Research on determinants of transit ridership

Graduate School of Human-Environment Studies

Kyushu University

Author: Chen Qi

Supervisor: Prof. Zhao Shichen

Last Edited: November 20, 2018

STATEMENT

This academic dissertation is independent research work conducted under the guidance of the supervisor. Except for the quoted content, this dissertation does not contain any research result that has been published by other individuals or groups.

Name: Chen Qi
E-mail: chenqi.clever@gmail.com
QQ: 331730634

Contents

List of Figures

List of Tables

Acknowledgments

The journey of PhD has both pleasure and pain, but I deeply trust that all I went through today, either good or bad, will be the valuable life experience in the future. If mentioning the most memorable thing during my PhD, maybe I would say the process of shifting the way of thinking from a common people to a real researcher, even I can say this is the most valuable thing that I obtained after rather than the degree certificate during the more than 4 years .

I am not willing to describe the period of PhD in a suffering way, but frankly, I have to say that without the support and help from various people around me, I should have already stopped my feet underway. There are so many thanks I would like to say, please let me express my appreciation one by one.

First and foremost, I would like to thank Prof. Zhao Shichen, my supervisor, for the guidance and instruction not only in my studies but more in work and life attitude. The time of drinking with Prof. Zhao is also joyful and memorable, that chance of drinking together also shortened the distance between us students and professor, for which I appreciated a lot. It's my luck that I came to Zhao laboratory doing research, thank to Prof. Zhao again for teaching me so many things.

I would like to express my appreciation to my co-supervisors, Prof. Sakai Takeru and Prof. Divigalpitiya Prasanna, for their instruction in the designing workshop. Even though I am not working on the architecture design, the sense of design they inspired me must be a valuable wealth in the future.

I feel very grateful that I met many nice senior fellow students when I just entered our laboratory. They taught me so many tips and experiences in both daily life and research, which helped me a lot in my later days. I would like to thank my brother Liu Ze, who should also be my teacher, for giving me so many comments and advices on my research.

I would like to thank the other fellow students in our laboratory for their feedback on my papers. Communication is an important segment of research, a research without readers has no meaning. Thanks to everyone for the their patients. Especially I would like to thank Li He, Chen Siting, and Sun Xia for helping a lot in checking my messy English also in organizing the logic of my papers.

Also, I have to say thanks to the cute Japanese fellows, although not for helping me in research. They talked with me, taught me language and culture, made me adapt Japanese life soon. My period in Japan is so enrich and interesting, a large part of which can be attributed to them. In particular, I would like to say thanks to Yoshioka Daiki, Ikutake Shin, and Higuchi Go, thanks for their accompany in countless overnights.

One more thing, as we know relax is also important especially during the relatively arduous PhD period, how to spend the boring time and adjust the pace of work is always a serious problem puzzling PhD students. Dr. Han Guofeng, my ex-roommate at Kyushu University International House, and his current-roommate provided us the party place in his apartment, where we have hotpot party and even playing mahjong in leisure time. Thanks to my mahjong partners, together with them a lot of happiness is created, boring time is spent, thus accomplishing research with a better status.

The most important thing always come last, all the expense during the 4 years PhD period is sponsored by China Scholarship Council (CSC). With this money, I clearly know that my country is standing behind me, she gives me confidences to step into a new environment, to face the world. Thanks to my country.

Abstract

Urbanization caused many social problems related to urban traffic in both developing and developed countries, of which although the problems are manifested in various aspects, in essence, they can be attributed to the demand for sustainability. Giving priority to the development of public transport has been widely accepted as the effective means to deal with the problem in the sustainability of urban traffic. To decrease the share of motorized travel and increase the willingness of using public transit, a lot of measures aimed at attracting people to use public transit were proposed and implemented. However, the implementation of these measures is not as simple as building rail transit stations, then waiting for people to come. Understanding the factors influencing rail transit ridership is always central and fundamental to decisions on this issue.

With the requirement of sustainable development, the primary purpose of this research is carried out: analyzing the determinants of transit ridership, thus providing references for the coordinated planning of urban and transit. Corresponding to the topics and previous studies, the specific research questions are proposed as follow: 1. to explain walking preference to transit stations in terms of socio-Demographic characteristics thus exploring the determinants of catchment areas; 2. to explore the characteristics and trends of changes in transit ridership in terms of land use and facilities around the stations; 3. to find an approach for screening effective influencing factors; 4. to describe the relationship of transit ridership between stations and stations.

On the first question, this research examines the feature distribution of individual characteristics at several thresholds of walking duration, thus trying to explore the willingness of walking duration to transit stations. The model of random forest decision is introduced to explain the relationship from the perspective of probabil-

ity. The results showed a quantitative relationship of surveyed walking durations and passengers' individual characteristics, which is expected to be used in planning catchment areas of rail transit station or estimating the catchment areas of existing stations.

As to the second question, the annual change of transit ridership in the past ten years is investigated, also the changes in land use around the transit stations during the same period are also examined, based on which the stations are classified into 5 types. On the basis of mastering the characteristics of change in both transit ridership and land use, a preliminary analysis of the determinants of transit ridership is conducted.

The third question is answered through the exploratory regression decreasing the risk of type I and type II errors in statistic when identifying the valid variables that should be put into the regression model. The specific influence of screened valid variables is then estimated using mixed geographically weighted regression. The results gave the factors actually playing a role on influencing transit ridership, and indicate the influence quantitatively.

For the fourth question, the relationship of transit ridership between stations and stations is analyzed from the perspective of TOD considering the land use within the station catchment area. This question is converted into a binary choice problem, which is estimated by the logistic regression model. The results showed that for passengers the variations in land use in departure stations can lead to variations in the choice of destination stations.

Finally, several recommendations are proposed: 1. before estimating transit ridership a general process of screening valid factors is necessary; 2. transit station is not an independent station but one station in a system, the transit ridership should be examined from the perspective of line or network level. 3. accuracy estimation for catchment area needs further exploration.

Keywords: Urban planning; Rail transit ridership; Catchment area; Land use; Individual characteristic; Walking duration

Chapter 1

Introduction

1.1 Background

1.1.1 The requirement of sustainability caused by urbanization

With the continued urbanization, people keep on migrating from rural areas to urban areas. As a result of this, 54% population in the world lives in urban areas by 2014, this population shift will continues and is predicted to reach 66% by 2050 (United Nations & Social Affairs, 2014). Living in cities is considered much more efficient and convenient than that in rural areas in a variety of ways, as it is easier to provide services when people live closer together. However, cities also change the way that people interact with each other and the environment, which often causes multiple problems. The second UN Conference on Human Settlements in 1996 came to the conclusion that the cities all over the world are facing problems due to urbanization.

