A Unified Framework to Predict Movement

Olga Gkountouna[1], Dieter Pfoser[1], Carola Wenk[2], and Andreas Züfle[1]

[1] George Mason University
{ogkounto, dpfoser, azufle}@gmu.edu
[2] Tulane University
cwenk@tulane.edu

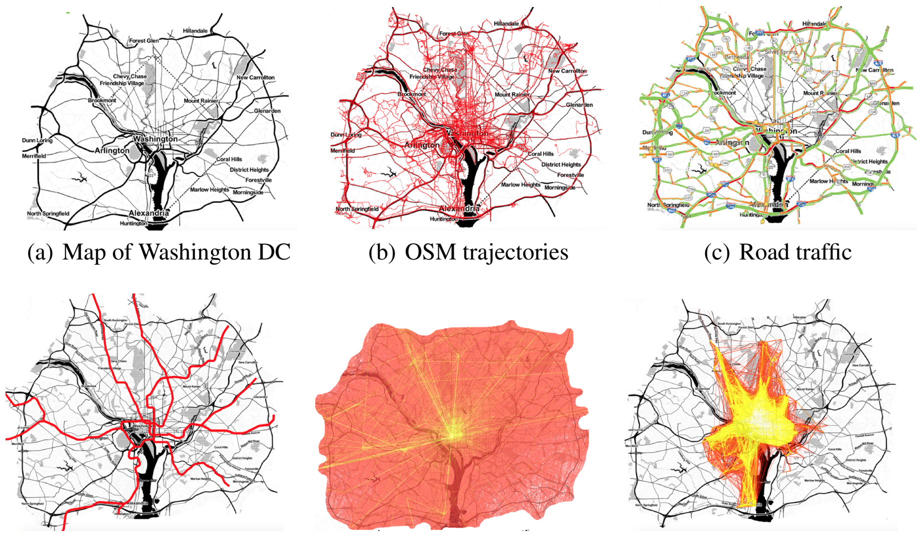
Abstract. In the current data-centered era, there are many highly diverse data sources that provide information about movement on networks, such as GPS trajectories, traffic flow measurements, farecard data, pedestrian cameras, bike-share data and even geo-social movement trajectories. The challenge identified in this vision paper is to create a unified framework for aggregating and analyzing such diverse and uncertain movement data on networks. This requires probabilistic models to capture flow/volume and movement probabilities on a network over time. Novel algorithms are required to train these models from datasets with varying levels of uncertainty. By combining information from different networks, immediate applications of such a unifying movement model include optimal site planning, map construction, traffic management, and emergency management.

1 Introduction

Numerous data sources exist that *capture movement* on (movement) networks relating to the same geographic area. The most popular movement data sources are GPS trajectories, which are readily available from a myriad of smartphone apps. In the case of social media data, movement is often captured implicitly, as trajectories are derived from geocoded tweets, Flickr images, or Foursquare check-ins. Such data is considerably less frequently sampled, i.e., the time between samples may range from minutes to hours. Additional movement data sources include farecard, bikeshare, and taxi trip data. Figure 1 shows various movement data sources for the greater Washington DC metropolitan area including GPS traces uploaded by users to Openstreetmap[3], a colored road network to symbolize traffic conditions derived from measuring stations, public transit fare card data, traces of Twitter users, and bikeshare data. Existing solutions to predict traffic in networks [1, 6–9, 11] focus on a single network, and assume that positional data is fairly accurate and frequent, such as GPS data.

Our vision is to *use this multitude of available data sources to generalize traffic prediction towards movement prediction for known and unknown movement networks*. We anticipate that meaningful movement information from unconventional data sources can be extracted by utilizing knowledge from a variety of data sources with varying degrees of uncertainty. The goal is to use *all* of these highly diverse and uncertain movement data sources *together* to enable movement prediction. This will allow us to estimate and to predict movement even in locations where no authoritative data is available, e.g., in the extreme case, to estimate traffic on road networks solely based on social media data.

[3] http://www.openstreetmap.org.



(a) Map of Washington DC (b) OSM trajectories (c) Road traffic

(d) Public transport (e) Social media trajectories (f) Capital bikeshare

Fig. 1. Heterogenous sources of movement network data.

2 Challenges

To develop a unified model for probabilistic movement prediction among multiple transportation networks, the following three major challenges need to be addressed. In the following, we describe these challenges and their feasibility in more detail.

Making Sense of Extreme Uncertainty: Techniques have been developed [1, 7–9] that yield accurate models to probabilistically predict trajectories in transportation networks. These models work well in the case when trajectories are given at high-frequency, thus narrowing down the space of possible locations for a user given their previous location. However, in cases of extreme uncertainty where observations are hours apart from each other, predicting possible locations of individuals becomes futile. The space of possible locations that could be reached in an hour becomes too large to make reliable predictions. Yet, we can use such uncertain observation data to estimate traffic density/volume in the network, rather than estimating the location of individuals. For instance, if we have a large number of users that have a very small probability of being in a specific area, then Poisson's Law of Small Numbers states that the number of individuals in this area is Poisson distributed, and thus, can be predicted accurately.

Correlation of Different Networks: Another research challenge is the problem of matching and correlating data from different movement networks, such as road-traffic loops data, public-transport data and pedestrian GPS data. These datasets stem from different modes of transportation. They may vary in size and may include different attributes (such as vehicular speed, volume, traffic density, etc). Furthermore, they may be supplied to us in an aggregated form or as raw data, and their measurements may be taken at different sets of timestamps. Finally, they may consist exclusively of observed data points, or may include interpolated values as well.

