MovementFinder: Visual Analytics of Origin-Destination Patterns from Geo-tagged Social Media

Siming Chen
Peking University

Cong Guo
Peking University

Xiaoru Yuan*
Peking University

Jiawan Zhang†
Tianjin University

Xiaolong(Luke) Zhang‡
Pennsylvania State University

Abstract

Geo-tagged social media data can be viewed as sampling of people's trajectories in daily life. It consists of people's movements and embeds the semantics of movements. However, it is challenging to reveal patterns from the sparse and irregular sampling data. We proposed an interactive multi-filter visualization approach to analyze the spatial-temporal movement pattern in people's daily life. People's trajectories are visualized on the map with multiple functional layers. With our visual analytics tools, users are able to drill down to details, with the awareness of the origin-destination flow patterns of spatial, temporal, and semantic meaning.

1 Introduction

Popular social media often embed rich data about people's behaviors. Microblogging services (e.g., Twitter, Sina Weibo), for example, let people include location and time in their microblogs. Such spatial and temporal information, as well as microblog contents, allows the construction of users' trajectories and the exploration of semantics of movements (e.g., why a trip was made). Rich social media data offer new opportunities for in-depth analysis of people's activities in innovative ways. One of the areas that can benefit from such data richness is the origin (O) and destination (D) analysis.

OD analysis is an important method to understand the transportation and movement patterns of a certain region. Travel behaviors and demands revealed by studying the patterns could potentially improve decision making in areas such as transportation planning, traffic management, and resource allocation. However, extracting the pattern from geo-tagged social media and constructing the semantic meaning are challenging. We surveyed literature in transportation and traffic management research [6] and identified the tasks usually concerned with in OD analysis:

- **T1** where people start and end their trips;
- **T2** how long these trips are;
- **T3** within a given region, what locations are popular origins and destinations;
- **T4** whether OD flows within a region exhibit different patterns at different time of a day;
- **T5** if more trip information is available, what the purposes of these trips are.

Our research extracted OD information from geo-tagged social media and addressed the above analysis tasks. Our visual analytics system, MovementFinder, supports the exploration of OD information from different perspectives independently and jointly. Users can control the spatial scale of the analysis, and look into specific regions of interest. Users can also choose different temporal scopes in analysis. By combining spatial and temporal features, users can interactively build filters to analyze OD information.

*e-mail: {csm,cong.guo,xiaoru.yuan}@pku.edu.cn
†e-mail: jwzhang@tju.edu.cn
‡e-mail: lzhang@ist.psu.edu

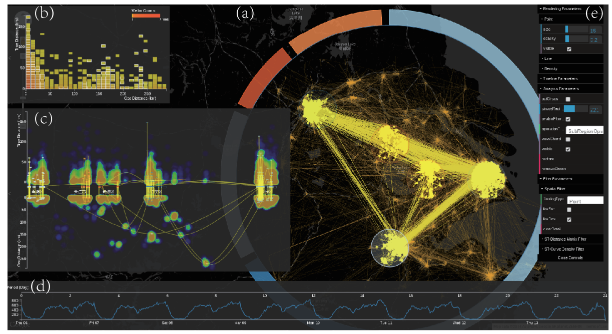

Figure 1: System Overview. (a) Spatial view, (b) ST Matrix View, (c) OD Detail View, (d) Timeline View, (e) Control Panel.

In general, designs for OD data visualization can be classified into three categories: flow map [4], OD matrix [1], and OD map [5]. However, the limitation of these methods, which includes cluttering in flow map, lack of spatial information in OD matrix and mismatching in OD map's grid and real world regions, encourages the development of new visual analytics tools for OD analysis. From the view of geo-tagged social media, Gabrielli et al. [3] used Twitter geo-tags to detect what types of locations had been involved in people's movements. Andrienko et al. [2] defined this type of data as "Episodic Movement Data", which needs aggregation and filtering tools for further analysis. Currently, sophisticated visual analytics tools for understanding spatial, temporal, and semantic features of OD flows constructed from social media data are rare.