Although the problems commonly exist all over the world, the type and scale differ from different stages of urbanization. As to developing countries who are now experiencing a rapid change cased by the rapid migrants from rural areas to urban areas, the problems mainly include traffic congestion, disorganization, and pollution. Essentially, most of the problems are caused by the imbalance of demand and supply in land use, resource, and infrastructure. How to satisfy this demand is an important issue that the governments in developing countries have to consider and address. Different from the situation in developing countries, population aging and low birthrate has become the serious problems in developed countries. With a rapidly increasing proportion of aged people in the population, governments are forced to increase expenditure on social security, adding to that the low birthrate, further intensifies the shrinking of working-age population, thus increasing the fiscal burden of governments. How to reduce the public financial expenditure and improve the efficiency of social operation has become a problem that the governments in developed countries have to face.

In some ways, the problems of developing country and developed country are not exactly the same, but in essence, either of them can be viewed as the problem of sustainability. Governments in the developing countries mainly focus on trying to increase the supply of services to satisfy the increasing population, while those in the developed countries are busy coping with economic slowdown due to decentralization and changing working patterns. Under such a reality, the sustainability of urban development has become a severe topic of which the environment-friendly and efficient travel mode is widely regarded as one of the most important issues.

1.1.2 The role of public transit in sustainable development

Giving priority to public transit is one of the important ways to help create sustainable cities since it's characteristics of large-capacity, high-speed, and low-emission. Increasing the share ratio of public transit has been widely accepted as the main measure for reducing per capita energy consumption and promoting sustainability. But the planning and construction of public transit are not as simple as building a station there and waiting for passengers come. Lots of factors can influence the use of public transit, especially for factors within the catchment area of transit stations. It is generally accepted that a development of land use and facilities coordinated with the transit station plays the key role in attracting people to use the public transit. Several common elements of a coordinate development of land use and facilities have been discussed in the previous studies (Boarnet & Crane, 1997; Bernick & Cervero, 1997; Megally, Silva, & Seible, 2002; Cervero, 2004).

- Mixed-use development
- Rail transit stations as cores
- Compactness
- Pedestrian-friendly

The guidelines summarized above has received a lot of attention by the governments which are regarded as the means of mitigating most of the common urban problems, such as traffic congestion, air pollution, and incessant sprawl (Cervero, Ferrell, & Murphy, 2002). In a common sense, an area taking the station as the core typically has a central transit stop (such as a train station, or light rail or bus stop) surrounded by a high-density mixed-use area, with lower-density areas spreading out from this center. The densest areas are normally located within a radius of 400 meters to 800 meters (5 minutes to 10 minutes walking duration, varying in terms of different kinds of transit station) around the central transit stop, which are determined by the acceptable walking distance for pedestrians, thus helping solve the last mile problem.

With the increase in demand for speed, punctuality and environment protection, urban rail transit has gained popularity by governments, especially in metropolises with high compactness and population density. Add to the continuous extension of city scale and the growth in travel demand, more and more newly developed areas are planned taking rail transit station as the core. With this popularity, it is probably easy to enter a misunderstanding that if the transit station is built, people will come, but the reality is not that simple. Because of this, understanding the determinants of rail transit ridership has become central and fundamental to decisions on urban planning and management.

1.2 Research purpose

How to encourage people to take rail transit? What explains rail transit ridership? How to increase the rail transit ridership? As interpreted before, these questions is placed in front of us. The answer seems to be both obvious and complex. Every element existing in the catchment area of stations associate with ridership, population, road network, parking, income, transit network, building density etc., all the factors surely play a certain role. But the relative importance of these various factors and the internal relations among them are much more complex, and still not well understood (Taylor & Fink, 2003). Besides, as transit station is a part of transit network but not an independent existence, once the transit ridership of one station varied, transit ridership of all the other stations connected to that station should vary as well. But the interaction among catchment areas of the stations in the same transit network remains unclear to date.

We mentioned the term of catchment area above, however, what does it mean in transit ridership analysis? As a general definition, a catchment area is the area from which a city, service or institution attracts a population that uses its services. In the issue of transit ridership analysis, it means a station's primary service area, within which people are willing to use the station, and beyond which land use, travel mode choice, population distribution are unlikely to be influenced by this transit station. The catchment area is thought to be largely determined by people's willingness of walking to transit stations, which, however, vary by people's individual characteristics including trip purpose, age, gender etc. (Guerra & Cervero, 2013). Because of this, the transit catchment area is important and fundamental to the ridership analysis, even the most.

Based on the interpretation of the real issue above, the research goal of the whole dissertation is defined to ***explore the determinants of rail transit ridership***. Three main questions are to be answered in this dissertation.

1. What determines catchment area?
2. What factors explain transit ridership at station level?
3. What factors influence transit ridership among stations?

1.3 Literature review

From the research purposes, the literature review is to be extended in terms of method, influencing factor, and catchment area respectively. Table 1.1 gives a brief summary for the literature about transit ridership analysis.

1.3.1 Catchment area

An important precondition for investigating factors influencing transit ridership is the definition of the catchment area of a station, which has been interpreted in the part of research purpose. Historically, the determination of transit catchment area was mainly implemented in GIS by creating a distance buffer around the transit station (O'Neill, Ramsey, & Chou, 1992; Hsiao, Lu, Sterling, & Weatherford, 1997; Ayvalik & Jotin Khisty, 2002; Peng, Dueker, Strathman, & Hopper, 1997). How-

Table 1.1: Summary of previous studies

Year		2004	2004	2009	2010	2011	2012	2013	2013	2015
Author		Chu	Kuby et al.	Taylor et al.	Sohn and Shim	Gutiérrez et al.	Cardozo et al.	Chakraborty et al.	Zhao et al.	Jun et al.
Catchment		1/4 mile (400m) walking distance	Half mile (800m) walking distance	N/A	N/A	Distance-decay 800m buffer	800m walking distance	N/A	800m radius	300m, 600m, 900m radius
Method		Poisson Regression	WLS	2SLS	OLS, SEM	OLS	OLS, GWR	OLS, SEM	OLS	OLS, MGWR
Sample Size		2568	268	265	251	158	190	900	55	442
Number of Valid Indicator		15	11	8	7	9	4	9	11	11
Coefficient of determination (Adjusted R2)		0.54	0.71	0.91	0.6	0.73	0.56	0.69	0.95	0.77
Land use factors	Building area				•	•		•	•	•
	Hospital								•	
	School/University				•				•	
	CBD		•						•	
	Land use mix				•	•	•			•
	Other infrastructures		•						•	•
Transit-related factors	Accessibility of pedestrian	•						•		
	Accessibility of transfer		•		•	•	•	•	•	•
	Road coverage			•	•				•	
	Parking		•						•	
	Service level of public transit	•	•	•		•	•		•	•
	Locational factor		•		•				•	
Demographic and socioeconomic environment factors	Population		•	•	•		•	•	•	•
	Employment	•	•	•	•	•	•	•	•	•
	Age	•		•						•
	Tenant proportion		•							•
	Race	•	•	•		•				
	Income	•						•		
	Vehicles holdings	•					•	•		
	Fare			•						

ever, since a transit catchment area is determined by the common walking distance from the transit station, which is also called pedestrian catchment area (Abbreviated as PCA), a circle buffer area is not adequate for presenting the catchment area. Recent decades, with the development in GIS and the richness in statistics, the application scenario of GIS has progressed far beyond the simple buffering operation (Biba, Curtin, & Manca, 2010; Wu & Hine, 2003; Jiang, Zegras, & Mehndiratta, 2012). Also, many efforts have been made to estimate the real walking distance by importing the data of real road network. According to the existing studies, the PCA generally range from 400 meters to 1000 meters due to different station types, city forms, also travel preference (Alshalalfah & Shalaby, 2007; Guerra, Cervero, & Tischler, 2012; Keijer & Rietveld, 2000; Murray, Davis, Stimson, & Ferreira, 1998; O'Sullivan & Morrall, 1996; F. Zhao, Chow, Li, Ubaka, & Gan, 2003).

