ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 5, May 2016

RECOMMENDER SYSTEM BASED on LIFESTYLES of USERS

Tofik R. Kacchi, Prof. Anil V. Deorankar

***Abstract*— Recommender systems or recommendation systems are a subclass of information filtering system that search for to predict the 'preference' or 'rating' that a user would give to an person/item/place/thing. Depending on these, recommendation system is classified into two subclasses: first one is depending on item i.e. Object recommendation while the second one is depending on person i.e. person to person link recommendation. In this paper, we have chosen second approach of link recommendation to deploy the system to be used for health application by using the different parameters habits or daily-activities, blood-group, location, etc. We have also presented the mathematical model used in this paper with the help of example for analysis purpose.**

***Index Terms*— Activity-bag representation, hadoop platform, health application, recommender system.**

I. INTRODUCTION

Twenty years ago, people typically made friends with others who live or work close to themselves, such as neighbours or colleagues. With the rapid advances in social networks, services such as Facebook, Twitter and Google+ have provided us revolutionary ways of making friends. According to Facebook statistics, a user has an average of 130 friends, perhaps larger than any other time in history. One challenge with existing social networking services is how to recommend a friend to a user according to need.

Most of them rely on pre-existing user relationships to pick friend candidates. For example, Facebook relies on a social link analysis among those who already share common friends and recommends symmetrical users as potential friends. Unfortunately, this approach may not be the most appropriate based on recent sociology findings.

Traditional way of making friends (G-Friend)

- Geographical location based: Neighbours, colleagues
- Pros: be familiar with each other
- Cons: No. of friends is limited

Emerging social networks

- Facebook, Twitter, Google+, etc
- Pros: unlimited no. of friends
- Cons: "Friends" are not the expected friends

What's the fact?

- People tends to make friends with people having similar interests

Existing system recommends based on social graphs which may not be the most appropriate to reflect a user's preferences. So, a new method is devised which recommends friends based on lifestyles of users which may include the following parameters:

1) daily-work/activities 2) blood group 3) location, etc.

Different Problems are:

- How to identify friend candidates based on interests rather than pre-existing relationships?
- How to automatically get one's interests without one's specification?
- How to help people find friends at any time and any place?
- How to measure the similarity of interests between different users?

So, in this paper we have given solution to all above issues. The aim of this paper is: To build Friend Recommendation System by using lifestyles of users to suggest friend as required & system should be used for health application.

Different objectives achieved in this paper are as follows:

- Designed and implemented a friend recommendation system that allows users with similar interests to be quickly identified and recommended.
- To automatically get one's interest without one's specification.
- To quickly find person according to matching required blood group and nearby location.

II. PROPOSED METHODOLOGY

In this system, five modules are designed to fulfil the task of friend recommendation. They are as follows:

1) Data Collection:

Data collection module collects the profile data of users from a file(I/p dataset) which is a semi-structured formatted file. The life styles of users are extracted by these module. The Screenshot of the implementation results of this module is as follows:

***Tofik R. Kacchi**, CSE, Govt. College Of Engineering,.Amravati,India 0 7276 55 7138*

***Prof. Anil V. Deorankar**, CSE, Govt. College Of Engineering,Amravati, India,0 8793 31 6581.*

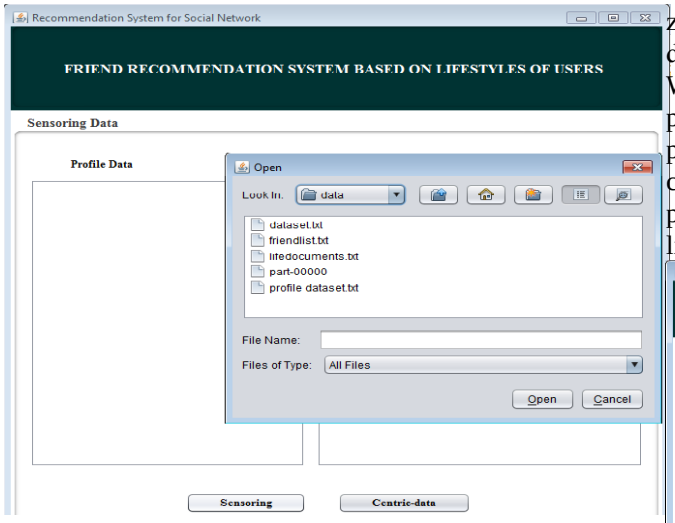

Fig. 1: System collecting sample data

It uses hadoop technology because type of file is semi-structured and for fast processing, as tokenizing and extracting the data from a file is easy & optimal in hadoop using the Map/Reduce tool.

