Article ID: 1672-5328(2017)04-0001-08 CLC number: U491.1+2 Document code: A DOI:10.13813/j.cn11-5141/u.2017.0407-en

Multi-Source Data in Analyzing Urban Travel Patterns

Wu Zixiao, Fu Lingfeng, Zhao Yixin
(China Academy of Urban Planning & Design，Beijing, 100037)

Abstract: As the information technology advancing, new data sources are constantly emerging in the urban transportation. Effectively utilizing these data helps us to better understand urban functionality and transportation characteristics. By reviewing the commonly used multi-source data in urban travel analysis and modeling, this paper elaborates the typical applications of the data, including data from mobile phone, transit system, vehicle GPS, traffic detection and sale of transit fare system. Based on the data demand in the traditional urban transportation analysis framework, the paper discusses the interconnection among the multi-source data sources and the combined applications. The paper emphasizes that replacing the traditional data with new information data is neither practical nor necessary. The value of the information data depends on the innovation of the travel analysis and modeling technology in the future. Finally, the paper outlines the future development and application of multi-source data.
Keywords: transportation survey; urban transportation characteristics; travel analysis and modeling; multi-source data; combined application

Received Date: 2017-05-02
About the Author: Wu Zixiao(1970—), Male, from Qishan Shanxi, Professor-level Senior Engineer, Research Focus Areas: Transport Modeling, E-mail: wuzx@caupd.com

0 Introduction

The research on urban traffic patterns is usually driven by traffic data. Based on the available traffic data, traffic analysis and modeling theory are developed. In turn, more advanced analysis and theory would create the demand for new data to serve traffic model calibration and validation. The various ways of data collection differ in scales, frequencies and costs, which could affect the development of theories of traffic analysis.

Smart transportation systems and new devices are widely used due to the fast advancement of information technology in various fields. The application of these devices also generates a significant amount of new data, that is, information data. Information data captures travel activity information as well as the key characteristics of an urban transport system. The analysis on such type of data would help to reveal urban travel patterns and characteristics as well as the functions of a city. Besides, the emergence of new data could enable future innovation in the techniques of transportation analysis and modelling.

1 The Data for Transportation Analysis and Modelling

1.1 Travel survey data

Travel survey data is the most direct and traditional data source for transportation analysis and modelling. The survey arose alongside the emergence of urban transport planning, in America, which published the Standardized Procedures for Personal Travel Surveys in 1944[1]. In the same year, such a survey method was brought into practice in Salta and New Orleans. Travel survey is costly in terms of time and resources. Therefore, it usually relies on statistical theories to generate sophisticated sampling design and identify a survey sample. Different types of survey could manifest the characteristics of urban travel from different perspectives (see Tab.1).

1.2 Census and statistical data

Census and statistical data include demographic census data (census survey in every certain years and sampled survey in the middle year between any two census surveys), economic census data, vehicle ownership data and transit passenger flow data etc. Census and statistical data could offer the essential data for transportation analysis (i.e. demographics and employment information). It could also serve as a basis for the sampling design and data processing in a travel survey.

1.3 Information data

Information data is a type of trace data that captures people's daily activities (traveling, shopping, payment and internet safari) with the help of the usage of electronic devices (mobile phones, traffic cameras etc.) (see Tab.2).

Information data is superior to the traditional transport analysis data such as travel survey data, census and statistical data in terms of data volume, renew frequency and data variety. However, as being a trace data assembled from different operation systems, information data is still raw in its structure and content, and cannot be used directly by transport analysis. Multiple steps of data mining and processing would be needed.

Tab.1 Usages of different urban transportation survey

Survey Type	Survey Target	Common Modeling Uses of Data
Household travel surveys	Households within a pre-specified study area	Trip generation, trip distribution, mode choice, time-of-day of travel, model validation
Road traffic surveys	Vehicle-trips and person-trips on road segments	Trip distribution, model validation
External cordon surveys	Vehicle-trips and person-trips at external cordon	Trip distribution, model validation
Public transport surveys	Transit passengers	Mode choice
Taxi surveys	Taxi drivers	Taxi trips (generation, distribution, time-of-day)
Freight survey	Freight vehicle drivers, vehicle trips at freight stations	Freight vehicle trip (generation、distribution、time-of-day)
Generator surveys	Employees and visitors of pre-specified establishment, such as transportation hubs	Trip attraction models, parking cost
Parking surveys	All vehicles parked at pre-specified locations	parking cost (for mode choice), trip distribution
Visitor travel surveys	Hotel guests or visitors at pre-specified establishments	Visitor models (generation, distribution、time-of-day)

Source: reference [2−3].