Since difficulties in estimating the accuracy catchment area are still not well addressed, a general 800 meters (half-mile) walking distance, which has been verified to be suitable for most cases, has been widely accepted as a principal reference of the catchment (Kuby, Barranda, & Upchurch, 2004; Gutiérrez, Cardozo, & García-Palomares, 2011; Cardozo, García-Palomares, & Gutiérrez, 2012; J. Zhao, Deng, Song, & Zhu, 2013). This 800 meters, however, is just a general distance for rail transit stations without considerations of people's individual characteristics and travel preference. The catchment area is mainly determined on the base of questionnaires by asking how far/long did passengers walk to stations, or how far/long are passengers willing to walk to stations (Keijer & Rietveld, 2000; F. Zhao et al., 2003; García-Palomares, Gutiérrez, & Cardozo, 2013), while the answers are usually given in a loose way like I remember I walked... or I think I prefer... What's more, the same reasoning has been used to justify other rail transit catchment areas and even in different cities and countries. Till now, restricted by analytical method and data, there are no better options than this general 800 meters walking distance in determining rail transit catchment area, unless doing a large-scale rail transit travel survey which is also thought to be impossible due to excessive costs.

1.3.2 Method

From the view of methodology, exploring and estimating the transit ridership can be treated as part of travel demand forecast (Miller, 1999; Boyce, Zhang, & Lupa, 1994). There are few fields in urban planning paying more attention to the statistical model for looking into the future than transportation. To date, a host of models have been developed and practiced for the issue of transit ridership, of which the most widely used are the activity-based four-step model and the direct model (McNally, 2007; Ewing & Cervero, 2010). While either of them works on travel demand forecast, they have different application scenario.

Traffic Analysis Zones (TAZs), which is a definition in four-step models, range in size from block groups to census tracts, commonly working at macro scales: corridors, subregions, and stats. The resolution of four-step models tends to be too gross to deal with the issues at neighborhood-scale (Cervero, 2006). Even though the four-step model has enjoyed widespread support from decades of use, it was never meant to predict the travel demand at neighborhood-scale taking transit stations as cores, not to mention the estimation of influencing factors on the transit ridership (Cervero, 2006; Chu, 2004; Duduta, 2013). Additionally, in the four-step approach, trip generation stage is usually conducted based on an empirical model using some socioeconomic variables like population, employment, auto ownership, however, in different city cases, the impact of traffic indicators will not be exactly the same (Jones & Nichols, 1983). It is also unclear whether these selected indicators are significantly related to traffic volume, or whether there is an indeed linear relationship between traffic volume and these selected indicators.

Direct models estimates ridership as a function of station environments and transit service features based on observed ridership and statistical data (Cervero, 2006). With the development of geographic information system (GIS) technology and the richness of digital statistic, direct models have gained popularity in estimating travel demand at neighborhood-scale. The advantages of direct models firstly stem from the ease of estimation with data that are readily available to transit agencies. The only critical requirements are GIS, statistics for built-environments around stations, and the corresponding transit ridership (Guerra et al., 2012). The advantages are also reflected in the accuracy of prediction, that much easier and faster than other travel demand models, direct models can also provide strong predictive power (Lane, DiCarlantonio, & Usvyat, 2006). Compared with the disadvantage mentioned in four-step models, since the estimation in direct models is based on the historical statistics, it can justify the importance and validity of indicators that are expected to influence transit ridership (Walters & Cervero, 2003).

1.3.3 Influencing factor

Transit ridership is generally thought to be related to land use, transportation environment, or travel preferences (Thompson, 1997). In 1997, Cervero proposed a three dimension index system (Density, Diversity, and Design) to examine the ridership of transit (Cervero & Kockelman, 1997), which has been generally accepted as a basic principle. In addition, many extensions have also been added to the 3D theory, such as accessibility to the station, connectivity of line, and capacity of station (Beimborn, Greenwald, & Jin, 2003; García-Palomares et al., 2013). In this study, all the candidate factors expected to influence transit ridership can be classified into three main categories: a. land use factors; b. transit-related factors; c. demographic and socioeconomic environment factors.

Land use includes the buildings or facilities that provide the setting for human activity, and it has been widely proved to have a strong relationship with ridership. Also, land use diversity has a significant effect on ridership since it reflects the balance between traffic demand and supply within the catchment area. Although the definitions of land use diversity are not the same according to different researchers, it is widely accepted that higher diversity tends to result in less transit ridership (Cardozo et al., 2012; Choi, Lee, Kim, & Sohn, 2012; Gutiérrez et al., 2011; Jun, Choi, Jeong, Kwon, & Kim, 2015; Sohn & Shim, 2010; Sung, Choi, Lee, & Cheon, 2014).

Transit-related factors are important for passengers going to take public transit. Better accessibility is thought to be attractive for passengers living further. The factors for accessibility are commonly described as the number of transfers, network density, number of parking facilities and walking convenience (Kuby et al., 2004; Sohn & Shim, 2010; Taylor, Miller, Iseki, & Fink, 2009; J. Zhao, Deng, Song, & Zhu, 2014; Chu, 2004). Also, the type and location of a station can affect accessibility as well. Terminal stations are more attractive for passengers because people can accept to spend more time on getting to a terminal station which is easier to transfer to other line or another mode of transportation (O'Sullivan & Morrall, 1996).

The demographic and socioeconomic environment is an important factor which can reflect the travel preference. Obviously, the resident population and employment-population within the catchment area are crucial factors on ridership of the subway station. Besides, the economic factors also play an important role in ridership. For example, in the area where the car ownership is higher, people are more likely to choose private car than public transit (Chiou, Jou, & Yang, 2015; F. Zhao, Chow, Li, & Liu, 2005); also the higher the percentage of low-income household is, the more likely people tend to take public transit (Thompson, Brown, & Bhattacharya, 2012). Furthermore, the ratio of apartments and rental house within catchment have been verified being relevant with ridership in some degree (Jun et al., 2015).

