IPHV8I1025X
International Journal Of Advanced Research and Innovation -Vol.8, Issue .I
ISSN Online: 2319 – 9253
Print: 2319 – 9245

FRIEND BOOK: PRIVACY ORIENTED FRIEND MATCHING BASED ON SHARED INTERESTS

[#1]G.ASHOK – Pursuing M.Tech,
[#2]A.KARTHEEK REDDY –Assistant Professor,
Dept of CSE,
SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, T.S., INDIA.

Abstract: Online friend recommendation is a rapid developing topic in web mining. Current social networking servicing recommend friends to users based on their social graphs and mutual friends , which may not be the most appropriate to reflect a user’s taste on friend selection in real lifetime .Existing social networking services recommend friends to users based on their social graphs, which may not be the most appropriate to reflect a user’s preferences on friend selection in real life. In this paper, we present Together4ever, a novel social network with a social-life based friend recommendation system, which recommends friends to users based on their life styles instead of social graphs. In this paper propose a system that recommends friends based on the daily activities of users. Here a semantic based friend recommendation is done based on the users life styles. By using text mining, we display a user's everyday life as life archives, from which his/her ways of life are separated by using the Latent Dirichlet Allocation algorithm. At that point we discover a similarity metric to quantify the similarity of life styles between users, and ascertain users effect as far as ways of life with a similarity matching diagram. At last, we incorporates a feedback component to further enhance the proposal precision.

Keywords: Mesh Topology, Multipoint Relay, Stateless Multicast Protocol, Virtual MIMO, Wireless Sensor Networks.

I.INTRODICTION

A social network is a system where clients (nodes) are joined with one another by relationship (edges). The edges are undirected and the quantity of edges demonstrates the quantity of companions a client has. A percentage of the remarkable interpersonal organizations are Facebook, Google plus LinkedIn and so forth. Each client keeps up a profile. There are numerous properties in the profile which can be utilized to anticipate the quality of ties between diverse clients.

The vast majority of the friend recommendations system depends on previous client connections to pick friend candidates. For example, Facebook depends on a social connection examination among the individuals who as of now impart basic friends and suggests symmetrical clients as potential friends. Existing social networking services prescribe friends to users based on their social graphs, which may not be the most appropriate to reflect a users preferences on friend selection in real life. With the quick advancement of social network, recommendation systems in different fields rose. A decent suggestion framework ought to consolidate different sorts of suggestion impacts and assurance differences on the base of exactness, in order to fulfil some disagreeable tastes.

One test with existing social networking services is the way to prescribe a good friend to a client. Most of them depend on previous user connections to pick friend candidates. In our ordinary lives, we may have several activities, which structure important groupings that shape our lives. In this paper, we utilize the word activity to explicitly refer to the actions made in the order of seconds, for example, "running", "strolling", or "perusing", while we utilize the expression way of life to allude to more elevated amount reflections of day by day lives, for example, "office work" or "shopping". For example, the "shopping" way of life basically comprises of the "strolling" movement, however might likewise contain the "standing" then again the "sitting" exercises.

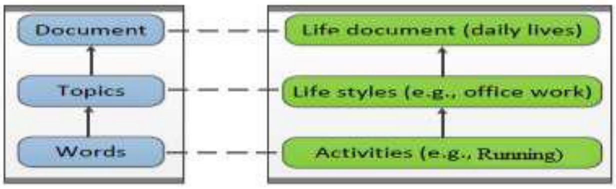

Fig. 1: A Relationship Between Word Archives And Individuals' Everyday Lives

The commitments of this work are summarized as follows:

- Friend recommendation is done based on life style of users.
- We display the everyday lives of clients as life reports by collecting activities and use the probabilistic topic model to extract life style data of clients.
- Then using similarity metric and calculate the similarity between users and constructing friend matching graphs.
- A user feedback mechanism for improving accuracy and based on that decide optimum threshold value.

To model daily lives properly, we draw an analogy between people's daily lives and documents, as shown in Figure 1. Previous research on probabilistic topic models in text mining has treated documents as mixtures of topics, and topics as mixtures of words. Inspired by this, similarly, we can treat our daily lives (or life documents) as a mixture of life styles (or topics), and each life style as a mixture of activities (or words). Observe here, essentially, we represent daily lives with "life documents", whose semantic meanings are reflected through their topics, which are life styles in our study Just like words serve as the basis of documents, people's activities naturally serve as the primitive vo-cabulary of these life documents. Our proposed solution is also motivated by the recent advances 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. In spite of the powerful sensing capabilities of smart phones, there are still multiple challenges for extracting users' life styles and recommending potential friends based on their similarities.

First, how to automatically and accurately discover life styles from noisy and heterogeneous sensor data? Second, how to ensure the similarity of users in terms of life styles?

