



# FRIENDBOOK: A SCALABLE AND EFFICIENT FRIEND RECOMMENDATION USING INTEGRATED FEEDBACK APPROACH

T. Gayathri Devi and R. Lakshmi

Department of Information Technology, MNM Jain Engineering College, Thoraipakkam, Chennai, India

E-Mail: [tgdevi05@gmail.com](mailto:tgdevi05@gmail.com)

## ABSTRACT

Friend book is a novel semantic-based friend recommendation system for social networks, based on their life styles instead of social graphs which recommends friends to users. Friend book discovers life styles of users, measures the similarity of life styles between users, if their life styles have high similarity it recommends friends to users. User's daily life is modelled as life documents, from which users life styles are extracted by using the Latent Dirichlet Allocation algorithm; Similarity metric to measure the similarity of life styles between users, user's impact is calculated in terms of life styles with a friend-matching graph. A linear feedback mechanism is integrated that exploits the user's feedback to improve recommendation accuracy.

**Keywords:** friend book, activity recognition, weighted page rank algorithm.

## 1. INTRODUCTION

Mobile Computing is a technology that allows transmission of data, voice and video via a computer or any other wireless enabled device without having to be connected to a fixed physical link. Mobile computing is human-computer interaction by which a computer is expected to be transported during normal usage. Mobile computing involves:

- Mobile communication
- Mobile hardware
- Mobile software

### a) Mobile communication

The mobile communication in this case, refers to the infrastructure put in place to ensure that seamless and reliable communication goes on. These would include devices such as Protocols, Services, Bandwidth, and Portals necessary to facilitate and support of the stated services. The data format is also defined at this stage. This ensures that there is no collision with other existing systems which offer the same service. Since the media is unguided/unbounded, the overlaying infrastructure is more of radio wave oriented. That is, the signals are carried over the air to intended devices that are capable of receiving and sending similar kinds of signals.

### b) Mobile hardware

Hardware includes mobile devices or device components.

### c) Mobile software

Mobile software deals with the characteristics and requirements of mobile applications. Many studies have reported that people typically made friends with others who live or work close to themselves, such as neighbors or colleagues. We call friends made through this traditional fashion as G-friends, which stands for geographical location-based friends because they are influenced by the geographical distances between each other. With the rapid advances in social networks, services

such as Face book, Twitter and Google+ have provided us revolutionary ways of making friends. One challenge with existing social networking services is how to recommend a good friend to a user. In this paper, we present Friend book, a novel semantic-based friend recommendation system for social networks, which recommends friends to users based on their life styles instead of social graphs. Upon receiving a request, Friend book returns a list of people with highest recommendation scores to the query user. In our proposed method, we use incremental computation of Page Rank can be implemented incrementally (or) distributive for large scale evolving graphs. Additionally, we propose a novel algorithm, Weighted Page Rank algorithm which distributes rank score based on popularity of the pages and we set threshold for each edge & it can represent the similarity relationship of friend-matching graph. Friend book is a semantic-based friend recommendation system for social networks. Different from the friend recommendation mechanisms relying on social graphs in existing social networking services, Friend book extracted life styles from user-centric data collected and recommended potential friends to users if they share similar life styles.

### Challenges

- System to take user feedback into consideration to improve the accuracy of future recommendation.
- Becoming increasing infeasible when size of system becoming very large.
- In this paper, similarity threshold used for friend-matching graph is fixed.

## 2. RELATED WORK

E. Miluzzo *et al.* [4] enables technology for mobile phone sensing to know about human behavior and context on mobile phones which uses combination of collaborative sensing and classification techniques. This is the first system that applies distributed machine learning



techniques and collaborative inference concepts to mobile phones. To achieve better interference accuracy we are using collaborative sensing. Machine learning techniques specifically designed to run directly on sensor enabled mobile phones. This paper is an automated approach to updating models over time such that the classifiers are robust to the variability in sensing conditions and settings common to mobile phones.

#### Advantages

- The classifiers are robust.
- The classifier methods are automated to update the models.
- Collaborative inference is implemented to achieve better accuracy.

K. Farrahi *et al.* [3] investigate probabilistic topic models as unsupervised machine learning tools for large-scale socio-geographic activity mining. They propose a methodology based on Latent Dirichlet Algorithm (LDA) for the discovery of dominant location routines. In this paper they used two Probabilistic model namely Multi-Level Topic Model and Pair wise –Distance Topic Model. First they propose a Multi-Level Topic Model as a method to incorporate multiple time duration sequences into a probabilistic generative topic model. And then they propose the Pair wise-Distance Topic Model as an approach to address the problem of modeling long duration activities with topics. Overall, this thesis addresses Investigations principled on mathematical models and multiple types of mobile phone sensor data are performed to mine real life human activities in large-scale scenarios.

#### Advantages

- Used for computing large scale data.
- Removes the need for coarse time-slot.

T. Huynh *et al.* [2] develop an unsupervised methodology based on two differing probabilistic topic models and apply them to the daily life. They have proposed a method to represent location sequences, and incorporated this into the LDA and ATM topic models. The resulting distributions of words for latent topics, as well as topics given days, and topics given users, re-veal hidden structure of routines which use to perform varying tasks, including finding users or groups of users that display given routines, and determining times.

#### Advantage

- Used to compute large amount of data easily.

L. Bian *et al.* [1] present Matchmaker, a collaborative filtering friend recommendation system based on personality matching. The goal of Matchmaker is to leverage the social information and mutual understanding among people in existing social network connections, and produce friend recommendations based on rich contextual data from people's physical world interactions. Matchmaker allows users' network to match them with similar TV characters, and uses relationships in

the TV programs as parallel comparison matrix to suggest to the users friends. The system's ranking schema allows progressive improvement on the personality matching consensus and more diverse branching of users' social network connections.