1.4 Primary research questions

A lot of progress has been made in the research field of transit ridership forecasting, however, some work is still inadequate in the specific problems of the catchment area, model, and influencing factors. Till now, even though there have been a lot of research and investigation on catchment areas, the key problems are still not well addressed. The only thing that can be determined is a general scope of catchment area which is from 400 meters to 1000 meters, but this scope varies too much in terms of different cases. A comprehensive and systematical methodology for estimating catchment area is needed, either the research perspective or the methodology needs to be further discussed. For the topic of forecasting rail transit ridership, most studies are keeping on mining the problems at station level, but rarely exploring the problems from perspectives of rail transit lines and networks. Specifically, in terms of the model, little progress has been made in the direct model, most existing studies are conducted based on the improvement of regression model but without essential progress. As to influencing factors, scholars have done a thorough job of summary and classification, nevertheless factors behave differently in different cases, the approaches for selecting valid factors towards specific cases are still insufficient.

On the base of fully understanding the existing research, this dissertation will be expanded from the three specific research points corresponding to the research purposes mentioned before.

1. Exploring the relationship between walking distance and walking preference, and trying to explain the catchment area.
2. Analyzing the characteristics in the variation of transit ridership.
3. Exploring the approach for screening effective influencing factors.

4. Shifting the research perspective, trying to describing the relationship between stations and stations.

1.5 Dissertation organization

The dissertation uses the rail transit system of Fukuoka City, which is the sixth largest city in Japan having a more than 1.5 million population, as the study case. The rail transit system of Fukuoka has 78 rail transit stations, including 27 JR Kyushu stations, 35 subway stations, and 16 West Japan Railway stations. Figure 1.1 presents the research area and the distribution of rail transit stations. The rail transit system of Fukuoka carries a daily average of more than 0.6 million passengers by 2015 that accounting for about 20% in total motorized travel (Ministry of Land, Infrastructure n.d.). The rail transit system of Fukuoka is not a large-scale one, the share ratio of rail transit is staying at a lower level. Like many local central cities in Japan, Fukuoka city also has the problems in urban traffic congestion and financial press on public transit due to the tendency of using private cars and the huge operating costs on the subway system. Improving subway ridership has been an urgent demand for the Fukuoka government.

The content of this dissertation is arranged as follow. Figure 1.2 gives a brief flowchart of the organization.

- ***Chapter 1*** gives a general introduction to the entire logic and flow with the support of a comprehensive literature review.
- ***Chapter 2*** explores the correlation between surveyed walking duration and people's individual characteristics, thus trying to explain the catchment area.
- ***Chapter 3*** describes the characteristics of transit ridership and land use, and explores the overall relationship between them.
- ***Chapter 4*** proposes an approach for selecting the valid indicator used for explain transit ridership, and makes some improvement in the model.
- ***Chapter 5*** explains the influence of land-use patterns on transit ridership at station-to-station level.
- ***Chapter 6*** makes a conclusion for the three research point, and prospects the

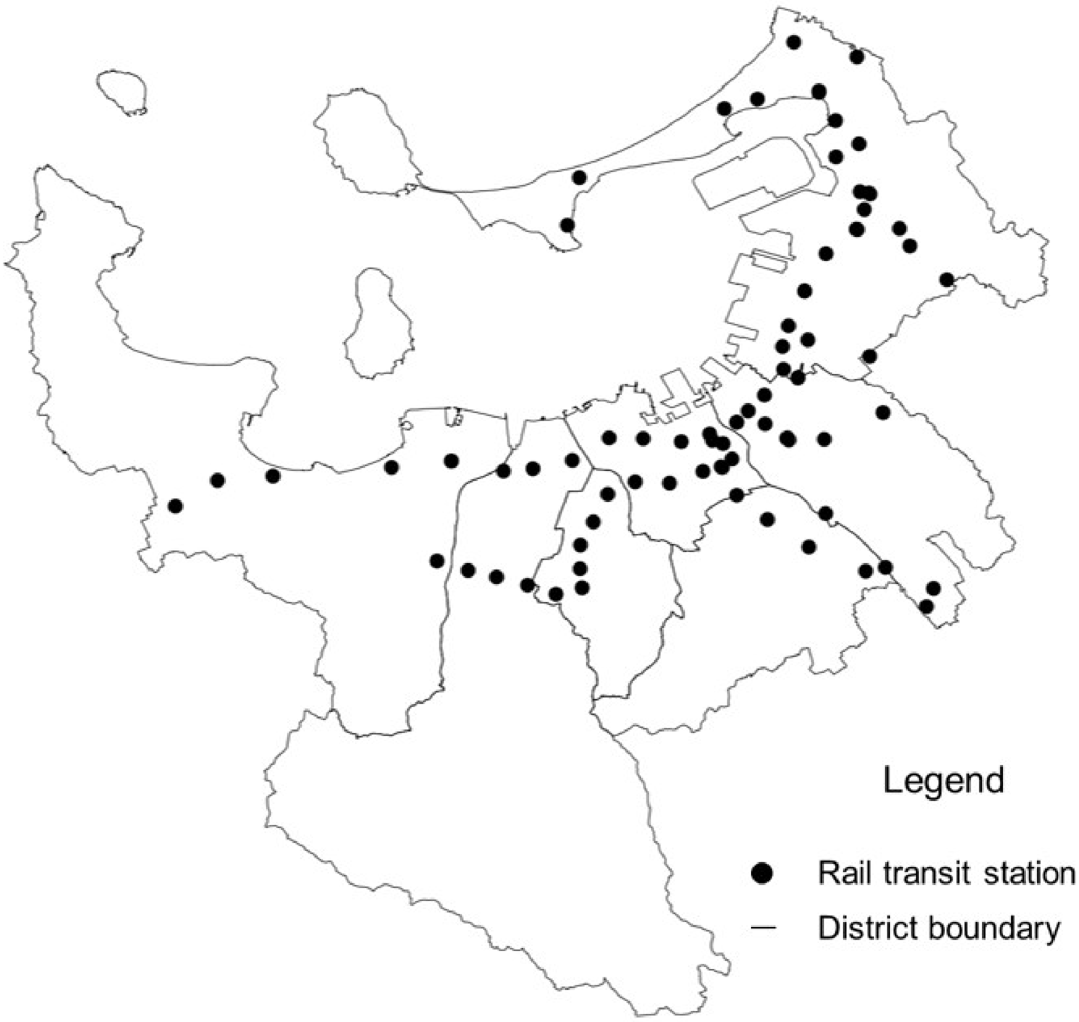


Figure 1.1: Research Area

next stage of this research.

