A Profile-based Friend Detection Scheme for Online Social Networks

Vandana V

Final Year M.Tech Degree, Dept. of Computer Science and Engineering
Cochin Institute of Science and Technology, Muvattupuzha, Kerala, India

Abstract - In past few years, users created friends with every other World Health Organization live or work near them, like neighbor or work place etc. this kind of friendly relationship is named as ancient means of constructing friends or G. Friend referred to as geographical primarily based friends. Now social networking services recommend friends to users based on their social graphs, which may not be the most accurate in friend selection. An efficient friend recommendation system for social networks, which recommends friends to users based on their life styles instead of social graphs. It can be applicable in smart phones, discovers life styles of users from user-centric sensor data, measures the similarity of life styles between users, and recommends friends to users if their life styles have high similarity. We adopt Latent Dirichlet Allocation algorithm for lifestyle extraction. Similarity metric to measure the similarity of life styles between users, and calculate user’s Impact in terms of life styles. Also have a feedback mechanism to improve recommendation accuracy.

In this paper, we propose a trust-based friend recommendation scheme for OSNs, where OSN users apply their attributes to find matched friends, and establish social relationships with strangers via a multi-hop trust chain. Based on trace-driven experimental results and security analysis.

Keywords – Friend recommendation, mobile sensing, social networks, lifestyle, trust, OSN-online social networks.

I. INTRODUCTION

Friend suggestions mechanism mostly relies on pre-existing user relationships to pick friend candidates. For instance, Facebook relies on a social link analysis among those who already share common friends and recommends users with some sort of similarities as potential friends. Habits or life style, Attitudes, Tastes, Moral standards, Economic level; and People they already know are the common rules to group people together.

Recommender systems (RS) are the ones used to recommend friends to the users based on some criteria. They are normally used to handle and solve the information overload. The existing social networking system includes Netflix for recommending movies, Facebook for recommending friends, Foursquare for recommending places etc. The existing social networking services find it challenging as how to recommend a good friend to a user, as most of them rely on the pre-existing user relationships to pick the friends. There are typically two types of algorithms for recommender systems:

1. content-based methods- measure the similarity of the recommended item to the ones that a target user likes or dislikes based on the item attributes.
2. collaborative filtering method- that finds users with tastes that are similar to the target users based on their past ratings. This can then make the recommendations to the target user based on the opinions of similar users.

For more than a decade, RSs have been proposed to overcome the information overload and many algorithms and systems have been developed for the same. Still, they face the challenge of cold start users and data sparsity in case of Collaborative filtering technique. The recent emergence of online social networks provides us with enormous amount of information related to user behaviour and friend interactions have demonstrated their importance to develop an efficient RS in this field. Facebook's friend recommender system is based on the concept of social graphs. It features people as the people you may know through connections on the user’s profile i.e. mutual friends based on their work, education details, networks etc. Finding mutual friends is generally by means of finding the similarity in the case of school, college or work place. Facebook guesses friend relationship as the number of mutual friends.

II. RELATED WORKS

[Katayoun Farrahi and Daniel Gatica-Perez] In this paper they discover the daily location-driven routines which are contained in a massive real life human data set collected by mobile phones. We aim to discover and analyse human routines which characterize both individual and group behaviors in terms of location patterns. We develop an unsupervised methodology based on two differing probabilistic topic models and apply them to the daily life of 97 mobile phone users over a 16 month period to achieve these goals. Topic models are probabilistic generative models for documents that identify the latent structure that underlies a set of words. Routines dominating the entire group’s
activities, identified with a methodology based on the Latent Dirichlet Allocation topic model, include “going to work late”, “going home early”, “working non-stop” and “having no reception (phone off)” at different times over varying time-intervals. They also detect routines which are characteristic of users, with a methodology based on the Author-Topic model. With the routines discovered, and the two methods of characterizing days and users, we can then perform various tasks. They use the routines discovered to determine behavioral patterns of users and groups of users. For example, they can find individuals that display specific daily routines, such as “going to work early” or “turning off the mobile (or having no reception) in the evenings”. We are also able to characterize daily patterns by determining the topic structure of days in addition to determining whether certain routines occur dominantly on weekends or week days. Furthermore, the routines discovered can be used to rank users or find subgroups of users who display certain routines. We can also characterize users based on their entropy. We compare our method to one based on clustering using K-means. Finally, they analyze an individual’s routines over time to determine regions with high variations, which may correspond to specific events.

