Spatio-Temporal Prediction of Social Connections

Guolei Yang
Iowa State University
Ames, Iowa, USA 50011
yanggl@iastate.edu

Andreas Züfle
George Mason University
Fairfax, Virginia, USA 22030
azufle@gmu.edu

ABSTRACT

TBD

ACM Reference format:
Guolei Yang and Andreas Züfle. 2016. Spatio-Temporal Prediction of Social Connections. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 2 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In the past decade, with the rise of Location-Based Social Networks (LBSN), huge amount of geo-spatial data is collected on a daily basis. For example, the Foursquare[7] dataset contains more than 30 millions of self-reported check-ins from thousands of user around the world. As a result, it becomes possible to mine spatio-temporal data and study human mobility pattern at unprecedented large scale.

It is long known that a user's mobility pattern can be affected by his social connections [1, 8]. For example, a group of close friends tend to check-in to the same locations at the same time period. As such, it is possible to predict a user's future movement by mining the historical trajectory of his friends on LBSN. In the past decade, making predictions with spatio-temporal data has been intensively studied. Existing research mainly focus on the prediction of future movements (e.g., [1–5]). To our knowledge, however, predicting a user's social connections with spatio-temporal data has not been studied in literate.

Towards the goal of a more thorough understanding of human mobility patterns, we propose to investigate the predictive power of spatio-temporal data in predicting a user's social connections. In particular, given the trajectories of two LBSN users u_i and u_j, we aim to predict the probability that u_i and u_j are friends on the LBSN. Social connection prediction is a long standing research topic, mainly used as a tool for friend recommendation on social networks. Most existing methods exploit a user's profile and existing social connections to make friend recommendations, but not the user's trajectory. Our research is not competitive, but supplementary to existing friend recommendation methods.

A straightforward way to predict the social connection, or the lack thereof, between two users is to exam the *spatio-temporal overlap* of their trajectories, i.e., how many time did the two users visit the same location at the same time on their trajectories. The assumption is, if two users frequently visits the same location during the same time peroid, they might be friend with each other. Thus the overlapping on trajectories probably reflects the time and locations when and where they were meeting. Algorithms such as co-location mining [6] can be used to discover such spatio-temporal overlap among users.

Although the above assumption is reasonable, this naive solution suffers from two problems. First, it treats all locations equally in predicting social connections, which is not realistic. For example, if two users frequently meet at private locations like someone's house, or a small coffee shop, it is very likely that they know each other. However, if they both check-in to the same Walmart supermarket after work, it might be just an coincidence simply because there it is the only supermarket near their home. Second, this method ignores the time difference of check-ins behaviours. If two users both check-in to a restaurant at 6:00pm, it is not as significant as two users visit the same location at 10:00pm. This is because most customer of the restaurant may choose to dine there around 6:00pm, but if two users both decide to check-in there at 10:00pm, the chance that it is purely an coincidence is relatively lower.

We propose to employ a more comprehensive methodology to study the social connection prediction problem. Unlike the naive solution, we assume different locations and different time slots have different predictive power. We propose a social connection prediction model that is able to capture the inherent difference among locations and times. Specifically, each location and time slot is assigned a weight which measures the significance of the location and time slot in predicting social connections. We then use the Foursquare dataset to learn the weights in the proposed model and make predictions using users' social connections on Foursquare as ground truth. We show that the proposed model significantly outperforms the naive algorithm that exams only trajectory overlapping. We summarize our contributions as follows:

- We study the predictive power of spatio-temporal data in predicting social connections. Given the trajectories of two LBSN users u and v, we aim to predict the probability that u and v are friends on the LBSN.
- We assume different locations and time may have different predictive power, which is in accordance with common sense. We propose a model that is able to capture this difference among locations and times.
- We demonstrate effectiveness of the proposed model using the Foursquare dataset. The result shows the proposed method significantly outperforms the naive trajectory overlap based solution in prediction accuracy.

The rest of the paper...

