Understanding Spatiotemporal Mobility Patterns related to Transport Hubs from Floating Car Data

Linfang Ding[1, 2], Mathias Jahnke[1], Shirui Wang[1], Katre Karja[1]

[1] Chair of Cartography, Technical University of Munich
[2] Group of Applied Geoinformatics, University of Augsburg

Abstract. Transport hubs such as airports and railway stations are places where plenty of passengers are exchanged between vehicles or between transport modes. Analyzing their mobility patterns, for instance traffic flows in/out of the transport hubs, helps better understanding passengers travel behaviors and improving the transportation planning. In this paper, we aim to visually present the spatiotemporal mobility patterns related to transport hubs using floating car data. Following a visual analysis workflow, which consists of computational algorithms, data mining approaches, and visualization techniques, we preprocess the raw data related to the transport hubs, derive relevant pick-up and drop-off events, cluster and aggregate those events, and visually analyze their spatiotemporal distributions using mainly proportional symbol mapping and pie-chart mapping techniques. We use one-week Floating Car Data (FCD) in Shanghai as our test dataset and select Hongqiao international airport as the test transport hub. The preliminary experiment results show that there are obvious temporal mobility patterns as well as significant hotspot places related to Hongqiao airport.

Keywords. Travel mobility, Spatial clustering, Spatial aggregation, Visual analysis, Floating car data

1 Introduction

With the advances in location positioning and communication technologies, there is an increasing trend collecting floating car data (FCD) or spatial trajectory data from moving vehicles (e.g. private cars, taxis, buses) for understanding urban motilities. Extensive work has been done to investigate general urban mobility patterns using FCD. For instance, (Liu et al. 2012) examined large amounts of floating car data to understand intra-urban human mobility patterns from taxi trajectory data. (Ding, Fan and Meng 2015a) investigated different mobility patterns of taxi drivers at two income levels and analyzed their distinct behaviors especially when there are no passengers.

In urban areas, transport hubs such as airports and railway stations, where plenty of passengers are exchanged between vehicles or between transport modes, are of particular importance and have great influence on urban traffics. Understanding spatiotemporal travel/mobility patterns related to (e.g. traffic flows in/out) transport hubs may help urban planners and traffic managers to better learn passenger behaviors and to improve transportation planning.

However, analyzing movement data is very challenging due to their large data volume, implicit spatiotemporal relationship, and uncertain semantics. To tackle the challenges, a wide range of researches have been conducted to extract geographic

LBS 2016
13th Conference on Location-Based Services
Vienna, 14–16 November 2016

Published in "Proceedings of the 13th International Conference on Location-Based Services", edited by Georg Gartner and Haosheng Huang, LBS 2016, 14-16 November 2016, Vienna, Austria.

knowledge from movement data. Zheng and Zhou (2011) systematically investigated spatial trajectories from a wide spectrum of perspectives and disciplines, e.g. spatial database, mobile computing and data mining. They specifically pointed out the importance as well as the challenges in applying visualization methods for understanding the uncertainties in the spatial trajectories, exploring original data to inspire new ideas, and presenting the computing results to help a decision making.

A straightforward way for visualizing the floating car GPS points is by dot mapping. For instance, (Stanica, Fiore and Malandrino 2013) used dot maps to show the velocities of the road traffic in Cologne. One problem related to dot mapping of the raw data (often numerous points) is the unavoidable point overlapping effect. Another popular visualization technique is 3-D spatiotemporal mapping by making use of the height component for the representation of certain attribute. For instance, (Grant et al. 2011) used the extrusion of lines or areas to show the transportation delay in the Salt Lake City region; (Keler, Ding and Krisp 2016) proposed a selection circle visualization method using the 3-D representations of both average speed and average density derived from floating car data to investigate the traffic congestion in Shanghai. At a more aggregated level, proportional symbol mapping techniques are popular to show the overview of movement patterns. For example, (Andrienko et al. 2013) partition the space to irregular polygons based on clustering results to show aggregated movement flows. In addition, density mapping is a highly effective way to show density differences in geographic distributions across a landscape. (Prasannakumar et al. 2011) used density maps to show the hot spots of road accidents based on spatiotemporal clustering of road accidents and furthermore to investigate whether spatial or temporal factors, such as the proximity to the school or the season, have influences to the road accidents. Interactive mapping techniques are also popular in supporting visual exploration and analysis of dynamic movement patterns, e.g. NYC Taxi Holiday Visualization System[1] allows users to explore traffic from JFK and LGA airports during the holiday season (Nov 15th to December 31st), and HubCab[2] system developed by MIT Senseable city lab allows users to get insight into the taxi mobility patterns at a fine granularity and supports future taxi sharing based on a model named 'shareability networks' (Szell and Groß 2014).

