

Social Network Friend Recommendation System Using Semantic Web

Pankaj Pingate¹, S. M. Rokade²

¹Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India

²Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India

Abstract: *In today's growing era of social networking services friend recommendations may not be most relevant to reflect users expectations on friend selection in real life because friend recommendations are based on social graphs which uses tastes and peoples are the basis for recommendations. In this paper, a personalised friendbook recommendation mobile application is presented, which is a novel semantic based friend recommendation system for social networking services. It recommends friend to system users based on their life styles, habits, locations or user profiles. This system measures the similarities of life styles, habits, locations or user profiles between users and recommends friends if their life styles, habits, locations or user profiles have higher similarities. Using text mining, daily activities of users are modelled as a life documents. This life document is used for extracting user's life styles or habits by using Latent Dirichlet Allocation algorithm. The similarity metric is proposed to measure similarity of life styles or habits between users and impact of life styles or habits is calculated using friend matching graph. Friendbook returns a list of users with highest recommendation score according to users request. After receiving recommendations user can provide feedback through feedback mechanism which is used to improve further recommendation accuracy. Finally, we implemented this system using Java, J2EE technology and android development kit to provide personalised friend recommendations.*

Keywords: friend recommendations, text mining, lifestyle or habits, LDA.

1. Introduction

Now a day social networking services are becoming more popular. Internet users prefer social network services to be in touch with their loved and dear ones. Many users of social networks find their so older friends on social network sites across country borders. In our model we have incorporated user preferences obtained from Social Networking Site. Social Networking sites are used intensively from last decade. According to the current survey, Social Networking sites have the largest data set of users. Each social networking site notes/records each and every activity of user (like: what user likes? what user is doing? what is user's hobby? Etc). Social Networking site will prove to be largest domain in understanding the user behavior. One of the best examples of social networking is FACEBOOK. According to current news FACEBOOK [2] is trying to develop algorithm, to understand user behavior.

Social Networking sites can help us in getting important information of users, such as age, gender, location, language, actives, likes etc. our model takes into account these parameters of the user to recommend books. Most of the friend suggestions mechanism relies on pre-existing user relationships to pick friend candidates. For example, Face book relies on a social link analysis among those who already share common friends and recommends symmetrical users as potential friends. The main challenge in developing relevant friend recommendations [3] is due to the dynamic nature of humans' perception of friendship, which constitutes a cause for heterogeneity in social networks. It is usual and frequent for humans to change their view of friendship. Further, this view varies from person to person in which a social network can undergo frequent and abrupt change over time even without the introduction of new nodes. Recommender

systems help users to identify their interests and sets of choices by predicting the usefulness degree of an item or group of items to these users. They are defined as a special type of information filtering that gives information about which items might be interesting to users.

Currently social network services recommend friends to users based on their social graphs, which may not be the most accurate to reflect a user's preferences on friend selection in real life. This paper focuses on, 'Social Network Friend Recommendation Using Semantic Web', which recommends friends to users based on their life styles, habits, locations or user profiles instead of social graphs. With the help of different sensors of Smart phones, Friendbook discovers life styles, habits, locations or user profiles from user-centric sensor data, measures the similarity of life styles between users, and recommends friends to users if their life styles or habits or locations have high similarities. Using text mining, we model a user's daily life as life documents, from which his/her life styles are extracted by using the Latent Dirichlet Allocation (LDA) [4] algorithm. We further implemented a similarity metric to measure the similarity of life styles between users. On the basis of higher recommendation score this system returns a list of people to the user. Finally feedbacks are collected from users to improve recommendation accuracy. We have implemented Friendbook on the Android-based smart phones, and evaluated its performance on both small scale experiments and large-scale simulations.

To extract proper life styles we tried to establish relation between people's daily activities and life style document. Previous researches on text mining treats life style document as mixture of topics and topics as mixture of words. Using same philosophy we treat our daily activities as a mixture of life styles and each life style is a mixture of habits. So we

represent daily lives as a life document whose meanings are reflected by their topics. Just like words serve as the basis of documents, people's activities naturally serve as the primitive vocabulary of these life documents. The rest of the paper is organized as follows. Section 2 covers Literature Review, Section 3 focuses on working of system, In Section 4 we evaluate the performance of Friendbook intensively with both simulations and real experiments. Finally, we conclude the paper in Section 5.

