Inferring Social Relationships from Mobile Sensor Data

Hsun-Ping Hsieh, Cheng-Te Li
Graduate Institute of Networking and Multimedia,
National Taiwan University, Taipei 106, Taiwan
sandoh714@gmail.com, reliefli@gmail.com

ABSTRACT
While mobile sensors are ubiquitous nowadays, the geographical activities of human beings are feasible to be collected and the geo-spatial interactions between people can be derived. As we know there is an underlying social network between mobile users, such social relationships are hidden and hold by service providers. Acquiring the social network over mobile users would enable lots of applications, such as friend recommendation and energy-saving mobile DB management. In this paper, we propose to infer the social relationships using the sensor data, which contains the encounter records between individuals, without any knowledge about the real friendships in prior. We propose a two-phase prediction method for the social inference. Experiments conducted on the CRAWDAD data demonstrate the encouraging results with satisfying prediction scores of precision and recall.

Categories and Subject Descriptors
H.2.8 [**Database Management**]: Database Applications – *Data mining*.

Keywords
Social network, social relationship inference, mobile sensor data.

1. INTRODUCTION
Nowadays, sensors are built-in modern mobile phones and can interact with mobile devices carried by others. Thus, mobile sensors provide the potential to investigate the human relationships in the real-life environment. In this work, we aim at inferring social relationships between individuals from sensor encounter data without prior knowledge about physical friendships. We assume that the social relationships explicitly exist, but they are either hidden or kept secret by the owners or the service providers. We claim that by exploiting encounter records from sensor data, it is feasible to predict the social relationships between users to construct the underlying social network.

An illustrated scenario about the encounter records could be that Bob works with John from 8AM to 5PM from Monday to Friday. After working, Bob often eats dinner with his wife, Mary from 6PM to 8PM. Moreover, Tom and John usually play together from 9PM to 11PM at weekends. Suppose sensors can detect one another within a radius of 10 meters. And the collected encounter data from daily sensor records are uploaded through the base stations to a central database. The problem we tackled is how to accurately infer the friendships that conforms to the real social network using only the sensor encounter records.

Inferring social relationships from encounter data is challenging. For example, if Bob and John are not familiar with each other and only work in the same place, how to avoid inferring them to be acquainted with one another? On the contrary, if Bob and Tom are friends but they have never encountered in the sensor data, how to infer them to have the friendship?

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577365

The benefit of inferring friendships from sensor data is three-fold. First, it can help improve the routing efficiency in a sensor system [1][2][4]. Second, it would facilitate social-network-based services such as customized advertisement since the inferred relationships can further capture the real-world physical interactions between people. Finally, the inferred social relationships can be used for the friend recommendation in social networking services.

In this work, we propose a two-phase method to infer social relationships from sensor data. The general approach is the supervised learning. The first phase is to predict the existence of relationships between individuals based on the sensor encounter records. The extracted features contain the meeting frequency and the cumulative time of being together in certain place, which are divided into two temporal variables including daytime/nights and weekends. The objective of the second phase is not only to refine the link prediction result but also predict connections between individuals who have not been directly observed to encounter each other. We boost the social links using graph-based features, where the graph is derived from the predicted links in the first phase. We compute some structural proximity in the first predicted graph to be the graph-based features. In the evaluation part, we use the existing social relationships from Facebook as the ground truth and utilize precision and recall to estimate the performance.

Related Work. Eagle et al. [3] collect the communication information, the location, and the proximity data from mobile phones over a period to extract temporal and spatial patterns to help manually infer friendships. Scellato et al. [5] propose a supervised learning framework which exploits the properties of the places visited by users to predict link among friends-of-friends and place-friends. Wang et al. [6] show that human mobility measures have surprising predictive power in calling social network. Comparing to their work, we use only the sensor encounter data and can automatically discover the underlying social relationships. We all demonstrate that our proposed method has great scores on the precision and recall in the experiment section. To the best of our knowledge, our work is the first attempt to predict the social network based on the encounter records from sensor data.

2. METHODOLOGY
The overview of our method is shown in

Figure **1**. We elaborate these two phases below.

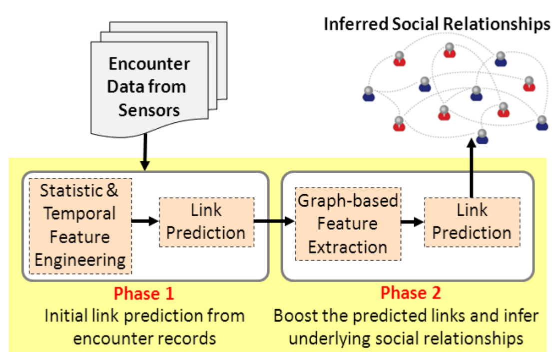

Figure 1: The overview of the proposed method.