[Jeff Naruchitparames, Mehmet Hadi Gunes, and Sushil J. Louis] Social networking sites employ recommendation systems in contribution to providing better user experiences. The complexity in developing recommendation systems is largely due to the heterogeneous nature of social networks. This paper presents an 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. Our research emphasizes that by combining network topology and genetic algorithms, better recommendations can be achieved compared to each individual counterpart. We test our approach on 1,200 Facebook users in which we observe the combined method to outperform purely social or purely network-based approaches. Our preliminary results represent strong potential for developing link recommendation systems using this combined approach of personal interests and the underlying network.

[Anil K.R and Gladston Raj S] The web friend recommendation system is the emerging trend in the social network and e-commerce. It plays an important role in the web based applications. The frequent pattern mining is one of the techniques used in mining applications and N-list is the efficient data structure representation used with this. The data pruning strategy has always been a challenge on improving the efficiency of N-list based system. In this paper we present a pre-rule based N-list which explores the power of association rule mining with pre-rule checked N-list. This technique enhances the speed and efficiency of the friend recommendation system. Thus the Pre-rule checking based algorithm has succeeded in establishing an improvement in the precision of recommendation systems.

**N-List**

N-list is data structure proposed to represent frequent patterns. The ability to mine patterns along the N-list nodes is made possible by maintaining a subset concept. N-list has a property called single path property which helps to mine data in a faster way. Recently proposed Pre-Post Code Tree algorithm for mining FP based N-list structure and children parent equivalence pruning is another high-performance technique for mining frequent item sets. It employs N-list to represent item sets and directly discovers frequent item sets using a set enumeration search tree.

**Draw Backs of N-List**

Implementation of N-list may be inappropriate in large data sets. The efficiency of N-list based algorithms is challenged when the consumption of memory is high and the processing speed is low. Also low support features are always generated in mined results, which caused unfavourable results.

**III. PROPOSED SYSTEM**

Existing friend recommendation system for social networks recommends friends to users based on their life styles instead of social graphs. We propose a system for recommending friends by using sensor-rich smart phones, and discovers life styles of users from user-centric sensor data, and finding the similarity of life styles between users, and recommends friends to users if their life styles have high similarity.

We find user’s daily life as life documents, from which his/her life styles are extracted by using the Latent Dirichlet Allocation algorithm. Similarity metric to measure the similarity of life styles between users, and calculate impact of users in terms of life styles with a friend-matching graph. We adopt an efficient feedback mechanism that exploits the user’s feedback to improve recommendation accuracy. Recommend potential friends to users if they share similar life styles. The feedback mechanism allows us to measure the satisfaction of users, by providing a user interface that allows the user to rate the friend list.
A. Finding Lifestyles and Activities

Finding lifestyles and activities are the first step in recommendation. Daily lives are recognized by recognizing lifestyles and activities. They can be treated as a combination of life styles and life styles as a combination of activities. This is equivalent to the treatment of documents as collection of topics and topics as collection of words. Development of text mining is advantageous in modelling the lifestyle. We represent 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 underlying “topics”. Therefore, we adopt the probabilistic topic model to determine the probabilities of hidden “lifestyles” from the “life documents”. Frequency of vocabulary is very important in probabilistic topic models, as different frequency of words denotes their information entropy variances. Following this observation, we propose the “bag-of-activity” model (Figure 1) to change 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 combination of activities.

![Fig.1 Bag-of-Activity modeling for life document.](image)

Life styles are usually reflected 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: 1. supervised learning and 2. unsupervised learning. In both case, many techniques have been developed and tested. In practice, the number of activities involved in the analysis is unpredictable and it is difficult to collect original data for each activity, which makes supervised learning algorithms inappropriate for our system. Therefore, we propose unsupervised learning approaches to recognize activities. Here, we adopt the K-means clustering algorithm to group data into clusters, where each cluster represents an activity. Mere activity recognition is not our main concern. Other more complicated clustering algorithms can definitely be used. K-means clustering algorithm is very simple and effective.

![Fig 2 flow chart of activity recognition.](image)

Original data collected on the smart phones are noisy, we first use a median filter with sliding windows to filter out the outliers of the noisy data. For improving the accuracy, features are extracted to characterize the data after preprocessing. We have tested some features, such as mean, standard deviation, correlation, and the combination of them, on the data and found that standard deviation is the most delegate feature for characterizing motion activities. Each smart phone could independently recognize activities based on the minimum distance rule and upload the activity sequence as an alternative of the raw data to the server.