Tab.2 Information data in urban travel analysis

Data Type	Content	Data Responsible Department
Mobile phone signaling data	Mobile phone traces(OIDD/WCDR)	Industry and Information Bureau, Operator
Traffic monitoring system data	Checkpoint flow data, checkpoint license plate data	Traffic Police Detachment
Highway tolling system data	Highway entrance flow data Entry and exit vehicle OD data	Highway Administration
Public transport information system data	Transit IC data Transit vehicle GPS data	Transportation Bureau, bus company
Metro operation system data	IC data Vehicle ATS operational data	Metro operation company
Taxi dispatching system data	Taxi GPS data	Transportation Bureau, taxi company
Freight vehicle GPS monitoring data	Freight vehicle GPS real-time monitoring data	Transportation Bureau
Railway waybill data	Shipping waybill data	Railroad Bureau
Highway, railway, air ticketing system data	Ticketing OD data	Transportation Bureau, Railroad Bureau, Civil Aviation Administration Bureau
Parking management system data	Parking charge data	Transportation Bureau, Traffic police detachment, parking management enterprise

Source: reference [4].

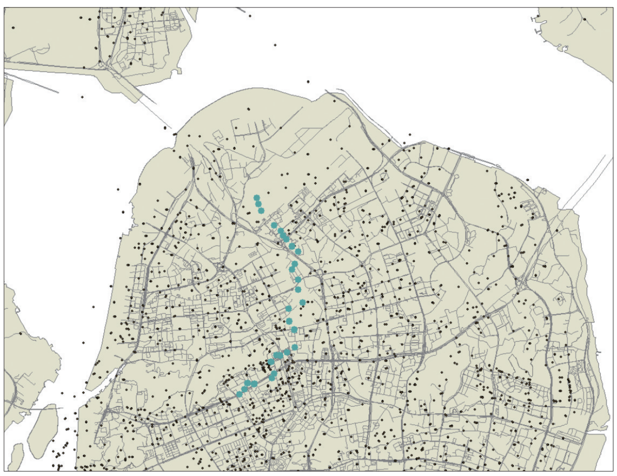
2 Typical Usage of Information Data in Urban Transportation Analysis

2.1 Analyzing urban travelling and activity characteristics using mobile phone traces

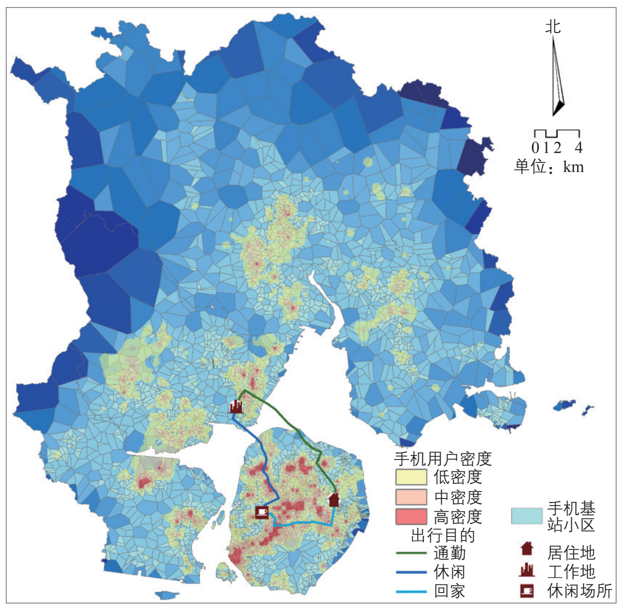
A typical data of information data is mobile phone traces, which could create a large sample containing the real-time information of individuals' space-time activities. Mobile phone traces can record the time and location (positioning precision depends on the coverage range of a base station; normally, radius of 100-500m in urban areas, 400-1000 m in rural areas) of individuals' each communication activity (text, call, data usage and location update).In average, a mobile phone per day would generate 60-80 space-time records, which is adequate for deriving an individual's complete activity trace.

