

## Dissertation Abstract

# **The Possibility of Big Data Spatio-Temporal Analytics for Understanding Human Behavior and Their Spatial Patterns in Urban Area**

Graduate school of  
Natural Science& Technology  
Kanazawa University

Major subject: Environment Design

Course: Environment Planning

School registration No.: 1424052011

Name: 李 苗裔 (LI Miaoyi)

Chief advisor: Prof. Zhenjiang Shen

## **Abstract**

The aims of this dissertation are two-fold: first, to develop a comprehensive framework to integrate multiple dimensions and scales, and systematically compare human behavior patterns and current urban form; second, to further the understanding of the role spatial and temporal effects might play in human behavior patterns analyzing. The objective of this dissertation is the application of spatio-temporal analysis employing big data on human behavior patterns for identifying urban problems hidden in current urban form. Under the comprehensive framework, this research will apply some new methods to analyze spatial distribution pattern of urban function and its mixture, spatio-temporal distribution pattern of population and correlation with urban function, spatiotemporal patterns and interactions between subway and taxi ridership.

As proved by published research articles, the application of the framework of spatio-temporal analytics employing big data we introduced in this PhD research would focus on two essential aspects of the urban system: urban form and human behaviors. In this dissertation, human behavior refers to the mobility of human beings taking advantage of various transportation modes in urban area, which distribution generates spatial patterns during their daily life. Their spatial patterns can reflect how the mobility of human beings relative to the urban form. We tested human behaviors using their moving tracks recorded in mobile devices and traffic cards, these types of datasets belong to spatio-temporal big data. The analysis results can provide support for planner and designer to find urban problems hidden in current urban form.

**Key Words:** Human Mobility; Urban Form; Urban Function Mix; Spatio-Temporal Distribution of Population; Transit Ridership; Spatio-Temporal Entropy Model; Spatio-Temporal Network Analysis

# **1. Introduction**

## **1.1 Existing Research for analyzing Human Behavior and Their Spatial Pattern**

There are many research reports on spatial analysis and modelling with Geographic Information System (GIS) in the world (Maguire et al, 2005). In this vein, many studies examined the capacity of GIS modelling in aid in understandings of household choices (Ettema, 2011; Gaube & Remesch, 2013; Q. Huang, Parker, Sun, & Filatova, 2013), residential demand (Fontaine & Rounsevell, 2009), and land-use changes (Jokar Arsanjani, Helbich, & de Noronha Vaz, 2013; Kocabas & Dragicevic, 2013). The applications of GIS modelling range from simulation to prediction to monitoring (Bernard & Kruger, 2000; Carrara, Guzzetti, Cardinali, & Reichenbach, 2000; Q. H. Huang & Cai, 2007) , from modelling human-related issues to modelling natural hazards (Al-Sabhan, Mulligan, & Blackburn, 2003; Perry, Sparrow, & Owens, 1999; Thornton, Pearce, & Kavanagh, 2011; Xu, Nyerges, & Nie, 2014), from modelling real-world problems to modelling virtual-world activities (Croitoru, Wayant, Crooks, Radzikowski, & Stefanidis, 2014; Turgeon, 2013), from using data collected by digitally instrumented devices (e.g. remote sensing images, GPS coordinates, sensor network data) to using data collected by humans (e.g. survey data, geo-social media data, mobile phone data) (Blaschke, Hay, Weng, & Resch, 2011; Liu, Sui, Kang, & Gao, 2014; Patino, Duque, Pardo-Pascual, & Ruiz, 2014; Widener & Li, 2014). In fact, the study of individual behaviors has a long history in GIS field. Survey data, such as the survey of people's daily activities, is initially exploited as the main data source (M. Kwan, 1998).