2 Data

The data we used in this research was extracted from Sina Weibo, the most popular microblogging service in China. Currently, our crawler collects 1 million new weibos per day. Data cleaning, keyword extraction and spatial-temporal aggregation for indexing are finished in preprocessing. We extracted the OD pairs from the sequence of each user's trajectory based on geo-tagged weibos in a chronological order. Although the location data based on the geo-tags from an individual user might be too sparse and irregular to sufficiently reflect the trajectory of the individual, aggregating such location information from massive number of users in a region provide opportunity to understand the collective movement patterns.

3 User Interface

Our system is web-based, and all functional layers fit into a single browser panel. It has five major components, all of which act as both display panel and interactive filtering panel (Figure 1).

View to Show/ Filter Location Clusters and OD Flows. The map view (Figure 1a) shows the region of interest, the location clusters within the region, and flow patterns among the clusters. The system would suggest the locations with largest amount of weibos as candidates of location clusters, and then users can create more filters (circle or polygon) to add, resize or delete the clusters for customized analysis. Color is identical to each cluster. The band outside the selected region is corresponding to each cluster, the length

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

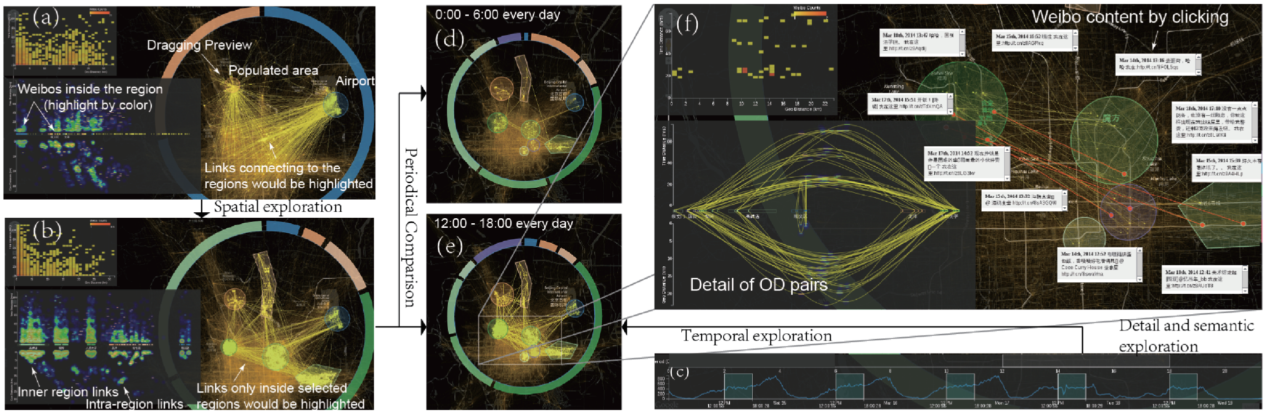

Figure 2: Exploring OD information in Beijing: a) a preview in general exploration, b) selecting a few location clusters, c) timeline used to do periodic filtering. d), e) OD flow patterns in two time periods, f) OD details of a specific region within a time period.

of which indicates the number of selected weibos in it. Keywords within selected location clusters could pop up on demand.

View to Show/ Filter OD Pairs based on Distance and Travel Time. The ST matrix view (Figure 1b) displays the spatial-temporal distribution of all OD locations within the region in a 2D histogram. The x axis of the diagram is for OD distance, and the y axis is for OD travel time. Based on the maximum and minimum distance and time values found in OD pairs, the system automatically generates distance-time bins, and constructs a 2D histogram. Rect brushing is supported to filter the OD pairs.

View to Explore Individual OD Pairs. In the OD detail view (Figure 1c), each OD pair can be defined as a four-point object: two spatial points for origin and destination, one point for distance, and one point for travel time. The location axis (x-axis) is the 1D spatial projection of each locations. Each distance and time point is connected to the related location points through a symmetric arc, and the height of an arc is determined by distance or time. It also provides heatmap to show the distribution of OD distances and travel times. A user can select a region of interest and use highlighting or dynamic zooming to further explore the data.