For visually exploring traffic flow patterns related to transport hubs, (Ding, Yang and Meng 2015b) proposed a visual analytics workflow and developed a web-based visual interface for the visual exploration of movement data at both an aggregated and an individual level. In their work, they revealed the significant places as well as their semantics (e.g. residential, commercial, industrial, public) related to the transport hubs. However, there is a lack of in-depth investigation of the temporal patterns related to transport hubs.

In this paper, following a similar visual analysis workflow in (Ding et al. 2015b), which incorporates computational algorithms, data mining approaches, and visualization techniques, we go one step further with the aim to visually analyze the spatial and temporal traffic flow patterns related to transport hubs using floating car data. More specifically, we firstly preprocess a large amount of movement data, reconstruct trajectories and extract the starting and ending points related to the transport hubs. Secondly, we apply hierarchical clustering methods to derive significant clusters related to transport hubs and aggregate them to significant places. Finally, we design appropriate spatial and temporal visualization techniques, e.g. dot maps, proportional symbol maps, pie chart maps, to visually analyze those significant places. We use one-week FCD in Shanghai as our test dataset and select Hongqiao international airport as the test transport hub. The experiment results reveal significant spatiotemporal traffic flow patterns related to Hongqiao airport.

[1] http://taxi.imagework.com/
[2] http://hubcab.org/#13.00/40.7219/-73.9484

2 The visual analysis workflow

In this section, we introduce the general workflow of visualizing spatiotemporal distributions of transport hub related traffics, which consists of mainly traffic-hub related trajectory reconstruction, data preprocessing, point clustering and aggregation, and visualization and analysis.

2.1 Data preprocessing and trajectory reconstruction

This paper focuses on the investigation of the taxi trajectories with passengers travelling from and to the transport hubs. We firstly filter out erroneous GPS points for instance, inappropriate GPS points with locations outside Shanghai and timestamps beyond the study time slots. Since the raw data are individual GPS points, we reconstruct from the FCD geospatial database the occupied trajectories (with passengers) by connecting the temporal sequences of GPS points with the “car status” attribute value of 1. After filtering out erroneous trajectories, we then extract occupied trajectories related to (i.e. from/to) the transport hubs, and their starting (pick-ups) and ending points (drop-offs).

2.2 Point Clustering and aggregation

In this work, we use spatial clusters to identify pick-up/drop-off hotspots. Spatial clustering can be used to gain insight into the distribution of data, to observe the characteristics of each cluster, and to focus on a particular set of clusters for further analysis (Han, Kamber and Tung 2001). Specifically, we use a hierarchical agglomerative clustering method (Kaufman and Rousseeuw 1990) to detect the significant areas or places where most of the pickup/drop-off events happen. The given set of data objects is hierarchically decomposed, forming a dendrogram - a tree that splits the database recursively into small subsets. The dendrogram can be formed either “bottom-up” or “top-down”. This work adopts the “bottom-up” way. The “bottom-up” approach, also called “agglomerative” approach, starts with each object forming a separate group. It successively merges the objects or groups according to some measures like the distance between the two group centers and this is done until a termination condition holds.

To perform agglomerative hierarchical cluster analysis on a data set, the following procedure is required: 1) Find the similarity or dissimilarity between every pair of objects in the data set. 2) Group the objects into a binary, hierarchical cluster tree. 3) Determine the threshold to cut the hierarchical tree into clusters.

In this work, the input of the clustering method is the extracted pickup/drop-off events. The parameters of the methods are normally a distance function and a linkage criterion. The parameter setting is largely dependent on the applications. The output is the clustered events indicating which cluster the event belongs to. To filter out the significant places, we set a significant threshold value for the minimum number of the cluster elements. If a cluster has a number of elements larger than the significant threshold value, then this cluster is considered as a significant cluster.

Based on the clustering results, a variety of aggregated values can be derived for the significant clusters, for instance, the occurrences of pick-up events, the occurrences of the drop-off events, or the total number of the pick-up and drop-off events. The cluster centroids will be calculated to represent the location of the aggregates, which are regarded as the hot pick-up/drop-off places related to the corresponding transport hubs.

2.3 Visual analysis

In our work, we mainly focus on presenting significant spatiotemporal travel patterns related to the transport hubs. To visualize the aggregated significant events, proportional symbol mapping and pie-chart mapping are mainly used to show the one-variate (e.g. the total number of the pick-up and drop-off events) and two-variate (e.g. pick-up and drop-off events) spatial distributions respectively. For showing especially temporal information, for instance distributions of events at different time slots during one day or different days, small multiples of proportional symbol maps and pie chart maps will also be applied.

