ISSN 2229-5518

A Novel Approach for Semantic Based Friend Recommendation System for Social Networks

Dr.V.Goutham[1], T.Ruthvik Manikanta [2]

[1,2]Department of Computer Science and Engineering,Teegala Krishna Reddy Engineering College,

Meerpet, Telangana, India

Abstract—The social media sites are apt for advertising activities and over the preceding years, users have seen an uptick in the amount of publicising on the sites. The present social networking services endorse friends to users established on their social graphs, which may not be applicable to reveal a user's inclinations on friend choice in real life. In this novel approach for semantic based friend recommendation system for social networks that is Friendbook endorses friends to users built on their life styles of users from user-centric sensor data, assess the resemblance of life styles between users, and acclaims friends to users if their life styles have high resemblance. Upon acceptance of a request, Friendbook proceeds a list of people with utmost sanction scores to the query user. Lastly, Friendbook assimilates a feedback mechanism to promote advance in the recommendation accurateness.

Index Terms— Friend recommendation, mobile sensing, social networks.

1 INTRODUCTION

The contest with present social networking services is how to acclaim a good friend to a user. Most of them rely on pre-existing user relationships to pick friend candidates. For example, Facebook dependent on a social link analysis among those who previously share common friends and acclaims proportioned users as potential friends[12],[2]. Inappropriately, this slant may not be the most accurately based on latest sociology outcomes. According to these studies, the instructions to group people collectively include: 1) habits or life style; 2) attitudes; 3) tastes; 4) moral standards; 5) economic level; and 6) people they already know. In present day lives, we may have hundreds of undertakings, which form meaningful sequences that shape our lives. In this Novel approach for semantic based friend recommendation system for social networks, we use the word activity to precisely state to the actions taken in the order of seconds, such as "sitting", "walking", or "typing", while we use the phrase life style to deliberate higher-level abstractions of daily lives, such as "office work" or "shopping". For this case, the "shopping" life style typically contains the "walking" activity, but may also encompass the "standing" or the "sitting" happenings. To model daily lives properly, we draw an analogy between people's daily lives and documents, as shown inFigure 1. Previous research on probabilistic topic models in text mining has treated documents as mixtures of topics, and topics as mixtures of words [10]. Inspired by this, likewise, we can treat our daily lives as a mixture of life styles, and[4] each life style as a mixture of activities. Observe here, basically, we signify daily lives with "life documents", whose semantic meanings are replicated through their topics, which are life styles in this study. Just like words Assist as the basis of documents, people's activities naturally serve as the primeval lexis of these life documents.

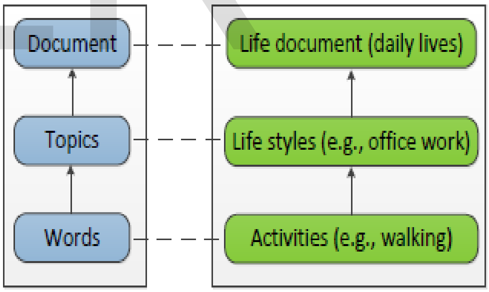


Fig. 1: An analogy between word documents and people's daily lives.

The projected elucidation is also driven by the topical progresses in smart phones, which have become more and more popular in people's lives. These smart phones (e.g., iPhone or Android-based smart phones) are equipped with a rich set of embedded sensors, such as GPS, accelerometer, microphone, gyroscope, and camera. Thus, a smart phone is no longer simply a communication device, but also a powerful and environmental reality sensing platform from which we can extract rich context and content-aware information. From this perspective, smart phones serve as the ideal platform for sensing daily routines from which people's life styles could be discovered.[9] The contributions of this work are summarized as follows:

_ Friend book is the first friend recommendation system exploiting a user's life style information discovered from smart phone sensors.

_Modelling the daily lives of users as life documents and use the probabilistic topic model to extract life style information of users.[8],[7]

_ A unique similarity metric to characterize the similarity of users in terms of life styles and then construct a friend-matching graph to recommend friends to users based on their life styles is introduced.