2.1 First Phase

The objective of the first phase is to fully exploit the sensor encounter records for the initial link prediction between individuals. We consider that if an individual stays with another for a long time, it will have high potential to acquaint with each other. In addition, if they stay together frequently in the space, they have higher chance to acquaint with one another. Therefore, we consider the meeting frequency and the cumulative time of staying together to be important indicators of inferring friendships from sensor data. We extract statistic features including the cumulative duration of encounter time (*CDT*), whether two individuals had ever met (*ever_encounterd*), and the total encountered times (*encounter_freq*). Moreover, we consider that staying together up to a couple of hours at weekends is different from in working days. People usually stay with their family or friends at weekends and work with their bosses or colleagues in working days. As a result, we investigate the temporal version of previous statistic features. The temporal aspect contains the time definitions of daytime/nights (separated at 6AM and 6PM) and workday/weekend. Based on statistic and temporal features, we train different classifiers to predict existence of social links between individuals. The instances we predict in classifiers are the pairs of individuals who have carried with mobile phone. Note that, in the first phase, since it is possible that for those pairs of individuals who never meets in space (derived from sensor data), the features are zero, and thus we ignore predicting such kind of pairs of individuals.

2.2 Second Phase

The second phase aims at correctly predict those pairs of individuals who have never encountered but indeed have social relationships. We boost the inference of social links from first phase considering some graph-based features. The graph-based features, due to the social correlation describing the friends of an individual tend to acquaint with one another, can help predict links between individuals who have not been directly observed to stay together. We extract three graph-based features, including the neighborhood-based features, path features, and the hitting time feature. The neighborhood-based features are the smaller degree of two individuals of a link (*max_degree*), the larger degree of two individuals of a link (*min_degree*), the common neighbors between individuals (*overlap*), the extent of a neighbor of x is to be a neighbor of y and vice versa (*Jaccard*), and the product of the neighborhood sizes (*productNeighbors*). The path features include shortest path length between x and y (*ShortestPath*) and the sum over all possible paths between x and y (*Katzβ*). The *hitting_time* feature is considered a random walk which starts at x and expected times of encountering y.

Based on aforementioned features, we train a series of classifiers to refine the social relationships between individuals. Finally, we ensemble these learning models built from different classifiers by the average voting strategy to have resulting inferred social relationships.

3. EXPERIMENTAL RESULTS

We use the CRAWDAD data http://www.crawdad.org/index.php, which is a wireless network data consisting of sensor mote encounter records and the corresponding social network data in Facebook of a group of participants at University of St Andrews. We regard the existing Facebook social relationships as the ground truth of the link structure that we aim to predict. This data contains 25 invent devices. Participants were asked to carry the devices over a period of 79 days. We adopt the metrics of *precision* and *recall* to be the evaluation criteria. We consider a rule-based method, which set a threshold on cumulative duration of encounter time (*CDT*), to be the baseline method. If the *CDT* value of a pair of individuals is larger than 300 sec, we infer such pair to have a social relationship. For the prediction setting, we randomly divide the individuals into training set (60%) and testing set (40%). We build models on training set and tune the parameters by performing 3 fold cross-validations, and then evaluate on testing set.

For the prediction result of the first phase, since classifiers cannot predict the pairs which have never encountered each other, the models often give very small recall. The results of in the first phase are shown in Table 2. The Adaboost performs better precision than other classifiers and LibSVM performs better recall than others.

Table 1: Prediction results in the first phase.

	Precision	Recall
Random Forest	68.3%	32.3%
Random Tree	62.9%	27.6%
LibSVM (nu-SVC)	62.3%	41.7%
Adaboost+RandomForest	71.8%	30.7%

In the second phase, since we have different feature set from first phase. We compare the performance of classifiers by cross-validation process to choose new classifiers for the prediction. The results of the second phase are shown in Table 3. We can find that based on the graph-based features, our method is able to further find more underlying social links that the first phase cannot reveal. The recall is significantly improved as well. Moreover, the results show that the precision of the baseline method reaches only 57.9%, far beyond the ensemble approach, whose precision is 73.3%. The recall of the baseline method is also far beyond our methods (43.3% vs. 71.4%).

Table 2. Prediction results in the second phase.

	Precision	Recall
Baseline	57.9%	43.3%
Naïve Bayes	68.4%	70.1%
Neural Network	72.5%	64.9%
LibSVM (C-SVC)	72.0%	70.1%
LibSVM (nu-SVC)	73.0%	70.1%
Voting (Ensemble)	73.3%	71.4%

4. CONCLUSION

We propose a two-phase supervised learning method to infer the social relationships from sensor encounter data, where the statistic and temporal correlation between people ever met are modeled in the first phase and the relationship prediction of two never encountered individuals are inferred. Experimental results show that it is promising to predict the social relationships based on the sensor encounter records and some graph-based features.

5. REFERENCES

[1] G. Bigwood, D. Rehunathan. M. Bateman, T. Henderson, S. Bhatti. Exploiting Self-Reported Social Networks for Routing in Ubiquitous Computing Environments. *IEEE WIMOB* 2008.

[2] E.Bulut, B.K. Szymanski. Friendship Based Routing in Delay Tolerant Mobile Social Networks. *IEEE GlobalCom* 2010.

[3] N. Eagle, A. Penland, D.Lazer. Mobile Phone Data for Inferring Social Network Structure, *P*roceedings of the National Academy of Sciences*(PNAS)* Vol. 106(36), 2009.

[4] P. Hui, J. Crowcroft, E. Yoneki. Bubble Rap: Social-based Forwarding in Delay Tolerant Networks. *ACM MobiHoc* 2008.

[5] S. Scellato, A. Noulas, C. Mascolo. Exploiting place features in link prediction on location-based social networks. *ACM KDD* 2011.

[6] D. Wang, D. Pedreschi, C. Song, F. Giannotti. A. -L. Barabási: Human mobility, social ties, and link prediction. *ACM KDD* 2011.