Semantic Understanding of Spatial Trajectories

Zhenhui Li

College of Information Sciences and Technology
Pennsylvania State University
jessieli@ist.psu.edu

1 Trajectory Mining without Contexts

The advances in location-acquisition technologies and the prevalence of location-based services have generated massive spatial trajectory data, which represent the mobility of a diversity of moving objects, such as people, vehicles, and animals. Such trajectories offer us unprecedented information to understand moving objects and locations that could benefit a broad range of applications. These important applications in turn calls for novel computing technologies for discovering knowledge from trajectory data.

Under the circumstances, trajectory data mining has become an increasingly important research theme in the past decade [10]. Extensive research has been done in the field of trajectory data mining with many interesting patterns have been proposed and studied. However, most existing studies focus on **only trajectory data** and did not consider **rich spatial-temporal contexts** that are associated with trajectories. As a consequence, trajectory patterns detected from existing methods could be trivial. Let's examine an example below.

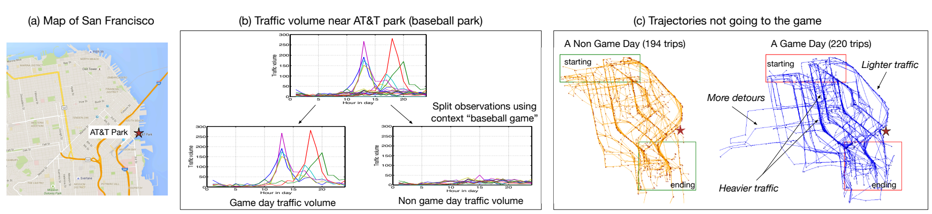

Fig. 1. Taxi trajectories in San Francisco. (a) San Francisco map and the location of AT&T park (a baseball stadium). (b) Recurrence pattern detected if we consider the context of baseball game near AT&T park. (c) Trajectories not going to the game will try to avoid the area near AT&T park. Such "anomalies" (e.g., heavier traffic and detours) are expected during the game time.

Example 1. Trajectory outliers or anomalies are usually defined as trajectory segments that are different from other trajectories in terms of some similarity metric. It can also be observations (represented by a collection of trajectories) that do not conform to an expected pattern (e.g., traffic jam). However, such anomalies might actually be expected if we consider the contexts. For example, as shown in Figure 1, it might be expected that, whenever there is a baseball game in San Francisco, the traffic is heavier than usual and a car not going to the game may choose a longer detour path. Such "anomalies" are actually normal under the condition of contexts (e.g., baseball game in this example).

2 Vision: Semantic Trajectories Understanding

The increasing availability of contextual information (e.g., venue information, local events, weather, and landscape) can potentially lead to a revolution in trajectory data mining. Mining trajectory data should no longer focus on trajectory only, but should also utilize the rich contexts from other data sources to provide a semantic understanding of trajectories.

Semantic trajectories with contexts could benefit a number of applications:

(1) User profiling. Contexts associated with trajectories will provide a more accurate profiling of users' interest, socioeconomic status, and health conditions. For example, a person frequently visits kindergarten and kids-friendly restaurants, he/she may have a young kid and we could recommend kids-related activities on weekend; if a person frequently visits fast food and convenience stores, and rarely visits recreational places (e.g., parks, fitness facilities), the person may live a poor lifestyle and might be at a greater risk of chronic health conditions.

(2) Intelligent transportation. A significant change in intelligent transportation system is to use more data collected from a variety of sources. Understanding traffic patterns with contexts, we could better predict the traffic and suggest the best route for drivers under different conditions of time, weather, and events.

(3) Ecology. Organism-environment interaction is a fundamental question in ecology that tells us how animals respond to the dynamic changes of environment and helps us predict how environmental change will impact animals' behaviors.

3 Research Challenges

Mining trajectory data with contexts is not simply using all the nearby contextual data or simply extending current data mining techniques with extra context information. To enable the power of contexts in trajectory mining, we need to understand how trajectories are associated with or impacted by the surrounding contexts. There are two key challenges we will face:

Implicit and complicated correlations. Since there are many surrounding contexts near a location, it is ambiguous which context correlates with the trajectory. For example, a person observed at Madison Square Garden (MSG) could be attending a concert in MSG, or could be transitioning at Penn Station which sits below MSG, or could be visiting a restaurant nearby. Moreover, the observed trajectories are impacted by many factors simultaneously, such as daily/weekly regularity, local events, weather, car accidents, and traffic jams. The impacts could be also at different scales from small farmer's market, to big football game, to extreme weather.