3 Experiment

3.1 Test data

The test dataset are raw GPS points collected from about 2000 GPS-enabled taxis within 47 days from 10th May to 30th June 2010, in Shanghai. The temporal resolution of the dataset is 10 seconds. Each position record has nine attributes, i.e. car identification number, company name, current timestamp, current location (longitude, latitude), instantaneous velocity, and car-status (meaning taxi occupied or empty). The detailed description of the fields is shown in Table 1. The data are stored in a MongoDB database.

Field	Example field value	Field description
Date	20100517	8-digit number, yyyymmdd
Time	235903	6-digit number, HHMMSS
Company name	QS	2-digit letter
Car identifier	10003	5-digit number
Longitude	121.472038	Accurate to 6 decimal places, in degrees
Latitude	31.236135	Accurate to 6 decimal places, in degrees
Velocity	16.1	In km/h
Car status	1/0	1-occupied; 0-unoccupied

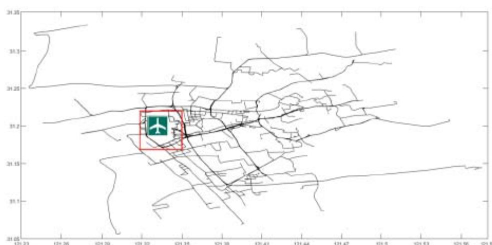
Table 1. The properties of the test data.

Figure 1 Shanghai Hongqiao airport area as study transport hub

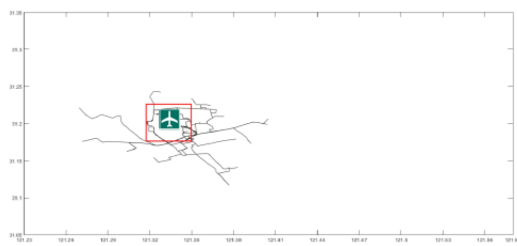
In this work, Hongqiao international airport is chosen as our study area. Figure 1 shows the location and extent of the Hongqiao international airport.

3.2 Data extraction and temporal division

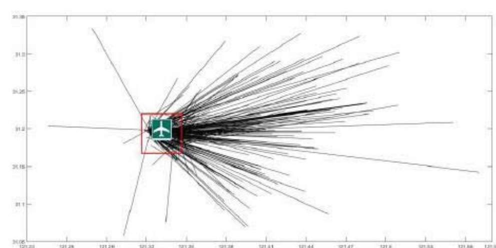
For illustration purpose, we extract one hour data at 6-7h (on 17th May 2010) from the database and reconstructed the GPS traces by connecting a temporal sequence of GPS points. Focusing on the occupied trajectories during this hour, we extract about 13,300 trajectories in the whole Shanghai area, of which around 300 trajectories are related to Hongqiao airport. Figure 2 shows the reconstructed occupied trajectories to/from Hongqiao airport (Figure 2(a) and 2(b)) and their origin-destination (O-D) trajectories (Figure 2(c) and 2(d)). Obviously, at 6-7 hour there are more taxis driving to Hongqiao airport than leaving the airport, which may indicate that many passengers take flight in the early morning at 6-7h.



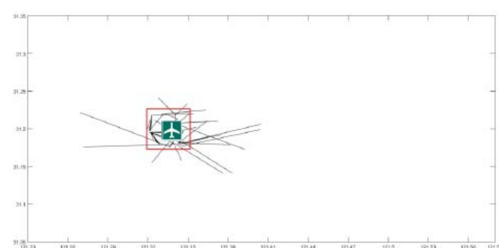
(a) Physical trajectories to airport



(b) Physical trajectories from airport



(c) O-D trajectories to airport



(d) O-D trajectories from airport

Figure 2 Spatial distribution of the occupied trajectories (a,b) and the corresponding origin-to-destinations (c,d) to and from airport at 6-7h.

To investigate hourly temporal mobility activities, we derive one-day trajectories to/from Hongqiao airport and divide the data to each hour. Figure 3 shows the hourly distribution of the number of trajectories from/to the airport using the bar charts. Red and blue represent trajectories to and from the airport respectively. The bar chart shows that: (1) there are temporal patterns like peaks and valleys for the traveling activities related to the airport during a day. For instance during the early morning (2-4h), there are much fewer trajectories related to the airport than other time periods, while at 6-7h, 11-12h, 23-24h there are more trajectories; (2) there are more trajectories to the airport at 4-10h than from the airport, approximately similar amount of trajectories from and to the airport at 11-16h, and increasing trajectories from the airport than to the airport at 17h-24h and 1-2h. This indicates that there are more passengers taking taxis in the morning to the airport, while in the late afternoon or during night there are more passengers taking taxis to the city centre.

