommendation, Life Style, Social Graphs, Social Networks*.

I. INTRODUCTION

Twenty years ago, people typically made friends only with the people who lived close to themselves such as neighbors or colleagues. The friends made through this fashion are termed as G-Friends, which stands for geographical location-based friends as they are influenced by the geographical distances between them. The rapid advances in the social networks, services such as Facebook, and Twitter have provided us revolutionary ways for making friends. According to the statistics of the Facebook, a user has an average of 130 friends [1].

One challenge residing in the existing social networking services is recommending a good friend to the user. Most of the existing friend recommending systems relies on pre-existing user relationships to suggest friend candidates. For example, Facebook relies on social graphs to recommend friends to the user, i.e. users who share same geographical location or same profession are recommended as friends to the user, which is not very appropriate in reflecting user's preferences in selecting a friend.

According to the studies [2] and [3], the basic rules for grouping people together are: 1) life styles;2) attitude;3) interests;4) moral standards;5) economic level;6) already known people. Most of the existing friend recommendation systems consider rule #3 and #6 as the main factors for recommending friends to users. Our proposed system considers rule #1, #3 and #6 as the main factors for recommending friends to users. Life styles are correlated with daily routines and activities performed by the people. The life style of the people comprises of activities such as shopping, travelling, playing sports, swimming, listening to music, watching TV etc. This proposed friend recommendation mechanism is deployed as an add-on to the existing social networking services, hence making it as a hybrid friend recommendation system which utilizes both the social graph feature of the existing social networking service and the similarity metric feature of the proposed system.

II. LITERATURE SURVEY

Recommendation systems that suggest items to the users have become popular in the recent years. For example, Amazon [4], recommends items to the user based on their previous visit and the items that are frequently visited by the other users. Netflix [5] and Rotten Tomatoes [6] recommend movies to the users' based on previous users' ratings and habits of watching.

Over the recent years, with the advances in the social networking services, friend recommendation has gained a lot of attention. The existing friend recommendation systems like Facebook and Twitter recommend friends to user based on their social relations.

In the meantime many other recommendation systems have been proposed by researchers. Bian and Holtzman [7] have presented a collaborative friend recommendation system called as MatchMaker that is based on personality matching. Kwon and Kim [8] have presented a friend recommendation system that is based on physical and social context. But the authors have not explained what a physical social context is and how to obtain that information.

These existing friend recommending systems are different from our proposed system. In our work, we exploit the recent sociology findings to recommend friends based on their similar life styles as well as social relations. The advance of smart phones enables activity recognition using the set of sensors on smart phones.

III. SYSTEM OVERVIEW

This section gives the high-level overview of the friend recommendation system. Fig. 1 shows the architecture of the proposed friend recommendation.

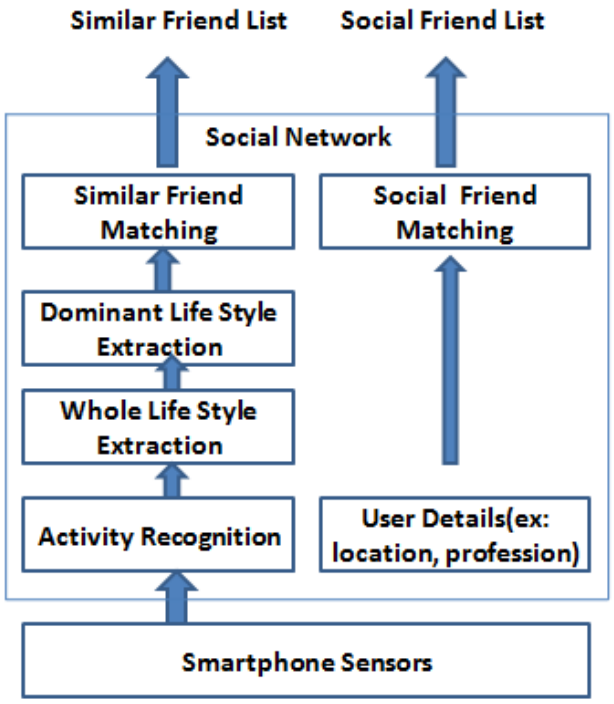

Fig. 1. System architecture of Friend Recommendation system.