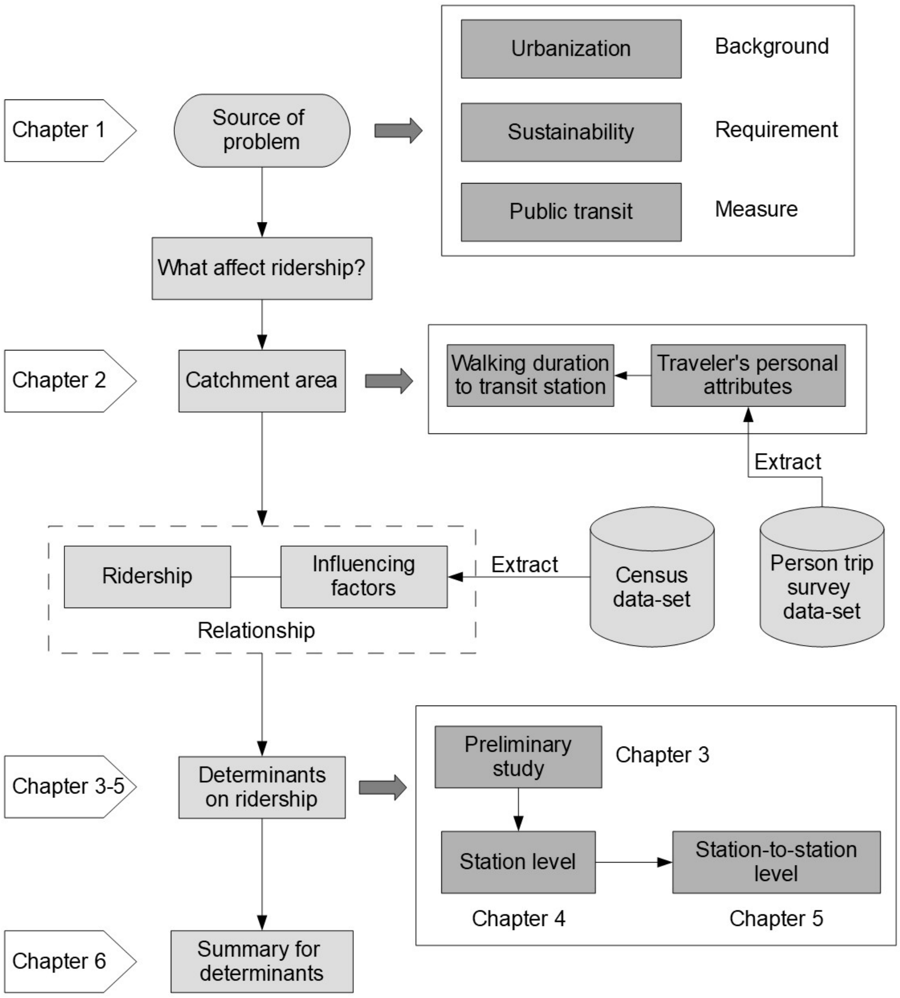


Figure 1.2: Flowchart of the Dissertation

References

Alshalalfah, B., & Shalaby, A. S. (2007). Case study: Relationship of walk access distance to transit with service, travel, and personal characteristics. *Journal of urban planning and development*, *133*(2), 114–118.

Ayvalik, C., & Jotin Khisty, C. (2002). Heuristic analysis of impacts of commuter rail station consolidation on pedestrian access. *Transportation Research Record: Journal of the Transportation Research Board*(1793), 47–54.

Beimborn, E., Greenwald, M., & Jin, X. (2003). Accessibility, connectivity, and captivity: impacts on transit choice. *Transportation Research Record: Journal of the Transportation Research Board*(1835), 1–9.

Bernick, M., & Cervero, R. (1997). *Transit villages in the 21st century*.

Biba, S., Curtin, K. M., & Manca, G. (2010). A new method for determining the population with walking access to transit. *International Journal of Geographical Information Science*, *24*(3), 347–364.

Boarnet, M., & Crane, R. (1997). La story: A reality check for transit-based housing. *Journal of the American Planning Association*, *63*(2), 189–204.

Boyce, D. E., Zhang, Y.-F., & Lupa, M. R. (1994). Introducing "feedback" into four-step travel forecasting procedure versus equilibrium solution of combined model. *Transportation Research Record*, *1443*, 65–74.

Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, *34*, 548–558.

Cervero, R. (2004). *Transit-oriented development in the united states: Experiences, challenges, and prospects* (Vol. 102). Transportation Research Board.

Cervero, R. (2006). Alternative approaches to modeling the travel-demand impacts of smart growth. *Journal of the American Planning Association*, *72*(3), 285–295.

Cervero, R., Ferrell, C., & Murphy, S. (2002). Transit-oriented development and joint development in the united states: A literature review. *TCRP research results digest*(52).

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219.

Chiou, Y.-C., Jou, R.-C., & Yang, C.-H. (2015). Factors affecting public transportation usage rate: Geographically weighted regression. *Transportation Research Part A: Policy and Practice*, *78*, 161–177.

Choi, J., Lee, Y. J., Kim, T., & Sohn, K. (2012). An analysis of metro ridership at the station-to-station level in seoul. *Transportation*, *39*(3), 705–722.

Chu, X. (2004). *Ridership models at the stop level* (Tech. Rep.). National Center for Transit Research, University of South Florida.

Duduta, N. (2013). Direct ridership models of bus rapid transit and metro systems in mexico city, mexico. *Transportation Research Record: Journal of the Transportation Research Board*(2394), 93–99.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American planning association*, *76*(3), 265–294.

García-Palomares, J. C., Gutiérrez, J., & Cardozo, O. D. (2013). Walking accessibility to public transport: an analysis based on microdata and gis. *Environment and Planning B: Planning and Design*, *40*(6), 1087–1102.

Guerra, E., & Cervero, R. (2013). Is a half-mile circle the right standard for tods?

Guerra, E., Cervero, R., & Tischler, D. (2012). Half-mile circle: Does it best represent transit station catchments? *Transportation Research Record: Journal of the Transportation Research Board*(2276), 101–109.

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography*, *19*(6), 1081–1092.

Hsiao, S., Lu, J., Sterling, J., & Weatherford, M. (1997). Use of geographic information system for analysis of transit pedestrian access. *Transportation Research Record: Journal of the Transportation Research Board*(1604), 50–59.

Jiang, Y., Zegras, P. C., & Mehndiratta, S. (2012). Walk the line: station context, corridor type and bus rapid transit walk access in jinan, china. *Journal of Transport Geography*, *20*(1), 1–14.

Jones, I. S., & Nichols, A. J. (1983). The demand for inter-city rail travel in the united kingdom: some evidence. *Journal of transport economics and policy*, 133–153.

Jun, M.-J., Choi, K., Jeong, J.-E., Kwon, K.-H., & Kim, H.-J. (2015). Land use characteristics of subway catchment areas and their influence on subway ridership in seoul. *Journal of Transport Geography*, *48*, 30–40.