2. Literature Review

In recent years, recommendation system becomes more and more popular in people because of different advancements in data mining and information retrieval (IR) strategies. Basically recommendation system offers potential items such as books, songs, products etc. by automatically analyzing individual's access patterns, habits and interests. For example, personalized learning systems recommends personalized topics or blogs to a user based on topics or blogs the user previously visited, and topics that other users are looking. Netflix [5] recommends musics to a user based on previously searched videos and watching habits. In today's era appropriate friend recommendation in social networking systems has a lots of attention. For example, Facebook, LinkedIn and Twitter, they share common friends according to their social relation. Meantime, the researchers also proposed other recommendation mechanisms.

For examples Bian and Holtzman [6] discovered a collaborative filtering friend recommendation system such as MatchMaker based on personality matching. Kwon and Kwim [7] presented friend recommendation using physical and social context. Yu et al. [8] by combining GPS information and social network structure, recommended geographically related friends. Hsu et al. Proposed a collaborative recommendation using the link structure and content based recommendation using mutual declared interest. Gou et al.[8] Proposed SFViz system based on visual structure that supports explore and finding of friends interactively. This system works under the context of interest. They reported a case study which is based on tagging behaviour of people in songs to explore the recommendation of friends. However all above mentioned recommendation systems are significantly different from our proposed system. Basically our focus is on recent advancement in findings to suggest a friends using similarities between lifestyles of users rather than their social relations.

Activity recognition is used as a basic building block for extracting high level daily routines from low level sensor data. This has been widely studied through different types of wearable sensors. Zheng et al. Proposes use of GPS data in transportation mode for understanding users traversal. Lester et al. [9] Provide Hidden Markov Model (HMM) for activity recognition using data collected by wearable sensors. Li et al.[10] Focuses a approach of using accelerometers and gyroscopes for finding of dynamic transitions and postures for recommendations. Smartphone's are equipped with rich set of sensors which provides flexible ways for activity recognitions. Reddy et al. Proposed the concept of detection

of individuals transportation mode with the help of built in GPS and accelerometer on the Smartphone's. CenceMe [12] focuses on multiple sensors on the Smartphone to capture routines, interest, patterns, surroundings. SoundSense [11] used microphone on smartphone to identify general sound types and discover user specific sound events. EasyTracker used data collected by smartphones to determines the routes for serving, stop locations, and schedules for vehicles with the help of GPS installed on transit vehicles.

Lot of Past research has been done on recognition of activities using smartphones. As compared to this very little work has been proposed for identification of daily routines using smart phones. The MIT reality mining project and Ferrahi and Gatica-Perez [13] tried to identify daily location driven routines using vast location data. They tries to find location routines such as from home to office, lunch at restaurant, jogging etc. but this research doesn't consider daily routines of people who are at same location or staying in home. This creates major difference in findings which affects potential recommendations. For example, when user stay at home his/her daily routines such as watching movies, cooking, eating, net surfing could not be identified only on location based information. In probabilistic mining of sociographic routines from mobile phone data, Ferrahi and Gatica-Perez used approach of combined location and physical proximity captured by mobile phones to overcome problem of people who stays at fixed location. Another more closer approach was presented by T.Huynh, M.Fritz et al. [14] In there paper discovery of activity patterns using topic models. In there paper they focuses on extraction of activity patterns from sensor data with the help of topic models. But this approach is based on two wearable sensor and it is not used with smart phones.

To discover lifestyles with the help of smartphones, we used probabilistic topic model. In this approach friend recommendation is based on patterns extracted from activities. This approach helps user to find friends who have similar lifestyle.

3. System Overview

This section gives complete overview of implemented system. Figure 1 shows the block diagram of implemented system which is based on client server architecture where each client is a smart phone carried by user and the servers are data centers.