An individual's residential and work places can be located in high accuracy by tracing the individual's activities in a certain period of time using mobile phone trace data (see Tab.1). Once the connection between a residential location and a workplace is captured, it is possible to develop a location matrix, which could support the in-depth analysis on commuter trips and residence-workplace relationships at various spatial scales. An individual's urban activity chain data can be derived from mobile phone traces after certain steps of data processing. The chain data consists of the location, arrival and departure times, and the purpose of each activity that an individual conducts in a day. Such a data could be used to study any urban activities (i.e. in any locations, at any times, with any purposes and by any demographic groups). It is possible to identify any special demographic groups, such as urban mobile population, by analyzing the activities conducted in multiple days and the phone number registration locations revealed by mobile phone data. As such, the mobile population can be better understood in terms of its size, origin places, living characteristics and urban activity patterns.

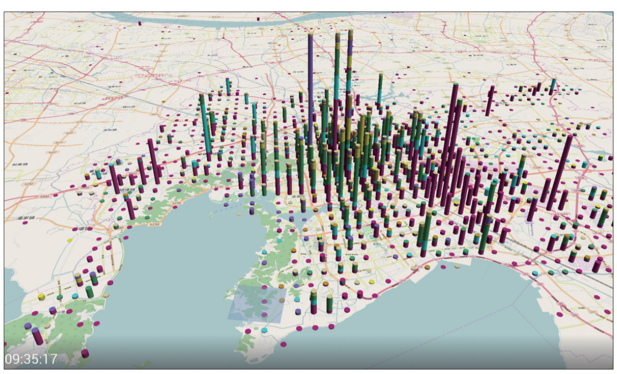
An analysis of population-specific activities at specific locations can provide insights into the range of radiation and the demand characteristics on transportation hubs, function centers and so on[5].



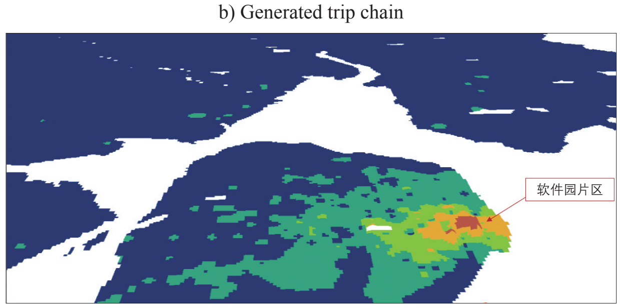
a) Mobile phone tracing point



b) Generated trip chain



c) Activity intensity analysis



d) OD analysis at pre-specified locations

Fig.1 Travel characteristics based on mobile communication data

Source: reference [4-5].

3.2 Analyzing passenger flow and system operation using bus information system data

Transit IC data records a passenger's trip information with sufficient details, including the card-swiping time, relevant line and vehicle information. Transit vehicle GPS data contains operation-related information, such as the arrival and departure times at a station. Since an individual travels continuously from one location to the next, it is possible to jointly use transit IC data and transit vehicle GPS data to generate a bus trip chain and trip OD[6-7] (see Fig.2; Tab.3 and Tab.4).

Transit vehicle GPS data space analysis technology can also be used to construct the urban bus line network space information, compute line network density, station coverage, line repetition coefficient and non-linear coefficient. Bus operation speed, actual frequencies and headways can also be obtained.

2.3 Analyzing taxi and freight vehicle operation using vehicle GPS data

The GPS data of taxi and freight vehicles share a similar structure. Both data captures a vehicle's lat-long coordinates and instantaneous speed at a particular moment. In addition, taxi status information such as occupied or non-occupied can be inferred based on taxi GPS data. The analysis of taxi GPS data could help to identify the pick-up and drop-off locations as well as the corresponding times for each ride and hence[8] the trip distance and the average speed can be derived. Many commonly used indices can also be calculated from taxi GPS data, such as number of occupied trips per day, operating mileage, non-occupied ration ratio and so on (see Fig.3). Besides, most of the taxi GPS data can be treated as floating cars and therefore can be used to analyze road traffic condition.

The analysis of freight vehicle GPS data could help to obtain the departure location, driving route and driving speed for each of the shipping tasks assigned to a vehicle. As a result, freight flow directions, urban freight transport corridor, parking lots and freight origin distribution can be acquired (see Fig.4).

