

# Friend Recommendation in Online Social Networks using LDA

**S. P. Adaikkammai**

*Department of Information Technology  
Karpaga Vinayaga College of Engineering &  
Technology*

**V. Amirtham**

*Department of Information Technology  
Karpaga Vinayaga College of Engineering &  
Technology*

**Mrs. M. Kanchana**

*Department of Information Technology  
Karpaga Vinayaga College of Engineering & Technology*

## Abstract

Existing social network services provide list of friends to users based on their request given. But it will not fulfil the user's preferences in real life. Due to overloaded memory of the server memory size increases and lacking its efficiency. By implementing the Latent Dirichlet Allocation Algorithm we extracting their lifestyles and sensing the similarity of lifestyles between users by using embedded sensors in the smartphones. Based on friend matching graphs we return a list of people with highest similarity of lifestyles. Feedback mechanism is integrated in this friendbook to get the results of users in choosing friends. We have implemented friendbook on the Android-based smartphones and evaluated its performance on both small scale experiments and large scale simulations. Finally, we reduce the memory size of the server and improving its performance.

**Keywords: Fluorescent Microscope Image, Adaptive Segmentation**

## I. INTRODUCTION

In olden days, people made friends in office (or) in working area and with their neighbours only because they get to know them by the Geographical distances between each other. We can't recommend a good friend user by using the existing social network services because fake people (or) duplication can be there and some people lifestyles will not match to the users. In that services, server will contain all the information about the users and get overloaded, while retrieving some information from the server it takes time. The main drawback is, it consuming much amount of time and lacking its efficiency.

## II. RELATED WORK

J.Gunes and M.H.Louis[1] proposed thatan approach to friend recommendation systems by using complex network theory, cognitive theory and a Pareto-optimal genetic algorithm in a two-step approach to provide quality, friend recommendations while simultaneously determining an individual's perception of friendship. J.Biagioni and J.Eriksson[2]GPS Traces collected from smartphones that are installed on transit vehicles to determined routes served,locate stops and infer schedules. K.Farrahi and D.Gatica-perez [3] proposed that they use the two wearable sensors but not smartphones to discover the daily routines. Guo,L.Zhang and C.Fang[4] Based on trace-driven experimental results and security analysis, they have shown the feasibility and privacy preservation of their proposed scheme.Based on trace-driven experimental results and security analysis, we have shown the feasibility and privacy preservation of our proposed scheme.Z.Wang,C.E.Taylor, [5]told that mobile social services utilize profile matching based on the preserving encryption to find friends with similar social attributes. But it is difficult to secure and information leakage problem also obtained. Personality plays an important role in advertisement and marketing environment but to detecting the personality from multiple social networks is difficult. K. Farrahi and D. Gatica-Perez[6] proposed a model, called bag of multimodal behavior, that integrates the modeling of variations of location over multiple time-scales, and the modeling of interaction types from proximity. Their representation is simple yet robust to characterize real-life human behavior sensed from mobile phones, which are devices capable of capturing large-scale data known to be noisy and incomplete. A.D.Sarma and A.R.Molla [7] Used iterative matrix vector multiplication method.X.Yu and A.Pan[8] uses GPS Data to understand the transportation mode of uses.B. Bahmani, A. Chowdhury, and A. Goel[9]considers fast incremental computations of SimRank on link-evolving graphs. .SimRank is an arresting measure of node-pair similarity based on hyperlinks. It iteratively follows the concept that 2 nodes are similar if they are referenced by similar nodes.

### III. PROPOSED SYSTEM

People lifestyles (e.g., attitude, taste, economic level, etc...) are extracted by implementing the latent Dirichlet Allocation Algorithm to create a document. We are using the embedded sensors such as GPS, Accelerometer, Microphone, Gyroscope and Camera in the smartphones. That sensor will measure the similarity of lifestyles between users. We are using the friend matching graph to return a list of friends with highest similarity of lifestyles in an order. In this friendbook there is a feedback mechanism to further improve the recommendation accuracy. Then we are compressing the document size with the help of Hadoop tool, to reduce the memory size of the server and increase the performance of the server. Fig.1. Explaining about the Architecture of the Proposed system that such that Collecting the Data from the user and creating a document, getting user query, comparing lifestyles of the users, provide ranking based on their lifestyles with the help of friend matching graph and providing feedback control.

#### A. Data Collection:

In this we are collecting the user's data such as their habits, lifestyles etc..., from the life document and processing in the server.