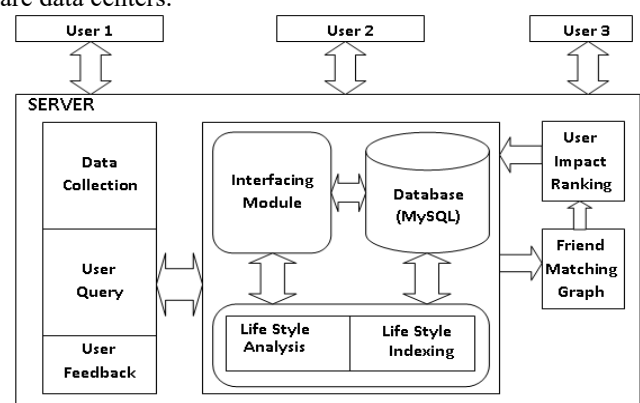


Figure 1: Block diagram of implemented system

On the client side each smart phone identifies user's activity, locations, users profile and captures this information. It prepares life documents from this recorded data and report it to server. Users may continually uses this system, he/she will contribute more and more activities in life document which is input for probabilistic topic model for discovery of lifestyles. Server side will be based on eight modules which fulfil the task of friend recommendation. Here data collection module is responsible for collection of life documents from client side. This collected data is passed to lifestyle analysis module via interfacing module which extracts lifestyle with the help of lifestyle probabilistic model. Lifestyle indexing module puts extracted lifestyles into the database in the format of (*lifestyle, user*). Friend matching graph module constructs lifestyle graph to represent the similarity relationships among different lifestyles. This information is passed to user impact ranking module to calculate the impacts of users. The user query module accepts user query and generates a ranked list of potential friends using impact ranking and sends it back as a response to user query. Finally propose system will accept users feedback through feedback control module to improve the accuracy of recommendation results ultimately which will improve friend recommendations.

3.1 Data Collection Module

This module collects data from users smart phone through our mobile app. Mobile app consist three different modules: life style, GPS tracker [8], profile. Life style module collects user life styles such as running, walking, moving. GPS tracker finds users location. User profile collects demographic information of user. All these modules submit this collected data to our server for further analysis.

3.2 Data Life Style Analysis and Indexing

After collecting data by data collection module passed to life style analysis and indexing module for life style modelling and activity recognition. This module works in three phases:

3.2.1 Life style modelling

This phase uses probabilistic topic model for discovery of probabilities of lifestyles from life document. Probabilistic topic model uses concept of bag-of-activities to replace the original sequences of activities recognized based on the raw data with their probability distributions. Thereafter, each user has a bag-of-activity representation of his/her life document, which comprises a mixture of activity words.

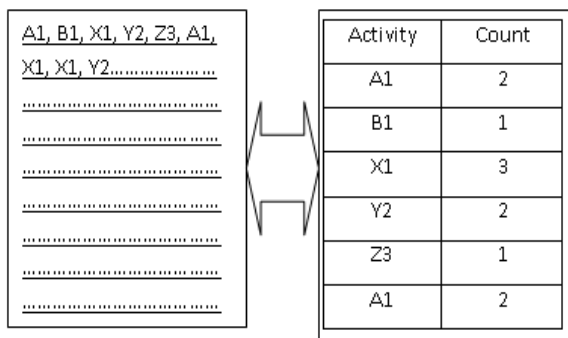


Figure 2: Bag-of-activity

3.2.2 Activity Recognition

To derive bag-of-activities we need to recognize activities. Life styles are nothing but motion activities with different occurrences. To capture these motion activities pedometer is used. Activities captured through pedometer are grouped by using K-means clustering algorithm where each cluster represents an activity.

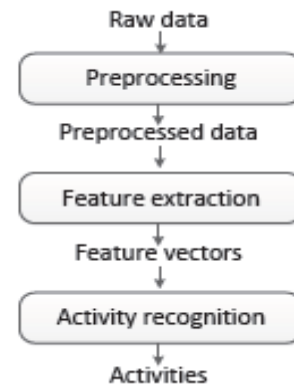


Figure 3: Phases of activity recognition

3.2.3 Life Style Extraction

For extracting life styles from activity clusters we used Latent Dirichlet algorithm. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.

The LDA algorithm shown below.

Input: Number of topics K

Corpus with M documents and N_d words in document d

Output: Model parameters: β, θ, z

initialize $\phi_{ni}^0 := 1/k$ for all i in k and n in N_d

initialize $\gamma_i := \alpha_i + N/k$ for all i in k

initialize $\alpha := 50/k$

initialize $\beta_{ij} := 0$ for all i in k and j in V

//E-Step (determine ϕ and γ and compute expected likelihood)

loglikelihood := 0

for $d = 1$ to M

 repeat

 for $n = 1$ to N_d

 for $i = 1$ to K

$\phi_{dni}^{t+1} := \beta_{iw_n} \exp(\Psi(\gamma_{di}^t))$

 endfor

 normalize ϕ_{dni}^{t+1} to sum to 1

 endfor

$\gamma^{t+1} := \alpha + \sum_{n=1}^{N_d} \phi_{dn}^{t+1}$

 until convergence of ϕ_d and γ_d

 loglikelihood := loglikelihood + $L(\gamma, \phi; \alpha, \beta)$ // See equation BLAH