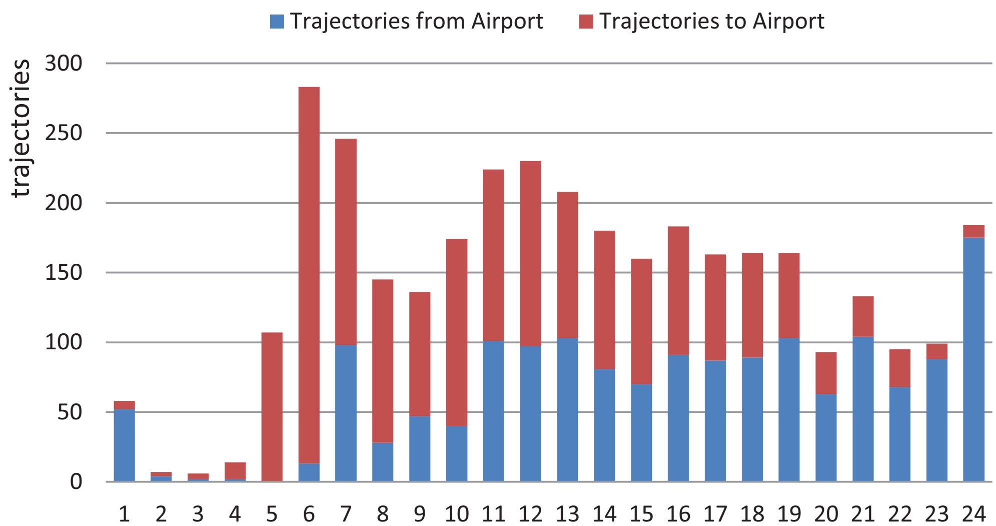
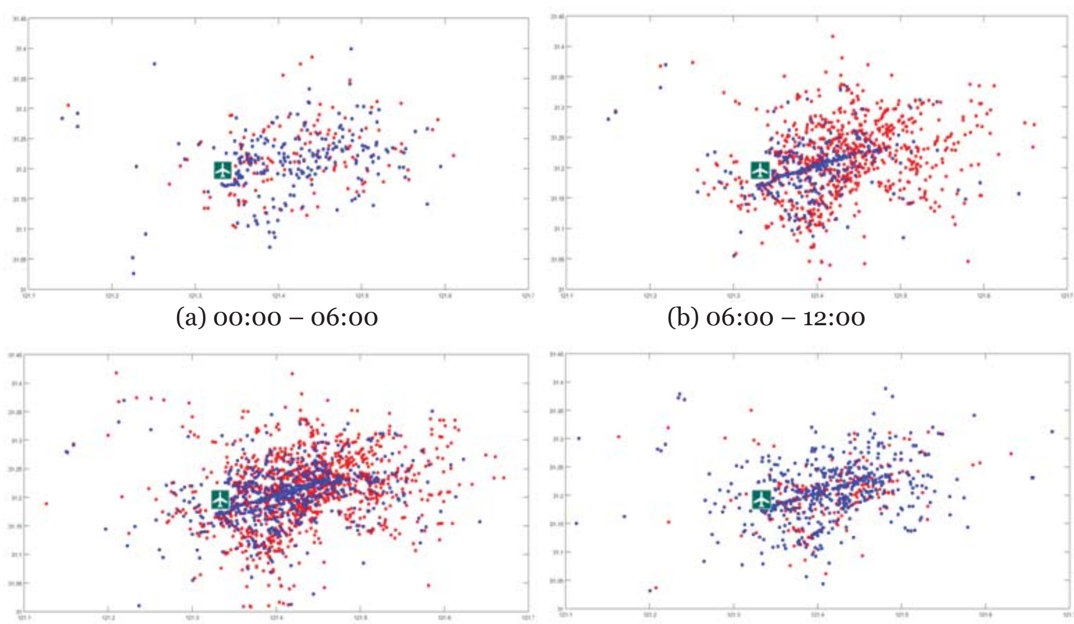


Figure 3 Hourly patterns of trajectories from and to the Hongqiao airport

Furthermore, we focus on analysing only the spatiotemporal distribution of the pick-up and drop-off points. We investigate the pick-ups and drop-offs distributions during four typical time-slots, i.e. 00:00-06:00, 06:00-12:00, 12:00-18:00, and 18:00-24:00. Figure 4 shows their spatial distributions at each time slot. Obviously, there are general distribution patterns can be observed, for instance, denser pick-up/drop-off activities at 06-12 in Figure 4(b) and at 12-18 in 4(c), and less events at 18-24 in Figure 4(d), especially at 00-06 shown in 4(a). In addition, we can also observe distinct patterns of pick-ups and drop-offs in each subfigure. For example, it seems that at 6-12 there are more drop-offs than pick-ups. However, due to the dot over-plotting problem, it is hard to detect hot spots of the two kinds of events.