Conference'17, Washington, DC, USA

DOI: 10.1145/nnnnnnn.nnnnnnn

2 RELATED WORK

Social connection prediction

Friend recommendation

3 OVERVIEW

3.1 Problem Statement

In this paper, we assume the precise check-ins, instead of geographic coordinates is given, as in the Foursquare dataset. Nevertheless, the proposed method is also applicable to coordinates...

We formally define the notion of *Check-in* and *User Trajectory* used in this paper.

We define a *Co-visitation* of two users as follows:

We formulate the social connection prediction problem as a classification problem, where the goal is to assign a given pair of users into one of the two classes: *Friends* and *Non-friends*.

3.2 Methodology

What model to use?? DNN? Linear?

Given the trajectories of two users u_i and u_j, we propose a three step method...

1. Feature extraction 2. Model learning 3. Prediction

Naive solution: Use average trajectory overlapping as threshold for prediction.

Hypothesis (For a full paper), Research questions 1. Active user vs Non-active users? 2. Percentage of friends can be explained? 3. Most significant types of locations? 4. Most significant time slots? 5. Geo-distance and predictive power?

4 SPATIO-TEMPORAL FEATURE EXTRACTION

5 MODEL LEARNING

6 EVALUATION

6.1 Dataset Description

We evaluate the proposed model on the widely-used Foursquare check-in dataset [7]. In our experiments, we mine the check-in data from two of the most popular cities, including New York City (NYC) and Tokyo. The dataset contains about 227,428 check-ins reported in NYC and 573,703 in Tokyo. The check-ins were collected for about 10 month. From each check-in, we extract the user ID, location ID, and a timestamp. Using the user ID or location ID, we retrieve the profile of that user or location on Foursquare. The user profile include the social connection between users ("follower - followee") and the location profile includes its category (*Food, Coffee, Nightlife, Fun, and Shopping*), coordinates, and user rating. The check-ins are grouped by user ID/location ID and sorted by their timestamps.

Select most active users...Ground truth...etc.

6.2 Experiment Design

For comparison purpose, we have implemented the following schemes:

- **Random** This scheme randomly assigns a user pair as friends or non-friends, each with a probability of 50%.
- **Naive** This scheme simple counts the number of co-visitations of two users. If the number is higher than a threshold, the two users are predicted to be friends, and otherwise non-friends. The threshold is set to be the average number of co-visitations of each pair of friends in the dataset we used.
- **Weighted** The proposed model with weighted co-visitation locations and time slots.

6.3 Results

7 CONCLUSION

REFERENCES

[1] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD conference on knowledge discovery and data mining*. ACM, 1082–1090.

[2] Huiji Gao, Jiliang Tang, and Huan Liu. 2012. Mobile location prediction in spatio-temporal context. In *Nokia mobile data challenge workshop*, Vol. 41. 44.

[3] Defu Lian, Vincent W Zheng, and Xing Xie. 2013. Collaborative filtering meets next check-in location prediction. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 231–232.

[4] Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. 2012. Mining user mobility features for next place prediction in location-based services. In *Data mining (ICDM), IEEE 12th international conference on*. IEEE, 1038–1043.

[5] Salvatore Scellato, Mirco Musolesi, Cecilia Mascolo, Vito Latora, and Andrew T Campbell. 2011. NextPlace: a spatio-temporal prediction framework for pervasive systems. In *International Conference on Pervasive Computing*. Springer, 152–169.

[6] Michael Weiler, Klaus Arthur Schmid, Nikos Mamoulis, and Matthias Renz. 2015. Geo-Social Co-location Mining. In *Second International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data*. ACM, 19–24.

[7] Dingqi Yang, Daqing Zhang, Longbiao Chen, and Bingqing Qu. 2015. NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. *Journal of Network and Computer Applications* 55 (2015), 170–180.

[8] Jihang Ye, Zhe Zhu, and Hong Cheng. 2013. What's your next move: User activity prediction in location-based social networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 171–179.