2 RELATED WORK

Commendation systems that try to recommend items (e.g., music, movie, and books) to users have become more and more popular in recent years. For instance, Amazon [1] recommends items to a user based on items the user previously visited, and items that other users are looking at. Netflix [3] and Rotten Tomatoes [4] recommend movies to a user based on the user's previous ratings and watching habits. Recently, with the advance of social networking systems, friend recommendation has received a lot of attention. Generally speaking, existing friend recommendation in social networking systems, e.g., Facebook, LinkedIn and Twitter, recommend friends to users if, according to their social relations, they share common friends. Meanwhile, other recommendation mechanisms have also been proposed by researchers. For example, Bian and Holtzman [8] presented Match Maker, a collaborative filtering friend recommendation system based on personality matching. Kwon and Kim [2] proposed a friend recommendation method using physical and social context. However, the authors did not explain what the physical and social context is and how to obtain the information. Yu et al. [2] recommended geographically related friends in social network by combining GPS information and social network structure. Hsu et al. [18] studied the problem of link recommendation in weblogs and similar social networks, and proposed an approach based on collaborative recommendation using the link structure of a social network and content-based recommendation using mutual declared interests. Gou et al. [7] proposed a visual system, SFViz, to support users to explore and find friends interactively under the context of interest, and reported a case study using the system to explore the recommendation of friends based on people's tagging behaviors in a music community. These existing friend recommendation systems, however, are significantly different from our work, as we exploit recent sociology findings to recommend friends based on their similar life styles instead of social relations. Activity recognition serves as the basis for extracting high-level daily routines (in close correlation with life styles) from low-level sensor data, which has been widely studied using various types of wearable sensors. Zheng et al. [3] used GPS data to understand the transportation mode of users. Lester et al. [11] used data from wearable sensors to recognize activities based on the Hidden Markov Model (HMM). Li et al. [12] recognized static postures and dynamic transitions by using accelerometers and gyroscopes. The advance of smart phones enables activity recognition using the rich set of sensors on the smart phones. Reddy et al. [6] used the built-in GPS and the accelerometer on the smart phones to detect the transportation mode of an individual. CenceMe [4] used multiple sensors on the smart phone to capture user's activities, state, habits and surroundings. Sound Sense [13] used the microphone on the smart phone to recognize general sound types (e.g., music, voice) and discover user specific sound events. Easy Tracker [7] used GPS traces collected from smart phones that are installed on transit vehicles to determine routes served, locate stops, and infer schedules. Although a lot of work has been done for activity recognition using smart phones, there is relatively little work on discovery of daily routines using smart phones. The MIT Reality Mining project [12] and Farrahi and Gatica-Perez [14] tried to discover daily location-driven routines from large-scale location data. They could infer daily routines such as leaving from home to office and eating at a restaurant. However, they could not discover the daily routines of people who are staying at the same location. For instance, when one stays at home, user daily routines like "eating lunch" and "watching movie" could not be discovered if only using the location information. In [13], Farrahi and Gatica-Perez took a step further and overcame the short-coming of discovering daily routines of people staying in the same location by considering combined location and physical proximity sensed by the mobile phone. Another closely related[9] work was presented in [9], which used a topic model to extract activity patterns from sensor data. However, they used two wearable sensors, but not smart phones, to discover the daily routines. In our work, we attempt to use the probabilistic topic model to discover life styles using the smart phone. We further utilize patterns discovered from activities as a basis for friend recommendation that helps users find friends who have similar life styles. Note that the work in this Novel approach for semantic based friend recommendation system for social networks is significantly different[12] from our preliminary demo work of Friendbook [8] that recommended friends to users based on the similarity of pictures taken by users.

3 SYSTEM DESIGN

Here, the system has designed a client-server mode where each client is a smart phone carried by a user and the servers are data centers or clouds. On the client side, each smart phone can record data of its user, accomplish real-time activity recognition and report the generated life documents to the servers. It is worth noting that an offline data collection and training phase is desired to build a suitable activity classifier for real-time activity recognition on smart phones. We finished three months