 endfor

 normalize ϕ_{dni}^{t+1} to sum to 1

 endfor

$\gamma^{t+1} := \alpha + \sum_{n=1}^{N_d} \phi_{dn}^{t+1}$

 until convergence of ϕ_d and γ_d

 loglikelihood := loglikelihood + $L(\gamma, \phi; \alpha, \beta)$ // See equation BLAH

endfor

//M-Step (maximize the log likelihood of the variational distribution)

for $d = 1$ to M

 for $i = 1$ to K

 for $j = 1$ to V

$\beta_{ij} := \phi_{dni} w_{dnj}$

```

    endfor
    normalize  $\beta_i$  to sum to 1
  endfor
endfor
estimate  $\alpha$  via Eq. (8)

if loglikelihood converged then
  return parameters
else
  go back to E-step
endif
    
```

3.3 Friend Matching Graph

This section consist a friend matching graph to identify closer similarity among life styles of user and how these similarities influence other users in the graph. This approach represents the similarity of lifestyles with the help of link weight between two users. The friend matching graph obtains users affinity to show likeliness of particular user will be chosen as friend for another user in the network. Here new similarity metric is used to measure the similarity between two lifestyle vectors. Similarity calculation is not only affected by vectors but also most important lifestyles known as dominant lifestyles. The real life relations among users are modelled as friend matching graph on the basis of similarity metric. The friend matching graph is a weighted undirected graph $G=(V,E,W)$, where V is a set of users, E is a set of links between users, and W is a set of weights of edges. The edges linking users if and only if there similarities are greater then predefined similarity threshold. The weight of that edge is represented by the similarity, i.e. $W(i,j)=S(i,j)$.

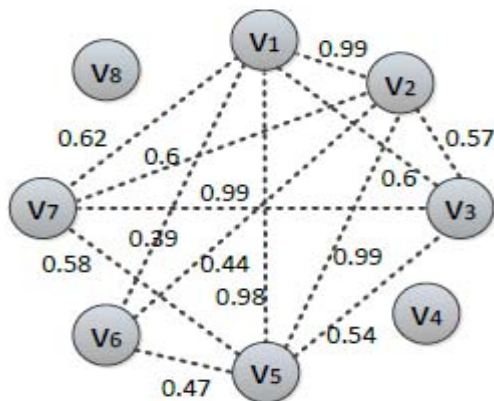


Figure 4: An example of Friend matching graph

3.4 User Impact Ranking

The ranking of user provides direction to the people who receive the recommendation list. The ranking is independent from the query means it is only depended on graph structure which contains two aspects : how the edge are connected, how much weight is on every edge. This ranking is used together with similarity score between user query and potential friend candidates.

3.5 Query Evaluation and Friend Recommendation

Before a user initiates a request, he/she should have accumulated enough activities in his/her life documents for efficient life styles analysis. The period for collecting data usually takes at least one day. Longer time would be expected if the user wants to get more satisfied friend recommendation results. After receiving a user's request (e.g., life documents), the server would extract the user's life style vector, and based on which recommend friends to the user. This module provides recommendations to the users according to following algorithm.

Algorithm Friend recommendation

Input: The query user i , the recommendation coefficient β and the required number of recommended friends from the system p .
Output: Friend list F_i .

- 1: $F_i \leftarrow \emptyset, Q \leftarrow \emptyset$
- 2: extracts i 's life style vector L_i using the LDA algorithm.
- 3: **for** each life style z_k the probability of which in L_i is not zero **do**
- 4: put users in the entry of z_k into Q
- 5: **end for**
- 6: **for** each user $j \notin Q$ **do**
- 7: $S(i, j) \leftarrow 0$
- 8: **end for**
- 9: **for** each user j in the database **do**
- 10: $R_i(j) = \beta S(i, j) + (1 - \beta)r_j \kappa$
- 11: **end for**
- 12: sort all users in decreasing order according to $R_i(j)$
- 13: put the top p users in the sorted list to F_i

4. Evaluation

After the successful implementation and testing of the system, next important move is the final results given by system and analysis of those results. So, this section focuses on analysis of results given by friend recommendation system.