In the activity inference phase, the activity of each user is recognized that is collected from the smartphones. The activities of the users are collected for a certain period of time. In the life style extraction phase, the users‘ whole life style and the dominant life style are extracted. From the activities recognized in the activity inference phase, the whole life style of the users are extracted i.e. the set of activities that are performed both frequently and infrequently in a given period of time, and are added to the MySQL database.

The whole life style activities are then given as input to the apriori algorithm which then computes the frequently performed activities that represents the dominant lifestyle for the given user and . In the friend matching phase, the dominant life of the query user is compared with all the other users and the no. of matching activities are compared, and using the proposed similarity metric a friend recommendation score is computed.

The computed friend recommendation score for each user exceeding the defined threshold value represents a friend to the query user with high similar life style. In the social friend matching phase, the profession and the geographical location details of the query user is compared with other users, the users details matching with the query user are recommended as social friends to the query user as they are social related. The following sections will elaborate on all the modules of the proposed system.

ACTIVITY RECOGNITION

The life styles are a mixture of motion activities performed by the user in the daily life. The sensors on the smart phone are used for inferring user‘s motion activity. Since the number of activities involved in the analysis is unpredictable, unsupervised learning approach is used for organizing the activities. K-means clustering algorithm is used for grouping data into clusters, each cluster representing an activity. Since the raw data collected by the smart phones are noisy, median filter is used for filtering the noisy data. The cluster centroids are calculated and distributed to the smartphones. The smartphones then recognize the activity based on the minimum distance rule and uploads the sequence of activity to the server instead of raw data.

We have the implemented the activity recognition phase of the proposed as a website consisting of several urls‘. Here the urls‘ represent the activities performed by the user in the daily life. Here we have considered activities like shopping, travelling, listening to music, watching TV, cooking etc. Each url is represented using a integer. The url and its associated integer value is added to MySQL database. The users registered with application can login to this website. Once the user logs into the website, he/she visits the url of his/her choice. An activity of the user is recognized when he/she visits the url, representing an activity or set of activities performed by the user in his/her daily life.

The following table shows how the url and its associated integer value is stored in database.

Table. 1. Activities and their corresponding id‘s stored in the database.

Activity_id	Activity
1	http://www.soundcloud.com
2	http://www.imdb.com
3	http://www.google.com/shopping
4	http://www.booking.com
5	http://www.webmd.com
6	http://www.howstuffworks.com
7	http://www.wired.com
8	http://www.yahoo.com/tech
9	http://www.indeed.co.in
10	http://www.bigfishgames.com
11	http://www.allrecipes.com
12	http://www.break.com

WHOLE LIFE STYLE EXTRACTION

Since life style is a combination of activities performed by the user in his/her daily, in our implementation urls‘ visited by the user in the given session represents the life style of the user. In real life, the activities of the user are observed for certain number of days. In our implementation, the activities are tracked for many sessions, so that the life style of the user can be predicted accurately. The urls‘ representing the activities of the user, when visited by the

user is added to the database along with its session id. This is done for all the users for tracking their life style. The life styles tracked in the above specified way are termed as whole life style of the user, as they are a combination of both frequently and infrequently performed activities.
The following table shows how the whole life style of each user is stored in the database.

Table. 2. Whole life style activities of each user stored in the database.