There are many national surveys on industrial activities and commercial activities and housing in urban spaces. But conventional surveys are impossible to show the tracks of all urban activities. Big data including limited attributes of individual person and his/her behavior's, helps us to understand the diverse urban activities and complicated urban structures in urban spaces (Nabian, Offenhuber, Vanky, & Ratti, 2013). Some initiatives in this area include studying Beijing's function zones using human mobility and POI

data(Yuan et al., 2012) and evaluating the effectiveness of a future road plan based on taxi trajectory (Zheng, Liu, Yuan, & Xie, 2011). Recently, many reports attempted to identify, characterize, and develop geospatial frameworks and algorithms for individual behavioral models for spatio-temporal knowledge discovery. Geospatial frameworks were developed to analyze people's activities patterns over time and geography (M. Kwan, 2004). Geo-visualization methods were introduced to aid in understandings of those patterns by delivering interactive visual outputs (M. Kwan, Lee, & Oxford, 2003). More recently, the increasing use of volunteered geographic information (VGI) and open data has further spurred the study of individual behaviors. In this thread, researchers built upon the previous work of spatio-temporal data mining, and developed methods to discover spatio-temporal patterns of human activities from geotagged photos and mobile phone data.(Andrienko, Andrienko, Mladenov, Mock, & Pölitiz, 2010; Sagl, Delmelle, & Delmelle, 2014; YT Zheng, Li, Zha, & Chua, 2011).

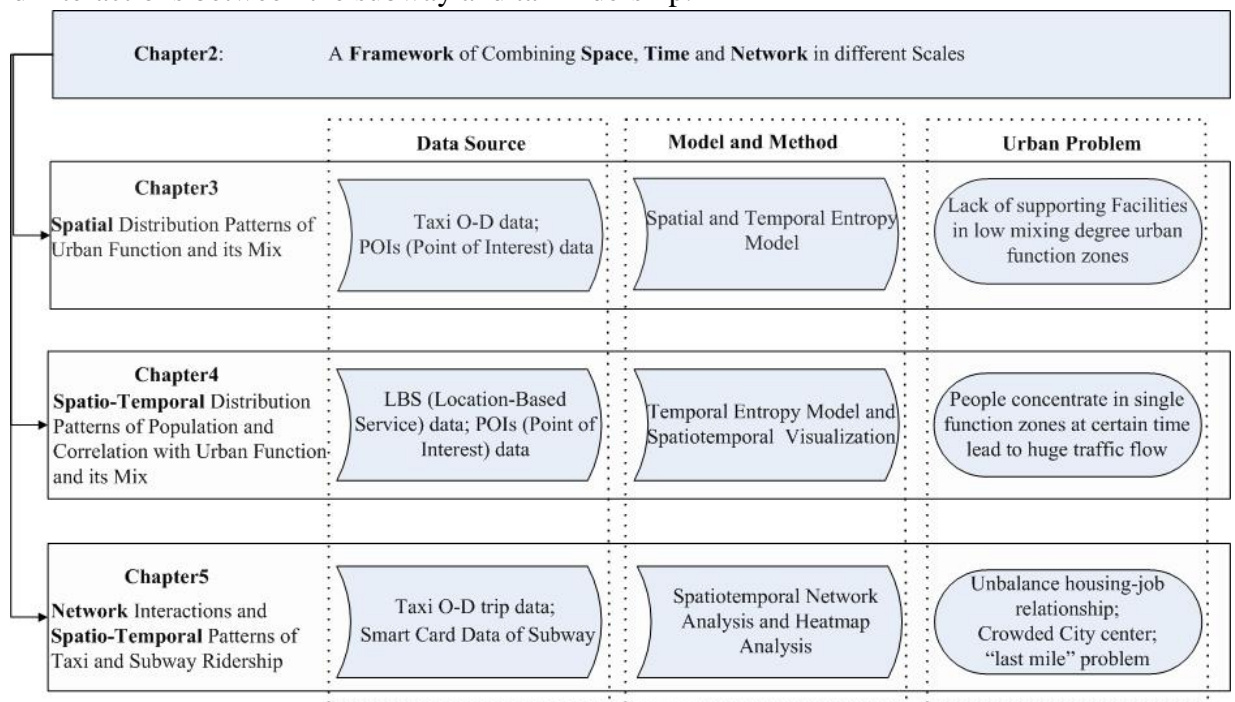
Through reviewing the rich literatures, the main research question of this dissertation will be raised: how to build a comprehensive framework for comparative spatio-temporal data analysis for analysing spatial pattern of human behavior? To illustrate the function of this comprehensive framework, some spatio-temporal analysis methods will be developed to compare spatio-temporal pattern of human behaviors and current urban form. Through the above comparative analysis, we argue that it is possible to identify problems hidden in urban form.