Keijer, M., & Rietveld, P. (2000). How do people get to the railway station? the dutch experience. *Transportation Planning and Technology*, *23*(3), 215–235.

Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the united states. *Transportation Research Part A: Policy and Practice*, *38*(3), 223–247.

Lane, C., DiCarlantonio, M., & Usvyat, L. (2006). Sketch models to forecast commuter and light rail ridership: Update to tcrp report 16. *Transportation Research Record: Journal of the Transportation Research Board*(1986), 198–210.

McNally, M. G. (2007). The four-step model. In *Handbook of transport modelling: 2nd edition* (pp. 35–53). Emerald Group Publishing Limited.

Megally, S. H., Silva, P. F., & Seible, F. (2002). *California department of transportation* (Tech. Rep.).

Miller, H. J. (1999). Potential contributions of spatial analysis to geographic information systems for transportation (gis-t). *Geographical Analysis*, *31*(4), 373–399.

Murray, A. T., Davis, R., Stimson, R. J., & Ferreira, L. (1998). Public transportation access. *Transportation Research Part D: Transport and Environment*, *3*(5), 319–328.

O'Neill, W. A., Ramsey, R. D., & Chou, J. (1992). Analysis of transit service areas using geographic information systems. *Transportation Research Record*(1364).

O'Sullivan, S., & Morrall, J. (1996). Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*(1538), 19–26.

Peng, Z.-R., Dueker, K. J., Strathman, J., & Hopper, J. (1997). A simultaneous route-level transit patronage model: demand, supply, and inter-route relationship. *Transportation*, *24*(2), 159–181.

Sohn, K., & Shim, H. (2010). Factors generating boardings at metro stations in the seoul metropolitan area. *Cities*, *27*(5), 358–368.

Sung, H., Choi, K., Lee, S., & Cheon, S. (2014). Exploring the impacts of land use by service coverage and station-level accessibility on rail transit ridership. *Journal of Transport Geography*, *36*, 134–140.

Taylor, B. D., & Fink, C. N. (2003). The factors influencing transit ridership: A review and analysis of the ridership literature.

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? analyzing the determinants of transit ridership across us urbanized areas. *Transportation Research Part A: Policy and Practice*, *43*(1), 60–77.

Thompson, G. (1997). Achieving suburban transit potential: Sacramento revisited. *Transportation Research Record: Journal of the Transportation Research Board*(1571), 151–160.

Thompson, G., Brown, J., & Bhattacharya, T. (2012). What really matters for increasing transit ridership: Understanding the determinants of transit ridership demand in broward county, florida. *Urban Studies*, *49*(15), 3327–3345.

United Nations, D. o. E., & Social Affairs, P. D. (2014). *World urbanization prospects: The 2014 revision, highlights*. United Nations.

Walters, G., & Cervero, R. (2003). Forecasting transit demand in a fast growing corridor: The direct-ridership model approach. *Fehrs and Peers Associates*.

Wu, B. M., & Hine, J. P. (2003). A ptal approach to measuring changes in bus service accessibility. *Transport Policy*, *10*(4), 307–320.

Zhao, F., Chow, L., Li, M., & Liu, X. (2005). A transit ridership model based on geographically weighted regression and service quality variables. *Lehman Center for Transportation Research, Florida International University, Miami, Florida. http://lctr. eng. fiu. edu/re-project-link/finalDO97591_BW. pdf (accessed December 12, 2010)*.

Zhao, F., Chow, L.-F., Li, M.-T., Ubaka, I., & Gan, A. (2003). Forecasting transit walk accessibility: Regression model alternative to buffer method. *Transportation Research Record: Journal of the Transportation Research Board*(1835), 34–41.

Zhao, J., Deng, W., Song, Y., & Zhu, Y. (2013). What influences metro station ridership in china? insights from nanjing. *Cities*, *35*, 114–124.

Zhao, J., Deng, W., Song, Y., & Zhu, Y. (2014). Analysis of metro ridership at station level and station-to-station level in nanjing: an approach based on direct demand models. *Transportation*, *41*(1), 133–155.

Chapter 2

Analyzing Willingness of Walking Duration to Transit Stations Using Socio-Demographic Characteristics

2.1 Introduction

2.1.1 Background

At present, the 800 meters (half-mile) walking distance has been widely accepted as a principal reference of the catchment area for the planning of TOD (Kuby, Barranda, & Upchurch, 2004; Gutiérrez, Cardozo, & García-Palomares, 2011; Cardozo, García-Palomares, & Gutiérrez, 2012; J. Zhao, Deng, Song, & Zhu, 2013). Planners and researchers also use rail transit catchment areas to make prediction of ridership. However, this 800 meters walking distance is loosely obtained from the sampling survey by asking how far people are willing to walk to rail transit stations, and the same reasoning has been used to justify other rail transit catchment areas and even in different cities and countries. As to passengers, the acceptable walking distance/duration should not only relate to the features of that, for example, the scale of stations and land use around stations, but also relate to the features of passengers' individual characteristics, such as occupations and trip purposes.

2.1.2 Research objective

Rail transit provides a cheaper and environment-friendly way of transportation to passengers. Like any other service and commodities, it needs to face the market and should be transacted at the cost acceptable to consumers. As a result, the distribution of transaction cost can reflect the consumption-ability of consumers in the market. Obviously, it is important for rail transit operators to know how much the cost that consumers are willing to pay, based on which thereby making rail transit more attractive. But a little different from general service and commodities, for passengers, the cost refers to the convenience rather than the fare, because the cheap enough fare is already not the important element for passengers to decide whether using rail transit, while the walking duration to stations become the important determinant.

In the case of rail transit, before a potential passenger makes the decision of walking to a station, the expected walking duration, which is the ”price” for this potential passenger, is just the reflection of the distance between departures and stations; while if this expected walking duration can be accepted by this potential passenger, this trip will happen and be surveyed. Based on the analogy given before, it’s easy to think of that people with different individual characteristics generally should have different willingness towards walking duration to stations, and the surveyed walking durations can be viewed as the transaction price which have been accepted by passengers. It follows that the willingness of using rail transit will increase if the expected walking duration is less than that of acceptant range, otherwise, the willingness will decrease, which reflected in the survey data will be that the records with shorter walking duration are more than that with longer walking duration.

2.1.3 Research purpose

As interpreted above, although the walking duration just presents the distance between departures and stations, it should have relation with individual characteristics. Thereby, this study uses Fukuoka, Japan, as the case to explore this relation, trying to explore how long passengers with have different individual characteristics are willing to take for walking to stations. The sample mainly includes people’s individual characteristics and trip chain information, for example gender, age, occupation, departures, destinations and travel time. This paper is constructed as follow: section two reviews the previous research on the walking duration/distance to transit station; section three describes the data used in this study, and presents methods and necessary assumptions; section four and five give the process of analysis; section six discusses the results and draws conclusions.