2.4 Analyzing road traffic using transport detection data

Road traffic operation analysis can be based on highway toll system, data from the road gate license plate recognition system, and intersection traffic detection data. The data usually covers freeways, external highways, urban roads and interchanges, ramps and important intersections. Comprehensive analysis of the data not only can obtain the traffic information of highway external traffic, screenlines and main intersection, but can also provide insights on the characteristics of traffic flow direction and automobile OD at the entrance of a city by tracking vehicle ID.

2.5 Analyzing inter-hub passenger flow characteristics using the ticketing system data of road, rail and air transportation

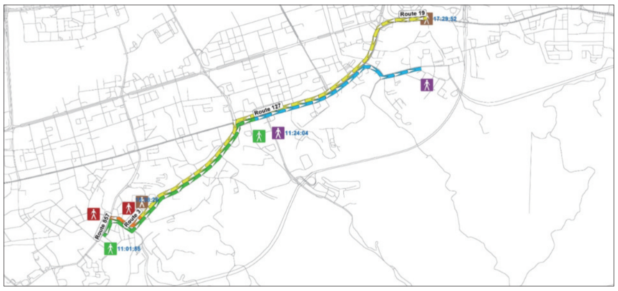


Fig.2 Bus OD analysis technology

Source: reference [4].

Tab.3 Bus operation data

			Line Network Information			Vehicle Operation Information (timetable)	
Route No.	Direction	Stop No.	Station	Longitude	Latitude	Vehicle No.	Arrival Time
857	0	16	Experimental Primary School	118084292	24457410	DZ3865	11:01:17
857	0	17	Zhongshan Park	118084916	24458839	DZ3865	11:02:59
857	0	18	No. 1 Middle School	118088411	24458708	DZ3865	11:04:03
857	0	19	174 Hospital	118092416	24461818	DZ3865	11:05:48
857	0	20	General Temple	118095395	24463655	DZ3865	11:07:21
857	0	21	Wenzao	118101092	24469189	DZ3865	11:11:47
959	1	26	Hubinzhong Road	118097355	24472170	DZ3208	11:21:46
959	1	27	Intersection of Wenyuan	118098569	24467076	DZ3208	11:24:17

Source: reference [4].

The ticketing system data of road, rail and air transportation records each individual passenger's transport carrier number, arrival and departure times, and destination information. It is approximate to population data and can provide insights on the characteristics of inter-city travelling direction, mode choice and modal split at every single hub. Traveler interview survey can offer information of socio-economic attributes and willingness to travel while bus, taxi and mobile phone data can reveal traffic distribution characteristics at hubs. A joint analysis based on such information can regenerate the traffic patterns and passenger flows at city hubs (see Fig.5).

3 Combined Usage of Multi-source Data under the Traditional Framework of Urban Transportation

Tab.4 Bus passenger OD

Category			Sequence					
			1	2	3	4	5	6
Passengers' IC card information		Card No.	0057980939	0057980939	0057980939	0057980939	0057980939	0057980939
		Swiping time	8:38:26	10:58:36	11:02:55	11:24:04	16:40:12	17:29:52
		Respective vehicle	DZ3767	Y8230	DZ3865	DZ3208	9036	1172
		Cost	0.8	0.8	0.8	0.8	0.8	0.8
TransitOD	Board	Line No.	3	87	857	959	127	19B
		Direction	0	1	1	0	1	0
		Stop name	13	12	16	26	7	17
		Arrival time	8:38:21	10:58:11	11:01:17	11:21:46	16:40:04	17:29:45
	Alight	Stop name	14	13	21	27	12	26
		Arriving time	8:40:57	11:00:43	11:11:47	11:24:17	16:55:09	17:48:25
	Transfer	Time costs/min	7.2	1.6	10.0	15.8	34.6	
		Transfer mark	1	2	2	2	3	4