Sparse and noisy data. Observations on trajectories and contexts are often quite sparse in real applications. For example, we may only have sporadic observations on individual data if the data collection mechanism requires users to voluntarily contribute data; some trajectory datasets, such as taxi data, only reflect a biased and incomplete version of the overall mobility density. It is also not realistic to obtain all the context information that impact trajectories. In addition, the data we obtained could also be noisy and imprecise. GPS positioning often has errors that vary from a few meters to hundred of meters, depending on the sensing equipments and atmospheric effects. For an event obtained from the news article with a description as "football game at 12 p.m. on Saturday", the game was probably from 12 p.m. to 2 p.m. In order to capture more local events, we may even need to extract the contexts from the noisy raw data (e.g., extracting events geo-tagged tweets).

4 Preliminary Studies and Future Directions

Recent studies have realized the importance of utilizing external context data to enrich the semantics of mobility patterns. However, most of these studies assume that the contexts are already associated with mobility records (e.g., check-in data). To bridge the gap between raw trajectory data and contexts, we need to associate the trajectories with the corresponding contexts. Various methods have been proposed to annotate the mobility records with landmarks [1], landscapes [4], land-use categories [7][8], geo-tagged tweets [6], and POI [8][5]. These methods can be generally classified into two categories. The first category [1][4][7][6][2][3] is to consider each mobility record separately and to annotate each record independent of other records. For example, the most common approach is to attach the closest context to a mobility record. The second category [9][8][5] considers the dependency among records. For example, Yan et al. [9][8] propose a hidden Markov model to consider the transition dependency in individual movement. Wu and Li [5] propose to use a Markov Random Field to consider the consistency in individual preference.

However, many challenges have not been well addressed by these preliminary studies. Here we discuss a few potential future research topics.

First, the context data are often messy and ambiguous. For example, there could be multiple duplicate POI entries corresponding to the same POI entity because POI entries are often generated by the crowd. Also, even though events could have a significant impact on trajectories, there is no such a good data source documenting all the events in a city. All the existing studies have been assuming the context data are clean and contain no ambiguity.

Second, depending on the data collection mechanism, the spatial trajectories could be in different forms. Constant GPS tracking may give a complete trajectory, but will require pre-processing to extract the meaningful location records. Data collected by social media or smartphone applications are often very sparse. Such sporadic location data are only collected when users use the applications. In addition, due to privacy concern, sometimes we may only have the crowd information, such as taxi pick-ups and drop-offs without knowing the passenger identities. Different data properties will require different methods for semantic understanding.

Third, it remains challenging how to evaluate the semantic patterns. How do we know whether the annotated venues are the true destination venues of a user? How do we know that an event is the cause of a person visiting a location or a person happens to locate at that venue during the event time? It will be valuable to generate a benchmark dataset that people can evaluate their methods on semantic trajectory mining.

Acknowledgements

This work was supported in part by NSF awards #1618448, #1652525, #1639150, and #1544455. The views and conclusions contained in this paper are those of the author and should not be interpreted as representing any funding agencies.

References

1. L. O. Alvares, V. Bogorny, B. Kuijpers, J. A. F. de Macedo, B. Moelans, and A. Vaisman. A model for enriching trajectories with semantic geographical information. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems (GIS'07)*, page 22. ACM, 2007.
2. R. Fileto, C. May, C. Renso, N. Pelekis, D. Klein, and Y. Theodoridis. The baquara 2 knowledge-based framework for semantic enrichment and analysis of movement data. *Data & Knowledge Engineering*, 98:104–122, 2015.
3. L. Ruback, M. A. Casanova, A. Raffaetà, C. Renso, and V. Vidal. Enriching mobility data with linked open data. In *Proceedings of the 20th International Database Engineering & Applications Symposium*, pages 173–182. ACM, 2016.
4. S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot. A conceptual view on trajectories. *IEEE Trans. on Knowledge and Data Engineering (TKDE)*, 65(1):126–146, 2008.
5. F. Wu and Z. Li. Where did you go: Personalized annotation of mobility records. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM'16)*, pages 589–598. ACM, 2016.
6. F. Wu, Z. Li, W.-C. Lee, H. Wang, and Z. Huang. Semantic annotation of mobility data using social media. In *Proceedings of the 24th international conference on World Wide Web (WWW'15)*, 2015.
7. Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer. Semitri: a framework for semantic annotation of heterogeneous trajectories. In *Proc. 14th Int. Conf. Extending Database Technology (EDBT'11)*, pages 259–270. ACM, 2011.
8. Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer. Semantic trajectories: Mobility data computation and annotation. *ACM Trans. TIST*, 4(3):49, 2013.
9. Z. Yan, N. Giatrakos, V. Katsikaros, N. Pelekis, and Y. Theodoridis. Setrastream: semantic-aware trajectory construction over streaming movement data. In *Proc. 2011 Int. Symp. Spatial and Temporal Databases (SSTD'11)*. Springer, 2011.
10. Y. Zheng. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):29, 2015.