## **1.2 Research Purpose**

In this dissertation, spatio-temporal big data refers to a large dataset which stores spatially referenced time-series with single or multiple thematic attributes, such as cell phone trip data, taxi O-D trip data, location-based service (LBS) data and so on. Thus, human behavior refers to the mobility of human beings taking advantage of various transportation modes in urban area, which distribution generates spatial patterns during their daily life. Their spatial patterns can reflect how the mobility of Human beings relative to the urban form. This PhD research aims to discuss the possibility of spatio-temporal analytics employing big data could contribute to analyzing human behavior pattern dynamically and comparatively for identifying urban problems hidden in current urban form that otherwise would be very difficult to detect.

Firstly, we develop a framework which can integrate spatio-temporal and network

considering different scale, it sheds new insights on possible data fusion and comparative analysis. It can be applied to different data sources and methods to visualize human mobility reflected by trajectory data which exists in the current urban form. Secondly, according to this framework, we continue to mainly employ spatio-temporal big data which stores rich human mobility information to analyze spatial distribution pattern of urban function and its mixture, the correlation between the spatio-temporal distribution of population and urban function, spatiotemporal patterns and interactions between the subway and taxi ridership.



**Fig 1.1** The framework of this Ph.D. research contents

In the following chapters, the framework of main contents is shown in Figure1.1, spatial and temporal entropy model, spatio-temporal network analysis, spatio-temporal visualization and heatmap analysis methods will be set as the specific analysis methods under the analytics framework, and the urban trajectory dataset will mainly be used for case studies of different cities in China. They are Taxi O-D trip data which is extracted from Taxi GPS dataset, Cell Phone trip data and smart card data of subway. Meanwhile, we also combine POIs (Point of Interest) and LBS (Location-Based Service) data into our research work. It is important to understand the methodology of how to check the human mobility in urban area through the spatio-temporal big datasets.

## 2. A Framework for Comparative Spatio-Temporal Analysis

Chapter 2 will stress the need to study multiple dimensions and multiple scales underlying spatio-temporal income big datasets. We establish an analytical framework that considers three dimensions and four scales. Table 2.1 conceptualizes the according to 12 basic analysis units (A1-A12).

**Table 2. 1** The framework for spatio-temporal analysis

		Scales			
		Individual	Local	Meso	Global
Distributions (Dimensions)	Space	A1	A2	A3	A4
	Time	A5	A6	A7	A8
	Network	A9	A10	A11	A12

By customizing the combination of analysis units, the application of a comparative exploratory analysis framework will be suggested for use in analysis of spatio-temporal big data series to measure human mobility pattern reflected by trajectory data dynamically and comparatively. Two case studies of Chinese cities are carried out to evaluate the usefulness of proposed conceptual framework. In our case studies, we extracted human mobility patterns and incorporate them with the analysis of urban structure through an integrated spatio-temporal network analysis framework. Our results suggest that the proposed framework can comprehensively quantify the variation of human mobility across various scales and dimensions, it also can expose some hidden patterns and trends that otherwise would be very difficult to detect.

*Publication1: Miaoyi Li, Xinyue Ye, Shanqi Zhang, Xiaotong Tang, Zhenjiang Shen. A Framework of Comparative Urban Trajectory Analysis[J]. Environment and Planning B: Urban Analytics and City Science. (Onlinefirst; SSCI IF:1.527)*

### **3. Evaluating Urban Function Mix based on POIs and Taxi**

#### **O-D Data**

Urban functions exist in urban form, it also occur interaction with human behaviors. That means identifying different functional zones and understanding their spatial pattern within the city is benefit for finding out some issues taking place in current urban form. In this Chapter, we will analyze the spatial distribution patterns of urban functions based on the data of points of interest (POIs). In general, the profit-oriented functions are more likely to aggregate in a heterogeneous way, of which the large-scale functions tend to be heterogeneous, while the small-scale functions are more likely to be more homogenous and hierarchical; public-service-oriented functions incline to present homogeneous and discrete pattern, of which the partial profit-oriented functions are gathered in a hierarchical and heterogeneous way and in small scales.