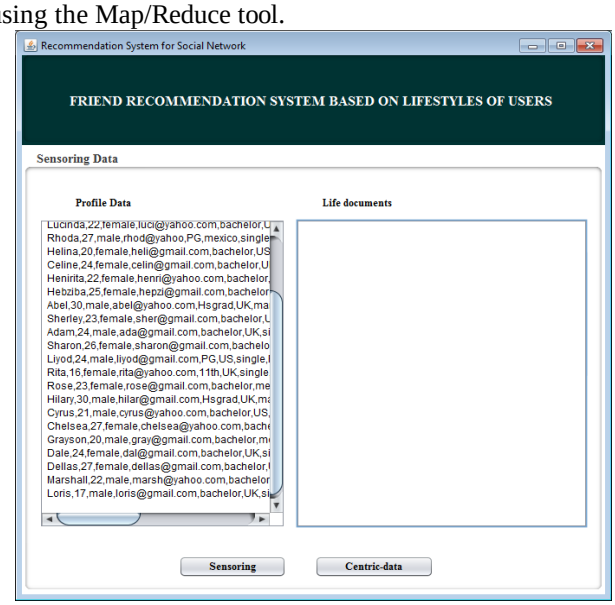

Fig. 2: Sample data collected is retrieved and sent to hadoop platform for life style extraction

2) Data Analysis:

Life document of user is given as input to these module. This module processes the life doc using hadoop technology and prepares its Activity-Bag with the help of knowledge of probability distribution.

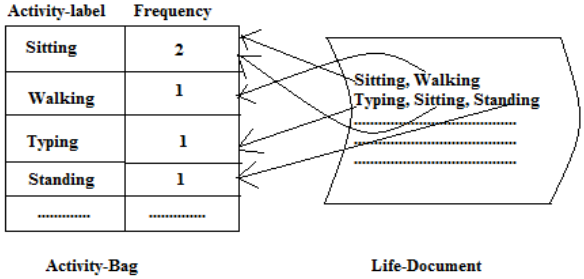

Fig. 3: Activity-Bag representation

Let w = [w1, w2, …….. , w_W] be the set of 'W' Activities

z = [z1, z2, ……. , z_Z] be the set of 'Z' Lifestyles

d = [d1, d2, ……. , d_n] be the set of life-documents

Where n = Number of users and

$p(w_i | d_k)$ is the Probability of activity w_i in a life-doc d_k

$p(w_i | z_j)$ is the Probability of how much the activity w_i contributes to the lifestyle z_j

$p(z_j | d_k)$ is the Probability of lifestyle z_j embedded in a life-document d_k

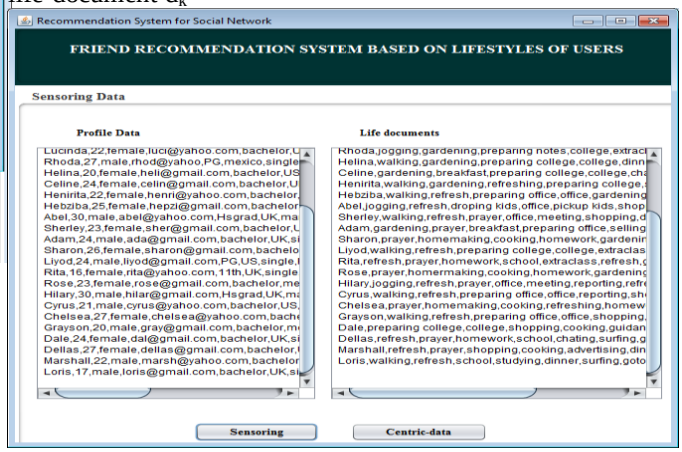

Fig. 4: Life-styles of users are extracted

By observing the following screenshot, we will automatically get one's interest without one's specification.

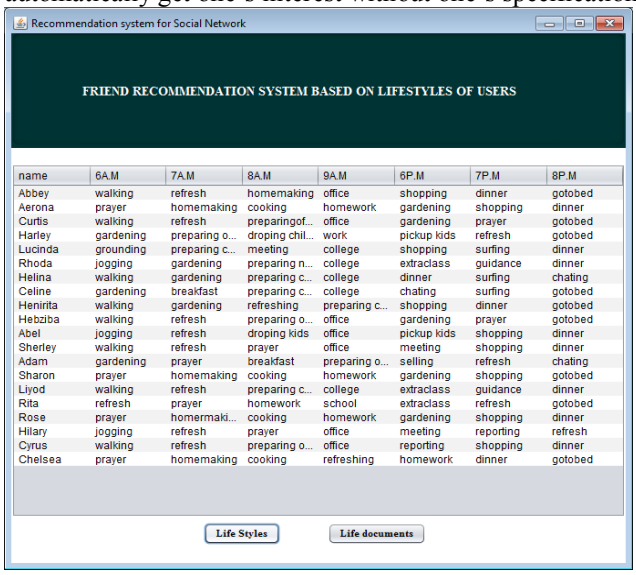

Fig. 5: Displaying Lifestyles of users in structures format

Let $f_k(w_i)$ be the frequency of occurrence of activity w_i in life-doc d_k. Therefore, we can write:

$p(w_i | d_k) = f_k(w_i) / \sum_{i=1}^{W} f_k(w_i)$

3) Indexing:

Then the life style indexing module puts the life styles of users into the database in the format of (life-style, user) instead of (user, lifestyle). The screenshots of the bag of activity representation of given life-document of users is as follows:

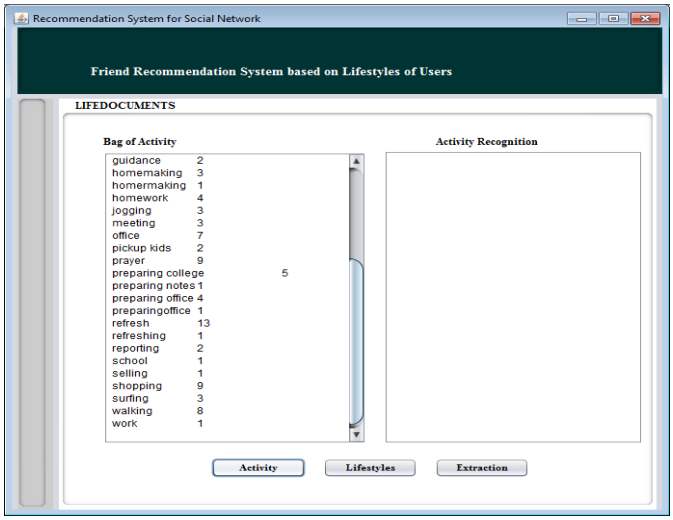

Fig. 6: Created the Activity-Bag Representation

4) Friend-Matching Module:

A friend-matching module can be constructed accordingly with the help of recommendation algorithm presented here to represent the similarity relationship between users' life styles and suggest the list of similar users.

Algorithm Recommendation(L_i , L_j , λ)
{
Step 1: Design Activity-Bag modeling of the user

Step 2: Use the knowledge of probability distribution and calculate lifestyle vectors

$L_i = [\ p(z_1 / d_i), p(z_2 / d_i), \ldots.., p(z_Z / d_i)\ ]$

Similarly we can construct L_j.

Step 3: Calculate cosine similarity between user 'i' and user 'j'.

$Sc(i,j) = \cos(L_i , L_j)$

$= L_i . L_j / |L_i| * |L_j|$

According to these similarity results, it will recommend all user 'j' to the user 'i' depending on the given threshold. This will provide us the list of similar users to a particular user.

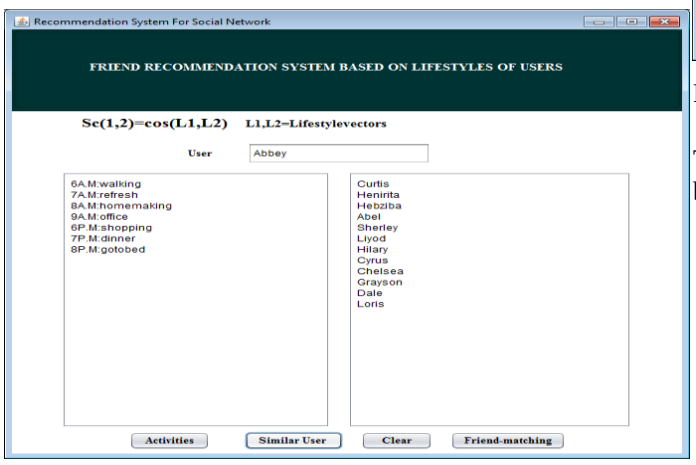

Fig. 7: Friend Matching Module showing Similar Users

Step 4: Sort L_i and L_j in descending order for determining dominant lifestyles.
$Di = \{zi_1, zi_2, \ldots.. , zi_{qi}\ \}$

s.t. $q_i = \arg\min_q (\sum_{k=1}^{q} p(zi_k | d_i) >= \lambda)$ is the size of dominant lifestyles.
Similarly we will find Dj.