User_id	Tracked_activities	Session_id
22	1-6-3-4-5	00233A78EE888888B88113
31	1-9-2	0456CC899AABB22899EE81
28	2-4-5	06445EE733666FF556A77BC
24	2-6-7-8	100125672891AADDD7777B
30	1-2-4-5	167755AABBEE8902263772
22	2-3-4	186777DDBBA8999E77655F
23	1-3-4	198767E5678F7666A77889B
25	2-5-8-9	20008953738EFD67A89B998
22	1-4-5-6	2001037467624882784FFED
26	2-4-5-6	21763758632BAECFFDE8488
25	1-3-5-7-8	222873485678EFAD8745876

DOMINANT LIFE STYLE EXTRACTION

To calculate the similarity of life styles between the users, only the whole life style activities of the user cannot be used, as they are a combination of both frequently and infrequently performed activities. To determine the dominant life style of the user, only the activities performed frequently by the user must be considered. Hence the dominant life style of each must be computed. Once the whole life style of the user is obtained, those set of activities are given as input to the Apriori algorithm. The application of the Apriori algorithm is to compute the frequent set of items i.e. the set of items occurring frequently for the given set of items. In the proposed system, the whole life style is treated as the given set of items, which then computes the frequently occurring item sets i.e. in the proposed system the algorithm computes the activities that are frequently performed in a given period of time. The set of frequently performed activities obtained represent the dominant life style of the user. We have considered a support of 30% in algorithm for computing the frequent item sets i.e. the frequently performed activities.

i. Apriori Algorithm

- For each item,
  - Check if it is a frequent itemset
  //appears in > minimum support transactions
  - k=1
  - repeat
    //iterative level-wise identification of frequent itemsets.
  - For each new frequent itemset Ik with k items
    //level k+1
  - Generate all itemsets Ik+1 with k+1 items, Ik is a subset Ik+1
  - Scan all the transactions once and check if the generated k+1 itemsets are frequent
  - k=k+1
  - Until no frequent itemsets are identified.

The following screenshots show how the dominant life style for one user is calculated.

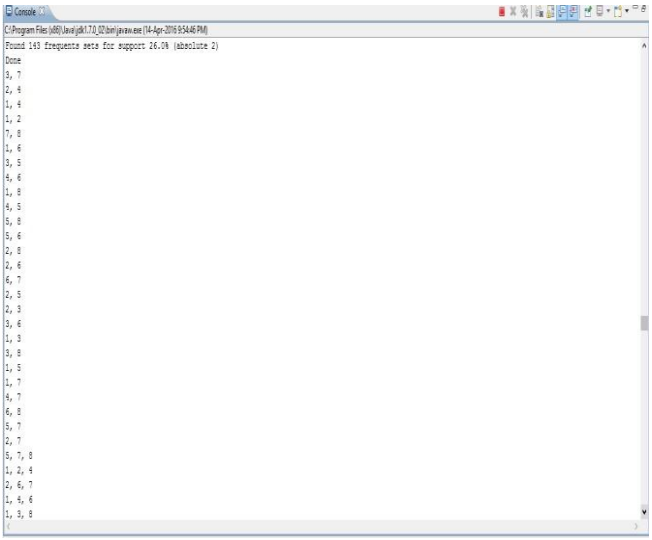

Fig. 2. Frequent set of activities being computed using Apriori algorithm.

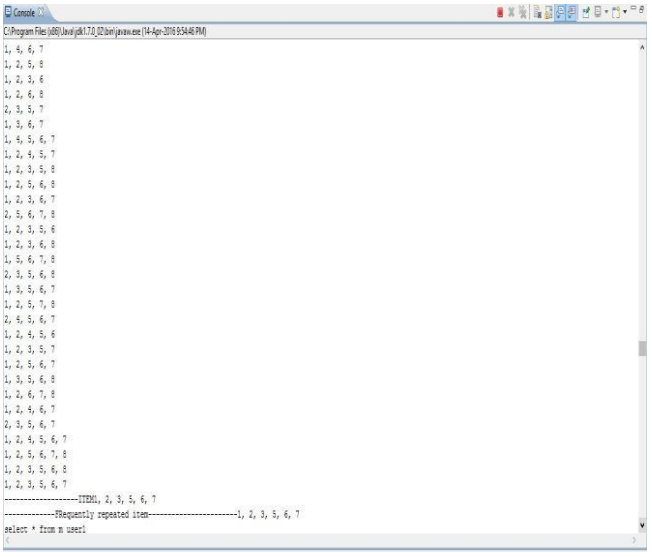

Fig. 3. Dominant life style computed for the given user using Apriori algorithm.