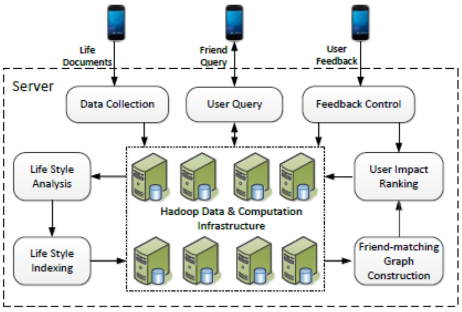


Fig 2: real-time activity recognition

on amassing raw data of 8 volunteers for erection a large training data set. As each user naturally spawns around 50MB of raw data each day, we indicate My SQL as the low level data storage platform and Hadoop Map Reduce as computation organization. After the activity classifier is built, it will be dispersed to each user's smart phone and then activity recognition can be accomplished in real-time manner. As a user repeatedly uses Friendbook, user will collect more[10] and more activities in user's life documents, based on which, we can ascertain user life styles using probabilistic topic model. On the server side, seven modules are designed to fulfil the task of friend recommendation. The data collection module collects life documents from users' smart phones. The life styles of users are extracted by the life style analysis module with the probabilistic topic model. 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, life-style). A friend-matching graph can be constructed consequently by the friend-matching graph construction module to signify the resemblance relationship between users' life styles. The impacts of users are then calculated based on the friend-matching graph by the user impact ranking module. The user query module takes a user's query and sends a ranked list of potential friends to the user as reply. The system also countenances users to give feedback of the recommendation results which can be managed by the feedback control module. With this module, the correctness of friend recommendation can be upgraded.

3.1 Life styles and activities:

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 combination of activities. This is comparable to the treatment of documents as ensemble of topics and topics as ensemble of words. By taking advantage of recent developments in the field of text mining, we model the daily lives of users as life documents, the life styles as topics, and the activities as words. Given "documents", the probabilistic topic model could discover the probabilities of essential "topics". Therefore, we adopt the probabilistic topic model to discover the probabilities of hidden "life styles" from the "life documents". In probabilistic topic models, the frequency of vocabulary is predominantly important, as dissimilar frequency of words signifies their information entropy alterations. The "bag-of-activity" model to replace the original sequences of activities recognized based on the raw data with their probability distributions. Subsequently, each user has a bag-of-activity representation of his/her life document, which embraces a mixture of activity words.

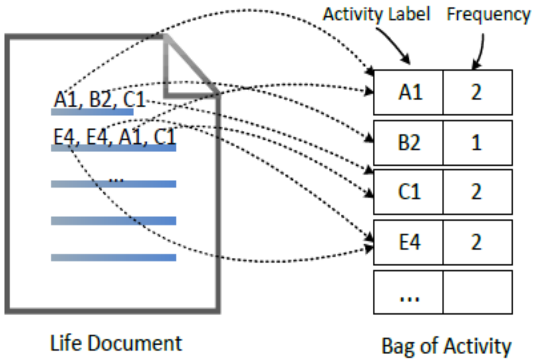


Fig. 3: Bag-of-Activity modelling for life document.

3.2 Activity Recognition

To derive p, we need to first classify or identify the activities of users. Life styles are generally redirected as a combination of motion activities with different occurrence probability. Therefore, two motion sensors, accelerometer and gyroscope, are used to infer users' motion activities. Generally speaking, there are two mainstream approaches: supervised learning and unsupervised learning. For both approaches, mature techniques have been developed and tested. In practice, the number of activities tangled in the analysis is changeable and it is problematic to accumulate a large set of ground truth data for each activity, which makes supervised learning algorithms unsuitable for our system. Therefore, we use unsupervised learning approaches to recognize activities. Here, we adopt the popular K-means clustering algorithm [9] to group data into clusters, where each cluster represents an activity. Here it is chosen K-means for its simplicity and effectiveness. Raw data→Pre processing→Feature->extraction→Activity recognition→Preprocessed->data→Feature vectors→Activities

4 PROPOSED SYSTEM: QUERY AND FRIEND RECOMMENDATION

Before a user initiates a request, user should have gathered enough activities in user's 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.[10]

Algorithm 1 Computing users' impact ranking

Input: The friend-matching graph G.
Output: Impact ranking vector r for all users.