资料来源：文献[4]。

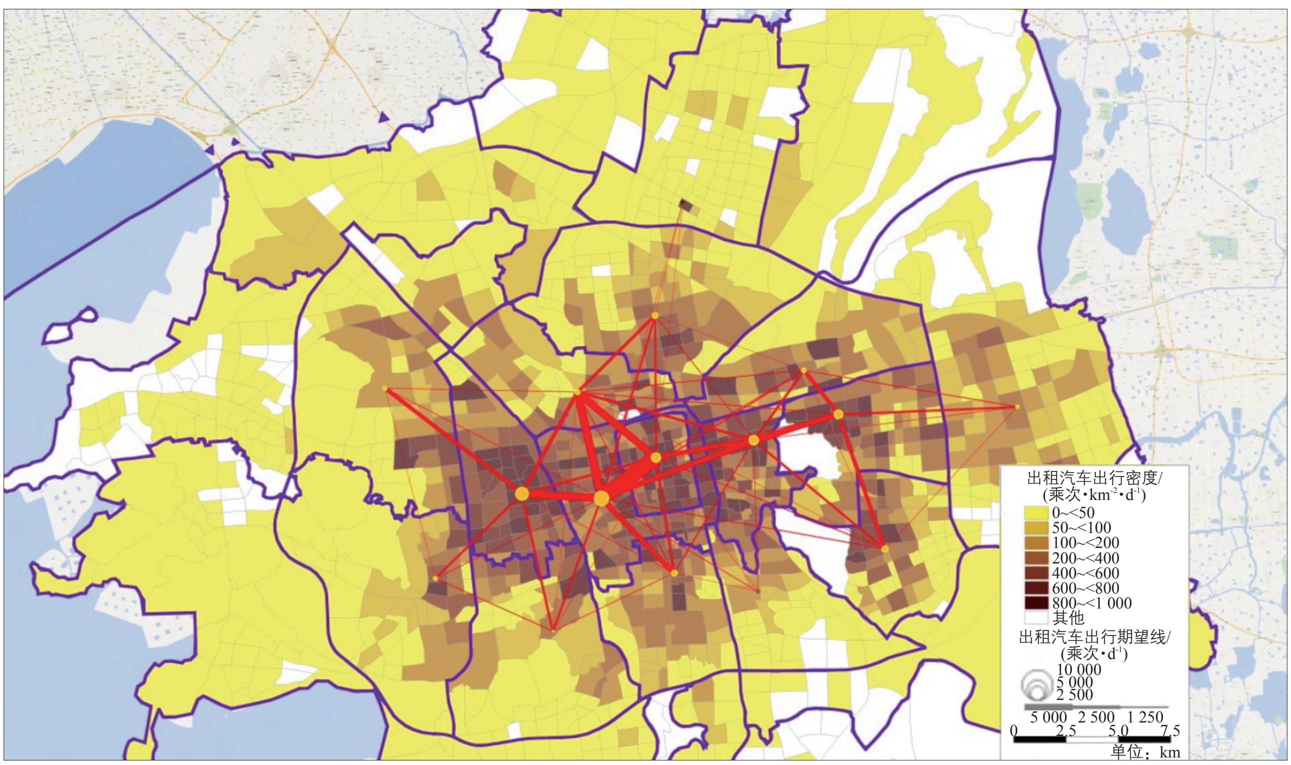


Fig.3 Characteristics of taxi passenger travel during weekday

Source: reference [8].

Analysis

3.1 Calibration across multi- dimensional data

The traditional data and information data both have their advantages and limitations. The traditional data is often collected by having sophisticated survey design. It is more effective in analyzing the characteristics and patterns of urban transportation; however, updating the data is a rather difficult task and normally requires a long interval before getting renewed in the next round of survey. For instance, the interval needed for updating urban transportation survey data, urban demographic and employment spatial distribution data (from census data) would normally be 5 years. In contrast, information data can be updated more frequently; however, its representativeness is often a problematic issue. Taking mobile phone data as an example, even if the total number of mobile phone users is close to the population of a city, it could still be a biased sample. This is because some people may not own any mobile phones while others may own two or more devices.

The demographic statistics would sometimes be different from the actual situation, especially in the newly developed areas of a city. For urban employment statistics, there would be a case that an individual's actual workplace is different from the registered address of the workplace (e.g. bus drivers). In such occasions, mobile data could be used to calibrate the original statistics, for example by identifying the true residential and working locations.