Third, who should be recommended to the user among all the friend candidates? To address these challenges, in this paper, we present Toghter4Ever, a social-life based friend recommendation system based on sensor-rich smart phones.

The contributions of this work are summarized as follows:

•To the best of our knowledge, Toghter4Ever is the first friend recommendation system exploiting a user's life style information discovered from smart phone sensors.

•Inspired by achievements in the field of text mining, we model the daily lives of users as life documents and use the probabilistic topic model to extract life style information of users.

•We propose 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.

•We integrate a linear feedback mechanism that exploits the user's feedback to improve recommendation accuracy.

•We conduct both small-scale experiments and large scale simulations to evaluate the performance of our system. Experimental results demonstrate the effectiveness of our system.

Advantages of social networks:

1.Social networking helps people stay in touch that might not do it otherwise.

2. Social networking can be used to help advertise goods and services.

3. Social networking can provide an extremely accessible medium for self expression to those with access to computer.

4. Social networking can help families torn apart by war, divorce, etc. stay in touch easier and quicker than by some other means.

5. Social networking can be a powerful engine for job searches.

6. Social networking can be used to find dating part-ners in a fractured society where healthy meeting plac-es are limited.

7. Social networking can be used to memorialize and honour dead persons to keep memories and history that would otherwise fade alive.

8. Social networking can be used to get difficult per-sonal issues out of the closet in front of others so they can be examined and evaluated, and people in trouble can garner support from their friends.

II.RELATED WORK

There is a broad class of Web applications that include anticipating client predicting user responses to options. Such a facility is called recommendation system. Recommendation systems can be separated into two areas of center: object suggestion and link recommendation. Organizations, for example, Amazon and Netflix stress object suggestion where items are prescribed to clients in light of past behavioral examples [2], [3], [4]. Person to person communication destinations for example, Facebook and LinkedIn concentrate on connection suggestion where companion suggestions are introduced to clients. The work we present in this paper mainlyfocuses on the latter, in which we develop friend Recommendation system within social networks. The recommendation systems employed by different sites are based on mutual friends. Friendbook [1], a novel semantic-based friend suggestion system for social networking communities, which prescribes friends to clients focused around their ways of life rather than social graphs. By exploiting sensor-rich cell phones, Friendbook finds ways of life of clients from client driven sensor information, measures the closeness of ways of life in the middle of clients, and prescribes friends to clients if their ways of life have high likeness.LDA is a probabilistic model for collecting distinct data with a three-level hierarchical Bayesian model, where each item of a collection is modelled as a definite mixture over an underlying set of topics or words. The data collection module gathers life reports from users post and other activities. The ways of life of clients are separated by the way of life analysis module with the probabilistic topic model. Latent Dirchlet allocation algorithm is a probability based entropy model with more accuracy. It extracts topics from set of words. Based on the similarity metric, we model the relations between users in real life as a friend-matching graph. Friend matching graph: It is a weighted undirected graph G=(V,E,W) where V represents users, Edge represents relationship and Weight represents similarity between users. At a time only 5-10 recommendation are made to the user. Suppose the user accept one of the recommendations or sends a friend request then a new recommendation from top of the k-neighbour list is suggested to the user. Also, changes are made to the „knearest" neighbour list which ensures the same people are not recommended again. To support performance optimization at runtime, we also integrate a feedback control mechanism. Based on the feedback from the user the threshold value can be set. Under different threshold value the algorithm is evaluated for the best results.

Netflix [4] and Rotten Tomatoes [5] suggest movies to a client focused around the client's past evaluations also viewing propensities. As of late, with the development of social networking systems, friend recommendation has gotten a ton of consideration. As a rule, existing friend recommendation in long range social networking systems, e.g., Facebook, LinkedIn and Twitter, suggest friends to clients if, as per their social relations, they impart common friends. In the interim, other proposal components have additionally been proposed via analysts.

Another Suggestion based on geologically related friends in social network by joining GPS data and social network structure[8]. The advancement of GPS-empowered mobile phones gives social network researchers a taste of digital cyber-physical social network in advance. Traditional link prediction methods are intended to discover companions exclusively depending on social network information. With area and direction information accessible, we can create more exact and topographically related results, and help web-based social service users find more friends in this present reality. Planning to suggest topographically related companions in interpersonal organization, a three-stage measurable proposal methodology is proposed for GPS-enabled digital physical informal community. By consolidating GPS data and informal community structures, we fabricate an example based heterogeneous information system. Links inside this system reflect both individuals' geographical information, and their social connections. Our methodology assessments join significance and discover promising geo-friends by employing a random walk process on the heterogeneous information network. Exact studies from both manufactured datasets and genuine dataset exhibit the force of fusing GPS information and social diagram structure, and recommend our technique outflanks different routines for companions proposal in GPS-based digital physical informal organization.