B. LDA for Life Style Extraction

For extracting life styles from activity clusters we used Latent Dirichlet algorithm. LDA is a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying collection of topics. Each topic is, in turn, modeled as an infinite combination over an underlying set of topic probabilities. In the context of text modeling, the topic probabilities provide an explicit representation of a document. We present efficient and approximate inference techniques based on variational methods and an EM algorithm for empirical Bayes parameter estimation. We describe results in document modeling, text classification, and collaborative filtering, comparing to a mixture of unigrams model and the probabilistic LSI model.
LDA Algorithm

**Input:** Number of topics K

**Output:** Model parameters $\beta, \theta$.

1. Initialize $\theta_0 = 1/k$ for all $i$ in $k$ and $n$ in $N_k$.
2. Initialize $\gamma_0 = n/N_k$ for all $i$ in $k$.
3. Initialize $\eta_0 = 50/k$.
4. For all $i$ in $k$ and $j$ in $V$.
   - $\phi_{i,j} \leftarrow \text{Sample from } \beta_{j,k}$
   - $\gamma_{i,j} \leftarrow \text{Sample from } \eta_{i,j}$

   Repeat for $s = 1$ to $N$:
   - For all $i$ in $k$ and $j$ in $V$.
     - $\phi_{i,j} \leftarrow \text{Sample from } \beta_{j,k}$
     - $\gamma_{i,j} \leftarrow \text{Sample from } \eta_{i,j}$
   - Normalize $\phi_{i,j}$ to sum to 1.
   - For all $i$ in $V$.
     - $\gamma_i \leftarrow \text{Sample from } \gamma_i$.
     - Update $\eta_{i,j} = \text{Sim}(\beta_{j,k}, \gamma_i)$.
   - Normalize $\gamma_{i,j}$ to sum to 1.
   - $\phi_{i,j} \leftarrow \text{Sample from } \beta_{j,k}$
   - $\gamma_{i,j} \leftarrow \text{Sample from } \eta_{i,j}$

**C.Constructing Friend Graph**

Friend graph is a weighted undirected graph. Let the graph $G=(V,E)$. Here $V$ is considered user and $E$ is the edge between two users with any relationship. $\text{Sim}(u_1,u_2)$ = $\sum_{i \in LV(u_1)} \sum_{j \in LV(u_2)} \text{Sim}(LV(u_1), LV(u_2))$ where $LV$ is the life style value.

*Fig.3.* An example of Friend-matching Graph for 8 users.

Following are the over all steps behind finding friends.

- Find current users list.
- Collect activities of each users in users list.
- Group these activities based on their lifestyle.
- Find life style values of each user that means number of actual life style per total number of life styles.

- Determine user to friend similarity.
- Construct friend graph.
- Find friends by finding path from each $u_i$.
- Sort these $u_i$ based on path weight in descending order.
- Select top $N$ users.

**IV. CONCLUSION AND FUTURE WORK**

Security is an important issue for different social networks, profile based friend recommendation system allows proper friend suggestions. The main purpose is to implement a system that can efficiently set up semantic-based friend recommendation system for social networks, which recommends friends to users based on their life styles instead of social graphs. It can be applicable in smart phones, discovers life styles of users from user-centric sensor data, measures the similarity of life styles between users, and recommends friends to users if their life styles have high similarity. We adopt Latent Dirichlet Allocation algorithm for lifestyle extraction. Similarity metric to measure the similarity of life styles between users, and calculate user’s Impact in terms of life styles. Also have a feedback mechanism to improve recommendation accuracy. Screening of friends by recognizing applications used by different users in their daily life. Inspired by 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 algorithm. We further propose a similarity metric to measure the similarity of life styles between users, and calculate users’ impact in terms of life styles. Beyond the current prototype, the future work can be as follows: 1: We would assess the system on massive scale field. 2: We would implement the un-varied matrix vector technique in user impact ranking, thus friend recommendation would be enlarge on larger basis systems. 3: The similarity used for locating the friend recommendation. It’d be attention-grabbing to additional expand the difference for every edge and check for higher illustration on the connection on friend matching graph. 4: Finally we tend to integrate more sensors on mobile phones into the system and conjointly further implement the data from wearable instrumentality.

**REFERENCES**


[4] Namrata M.Eklaspur and Anand S.Pashupatimath; SDMCET, Dharwad; A Friend Recommender System for
Social Networks by Life Style Extraction Using Probabilistic Method −“Friendtome”

[5] Nathan Eagle and Alex(Sandy) Pentland; Realitymining:sensingcomplexsocialsystems.