Inevitably, there could also be man-made errors in an urban travel survey. In a household travel survey, the survey participants may omit some trips that they conducted when completing a trip diary. Thus, the given trip rate in a survey would be lower than the actual trip rate. As a comparison, information data serves as a more objective data source. Specifically, mobile phone traces could be used to capture the omitted trips by revealing individuals' actual daily activity chains.

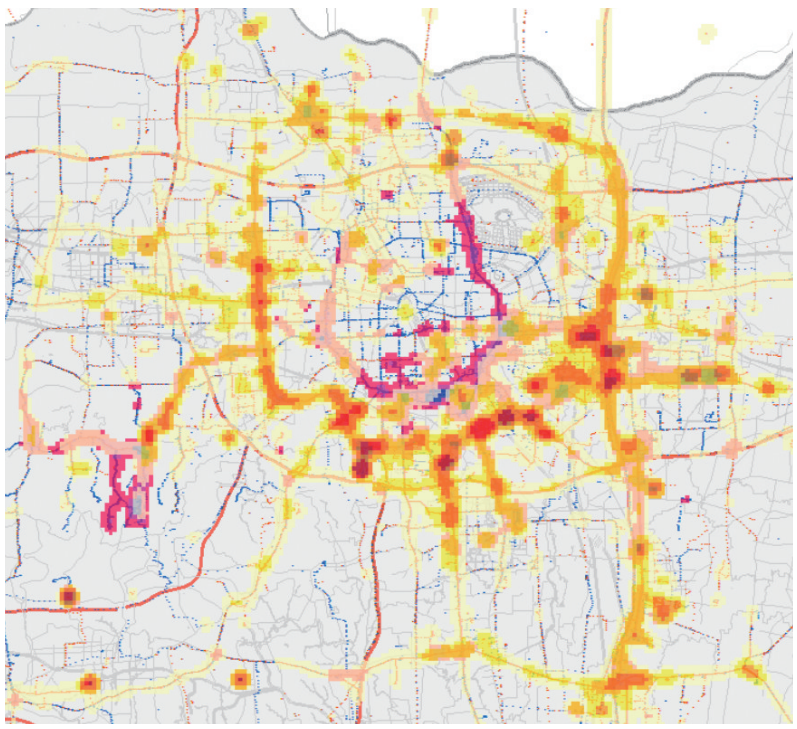
3.2 Reinforcement across multi- dimensional data

The household travel survey is the core part of an urban travel survey. It focuses on the residents' daily travelling activities (mainly short-distance trips) conducted in the urban area of a city. With the emergence of urban agglomeration and the collapse of the border lines between cities and towns, the scope of transportation analysis and modelling should extend and cover wider areas. A household travel survey can offer limited information on long-distance trips and inter-city trips; however, mobile phone data, highway toll data and ticketing system data (road, rail and air) can fill such a gap.

Urban transportation analysis and modelling have become more sophisticated. A key development lies on the tranship

a Freight flow direction

b Freight transport corridor



c Freight vehicle parking distribution

Fig.4 Freight vehicle GPS data analysis

Source: reference [7].

transport that is associated with metro stations and transport hubs. The traditional travel survey can hardly provide such information due to the constraints on survey detail and scale. However, mobile phone data, bus information system data and subway operation data can reveal more detailed trip information which would be the guidance for the modelling development to more sophisticated levels and the transport analysis at micro levels.

In the analysis and simulation for urban transportation, there could be more errors in simulating travel time and speed than in analyzing and simulating traffic volume. One of the causes is the lack of relevant data. Traffic camera data and taxi GPS data could make it possible to identify the function of flow delay with better richness and accuracy. Traditional vehicle speed survey can only offer the data on typical roads in certain circumstances. It cannot take into account the influence of different road parameters and traffic flow conditions on vehicle speed.

Under the traditional framework of transportation analysis and modelling, information data could extend further the analysis and modelling for the sub-systems. The sub-systems of bus services could be refined by using bus information system data; taxi GPS data and freight vehicle GPS data could enrich the database for the sub-systems of business vehicles. Mobile phone data could help to depict the distribution pattern and the trip characteristics of visitors, and hence improve the precision of sub-system analysis on the travelling activities conducted by visitors.

4 The Evolvement of Multi-source Data Usage

The data collection and handling techniques for information data have developed rapidly. As such, a variety of traditional travel survey projects have been replaced. For instance, urban road traffic survey data can be obtained via assembling traffic camera system data and taxi GPS data. Other travel survey projects would require much less data collection effort due to the information data becomes available. For example, the traditional transit survey can be simplified by having access to bus information system data.