(a) 00:00 – 06:00

(b) 06:00 – 12:00

(c) 12:00 – 18:00

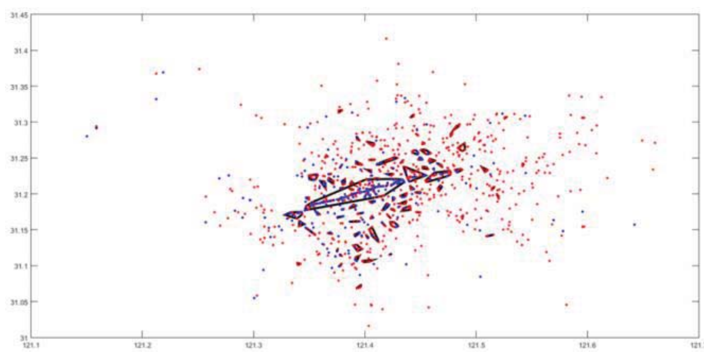
(d) 18:00 – 24:00

Figure 4 Pick-up and drop-off events in four time slots. Red represent pick-up events and blue drop-off events.

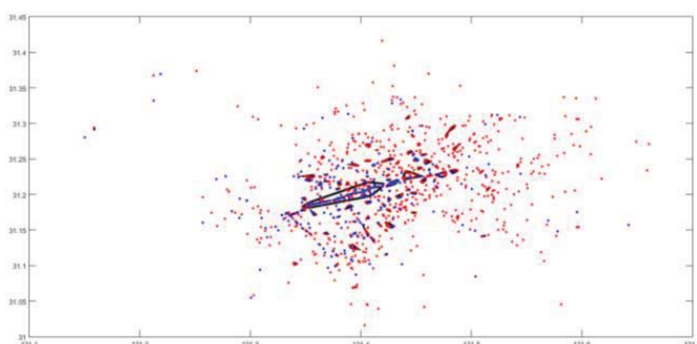
3.3 Pick-up/Drop-off clustering and aggregation

To solve the dot over-plotting issues and detect pick-up and drop-off hotspots, we use clustering methods to find dense pick-up and drop-off areas and aggregate the events to significant places, which has more mobility activities related to the airport.

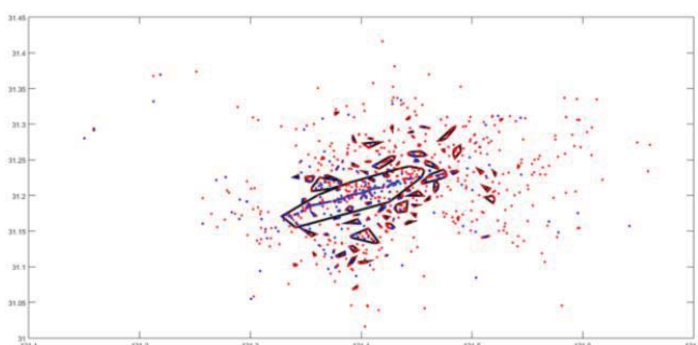
In our experiment, we use the agglomerative hierarchical clustering method to find dense areas. In order to define the significant places, we set a significant threshold value for the minimum number of the cluster elements. If a cluster has a number of elements larger than the significant threshold value, then this cluster is a significant cluster. Since the clustering method will assign each point to a cluster and here we are interested only in clusters with a higher density or a larger number of cluster elements, we select clusters with more than 5 points as significant clusters. Instead of trying different distance and linkage parameters, we used the hierarchical clustering methods provided by Matlab and specify arbitrary clusters using a parameter "maxclust". To find appropriate "maxclust", we try different values and the resulted clusters are illustrated in Figure 5.



(a)



(b)



(c)

Figure 5 Hierarchical clustering results of pick-up and drop-off events at 6-12 with different "maxcluster" parameter (a) 500 (b) 600 (c) 400. The convex hull shows significant clusters.

For each time slot, we select different parameter values for the clustering since the temporal distribution of the pick-up and drop-off events are different.

After we get the significant clusters, we aggregate the cluster elements and use the cluster centroids as the spatial locations of the aggregates. Each cluster has corresponding aggregated values, for instance, the number of pick-up events, the number of drop-off events, and the total number of pick-up and drop-off events. Those summarized values at the cluster centroids reflect pick-up and drop-off activity hot spots.

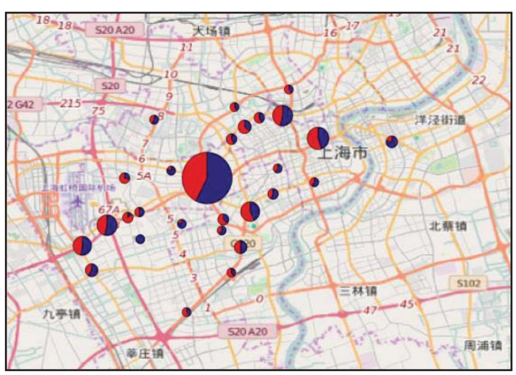
3.4 Visual analysis

To show the pick-up and drop-off hotspots related to the transport hub, we apply pie chart mapping and proportional symbol mapping techniques, which respectively represent the pick-up and drop-off distribution and the total number of the pick-ups/drop-offs.