Step 5: Now $S_d(i, j) = 2 * | D_i \cap D_i | / | D_i | + | D_j |$
It is nothing but the dominant lifestyle score.

5) Recommendation Module:

The user query module takes a user's query and handover it to the recommendation module to send a list of potential friends to the user as response.
This module is divided into two categories:
1) Recommendation by Similarity of lifestyles
2) Recommendation by using Similarity of blood-group & location

Step 6: So our similarity score is determined as follows:
$S(i, j) = S_c(i, j) \cdot S_d(i, j)$

Step 7: If these $S(i, j) >= S_{thr}$ then Recommend
Otherwise Does NOT Recommend.
}

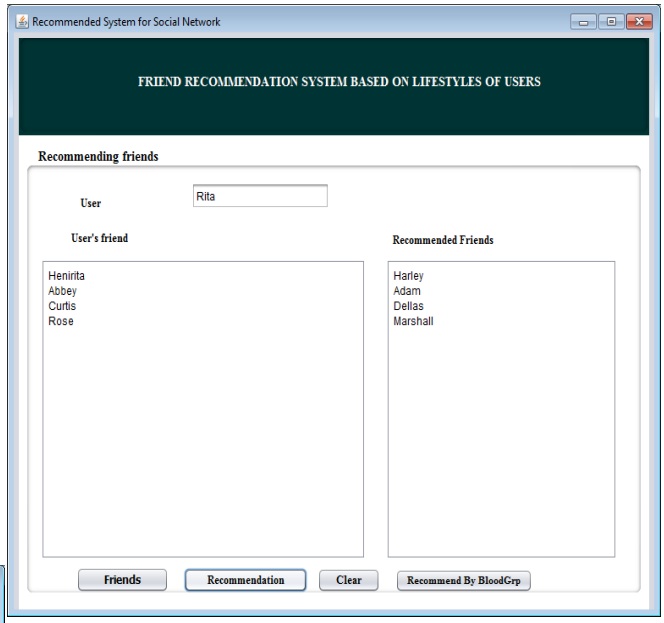

Fig. 8: Recommendation by Similarity of lifestyles

The result of recommendation by matching required blood-group & location is as follows:

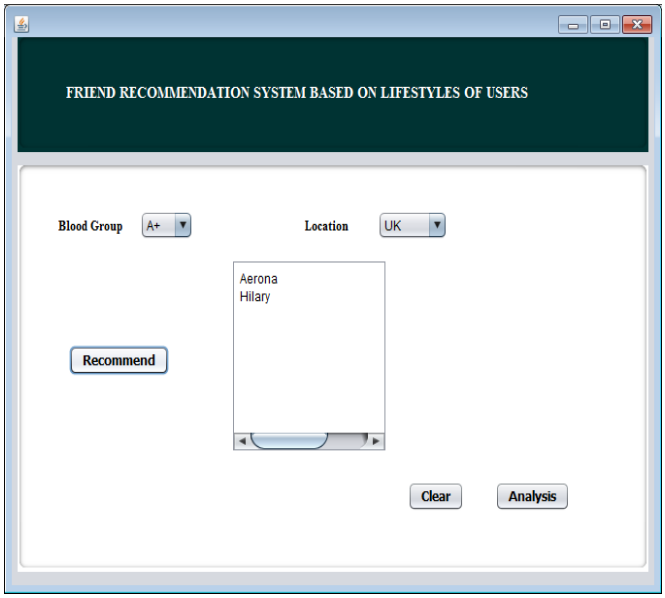

Fig. 9: Recommendation by using Similarity of blood-group & location

III. Mathematical Model and Analysis

In this section we will present the mathematical model and will do the analysis by considering the lifestyle vectors of two users i.e. Li and Lj.

Example:

$\widehat{L_i}$ = [0.3, 0.1, 0.2, 0.3, 0.1] // Life-style vector of user 'i'

$\widehat{L_j}$ = [0.2, 0.1, 0.4, 0.01,0.3] // Life-style vector of user 'j'

We have $Sc(i, j) = \cos(\widehat{L_i}, \widehat{L_j})$

$= \widehat{L_i} \cdot \widehat{L_j} \; / \; |\widehat{L_i}| * |\widehat{L_j}|$

$= 0.18 \; / \; (0.4899) * (0.5477)$

Therefore, $Sc(i, j) = 0.6708 >= 0.5$

So, user 'j' is a similar user to 'i'.