Meanwhile, some electronic devices have been used in traditional travel surveys making some survey data become "informational". Many cities in China have introduced electronic devices containing address information for their household travel surveys. As a result, trip locations can be coded by their longitudes and latitudes while allowing validity check and tracing the progress of data collection. In some other household travel surveys, volunteers were recruited and were asked to carry GIS devices or install car-based GPS devices for correcting any errors, such as omitted trips, that may occur during a survey process.

Nevertheless, it is neither practical nor necessary for information data to completely replace traditional data. As the core data source for transportation analysis, household trav-

Fig.5 Characteristics of passenger flow at multimodal terminal

Source: reference [4].

el survey has not incurred any significant changes of its elements during the past 70 years, so has not the theoretical framework of urban transportation analysis. When gravity models were massively adopted in forecasting trip distribution pattern, travel survey data was only used to calibrate the frequency distribution curve (i.e. the damping function) of trip lengths (normally expressed as travel time) under different trip purposes. Due to the exclusion of spatial distribution information of trips, the required sample size of a household travel survey dropped from the original level of > 25% to <5%. When deriving a trip OD, mobile phone data could create significant travel time errors due to its special data structure and generation process. Thus, the derived OD cannot be used to calibrate a gravity model. Moreover, the detected population who possess and use mobile phones might be a biased sample to the entire population which means how to correct such a potential bias could be another key challenge. On the other hand, however, exploring the usage of information data under the traditional framework of transportation analysis would normally reduce its power. Taking mobile phone data as an example, the quantity of urban activity chain data that it derives is close to the total number of urban activities; the spatial distribution pattern of trips that it reveals cannot be captured by a standard household travel survey. Apparently, when a huge amount of travel information is available, using a growth rate model to forecast future trip distribution pattern is a superior solution which could reproduce and inherit the spatial distribution pattern of trips in reality. Hence, the true value of information data lies upon its ability to promote future innovation in the techniques of transportation analysis and modelling.

5 Conclusion

Urban transportation analysis and modelling and the acquisition of transport data are heavily dependent upon each other. The emergence of information data significantly enriched the database for transportation analysis, especially by allowing the analysis to be conducted from various perspectives. In theory, information data could validate and reinforce traditional data and such a practice would promote the development of transportation analysis and modelling to more sophisticated levels. Specifically, information data could serve as the database for the innovation in the techniques of transportation analysis and modelling. Meanwhile, the innovation in the modelling techniques would in turn lead to revolutions in transport data perse[3].

References:

[1] Pisarski A E, Alsnih R, Zmud J P, etc. Standardized Procedures for Personal Travel Surveys[R]. Washington DC: Transportation Research Board, 2008.

[2] Ministry of Housing and Urban-Rural Development the People's Republic of China. Urban Comprehensive Travel Survey Code[R]. Beijing: Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2014

[3] Wu Zixiao, Fu Lingfeng. Regulation and Innovation of Urban Comprehensive Transportation Survey[J]. Urban Transport of China, 2016，14(2): 11−17.

[4] Fu Lingfeng, Qin Wei, Liu Yan, Wu Zixiao. Urban Comprehensive Transportation Survey in Wuxi[R]. Beijing: China Academy of Urban Planning & Design, 2017

[5] Fu Lingfeng, Wu Zixiao, Liu Yan. Technical Guidance of Person Trip Survey and Bus Passenger Flow in Xiamen in 2015[R]. Beijing: China Academy of Urban Planning & Design, 2016

[6] Wu Zixiao, Fu Lingfeng. Research and Development of Transportation Planning Models in Zhengzhou: Problem & Innovation[J]. Urban Transport of China, 2012, 10(1): 26−31.

[7] Wu Zixiao, Fu Lingfeng. Development Planning of Comprehensive Transportation Planning Models in Zhengzhou[R]. Beijing: China Academy of Urban Planning & Design, 2015

[8] Zhao Yixin, Cai Runlin, Wu Zixiao. Transport Development Strategic Planning and Monographic Study in Suzhou[R]. Beijing: China Academy of Urban Planning & Design, 2015.