Link recommendation [9] in weblogs and comparable social networks, and proposed an methodology focused around community suggestion utilizing the link structure of an social network and substance based proposal utilizing shared pronounced diversions.Kuan et al. proposes an algorithm to place gatherings utilizing a transitive extension based methodology [10]. This examination proposed represented the utilization of a 1.5cclique expansion technique to infer substructures, or groups, inside informal organizations. Results demonstrated that this strategy was genuinely compelling in finding group of friends. Be that as it may, this technique does not give knowledge into how these groups are structured. That is, it is noteworthy to comprehend what basic hobbies cause a arrangement in these groups. Activity recognition serves as the premise for separating abnormal state day by day schedules (in close connection with ways of life) from low-level sensor information, which has been broadly mulled over utilizing different sorts of wearable sensors.

Reddy et al. [14] utilized the inherent GPS and the accelerometer on the cell phones to discover the transportation mode of an single person. Cenceme [15] utilized numerous sensors on the cell phone to catch client's exercises, state, propensities and surroundings. Soundsense [16] utilized the amplifier on the cell phone to perceive general sound sorts (e.g., music, voice) and find client particular sound occasions. Easytracker [17] utilized GPS follows gathered from cell phones that are introduced on travel vehicles to focus courses served, find stops, and induce plans. The MIT Reality Mining project [18] and Farrahi and Gatica-Perez [19] attempted to find every day area driven schedules from huge scale area information. They could construe every day schedules, for example, leaving from home to office and consuming at a restaurant.

Collaborative Filtering (CF) based Recommender Systems are most essential procedures of prescribing things to the clients. The easiest and unique execution of this methodology prescribes to the dynamic client the things that different clients with comparative tastes enjoyed before. The likeness in taste of two clients is ascertained in view of the similitude in the rating history of the clients. Collaborative

Filtering (CF) frameworks work by gathering client criticism as appraisals for things in a given space and abusing similitudes in appraising conduct amongst a few clients in deciding how to prescribe a thing. Collaborative oriented Filtering (CF) routines can be further subdivided into neighbourhood-based and model-based approaches. Collaborative Filtering makes a gathering of clients with comparative conduct, and finds the things favoured by this gathering. Evaluations from client will be taken from client in two ways unequivocal rating and certain rating [5]. CF calculations are partitioned into two sorts, memory-based algorithm and model based algorithm. Memory-Based algorithm just stores all the client evaluations into memory. There are two variations of memory-based proposal and both are in view of the k-Nearest Neighbour calculation: client based sifting and thing based separating. In User – Based Filtering, Rating lattice is utilized to discover neighbouring clients for the dynamic client. This is carried out by utilizing cosine or Pearsons correlation matrix.

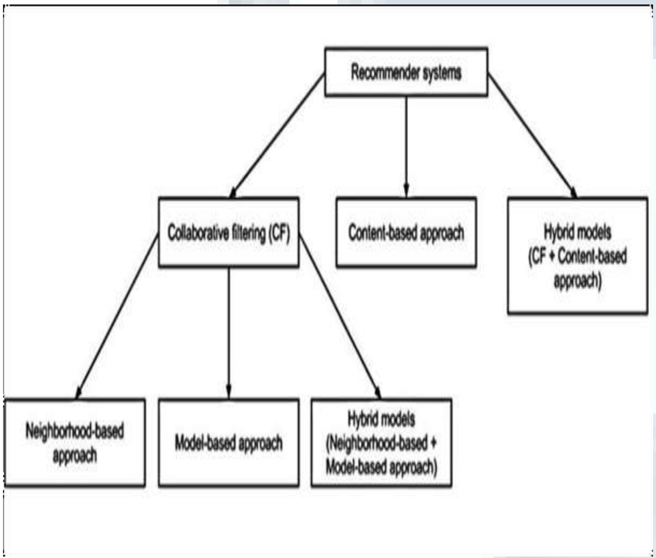

Fig. 2: Types of Recommender Systems

III.IMPLEMENTATION

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.

Friend-matching graph:

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.

User Impact Ranking:

The friend-matching graph has been constructed to reflect life style relations among users. However, we still lack a measurement to identify the impact ranking of a user quantitatively. Intuitively, the impact ranking means a user's capability to establish friendships in the network. 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. Inspired by Page Rank which is used in web page ranking, we form the idea that a user's ranking is reflected by his neighbors in the friend-matching graph and how much his neighbors endorse the user as a friend.

Query 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.

Feedback control :

To support performance optimization at runtime, we also integrate a feedback control mechanism into Togther4Ever. After the server generates a reply in response to a query, 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.