4.1 Performance Measures

Many different measures have been proposed for evaluating the performance of information retrieval systems. Typically the measures require a collection of documents and a query, whereas every document is either relevant or non relevant to a particular query. The following part describes some of the most common measures for performance:

(i) Precision:

$$precision = \frac{|\{relevant_documents\} \cap \{retrieved_documents\}|}{|\{retrieved_documents\}|}$$

(ii) Recall:

$$recall = \frac{|\{relevant_documents\} \cap \{retrieved_documents\}|}{|\{relevant_documents\}|}$$

4.2 Evaluation

For evaluating system we have used some performance measures such as **accuracy, precision, recall** in the context of relevant documents and irrelevant documents. Using this we have evaluated the performance of system with respect to user profile, overall performance of system, precision, and recall. Following are the some bar charts which show the evaluation results of system:

Table I: Profession of Users

User ID	1	2	3	4	5	6	7	8
Student	✓	✓	✓		✓		✓	
Waitress				✓				
Office Worker						✓		
Businessman								✓

Table 2: 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

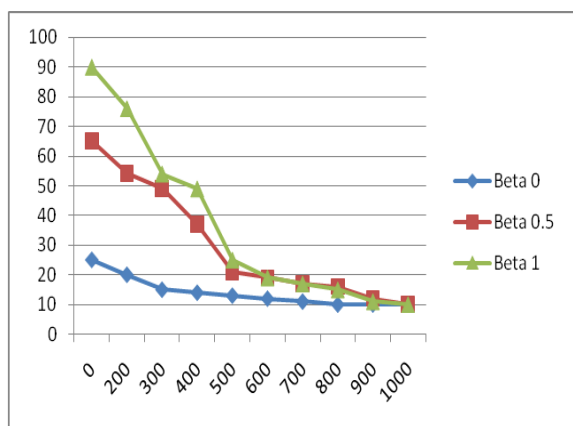


Figure 5: Precision Result

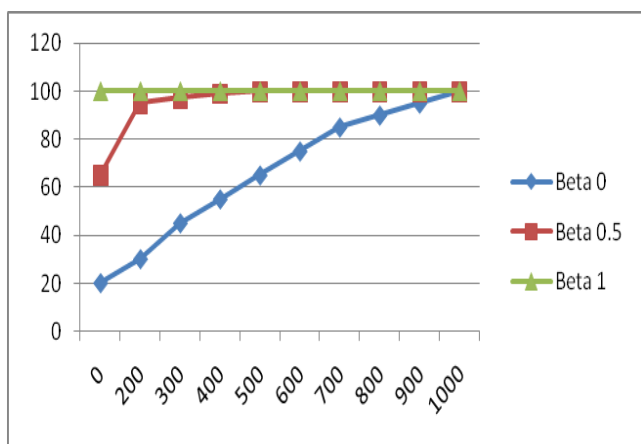


Figure 6: Recall Result

4.3 Screenshots

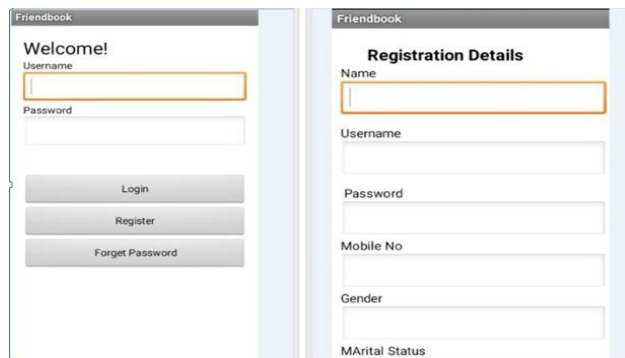


Figure (a) User Login

Figure (b) User Registration Form

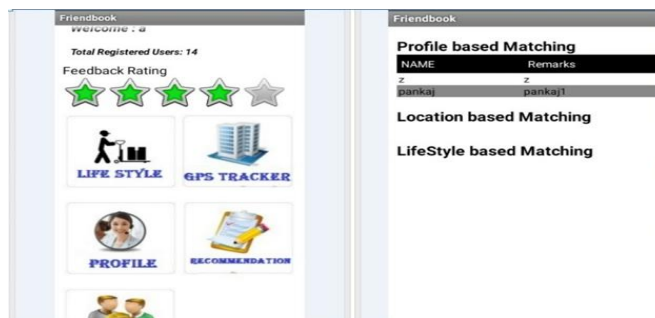


Figure (c) Calculate Life Style & Location

Figure (d) Recommendation

Figure 7: Friendbook App Screenshots

The above figure 7(a) shown the existing user can login using own user name and password. The new user can register with the system using touch on register button. Figure 7(b) shown user registration details which new user fill it and register with the system. Figure 7(c) shown life style, GPS Tracker and user profile. Life style used for collecting daily lives data using pedometer, GPS tracker used for tracking user's location and profile data used for recommendation friend. Figure 7(d) shown friend recommendation using profile based or location based or life style based.

