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 produced on a daily basis. It becomes possible to mine spatio-temporal data at unprecedented large scale. For example, the Foursquare[7] dataset contains more than 30 millions of self-reported check-ins from thousands of user around the world.

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, a user's future movement can be predicted with the historical trajectory of his friends on LBSN.

In this paper, we propose to investigate the inverses problem: How does a user's mobility pattern affects his social connections. Towards this goal, we are interested in studying the predictive power of spatio-temporal data in predicting a user's social connections. In particular, 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.

Making prediction with spatio-temporal data is intensively studied, but mostly for prediction of future movements (e.g., [1–5]). To our knowledge, however, making prediction on the user's social connects remains an open issue. We aim to complement existing research in this field by addressing the problem. Predicting social connections with spatio-temporal data also provides a supplement method for *friend recommendation* on LBSN. It can be used to discover users that potentially share the same interests for locations such as restaurants, shops, and museums.

A straightforward way to solve the problem is to compute the spatio-temporal overlap of two trajectories. The assumption is, if two users frequently visits the same location during the same time peroid on their trajectories, they might be friend with each other and were meeting at those locations. Algorithms such as co-location mining [6] can be used to detect such spatio-temporal overlap among users. However, 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 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 is not other nearby supermarket. 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 visit there at 10:00pm, it is more likely that they plan to meet with each other.

We propose a spatio-temporal social connection prediction framework that aims at address the above problems. Unlike the naive solution, we assume different locations and different time slots have different predictive power, which is modelled as a latent variable. As such, we first formulate the social connection prediction problem as a latent variable estimation problem, and then design a matrix factorization-based algorithm to estimate the variables and make predictions. We summarize our contributions as follows:

- We introduce the spatio-temporal social connection prediction problem. 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 formulate the social connection prediction problem as a latent variable estimation problem, and then design a matrix factorization-based algorithm address the problem.
- The proposed framework is implemented and evaluated on 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...

2 RELATED WORK

3 OVERVIEW

3.1 Problem Statement

3.2 Proposed Framework

4 PROPOSED METHOD

5 EXPERIMENTS ON REAL-WORLD DATA

5.1 Dataset Description

We evaluate the proposed techniques 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

Conference'17, Washington, DC, USA

DOI: 10.1145/nnnnnnn.nnnnnnn

(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.

5.2 Experiment Design and Results

6 CONCLUSION

In this paper, we study the predictive power of historical check-ins in predicting a user's future check-in time to a given location. Using historical location data to model human mobility pattern is a long-standing research topic. To our best knowledge, although check-in location prediction has been intensively studied, making prediction in the time-dimension remains an open issue. The scarcity of check-in data makes it challenging to accurately model the check-in time of a location. To address this problem, we propose a comprehensive framework that can learn model using both historical visitors and potential visitors to the location. Here, potential visitor is defined as a user who has not reported a check-in to the location before, but is likely to do so in the future. Standard regression models cannot use potential visitors. We circumvent this problem by first formulating our problem as a censored regression problem, and then propose a modified Cox-regression method which can take advantage of potential visitors. In our experiments, the proposed method demonstrates superior performance comparing with standard regression techniques with uses only historical visitor data. Moreover, our method is capable of predicting the time of several forms of complex check-in events.

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.