#### B. Lifestyle Analysis and Indexing:

We are extracting the lifestyles by implementing the latent dirichlet allocation algorithm and the user's lifestyle are analysed based on their queries. After analysing the lifestyle to be stored in the database. First the user can see the lifestyles of the another user and then only they can see the user name (or) any other detail.

#### C. Friend Matching Graph:

After extracting the lifestyles from the user. We represent the similarity of other user's one by one in a graph construction method.

#### D. User Impact Ranking:

Based on the graph method, ranking method will be displayed according to the user's expectation.

#### E. Feedback Control:

By implementing all those above activities at last we are adding a feedback mechanism to gather the user's query and further improve the friend recommendation process.

#### F. User Query:

In this we are taking the life documents of the other user based on the individual user query and send a list of friends in the ranking order to the individual user.

### IV. MODULES

- 1) Life Style Modeling
- 2) Activity Recognition
- 3) Friend-matching Graph Construction

#### A. User Impact Ranking:

##### 1) Life Style Modeling:

- Life styles and activities are reflections of daily lives at two different levels where daily lives can be treated as a mixture of life styles and life styles as a mixture of activities.
- This is analogous to the treatment of documents as ensemble of topics and topics as ensemble of words.

We represent,

$W=[w_1, w_2, \dots, w_w]$  denote a set of activities ( $W$  is the total number of activities)

$Z=[z_1, z_2, \dots, z_z]$  denote a set of life styles ( $Z$  is the total number of life styles)

$d=[d_1, d_2, \dots, d_n]$  denote a set of life documents ( $n$  is the total number of users)

According to the probabilistic topic model, we have

$$p(w_i | d_k) = \sum p(w_i | z_j) p(z_j | d_k)$$

Where,  $p(w_i | d_k)$  denote the probability of the activity  $w_i$  in a certain life document  $d_k$

$p(w_i | z_j)$  denote the probability of how much the activity  $w_i$  contributes to the life style  $z_j$

$p(z_j | d_k)$  denote the probability of the lifestyle  $z_j$  embedded in the life document  $d_k$

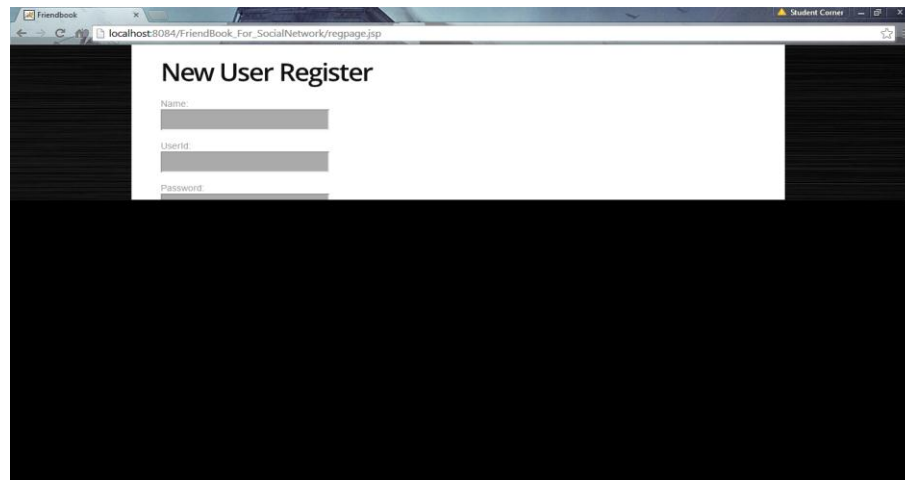


Fig. 1: New User Registration Form.

By using LDA we are extracting lifestyle,

$$p(w|d)=p(w|z)p(z|d)$$

Where,  $p(w|d)=[p(w|d_1),p(w|d_2),\dots,p(w|d_n)]$  is the activity-document matrix.

2) *Activity Recognition:*

We need to first classify or recognize the activities of users. Life styles are usually reflected as a mixture of motion activities with different occurrence probability.

Generally speaking, there are two mainstream approaches: supervised learning and unsupervised learning.

3) *Friend-Matching Graph Construction:*

To characterize relations among users, in this section, we propose the friend-matching graph to represent the similarity between their life styles and how they influence other people in the graph.

In particular, we use the link weight between two users to represent the similarity of their life styles.

Based on the friend-matching graph, we can obtain a user's affinity reflecting how likely this user will be chosen as another user's friend in the network.