Timeline View. The timeline view (Figure 1d) allows users to examine OD information based on time, and to filter data based on certain temporal constraints, such as a particular period of a day.

Control Panel. The filtering and control panel (Figure 1e) is used to control the workspace, ST matrix view, OD detail view, and timeline view. All these views are coordinated, so explorative analysis is supported by combining multiple filters.

4 Case Study: Exploration of ODs in Beijing

The data set used in this case study contains weibos with geo-tags in Beijing, posted between March 6 and March 13, 2014. There are 144,655 weibos and 54,686 OD pairs after preprocessing. To get a general sense of OD flow patterns, we can use the preview tool and move among different regions. In Figure 2a, we can find a place at the upper-right corner inside the preview circle with high volumes of weibo activities. By examining the location and the contents of weibos, we know that it is close to an airport (**T1**). Following the OD flows, we can find other areas that have heavy links with it.

In spatial exploration, we can get the location clusters with the largest amount of weibos recommended by MovementFinder. (Figure 2b), symbolized with circles. We also add two polygon filters for additional areas of interest. We can see the general OD flow patterns among these six clusters on the map. One cluster in the center seems to be getting lots of traffic from other clusters (**T3**). The ST matrix view shows high frequencies on the left side, so there are many relatively short OD pairs (**T2**). However, their travel time vary significantly, as shown by high frequency bins along the y axis. The OD detail view also provides rich information, including six dense areas and large amount of inner-region travels.

With the understanding of the above general flow patterns, we can explore the temporal features of OD pairs. Using the timeline, we can specify a period of a day for periodic filtering (Figure 2c). Figure 2d,e shows the flow patterns in two different time periods: 0:00 to 6:00, and 12:00 to 18:00 (**T4**). Difference in flow patterns between these periods can be clearly seen. Interested in the flow patterns between 12:00 to 18:00, we choose a particular area and see the OD pairs within it (Figure 2f). When clicking the location in the x-axis of OD Detail View, the map navigates to the corresponding place and shows the keywords. By clicking the OD pair on the map, we read the weibos of OD pairs of interest and know some people went to shopping while others usually went back to the university for dinner during that period (**T5**).

5 Conclusion

In this report, we presented a visual analytics system to support the analysis of origin and destination data. Our research emphasizes the use of geo-tags embedded in microblogs in the construction of origin and destination pairs and the development of interactive filtering tools to support in-depth analysis of movement patterns of people.of the movement data.

Acknowledgements

The authors wish to thank the anomynous reviewers for their valuable comments. This work is supported by NSFC No.61170204

References

[1] N. Adrienko and G. Adrienko. Spatial generalization and aggregation of massive movement data. *IEEE TVCG*, 17(2):205–219, 2011.

[2] N. Andrienko, G. Andrienko, H. Stange, T. Liebig, and D. Hecker. Visual analytics for understanding spatial situations from episodic movement data. *KI - Knstliche Intelligenz*, 26(3):241–251, 2012.

[3] L. Gabrielli, S. Rinzivillo, F. Ronzano, and D. Villatoro. From tweets to semantic trajectories: mining anomalous urban mobility patterns. In *Citizen in Sensor Networks*, pages 26–35. Springer, 2014.

[4] D. Phan, L. Xiao, R. Yeh, and P. Hanrahan. Flow map layout. In *Proc. IEEE InfoVis*, pages 219 – 224, 2005.

[5] J. Wood, A. Slingsby, and J. Dykes. Visualizing the dynamics of london's bicycle hire scheme. *Cartographica*, 46(4):239–261, 2011.

[6] L. Zhang, F. Southworth, C. Xiong, and A. Sonnenberg. Methodological options and data sources for the development of long-distance passenger travel demand models: a comprehensive review. *Transport Reviews*, 32(4):399–433, 2012.