- **for** $i = 1$ to n **do**
  - $\mathrm{r}_0(i) = \frac{1}{n}$
- **end for**
- $\delta = \infty$
- $\epsilon = e^{-9}$
- **while** $\delta > \epsilon$ **do**
  - **for** $i = 1$ to n **do**
    - $\mathrm{r}_{k+1}(i) = \sum_j \frac{1-\varphi}{n}\mathrm{r}_k(j) + \varphi\frac{\sum_j \omega(i,j)\cdot \mathrm{r}_k(j)}{\sum_j \omega(i,j)}$
  - **end for**
  - $\delta = \sum_{i=1}^{n} |\mathrm{r}_{k+1}(i) - \mathrm{r}_k(i)|$
- **end while**
- **return** r

The reference results are highly dependent on users' preference. Some users may prefer the system to recommend users with high impact, while some users may want to know users with the most similar life styles. It is also possible that some users want the system to recommend users who have high impact and also similar life styles to them. As the number of users increases, the overhead of[14] query and recommendation increases linearly. In reality, users may have totally different life styles and it is not necessary to calculate their recommendation scores at all. Therefore, in order to speed up the query and recommendation process, we adopt the reverse index table using (life-style, user) pair instead of (user, life-style) pair in the database. With the reverse index table, before calculating recommendation score for each user, the server first picks up all the users having overlapping life styles with the query user and sets the resemblances of rest users to the query user to 0. The server then checks all the users to calculate their endorsement scores. Although the complexity is still O(n), we can observe that the reverse index table reduces the computation overhead, the advantage of which is considerable when the system is in large-scale. Privacy is very important especially for users who are sensitive to information leakage. In our design of Friendbook, we also considered the privacy issue and the existing system can provide two levels of privacy[12] protection. First, Friendbook protects users' privacy at the data level. Instead of uploading raw data to the servers, Friendbook processes raw data and classifies them into activities in real-time. The predictable activities are labeled by integers. In this way, even if the documents comprising the integers are compromised, they cannot tell the physical meaning of the documents. Second, Friendbook protects users' privacy at the life pattern level.

Algorithm 2 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.

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

Instead of telling the similar life styles of users, Friendbook only shows the recommendation scores of the suggested friends with the users. With the recommendation score, it is almost impossible to deduce the life styles of recommended friends.

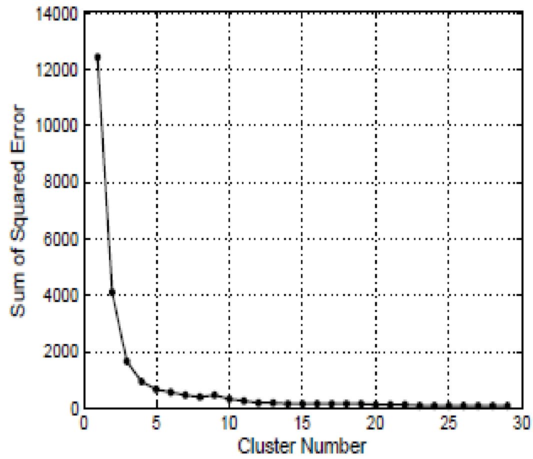
5 EVALUATION

We first assess the recital of Friendbook on small-scale experiments and then eight volunteers takes advantage and contribute data and assess the system. Most of them are students, while the rest include a businessman, an office worker, and a waitress. Each volunteer carries a Nexus S smart phone with Friendbook application installed priorly. They are required to start the application after they wake up and turn it off before they go to bed. Separately from this, we do not levy any extra requirement on the usage of the smart phone. For example, we do not require them to carry the smart phone all the time during the day or attach the smart phone to some special parts of the body.[13] It is worth noting that some of the eight users are already friends before experiments but some of them are not. In fact, some strangers within the group become friends afterwards. Nevertheless, strangers living far away from each other do not become friends although they choose each other as a friend at the friend recommendation phase. This also encourages the usage of GPS information into the system to progress the recommendation accurateness.