5. Conclusion and Future Work

In this paper, we have implemented the concept of friend recommendation service. We have also presented comprehensive steps of the common friend recommendation services of social networks. Recommender systems are powerful tools that overcome the information overload problem by providing users with the most relevant contents. The importance of contextual information has been recognized by researchers and practitioners in many disciplines including Ecommerce, personalized IR, ubiquitous and mobile computing, data mining, marketing and management.

Our future work will be (1) use of wearable sensor devices to capture more interesting and meaningful lifestyles. (2) To add MatchMaker module with existing friendbook application, where a collaborative filtering can be used for personality matching. For recommending friends personalities of each individuals can be matched. (3) For non mobile users we will provide a friend recommendations using complex network

theory and genetic algorithm on web based social networking system.

References

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993-1022, 2003.
- [2] P. Desikan, N. Pathak, J. Srivastava, and V. Kumar. Incremental page rank computation on evolving graphs. *Proc. of WWW*, pages 1094-1095, 2005.
- [3] N. Eagle and A. S. Pentland. Reality Mining: Sensing Complex Social Systems. *Personal Ubiquitous Computing*, 10(4):255-268, March 2006.
- [4] K. Farrahi and D. Gatica-Perez. Probabilistic mining of sociogeographic routines from mobile phone data. *Selected Topics in Signal Processing*, *IEEE Journal of*, 4(4):746-755, 2010.
- [5] K. Farrahi and D. Gatica-Perez. Discovering Routines from Largescale Human Locations using Probabilistic Topic Models. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1), 2011.
- [6] B. A. Frigiyik, A. Kapila, and M. R. Gupta. Introduction to the dirichlet distribution and related processes. Department of Electrical Engineering, University of Washington, UWEETR-2010-0006, 2010.
- [7] A. Giddens. *Modernity and Self-identity: Self and Society in the late Modern Age*. Stanford Univ Pr, 1991.
- [8] L. Gou, F. You, J. Guo, L. Wu, and X. L. Zhang. Sfviz: Interestbased friends exploration and recommendation in social networks. *Proc. of VINCI*, page 15, 2011.
- [9] W. H. Hsu, A. King, M. Paradesi, T. Pydimarri, and T. Weninger. Collaborative and structural recommendation of friends using weblog-based social network analysis. *Proc. of AAAI Spring Symposium Series*, 2006.
- [10] T. Huynh, M. Fritz, and B. Schiel. Discovery of Activity Patterns using Topic Models. *Proc. of UbiComp*, 2008.
- [11] J. Kwon and S. Kim. Friend recommendation method using physical and social context. *International Journal of Computer Science and Network Security*, 10(11):116-120, 2010.
- [12] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A Hybrid Discriminative/Generative Approach for Modeling Human Activities. *Proc. of IJCAI*, pages 766-772, 2005.
- [13] Q. Li, J. A. Stankovic, M. A. Hanson, A. T. Barth, J. Lach, and G. Zhou. Accurate, Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information. *Proc. of BSN*, pages 138-143, 2009.
- [14] E. Miluzzo, C. T. Cornelius, A. Ramaswamy, T. Choudhury, Z. Liu, and A. T. Campbell. Darwin Phones: the Evolution of Sensing and Inference on Mobile Phones. *Proc. of MobiSys*, pages 5-20, 2010.
- [15] E. Miluzzo, N. D. Lane, S. B. Eisenman, and A. T. Campbell. Cenceme-Injecting Sensing Presence into Social Networking Applications. *Proc. of EuroSSC*, pages 1-28, October 2007.
- [16] L. Page, S. Brin, R. Motwani, and T. Winograd. The Pagerank Citation Ranking: Bringing Order to the Web. Technical Report, Stanford InfoLab, 1999.

- [17] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using Mobile Phones to Determine Transportation Modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13, 2010.
- [18] I. Ropke. The Dynamics of Willingness to Consume. *Ecological Economics*, 28(3):399-420, 1999.

Author Profile



Pankaj L. Pingate has completed his B.E in Information Technology from University of Pune in 2009. Currently pursuing Master of Engineering from SVIT Chincholi, Nashik, India.



Prof. S. M. Rokade has completed his B.E in Computer Engineering from Pune University and M.E in Computer Science & Engineering from MGM, Nanded. He is presently working as an Associate Professor in SVIT Chincholi, Nashik India.