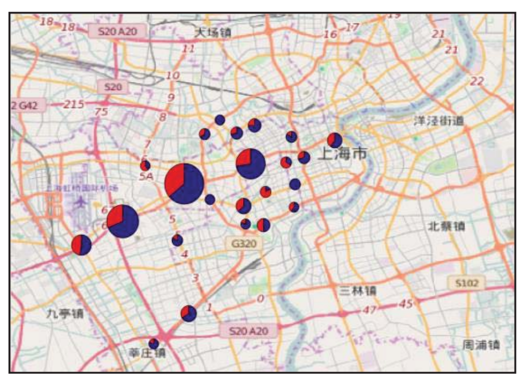
To get an overview of the temporal hotspots of the pick-ups and drop-offs respectively, we use multiple pie-chart maps at four time slots. Figure 6 shows the results of the pie-chart mapping results. The sizes of the pie charts are proportional to the total number of the pick-up and drop-off events. The red and blue pie sectors represent pick-up and drop-off events.

(a) 00:00 – 06:00

(b) 06:00 – 12:00



(c) 12:00 – 18:00



(d) 18:00 – 24:00

Figure 6 Multiple pie chart maps showing the pick-up and drop-off patterns. Red represents pick-up events and blue drop-off events.

From Figure 6, we can easily observe the spatial distribution of the hotspots of pick-up and drop-off events at the four time slots during a day. Most of the host spots are distributed to the east of the airport, which towards the city center area. Compared with the other time slots, at 00-06 (Figure 6(a)) there are obviously fewer and smaller hot spots. At 06-12, pick-up events are dominant, which indicates in the morning there are more passengers taking taxis at the hot spots (shown Figure 6(b)) and traveling to Hongqiao airport. Beside the hot spots eastwards of the airport, at this time slot there are also several hotspots southeast of the airport. While at 12-18 (Figure 6(c)) and 18-24 (Figure 6(d)), there is an increasing trend more passengers travel from the airport to the hotspots in the city center.

To visualize the temporal distribution of the total amount of travel activities (pick-ups and drop-offs) related to the airport during a day, we present a pie cart map in Figure 7. Different colors (black, white, light gray, and dark gray) represent different time slots (00-06, 06-12, 12-18, and 18-24). The size of each sector are proportional to the total number of pick-up and drop-off events at this time slot. The biggest pies represent the most active areas that traveling to/from the airport. The proportions of the pies reveal the temporal distribution of the travelling activities. There are a few black or white pies which indicate the corresponding hotspots are significantly active at early morning (00-06) or in the morning (06-12). There is a big dark gray circle in the center of the map and another one right on the south of the airport, which represent two hotspots that especially active in the evening (18-24) travelling to/from the airport. The pie charts with even temporal distributions indicate hotspots that are active for the whole day or most of the day.