Then according to taxi O-D trip which is extracted from the taxi GPS data, we identified different functional zones. Finally, we establish a spatial entropy and a temporal entropy based on POIs data and taxi O-D data which belongs to one type of spatio-temporal big data, using these methods to calculate the degree of urban function mix and analyze the spatial pattern of the degree of urban function mix. Beijing's the degree of urban function mix are featured by gradually decreasing from the 3rd and 4th Ring Roads out to the periphery, among which traditional core areas including the Old Dongcheng District, Old Xicheng District, the area within the 2nd Ring Road, and CBD with the highest mixing degree; followed by Zhongguancun and Wangjing; then Tiantongyuan and Huilongguan communities with the lowest mixing degree because of their dominant residential functions.

*Publication2: Miaoyi Li, Zhenjiang Shen\* et al. Application of Spatial and Temporal Entropy Based on Multivariate Data for Measuring the Degree of Urban Function Mix[J]. China City Planning Review, Vol.24, No.1, 40-48. (2015) (CSCD)*

## **4. Uncovering the Relationship between Spatio-Temporal Distribution of Population and Urban Function with Location-Based Service Data**

On basis of Chapter3, Chapter4 examine This research examines the spatio-temporal distribution of population in an urban area and its relationship with urban functions using an unprecedented high-resolution and broad-coverage-crowd LBS dataset from Tencent, one of the biggest Internet companies in China. By examining the distribution of the population during different time periods, different urban morphologies are observed. The analysis of the spatio-temporal population distribution based on temporal entropy indicates that population distributions of employment, commercial, and scenic areas have larger temporal fluctuations than those in residential and mixed-use areas. As for the Spearman correlation coefficient between urban functions and temporal population distribution based on 300\*300-meter grids, it is then measured to uncover the underlying reason for this spatio-temporal distribution of population.

The result demonstrates that as urban functions become more mixed, the temporal distribution of the population becomes more even. At a local scale, temporal population distribution in key areas shows that the location of people in a certain place is in accordance with human behavior. In addition, temporal entropy indicates the level of population concentration at different times. Areas with low temporal entropy may have a high temporal concentration, meaning that people are likely to come to these areas at the same time. This concentration is likely to affect the traffic nearby greatly.

*Publication3: Miaoyi Li , Zhengjiang Shen\* , Xinhua Hao. Revealing the relationship between spatio-temporal distribution of population and urban function with social media data[J]. GeoJournal, Volume 81, Issue 6, 919-935.(2016) (EI/SCImago)*



## **5. Examining Taxi Ridership Impacts from Newly Opening Subway Line with Taxi Trip Data**

In addition to urban function discussed above Chapter3 and Chapter4, Transportation Network in big cities is also a crucial part of the urban form. Nowadays, smart card and mobile device for transportation are widely used in many Chinese cities. We can try to use this type of spatio-temporal big datasets to visualize the human mobility of transportation network in urban area.

Chapter5 quantitatively examines the impact of the first and newly opened subway, using spatio-temporal network analysis based on the combination of emerging taxi trips and subway transaction records, which can lead to a deeper understanding of human mobility and sustainable urban development as well as provide informative insights in population mobility and urban configuration. We calculate the subway network and evaluate its impact on the surrounding taxi volume and analyze the influence of the subway on the taxi network structure (pick-ups and drop-offs) by defining an index to analyze the taxi imbalance issue within the city. Further, we divide the taxi trips into four modes to see where the subway impacted taxi ridership. Finally, from our findings we conclude that the farther away from the city center, the greater the impact; the closer to the subway station, the greater the impact; and we find that not all taxi ridership decreased but in fact increased in some locations, which was explained with analysis. From the result of taxi trips' O-D density in the travel distribution and the distance effect of distribution, we come up with the conclusion that Wuxi is a mono-centric city.

*Publication4:Miaoyi Li, Lei Dong, Zhenjiang Shen\*,Wei Lang, Xinyue Ye\*. Examining the interaction of taxi ridership and subway for sustainable urbanization[J]. Sustainability. 2017; 9(2):242. (SSCI/SCI IF:1.789)*

## **6. Conclusions**

### **6.1 Summary**

After development, planner and designer in Chinese cities have to pay more attention to find out the urban problem, understanding the reason of urban problem in current urban form is the start point for making the future plan or design of a city. For this, my PHD research work developed the theoretical framework of spatio-temporal analytics, according to it, we employ some specific methods based on spatio-temporal big datasets for identifying urban problems hidden in current urban form through analyzing human behaviors and their spatial patterns.