2.2 Review

This section reviews the literature on the issue of walking duration/distance to rail transit stations, some problems that are still not well addressed are summarized. The review is arranged into two parts, which are the disposal of the variable of walking duration/distance, and the effect of influencing factors.

For the walking duration / distance, which is the research object in this study, there is always a difficult point in obtaining the accuracy walking duration/distance by questionnaire because of the discrepancy between perceived values and objective values. Also, the observed walking duration/distance is not the reflection of how long people are willing to spend on walking to rail transit stations, but only the walking duration/distance between departures and stations. For this problem, some studies chose a different perspective trying to explain the walking duration/distance by introducing a threshold of walking duration (Besser & Dannenberg, 2005; McCormack, Cerin, Leslie, Du Toit, & Owen, 2008). They examined the differences in the distribution of influencing factors in terms of the specific threshold of walking duration. Indeed, using a threshold can decrease the discrepancy between observation and reality in some extent, but this disposal also brought some new problems in, for example, it may lead to a great loss of information in the raw data, and it is also difficult to decide the threshold of walking duration/distance.

Moreover, another point for this issue is whether to choose walking distance or walking duration as the threshold. To date, there are a lot of studies working on the relationship between walking distance and passengers' individual characteristics, some of them argued that an individual has a limited amount of time spending on traveling during a day, people tend to accept further walking distance as the speed of travel increases (Marchetti, 1994; Larsen, El-Geneidy, & Yasmin, 2010). With this reasoning it is easy to think of that passengers with the same individual characteristics may have the similar willingness of walking duration, but they generally have different willingness of walking distance due to different travel speed. This is also the reason why this study chooses the walking duration as the research object.

Among the influencing factors for walking durations, passengers' individual characteristics are generally thought to be the key that can affect walking distance (Besser & Dannenberg, 2005; Weinstein Agrawal, Schlossberg, & Irvin, 2008; Krygsman, Dijst, & Arentze, 2004; Yang & Diez-Roux, 2012; Daniels & Mulley, 2013; Guerra, Cervero, & Tischler, 2012), whereas, there are few studies having clearly verified the relationship between individual characteristics and walking distance/duration, even there is a study suggesting the walking distance should not be viewed as a function of socio-demographic characteristics (Krygsman et al., 2004). Several studies have confirmed the role of travel purposes in determining walking distance, the commute trip showed particularity from the other purposes, people with the purpose of commute tend to walk a longer distance to rail transit stations (Larsen et al., 2010). However, the definite relationship between trip purposes and walking distance is still unclear. The situation is the same with other categories of factors, such as the factors of transportation environment, land use, and willingness of passengers (Guerra et al., 2012; Krygsman et al., 2004; Weinstein Agrawal et al., 2008). The only thing that has been confirmed to date is that the walking distance/duration can be influenced by some specific kinds of factors, such as socio-demographic characteristics, trip purposes, and built-environment, but the problem is how and to what degree the walking distance/duration can be influenced.

It may be because of the problems in the disposal of walking distance/duration or the selection of influencing factors, most of the existing studies did not find the significant relation between the walking distance/duration and the influencing factors. The studies working on the qualitative description for the distribution of walking distance accounted for the majority although some of the existing studies applied regression model on this issue, the expected results were not obtained. To avoid the problems mentioned in the review, this study uses the thresholds of walking duration as the research object; for any given threshold of walking duration, the respondent who gives the answer greater than the given threshold can be viewed as this respondent can accept this threshold of walking duration; the surveyed walking duration is the reflection of passengers' willingness for using rail transit. For the analytical method, since various types of regression model has been verified unsuitable for this issue, instead of finding the linear relationship between dependent and independent variables, this study introduces the approach of machine learning into this issue, and try to explain the relation between walking durations and individual characteristics from the view of probability.

2.3 Data

2.3.1 Case introduction

All the rail transit stations and rail transit passengers within the urban area of Fukuoka are investigated as the research objective. The dataset is extracted from the Northern Kyushu Area Person Trip Survey, which is conducted about every 12 years, the latest available data is from the 4th survey surveyed by the year of 2005, and the 5th survey is already in preparation from September 2017. Figure 2.1 shows the research area and the distribution of rail transit stations. By the year of 2005 (the 4th Northern Kyushu Area Person Trip Survey was conducted), there are more than 70 stations located within the city area of Fukuoka, of which the number of JR Kyushu station is 27, Fukuoka Subway station is 35, and West Japan Railway station is 16. Now some new rail transit lines and stations are still under planning and construction.

2.3.2 Data description

The main purpose of person trip survey is to know the travel trends, thus making a better living environment and providing support for traffic planning. The original data covered the range of all the main cities in Northern Kyushu Area, which has more than 483,000 records of trip chaining behavior. The available data in this study mainly includes trip chaining behavior and socio-demographic characteristics, as shown in the Table 2.1.

To analyze the walking duration between departures and rail transit stations in Fukuoka, the first step is to extract the valid records of rail transit trip within the city area of Fukuoka from over 480,000 records in the dataset. The procedure of extracting the valid data is divided into 3 steps. Firstly, extracting all the person trip data that surveyed within the city area of Fukuoka; secondly, selecting the trip

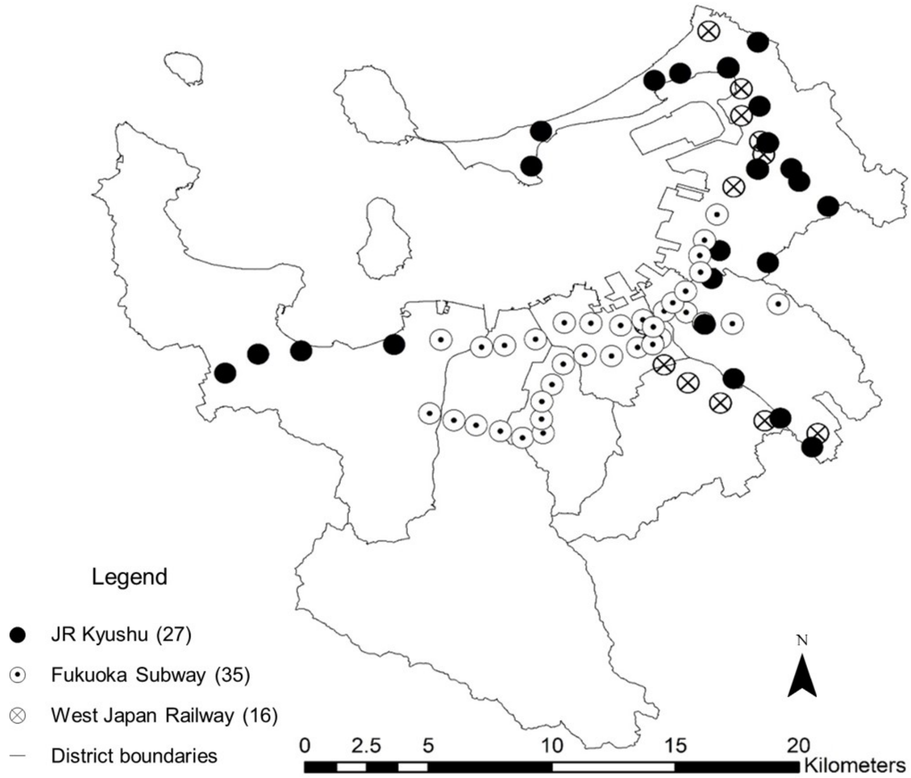


Figure 2.1: Distribution of transit stations

Table 2.1: Available data contents

Category	Feature
Trip chaining behavior	Departure location Departure time Destination location Arrival time Transport modes Time spent for each mode Location of bus stop or rail transit station
Socio-demographic attributes	Age Sex Occupation Trip purpose Vehicle/License holding Address