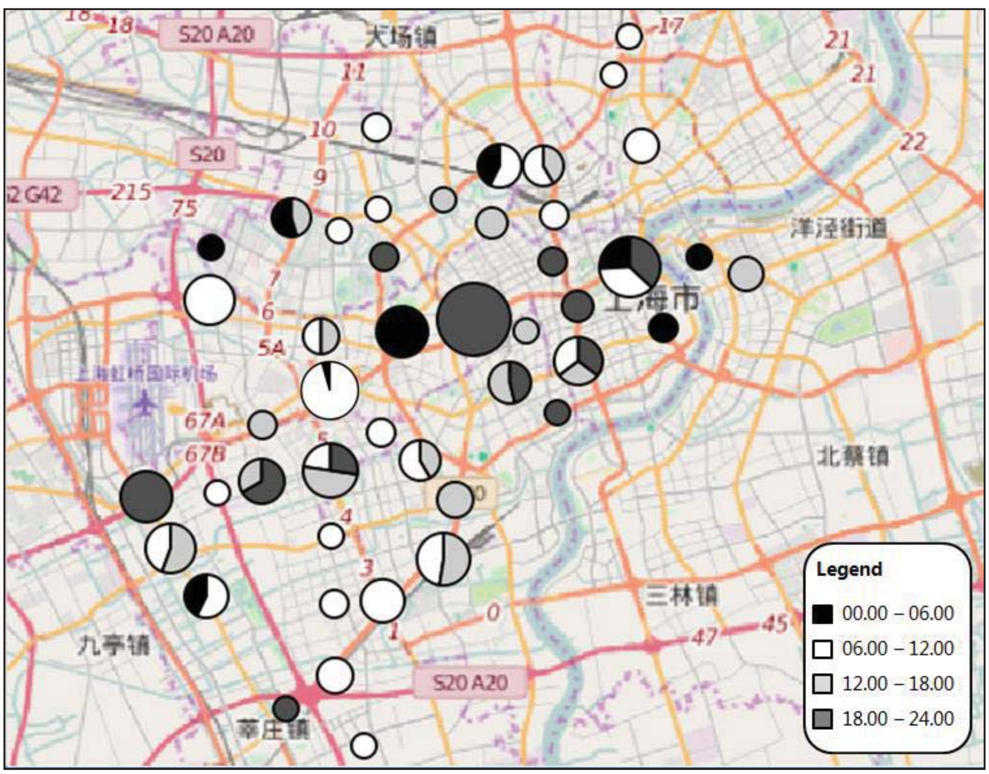


Figure 7 Temporal pie-chart mapping of the pick-up and drop-off events.

Moreover, we investigate one-week FCD data related to Hongqiao airport. The bar chart in Figure 8 shows a daily pattern in one week for trajectories to and from airport. In general, there are relatively more activities related to the airport from Monday to Thursday than on Friday and at weekend. In terms of the hourly pattern, similar to Figure 3, it shows active travels to the airport in the morning and afterwards an increasing amount of passengers traveling from airport to the city center.

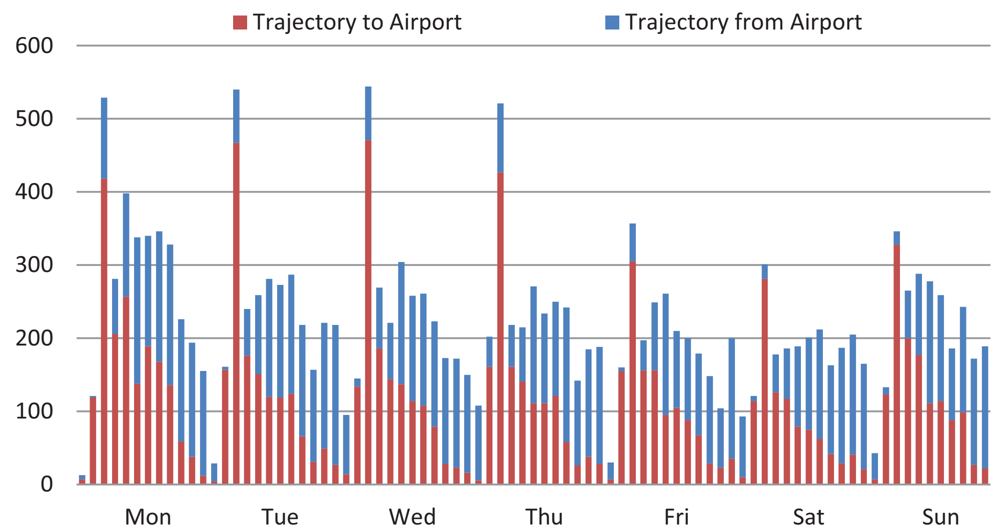


Figure 8 Daily-patterns of trajectories from and to the Hongqiao airport

In order to present the weekly data in one map, we firstly calculate for each day the significant clusters and for the whole week we aggregate the overlapped clusters into one big cluster with a recalculated centroid. Then we create a pie chart over each cluster to show the distribution of the number of events from Monday to Sunday in each significant place. As shown in Figure 9, red, orange, yellow, green, cyan, blue and purple color represent the sum of the pick-up and drop-off events on Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday respectively. The area of each pie sector are proportional to the total number of pick-ups and drop-offs at the corresponding day.