The following table shows the computed dominant life styles for all the users.

Table. 3. Dominant life style computed for each user stored in the database.

User_id	Dominant_life_style
22	1 ,2 ,5 ,6, 7
23	1 ,2 ,5
24	1 ,2 ,3 ,5 ,6
25	1 ,2 ,3 ,5
26	1 ,2 ,5 ,6, 7
27	2 ,6 ,7
28	1, 5, 6, 9
29	3, 5, 6, 7
30	2, 6, 7, 8
31	1, 3, 4, 5
32	8 ,9, 11

SIMILAR FRIEND MATCHING

Once the dominant life style of the all the users are obtained by the Apriori algorithm. The dominant life styles of all users are compared with query user‘s dominant life style. From the life style comparison, parameters like the no of activities matching with each user and total life style match value are obtained.

The proposed similarity metric computes the friend recommendation score for each user using the above values obtained on comparison. A threshold value is defined for the friend recommending system. The list of users whose friend recommendation scores exceed the predefined threshold value are recommended as friends sharing similar life style with the query user. Here we have defined the threshold value as 4. Hence all the users‘ friend recommendation scores exceeding 4 are recommended as friends sharing similar life styles. The friends‘ list contains only the names of the users, to preserve the privacy of the users‘ by not revealing the users‘ life style details.

The friend recommendation score is computed using the following the equation:

F_score=matching activities + whole life style match (1)

Where

- F_score: friend recommendation score
- Matching activities: no of activities between the query user and the user considered for friendship.
- Whole life style match: this value is 1 if all the activities match in the life style set matches otherwise zero.

The following table shows the friend recommendation scores that are computed for all the users.

Table. 4. Friend recommendation score computed for each user stored in the database.

User_id	No_of_activities_matching	Whole_life_style_match	Friend_recommendation_score
23	3	0	3
24	4	0	4
25	3	0	3
26	5	1	6
27	3	0	3
28	3	0	3
29	3	0	3
30	3	0	3
31	1	0	1
32	0	0	0

The following screen shots show how friends with similar life styles are recommended.

Fig. 4. Screenshot showing the friends‘ list sharing similar life style with the query user.

SOCIAL FRIEND MATCHING

Social graphs represent the social relationship existing between the people in the graph.The people who share social relations are termed as social friends. Social relations are based on the profession, geographical location, etc. that the people share with others. Already known people are also termed as social friends. Recommending friends to users based on the social relationships is the feature of the existing social networks. Facebook and Twitter also relies on the social graphs for suggesting friends to the users. Since we are incorporating the proposed friend recommending mechanism into the social networks, we are making use the existing social graphs feature for suggesting the social friends to the users along with the friends sharing similar life style. Hence the proposed system behaves as a hybrid friend recommendation system recommending both similar life style friends as well as social friends to the query user.

The following screenshots depict how the proposed system recommends social friends to the query user.

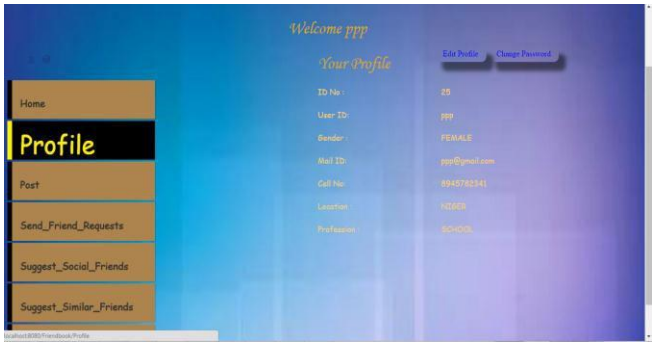

Fig. 5. Screenshot showing the profile details of the query user.

