Understand Movement Pattern: from Individual Trajectory to Aggregated Flow

Hongjian Wang hxw186@ist.psu.edu

1 Background

Millions of trajectory data are collected.

- Geo-tagged tweets.
- Taxi on-board GPS.
- LEHD social flow

2 Problem

How to interpret the movement that we have observed? This question can be further specified as two questions.

2.1 Given individual trajectories, answer the question why it moves like this.

We define simple pattern such as periodic pattern [1].

When other nearby individual's trajectories are available, we further study following pattern [2], attraction/avoidance [3], and many more.

Challenge 1: Most simple patterns are defined and mined on animal data. For human beings, the mobility pattern is more complicated to define and mine. How to correlate one's mobility with others is hard to answer.

2.2 Given an observed social flow, explore what are the factors that strongly influences this flow.

Examples are urban traffic flow prediction [4]. However, such work only interprets the flow with historical observation. We still cannot answer what factors are there that significantly influence the traffic.

To better understand the agreggated flow, we need extra information. Luckily, we have them. The urban data we are able to collect:

- POI
- Tweets (geo-tagged)
- Air pollution
- City noies

- crime

Some existing work has already demonstrated the correlation between traffic and other urban factors, such as air pollution [5] and city noise [6].

Challenge 2: how to combine different data types and model them simultaneously is difficult.

2.3 Link the individual trajecotry to aggregated flow pattern

The aforementioned two problmes should not be orthogonal. The individual level results or minined pattern, are suppoed to be able to naturally scale to aggragated pattern.

Take the geographical relationship strength for example. At individual level, we can propose pair-wise measure, which works pretty well. At aggregated level, we can use clustering to detect community. However, when the second method is applied on the first problem, it cannot outperform the algorithm dedicated for the first. Similarly, it is not efficient to apply the dedicated pair-wise measure for the first problem to the second one.

Challenge 3: propose a generalized framework to fill the gap between the first two problmes.

3 Methodology

Table 1: The category of research problems.

	Individual level	Aggregated level
Single domain correlation		
Cross domain correlations		

3.1 Measure Geographical Relationship Strength

The proposed measure consists of three different factors, which effectively outperforms the meeting frequency-based baseline method. Such a effective measure can benefit the recommendation and a series of other real applications.

- Location popularity
- User personal preference
- The temporal distribution of pair-wise co-location events.

3.2 Explore Aggregated Flow

Explore the correlation between community area flow transition and crime. To do so, we build a regression model to predict the crime count with traffic flow and many other features. The null hypothesis is that if the traffic flow is not correlated with the crime, adding them to the prediction model has no effect on the results. Through experiments on real data, we verify the correlation between taxi flow and crime propagation.

References

[1] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, "Mining periodic behaviors for moving objects," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1099–1108, ACM, 2010.

[2] Z. Li, B. Ding, J. Han, and R. Kays, "Swarm: Mining relaxed temporal moving object clusters," *Proceedings of the VLDB Endowment*, vol. 3, no. 1-2, pp. 723–734, 2010.

[3] Z. Li, B. Ding, F. Wu, T. K. H. Lei, R. Kays, and M. C. Crofoot, "Attraction and avoidance detection from movements," *Proceedings of the VLDB Endowment*, vol. 7, no. 3, pp. 157–168, 2013.

[4] P. S. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi gps traces," in *Pervasive Computing*, pp. 57–72, Springer, 2012.

[5] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1436–1444, ACM, 2013.

[6] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang, "Diagnosing new york city's noises with ubiquitous data," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 715–725, ACM, 2014.