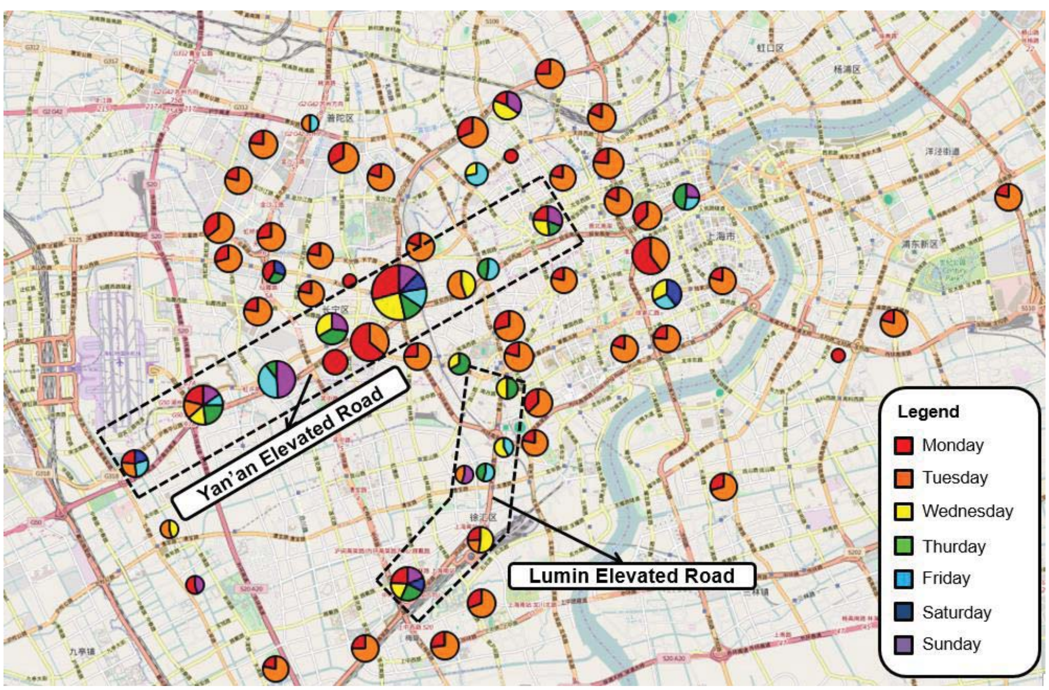


Figure 9 Pie chart map of Hongqiao airport related significant places from Monday to Sunday

It is evident from Figure 9 that the majority of the significant places based on the clustering and aggregation results have large portions of activities on Monday and Tuesday (red and orange), especially Tuesday (orange). In addition, we could see that a large amount of significant places located along several major roads. Along the

"Yan'an Elevated Road" and "Luming Elevated Road" the time distribution of significant places from Monday to Sunday appears to be more even, which indicate these two road are popular road where taxis pick up passengers to Hongqiao airport or drop off passengers from Hongqiao airport.

4 Conclusion & Outlook

This paper aims to analyze and visualize the spatiotemporal mobility patterns related to transport hubs using floating car data. By following a visual analysis workflow incorporating computational algorithms, data mining approaches, and visualization techniques, we investigated the spatiotemporal patterns of one-week FCD data. The experiment results show that there are obvious hourly and daily temporal mobility patterns and as well significant spatial hotspots related to transport hubs.

In the future, we plan to improve our work in the following aspects. Firstly, we would like to investigate the appropriate clustering parameter in an interactive manner, for instance, by implementing a slider bar for selecting different parameter values. Secondly, we will integrate this work to the web-based visualization system proposed in (Ding et al. 2015b) and improve the graphic user interface with powerful interactive functions to allow users explore the visualization results.

References

Andrienko, G. L., N. V. Andrienko, P. Bak, D. A. Keim & S. Wrobel. 2013. *Visual Analytics of Movement*. Springer.

Ding, L., H. Fan & L. Meng. 2015a. Understanding taxi driving behaviors from movement data. In *AGILE Conference on Geographic Information Science*. Lisbon.

Ding, L., J. Yang & L. Meng. 2015b. Visual Analytics for Understanding Traffic Flows of Transportation Hub from Movement Data. In *International Cartographic Conference 2015*. Rio de Janeiro, Brazil.

Grant, M., M. Day, R. Winick, A. Chavis, S. Trainor & J. Bauer. 2011. Showcasing Visualization Tools in Congestion Management. Federal Highway Administration (FHWA), U.S. Department of Transportation

Han, J., M. Kamber & A. K. H. Tung. 2001. Spatial Clustering Methods in Data Mining: A Survey. In *Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS,* eds. H. J. Miller & J. Han. Taylor and Francis.

Kaufman, L. & P. Rousseeuw. 1990. *Finding Groups in Data*. Wiley & Sons, Inc., New York.

Keler, A., L. Ding & M. J. Krisp (2016) Visualization of traffic congestion based on Floating Taxi Data. *Kartographische Nachrichten (KN)*, 2-13.

Liu, Y., C. Kang, S. Gao, Y. Xiao & Y. Tian (2012) Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems,* 14, 463-483.

Prasannakumar, V., H. Vijith, R. Charutha & N. Geetha (2011) International Conference: Spatial Thinking and Geographic Information Sciences 2011Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment. *Procedia - Social and Behavioral Sciences,* 21, 317-325.

Stanica, R., M. Fiore & F. Malandrino. 2013. Offloading Floating Car Data. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*, 1-9.

Szell, M. & B. Groß. 2014. Hubcab – Exploring the Benefits of Shared Taxi Services. In *Decoding the City How Big Data Can Change Urbanism*.