Weighted undirected graph  $G=(V,E,W)$

Where,  $V=\{v_1,v_2,\dots,v_n\}$  is the set of users and  $n$  is the number of users

$E=\{e(i,j)\}$  is the set of links between users

$W:E \rightarrow R$  is the set of weight of edges

**B. User Impact Ranking:**

- 1) The impact ranking means a user's capability to establish friendships in the network.
- 2) In other words, the higher the ranking, the easier the user can be made friends with, because he/she shares broader life styles with others.
- 3) Once the ranking of a user is obtained, it provides guidelines to those who receive the recommendation list on how to choose friends.

**C. User Viewing their Page:**

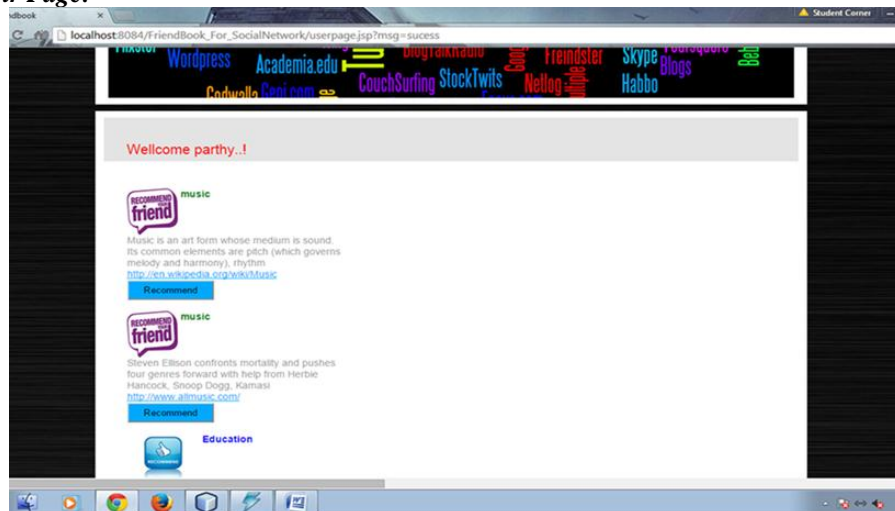


Fig. 4: User Viewing their Page

Table – 1  
User impact ranking of 8 users

Rank	User ID	Rank Score
1	1	0.133
2	7	0.127
3	4	0.125
4	8	0.125
5	5	0.124
6	2	0.123
7	6	0.123
8	3	0.118

## V. DATA FLOW DIAGRAM

- 1) The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
- 2) The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
- 3) DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
- 4) A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

## VI. CONCLUSION

In this paper, we presented the An Innovative Concept for finding friends in Social Networks Based on Their Lifestyles. One challenge with existing social networking services is how to recommend a good friend to a user. Most of them rely on pre-existing user relationships to pick friend candidates. Friend book is the first friend recommendation system exploiting a user's life style information discovered from smart phone sensors. We integrate a linear feedback mechanism that exploits the user's feedback to improve recommendation accuracy. Then we are compressing the document size with the help of Hadoop tool, to reduce the memory size of the server and increase the performance of the server.

## REFERENCES

- [1] Naruchitparames, J.Gunes, M.H.Louis: Friend recommendation in social networks using genetic algorithms and network topology.
- [2] J.Biagioni, T.Gerlich, T. Merrifield, and J. Eriksson. EasyTracker: Automatic Transit Tracking, Mapping, and Arrival Time Prediction Using Smartphones. Proc. of SenSys, pages 68-81, 2011.
- [3] T. Huynh, M. Fritz, and B. Schiel. Discovery of Activity Patterns using Topic Models. Proc. of UbiComp, 2008.
- [4] Guo, L.Zhang,C.Fang: A Trust based Privacy Preserving Friend Recommendation Scheme for Online Social networks.
- [5] Z. Wang, C. E. Taylor, Q. Cao, H. Qi, and Z. Wang. Demo: Friendbook: Privacy Preserving Friend Matching based on Shared Interests. Proc. of ACM SenSys, pages 397-398, 2011.
- [6] K. Farrahi and D. Gatica-Perez. Probabilistic mining of socio- geographic routines from mobile phone data. Selected Topics in Signal Processing, IEEE Journal of, 4(4):746-755, 2010.
- [7] A. D.Sarma, A. R.Molla, G.Pandurangan, and E.Upfal. Fast distributed computation. Springer Berlin Heidelberg, pages 11-26, 2013.
- [8] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma. Understanding Transportation Modes Based on GPS Data for Web Applications. ACM Transactions on the Web (TWEB), 4(1):1-36, 2010.
- [9] B. Bahmani, A. Chowdhury, and A. Goel. Fast incremental and personalized pagerank. Proc. of VLDB Endowment, volume 4, pages 173-184, 2010.