Algorithm in Use:

Introduction to latent Dirichlet allocation (LDA):

In natural language processing, latent Dirichlet allocation (LDA) is a generative model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar.

Suppose you have the following set of sentences:

•I like to eat broccoli and bananas.
•I ate a banana and spinach smoother for breakfast.
•Chinchillas and kittens are cute.
•My sister adopted a kitten yesterday.
•Look at this cute hamster munching on a piece of broccoli.

What is latent Dirichlet allocation? It's a way of auto-matically discovering topics that these sentences con-tain.

For example, given these sentences and asked for 2 topics, LDA might produce something like

•Sentences 1 and 2 100% Topic A

•Sentences 3 and 4 100% Topic B

•Sentence 550% Topic A, 40% Topic B

•Topic A 30% broccoli, 15% bananas, 10% breakfast, 10% munching, … (at which point, you could interpret topic A to be about food)

•Topic B 20% chinchillas, 20% kittens, 20% cute, 15% ham-ster, … (at which point, you could interpret topic B to be about cute animals) The question, of course, is how does LDA perform this discovery?

LDA: In more detail, LDA represents documents as mixtures of topics that spit out words with certain probabilities. It assumes that documents are produced in the follow-ing fashion when writing each document, you

•Decide on the number of words N the document will have (say, according to a Poisson distribution).

•Choose a topic mixture for the document (according to a Dirichlet distribution over a fixed set of K topics). For example, assuming that we have the two food and cute animal topics above, you might choose the docu-ment to consist of 1/3 food and 2/3 cute animals.

•First picking a topic (according to the multinomial distribution that you sampled above; for example, you might pick the food topic with 1/3 probability and the cute animals topic with 2/3 probability).

•Using the topic to generate the word itself (according to the topic's multinomial distribution). For example, if we selected the food topic, we might generate the word "broccoli" with 30% probability, "bananas" with 15% probability, and so on. Assuming this generative model for a collection of documents, LDA then tries to backtrack from the documents to find a set of topics that are likely to have generated the collection.

Example: Let's make an example. According to the above pro-cess, when generating some particular document D, you might

•Pick 5 to be the number of words in D.

•Decide that D will be 1/2 about food and 1/2 about cute animals.

•Pick the first word to come from the food topic, which then gives you the word "broccoli".

•Pick the second word to come from the cute animals topic, which gives you "panda".

•Pick the third word to come from the cute animals topic, giving you "adorable".

•Pick the fourth word to come from the food topic, giving you "cherries".

•Pick the fifth word to come from the food topic, giv-ing you "eating". So the document generated under the LDA model will be "broccoli panda adorable cherries eating" (note that LDA is a bag-of-words model).

IV.CONCLUSION

In this paper, we presented the design and implementation of Togther 4 Ever, a Social-Life based friend recommendation system for social networks. Different from the friend recommendation mechanisms relying on social graphs in existing social networking services, Togther4Ever extracted life styles from user-centric data collected from sensors on the smart phone and recommended potential friends to users if they share similar life styles. First, we would like to evaluate our system on large-scale field experiments. Second, we intend to implement the life style extraction using LDA and the iterative matrix-vector multiplication method in user impact ranking incrementally, so that Togther4Ever would be scalable to large-scale systems. Third, the similarity threshold used for the friend-matching graph is fixed in our current prototype of Togther4Ever. It would be interesting to explore the adaption of the threshold for each edge and see whether it can better represent the similarity relationship on the friend-matching graph.

REFERENCES:

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

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

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

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

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

[6] B. Bahmani, A. Chowdhury, and A. Goel. Fast incre-mental and personalized page rank. Proc. of VLDB En-dowment, volume 4, pages 173-184, 2010.

[7] J. Biagioni, T. Gerlich, T. Merrifield, and J. Eriksson. Easy Tracker: Automatic Transit Tracking, Mapping, and Arrival Time Prediction Using Smartphone's. 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] C. M. Bishop. Pattern recognition and machine learning. Springer New York, 2006.

[10] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. Journal of Machine Learning Research, 3:993-1022, 2003.

[11] P. Desikan, N. Pathak, J. Srivastava, and V. Kumar. Incremental page rank computation on evolving graphs. Proc. of WWW, pages 1094-1095, 2005.

[12] N. Eagle and A. S. Pentland. Reality Mining: Sensing Complex Cocial Systems. Personal Ubiquitous Computing, 10(4):255-268, March 2006.

[13] 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.

[14] 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.

[15] B. A. Frigyik, A. Kapila, and M. R. Gupta. Introduction to the dirichlet distribution and related processes. Department of Electrical Engineering, University of Washignton, UWEETR-2010-0006, 2010.