However, in this stage, we just proposed the spatio-temporal analytics framework and from the perspective of human behaviors according to this analytics framework to explain spatial distribution pattern of urban function and its mixture, spatio-temporal distribution pattern of population and correlation with urban function, spatiotemporal patterns and interactions between subway and taxi ridership. In this PhD dissertation, we have not done the comprehensive analysis of the urban problem and also haven't found solutions for urban problem. But this framework and analyzing methods can provide the good support for the urban planner and designer to consider the future plan of a city based on the findings of spatio-temporal analytics.

### **6.2 Future Work**

For future work, we should design the complete analyzing indicator putting into this framework to comprehensive analyze urban problem. And then, we could based on open source spatial analysis software such as Open Source Package STARS (Spatio-temporal Analysis of Regional Systems) and PySAL (Open to develop the Source Python Library for Spatial Analytical Functions) (Rey and Janikas, 2006; Rey and Anselin, 2007) to develop the new tools for analyzing and evaluating urban problem. It's easy to use for urban planning and urban design practices.

## References

- Andrienko G, Andrienko N, Demsar U et al., 2010. Space, time and visual analytics. *International Journal of Geographical Information Science*, 24(10): 1577–1600. doi: 10.1080/13658816.2010.508043
- Andrienko G, Andrienko N, Keim D et al., 2011. Editorial: Challenging problems of geospatial visual analytics. *Journal of Visual Languages and Computing*, 22(4): 251–256. doi: 10.1016/j.jvlc.2011.04.001
- Ettema, D. (2011). A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. *Computers, Environment and Urban Systems*, 35, 1–11. doi:10.1016/j.compenvurbsys.2010.06.005
- Feick, R., & Robertson, C. (2014). A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems*. doi:10.1016/j.compenvurbsys.2013.11.006
- Fontaine, C. M., & Rounsevell, M. D. A. (2009). An Agent-based approach to model future residential pressure on a regional landscape. *Landscape Ecology*, 24, 1237–1254. doi:10.1007/s10980-009-9378-0
- Gaube, V., & Remesch, A. (2013). Impact of urban planning on household's residential decisions: An agent-based simulation model for Vienna. *Environmental Modelling and Software*, 45, 92–103. doi:10.1016/j.envsoft.2012.11.012
- Huang, Q., Parker, D. C., Sun, S., & Filatova, T. (2013). Effects of agent heterogeneity in the presence of a land-market: A systematic test in an agent-based laboratory. *Computers, Environment and Urban Systems*, 41, 188–203. doi:10.1016/j.compenvurbsys.2013.06.004
- Jokar Arsanjani, J., Helbich, M., & de Noronha Vaz, E. (2013). Spatiotemporal simulation of urban growth patterns using agent-based modeling: The case of Tehran. *Cities*, 32, 33–42. doi:10.1016/j.cities.2013.01.005
- Kocabas, V., & Dragicevic, S. (2013). Bayesian networks and agent-based modeling approach for urban land-use and population density change: A BNAS model. *Journal of Geographical Systems*, 15, 403–426. doi:10.1007/s10109-012-0171-2
- Kwan, M., Lee, J., & Oxford, C. (2003). Geovisualization of Human Activity Patterns Using 3D GIS : A Time-Geographic Approach. In M. Goodchild & D. G. Janelle (Eds.), *Spatially Integrated Soical Science: Examples in Best Practices*. Oxford Unviersity Press.
- Kwan, M.-P. (2004). GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns. *Geografiska Annaler: Series B, Human Geography*, 86, 267–280. doi:10.1111/j.0435-3684.2004.00167.x
- Liu, Y., Sui, Z., Kang, C., & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PloS One*, 9(1), e86026. doi:10.1371/journal.pone.0086026
- Luo, F., Cao, G., Mulligan, K., & Li, X. (2015). Explore Spatiotemporal and Demographic Characteristics of Human Mobility via Twitter: A Case Study of Chicago. *Applied Geography*, 70, 11–25. <http://doi.org/10.1016/j.apgeog.2016.03.001>
- Robertson, C., & Feick, R. (2015). Bumps and bruises in the digital skins of cities: unevenly distributed user-generated content across US urban areas. *Cartography and Geographic Information Science*, <http://doi.org/10.1080/15230406.2015.1088801>
- Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In *ACM KDD'12*. Beijing, China.