Example:

$\widehat{L_i}$ = [0.3, 0.1, 0.2, 0.3, 0.1] // Life-style vector of user 'i'

$\widehat{L_j}$ = [0.2, 0.1, 0.4, 0.01,0.3] // Life-style vector of user 'j'

$\lambda = 0.8$

We have $Sc(i, j) = \cos(\widehat{L_i}, \widehat{L_j})$

$= \widehat{L_i} \cdot \widehat{L_j} \; / \; |\widehat{L_i}| * |\widehat{L_j}|$

$= 0.18 \; / \; (0.4899) * (0.5477)$

Therefore, $Sc(i, j) = 0.6708 >= 0.5$

So, user 'j' is a similar user to 'i'.

Now after sorting, $\widehat{L_i}$ = [0.3, 0.3, 0.2, 0.1, 0.1]

$\widehat{L_j}$ = [0.4, 0.3, 0.2, 0.1, 0.01]

Therefore, Di = { z_1, z_4, z_3 } as $0.3+0.3+0.2=0.8 >= \lambda$

Dj = { z_3, z_5, z_1 } as $0.4+0.3+0.2=0.9 >= \lambda$

We have

$Sd(i, j) = 2 * |Di \cap Dj| \; / \; |Di| + |Dj|$

Therefore, $Sd(i, j) = 2 * |\{z_1, z_3\}| \; / \; (|\{z_1, z_4, z_3\}|) + (|\{z_3, z_5, z_1\}|)$

$= 2 * (2) / (3) + (3)$

$= 4 / 6$

So $Sd(i, j) = 0.67$

Sc is calculated as = 0.6708

Sd is evaluated as = 0.67

So, S(i, j) = Sc(i, j) * Sd(i, j)

= (0.6708) * (0.67)

= 0.45 >= 0.4

So user 'j' is recommended as a friend to user 'i'. In this way, we can recommend number of friends to a user using lifestyles of users.

IV. Conclusion

Hence, the objectives that are achieved in this paper:

- To design and implement a friend recommendation system that allows users with similar interests to be quickly identified and recommended.
- To automatically get one's interest without one's specification.
- To quickly find person according to matching required blood group and nearby location.

In this proposed method, we have added one content feature of blood group, which is going to enhance the application of recommendation. In this method, we are going to match the blood group of the searched friend and this matching will lead to help the application search meticulously to stretch out to the desired person. In the future, we can add content features like DNA of a person. By this we aim at creating a revolution in the field of recommendations.

REFERENCES

[1] [1] Zhibo Wang, Hairong Qi, "Friendbook: A Semantic-Based Friend Recommendation System for Social Networks," IEEE Transactions on Mobile Computing, Vol. 14, No. 3, MARCH 2015

[2] Pankaj L. Pingate, S. M. Rokade, "A Survey of Friendbook recommendation Services " , International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064, Vol 3, Issue 11, November 2014.

[3] S.Adsure, A.Arane, A.Chavhan, R.Jagdhane, A.Pardeshi, "A Survey on Friendbook Using Semantic based Friend Recommendation System," IJARIIE-ISSN(O)-2395-4396, Vol. 1, Issue-4, 2015.

[4] B. Bahmani, A. Chowdhury, and A. Goel, "Fast incremental and personalized pagerank," Proc. VLDB Endowment, vol. 4, pp. 173– 184, 2010.

[5] 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.

[6] Z. Wang, C. E. Taylor, Q. Cao, H. Qi, and Z. Wang, "Demo: Friendbook: Privacy preserving friend matching based on shared interests," in Proc. 9th ACM Conf. Embedded Netw. Sens. Syst., 2011, pp. 397–398.

[7] Tom White, "Hadoop: The Definitive Guide," Vol. No. 2, OCTOBER, 2010, pp. 15-40.

[8] L. Page, S. Brin, R. Motwani and T. Winograd, "The Pagerank citation ranking: Bringing order to the web " , Stanford InfoLab, Stanford, CA, USA, Tech. Rep. 1999-66, 999

[9] T. Huynh, M. Fritz, ans B. Schiel,, "Discovery of activity patterns using topic model," in Proc. Int. Conf. Ubiquitous Comput., 2008, pp. 10-19.

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

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

[12] Netfix. (2014). [Online]. Available: https://signup.netflix.com/

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

[14] J. Biagioni, T. Gerlich, T. Merrifield, and J. Eriksson, "EasyTracker: Automatic transit tracking, mapping, and arrival time prediction using Smartphones," in Proc. 9th ACM Conf. Embedded Netw. Sensor Syst., 2011, pp. 68–81.

[15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, pp. 993–1022, 2003.

[16] K. Farrahi and D. Gatica-Perez, "Probabilistic mining of sociogeographic routines from mobile phone data," IEEE J. Select. Topics Signal Process., vol. 4, no. 4, pp. 746–755, Aug. 2010.