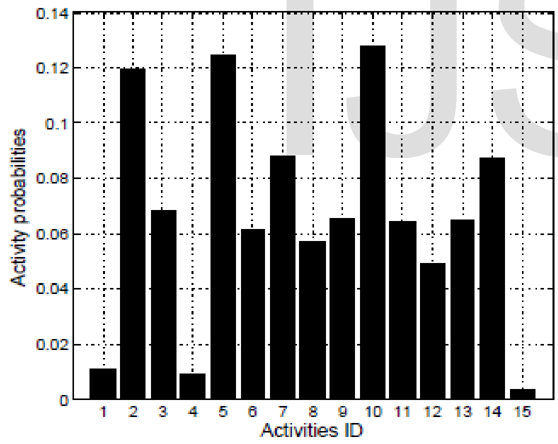
ISSN 2229-5518

TABLE 1: Profession of users

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



(a) Classification results

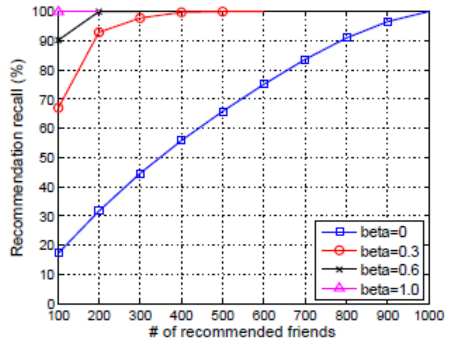


(b) Activity distribution

Fig. 8: Classification performance using the K-means clustering.

5.1 Effect of Similarity Threshold:

It firstly estimates the effect of the similarity threshold which is significant for friend-matching graph construction. The value of the similarity threshold for the friend-matching graph construction should be cautiously chosen.



(a) Recommendation recall

5.2 Resource Consumption

Evaluating the energy consumption enactment of the Friendbook client application is done and frequently, the user will not have the enticement to use the application if the battery runs out in less than 10 hours, which is the typical hour of usage for a day. Then, energy consumption is another vital metric that has to be measured. The energy consumption of the same smart phone under two modes: idle mode with Friendbook off and active mode with Friendbook on. Either mode is under a user's normal use such as[14] making phone calls, checking emails, sending SMS, etc. Friendbook drops the battery to 15% in about 13 hours. The evaluation shows that Friendbook achieves satisfactory results on the energy performance.

TABLE 3: Resources requirements comparison.

Aspect	Friendbook Client	Android Service	Google Maps
Size	76 KB	–	12 MB
CPU usage	<1%	13%	<1%
Data usage	<10KB	1.95MB	526KB
Runtime memory	4.2MB	33MB	6.9MB

6 CONCLUSION AND FUTURE WORK

In this Novel approach for semantic based friend recommendation system for social networks, the design and implementation of Friendbook, a semantic-based friend recommendation system for social networks is projected. Unlike from the friend recommendation contrivances depending on social graphs in present-day social networking services, Friendbook haul out life styles from user-centric data collected from sensors on the smart phone and suggested prospective friends to users if they stake related life styles. The future work can be to estimate the system on large-scale field experimentations.

REFERENCES

[1] Amazon. http://www.amazon.com/.

[2] Facebook statistics. http://www.digitalbuzzblog.com/ facebook-statistics-stats-facts-2011/.

[3] Netfix. https://signup.netflix.com/.

[4] Rotten tomatoes. http://www.rottentomatoes.com/.

[5] G. R. Arce. Nonlinear Signal Processing: A Statistical Approach. John Wiley & Sons, 2005.

[6] B. Bahmani, A. Chowdhury, and A. Goel. Fast incremental and personalized pagerank. Proc. of VLDB Endowment, volume 4, pages 173-184, 2010.

[7] 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.

[8] L. Bian and H. Holtzman. Online friend recommendation through personality matching and collaborative filtering. Proc. of UBICOMM, pages 230-235, 2011.

[9] 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.

[10] K. Farrahi and D. Gatica-Perez. Discovering Routines from

[11] 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. Largescale Human Locations using Probabilistic Topic Models. ACM Transactions on Intelligent Systems and Technology (TIST), 2(1), 2011.

[12] 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.

[13] X. Yu, A. Pan, L.-A. Tang, Z. Li, and J. Han. Geo-friends recommendation in gps-based cyber-physical social network. Proc. Of ASONAM, pages 361-368, 2011.

[14] 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.

AUTHORS

[1] Dr V. Goutham is a Professor and Head of the Department of computer Science and Engineering at Teegala Krishna Reddy Engineering College affiliated to J.N.T.U Hyderabad. He received Ph.d from Aacharya Nagarjuna University and M.Tech from Andhra University and B.Tech from J.N.T.U Hyderabad. He worked for various MNC Companies in Software Testing and Quality as Senior Test Engineer. His research interests are Software Reliability Engineering, software testing, software Metrics, and cloud computing.

[2] Mr T.Ruthvik Manikanta Department of computer Science and Engineering at Teegala Krishna Reddy Engineering College affiliated to J.N.T.U Hyderabad.