#### Advantages

- Collaborative filtering is used for large datasets. Ranking schema allows progressive improvement on the personality matching consensus.

Katayoun Farrahi *et al.* [5] proposed a method to recognize daily routines as a probabilistic combination of activity patterns. The use of topic models enables the automatic discovery of patterns in a user's daily routine. They report experimental results that show the ability of the approach to model and recognize daily routines without user. The conclusion of this paper, the approach used is highly appealing for the field of activity recognition. To overcome this, they used a technique namely unsupervised techniques such as clustering.

#### Advantages

- Recognizes the daily routines without user annotation.
- Enables scalability to long term recordings of activities.

Wenpu Xing *et al.* [6] present with the rapid growth of the Web, users get easily lost in the rich hyper structure. Therefore, finding the content of the Web and retrieving the users' interests and needs from their behavior have become increasingly important. In this paper, Web Mining is used to categorize users and pages by analyzing the users' behavior, the content of the pages, and the order of the URLs that tend to be accessed in order. Two Page Ranking Algorithm is used namely HITS and Page Rank. Both algorithms treat all links equally when distributing rank scores. Weighted Page Rank algorithm is used to improve the performance. The results of this simulation show that Weighted Page Rank performs better than the conventional Page Rank algorithm in terms of returning larger number of relevant pages to a given query.

#### Advantages

- Better performance than traditional Page Rank algorithm.
- Identifies a large number of relevant pages for a given query.
- Rank scores are based on the popularity of the pages.

### 3. MODULE DESCRIPTION

This section gives an overview of Figure-2.

#### a) 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. 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.

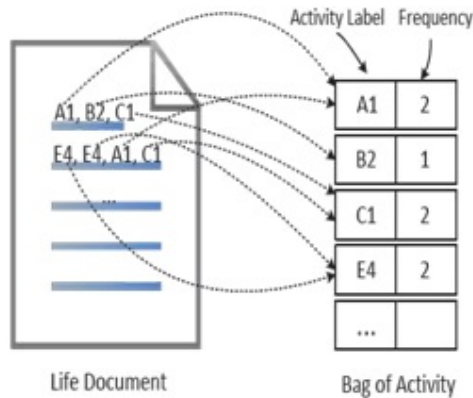


Figure-1. Bag of activity modeling.

#### b) 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. For both approaches, mature techniques have been developed and tested. In practice, the number of activities involved in the analysis is unpredictable and it is difficult to collect 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 to group data into clusters, where each cluster represents an activity. Note that activity recognition is not the main concern of our paper. Other more complicated clustering algorithms can certainly be used. We choose K-means for its simplicity and effectiveness.

#### c) 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. 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. We define a new similarity metric to measure the similarity between two life style vectors. Based on the similarity metric, we model the relations between users in real life as a friend-matching graph. The friend-matching graph has been constructed to reflect life style relations among users.

#### d) User impact ranking

The impact ranking means a user's capability to establish friendships in the network. Once the ranking of a user is obtained, it provides guidelines to those who

receive the recommendation list on how to choose friends. The ranking itself, however, should be independent from the query user. The ranking depends only on the graph structure of the friend-matching graph, which contains two aspects: 1) how the edges are connected; 2) how much weight there is on every edge. This can be achieved using Weighted Page Rank algorithm.

#### 4. PROPOSED SYSTEM

Our proposed system addresses the problem of computing large amount of data and scalability. In our proposed method, we use incremental computation of Page Rank, can be implemented incrementally (or) distributively for large scale evolving graphs. Additionally, we propose a novel algorithm, Weighted Page Rank algorithm which distributes rank score based on popularity of the pages and we set threshold for each edge & it can represent the similarity relationship of friend-matching graph. We also have implemented Hex Converter to efficiently secure the password of the users. Friend book returns a list of people with highest recommendation scores to the query user. Finally, Friend book integrates a feedback mechanism to further improve the recommendation accuracy.

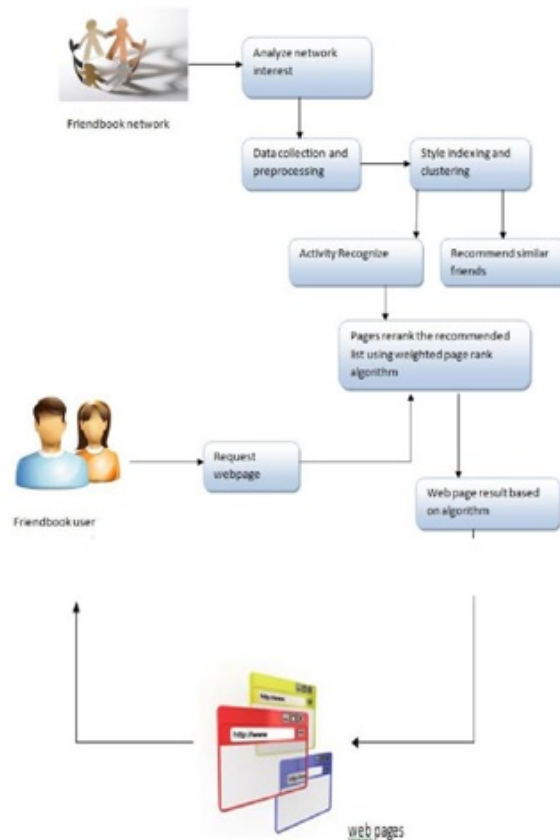


Figure-2. System architecture.