The correlation of separate movement networks used to be impossible due to the lack of open data sources. Using *all* available data sets discussed in Section 1, we are now able to align different networks to identify how individuals move and transition between them. Consider here for example a daily commute that includes driving by car to the metro station, taking the metro downtown, walking to an office building and then inside, directly to the coffee maker. By joining available movement data, we can identify transitions in between road networks, public transportation networks, pedestrian networks, and building floor plans. Considering this correlation between networks of varying degrees of uncertainty, we can further enhance movement flow predictions. For example, an inbound full metro train will lead to increased road traffic around metro parking lots.

Correlation of External Variables: The movement volume estimation can be improved using other parameters such as vehicular speed. If the number of cars observed on an edge is relatively low, but we notice that the traffic moves at a relatively slow speed, then this might be an indicator that the actual traffic volume is higher than the estimated movement flow. Other observable variables that can be used to infer movement flow are weather conditions, day-time, and event information (e.g., a football game). By taking into account such information, we can condition the model to a set of current variables, resulting in a model that is more tailored towards current conditions by exploiting learned implications of these variables. For this purpose, we need to learn the impact of such observable (and publicly available) parameters and correlate them to movement flow, to improve prediction. The main challenge is to assess the significance of each data source, which can be learned empirically from authoritative ground-truth traffic volume data measured by road-side detectors.

3 Applications of Multi-Network Movement Flow Prediction

Traffic Management: Given the capability to connect different networks for volume prediction, we can improve efficient decision making in traffic management. By simulating road closures, or closed public transportation lines, critical areas can be identified in other networks, which would be impacted by such an event. Furthermore, for a road closure that has just occurred, our vision will allow online simulation of the consequences, thus quickly predicting areas that will become congested soon, and so providing traffic management systems with decision criteria to reroute traffic.

Emergency Management: In the case of a large scale disaster or emergency such as an earthquake, a forest fire, or a terrorist attack, an optimal deployment of fire fighters and police forces is paramount to saving lives. In order to make the right decisions, emergency management personnel needs accurate information quickly and in the right form. They need predictions about what may happen and information about what has happened and where. In the report to the US Congress on Hurricane Katrina, Secretary of Homeland Security Michael Chertoff emphasized "the importance of having accurate, timely and reliable information about true conditions on the ground" and pointed out that the response efforts during Katrina "were significantly hampered by a lack of information from the ground" [2]. In addition to existing work [3, 4, 10], which uses only Twitter as an information source, our envisioned framework enables the fusion of streams of location and movement data such as tweets, OSM trajectories, traffic loop data and other sources, to model a traffic and population flow under the new conditions, and thus making this flood of geo-social data actionable for immediate decision making.

Managing the flow of pedestrian movement in the case of an emergency is vital to saving human lives. On 7/24/10, a crowd disaster at the Love Parade electronic dance music festival in Duisburg, Germany caused the death of 21 people from suffocation [5]. To prevent such disasters, we envision a powerful movement prediction model for both prevention and counter-measures. For prevention, we can simulate a disaster, showing potential bottlenecks. To enable counter-measures, we can use real-time social media data to quickly predict large crowd movements. This way, emergency responders can be alerted within minutes before a deadly stampede occurs.

4 Conclusions

Most sources of movement and traffic information are highly uncertain, due to many reasons. Existing work focuses on analyzing these sources individually. In this vision paper we propose to unify these sources into a single movement prediction framework, that is able to learn movement and traffic patterns from *all* data sources and movement networks, including very sparse and uncertain data, simultaneously. Such a framework will not only improve existing applications, but will also inspire entirely new research directions in applications where precise movement data is not available.

Acknowledgements

This research has been supported by National Science Foundation AitF grants CCF-1637576 and CCF-1637541.

References

1. A. Abadi, T. Rajabioun, and P. A. Ioannou. Traffic flow prediction for road transportation networks with limited traffic data. *IEEE ITS*, 16(2):653–662, 2015.
2. M. Chertoff. Statement by homeland security secretary Michael Chertoff before the United States house select committee on hurricane Katrina. *Washington, DC: Department of Homeland Security*, 2005.
3. B. De Longueville, R. S. Smith, and G. Luraschi. OMG, from here, I can see the flames!: A use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *LBSN*, pages 73–80. ACM, 2009.
4. M. F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.
5. D. Helbing and P. Mukerji. Crowd disasters as systemic failures: analysis of the love parade disaster. *EPJ Data Science*, 1(1):1–40, 2012.
6. H. Jeung, M. L. Yiu, X. Zhou, and C. S. Jensen. Path prediction and predictive range querying in road network databases. *The VLDB Journal*, 19(4):585–602, Aug. 2010.
7. H.-P. Kriegel, M. Renz, M. Schubert, and A. Züfle. Statistical density prediction in traffic networks. In *SDM*, volume 8, pages 200–211. SIAM, 2008.
8. J. Niedermayer, A. Züfle, T. Emrich, M. Renz, N. Mamoulis, L. Chen, and H.-P. Kriegel. Probabilistic nearest neighbor queries on uncertain moving object trajectories. *Proceedings of the VLDB Endowment*, 7(3):205–216, 2013.
9. D. Pfoser and C. S. Jensen. Capturing the uncertainty of moving-object representations. In *SSD'99*, pages 111–131. 1999.
10. S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *SIGCHI*, pages 1079–1088. ACM, 2010.
11. J. Zhang, Y. Zheng, and D. Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI 2017*, pages 1655–1661, 2017.