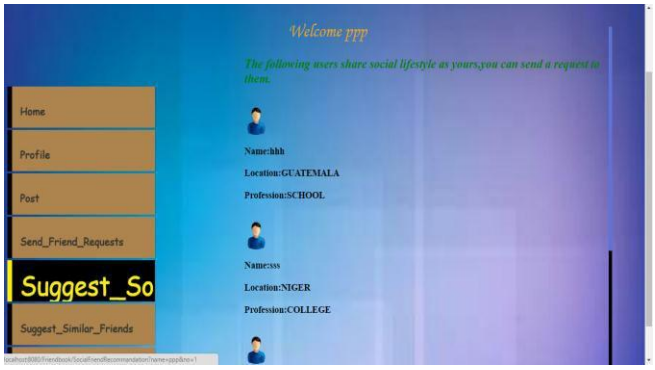

Fig. 6. Screenshot showing the social friends' list for the query user .

CONCLUSION

In this paper, we have presented the design and implementation of a friend recommendation system that is based on similarity metric and social graphs. The proposed system behaves as a hybrid friend recommendation system, recommending both social friends and friends' sharing similar life style to the query user, as it is incorporated in the social networking service. Hence the user is provided with a wide range of choices for selecting a friend for his/her preference. Also privacy is preserved, which is achieved by revealing only the names of friends in the friend list and not their life style details to the query user.

In future, the activities of the users representing their behavior can be kept tracked at the server/admin side. Therefore, if any user is involved in any activities such as crime, then it can be easily identified by their activities that are observed and stored at the server/admin side.

REFERENCES

[1] Facebook statistics. (2011). [Online]. Available: http: // www.digitalbuzzblog.com /facebook-statistics-stats-facts-2011/

[2] G. Spaargaren and B. Van Vliet,–Lifestyle consumption and the environment : The ecological modernization of domestic consumption,‖ Environ. Politic, vol. 9, no. 1, pp. 50–76, 2000.

[3] M. Tomlinson, –Lifestyle and social class,‖ Eur. Sociol. Rev.,vol. 19, no. 1, pp. 97–111, 2003.

[4] Amazon. (2014). [Online]. Available: http:// www.amazon.com/

[5] Netflix. (2014). [Online]. Available: https: //signup-netflix.com/

[6] Rotten tomatoes. (2014). [Online]. Available: http: //www.rottentomatoes.com/

[7] L. Bian and H. Holtzman, –Online friend recommendation through personality matching and collaborative filtering,‖ in Proc.5th Int. Conf. Mobile Ubiquitous Comput., Syst., Services Technol.,2011, pp. 230–235.

[8] J. Kwon and S. Kim, –Friend recommendation method using physical and social context,‖ Int. J. Comput. Sci. Netw. Security,vol. 10, no. 11, pp. 116–120, 2010

Authors Profile

Rashmi.J is currently pursuing M.Tech degree from Bangalore Institute of Technology, Bangalore. She has obtained her B.E degree from B.T.L. Institute of Technology and Management, Bangalore. Her research interests are Social Networking and Web Services.

Dr. Asha.T obtained her Bachelors and Masters in Engineering, from Bangalore University, Karnataka, India. She has her Ph.D from Visveswaraya Technological University under the guidance of Dr. S. Natarajan and Dr. K.N.B. Murthy. She has over 20 years of teaching experience and currently working as Professor in the Dept. of Computer Science & Engg., B.I.T. Karnataka, India. Her research interests are Data Mining, Medical Applications, Pattern Recognition and Artificial Intelligence.