chaining behavior which contains the rail transit mode; thirdly, filtering the invalid data that with null value and abnormal value. The procedure of data cleaning is shown in Figure 2.2, at last the valid dataset is reduced to a size of 4,254 trips.

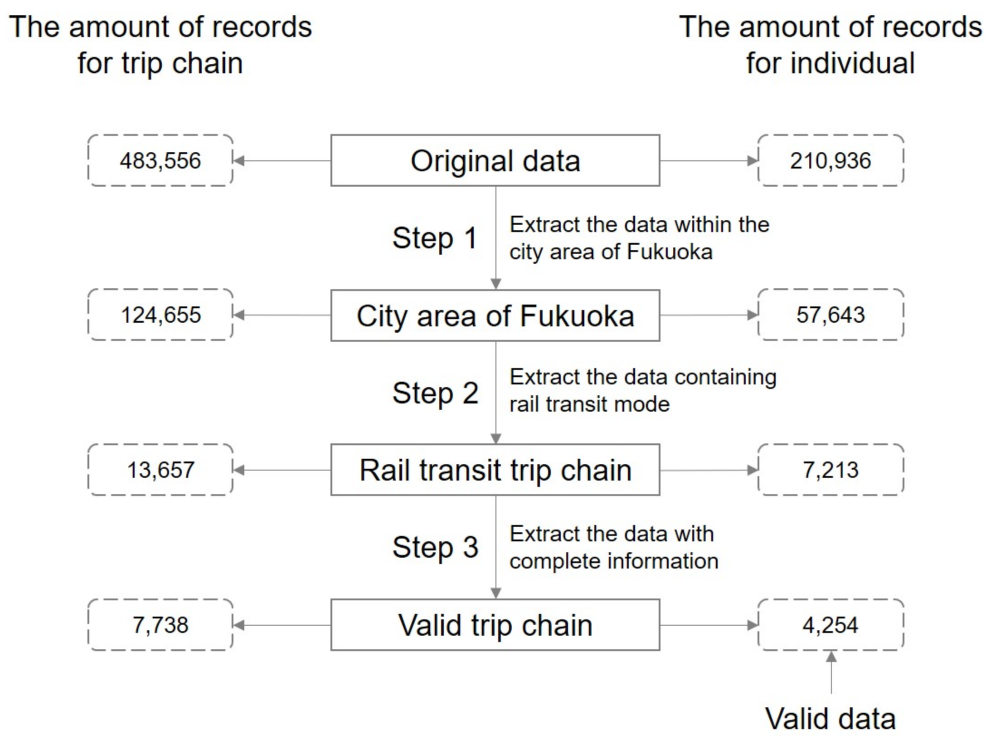


Figure 2.2: Process of data cleaning

Figure 2.3 shows the age distribution for walking trips to rail transit stations based on the finally valid dataset. The passengers aged from 25 to 55 account for the majority of the whole passengers, while schoolchildren aged under 15 rarely take rail transit. The distribution graph does not show significant peak values at any specific age group. Figure 2.4 shows the distribution of real walking durations to stations, it has a mean value of 8.32, and the standard deviation is 4.63. Notably, there are several peak values at the time of 5 multiples. It is speculated that the peak values may be caused by deviation occurred in the investigation. Since people's feeling about the specific time or number is inaccurate, they are inclined to reply a loose answer when they are asked some questions about the details of walking duration. This inclination will count some of the real walking duration that is near to 5 multiples as the 5 multiples and finally expressed in the result of investigation. Despite the bias between survey and reality, as the second assumption proposed before, passengers are viewed that they can accept the walking duration what they answered, therefore, the peak values are thought available in this study.

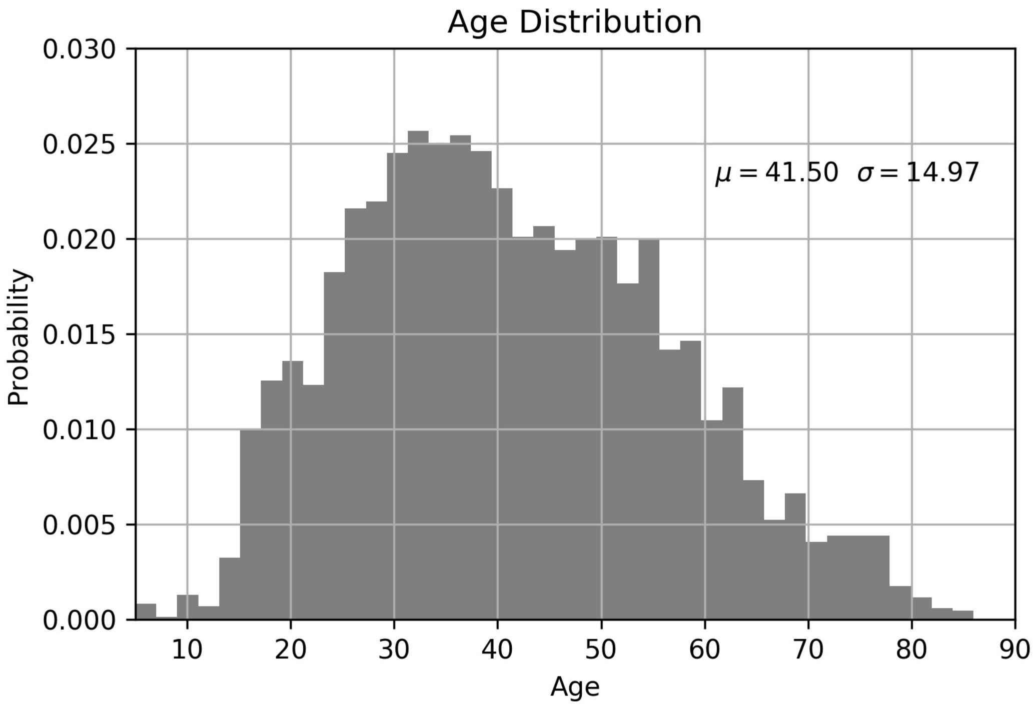


Figure 2.3: Age distribution of passengers

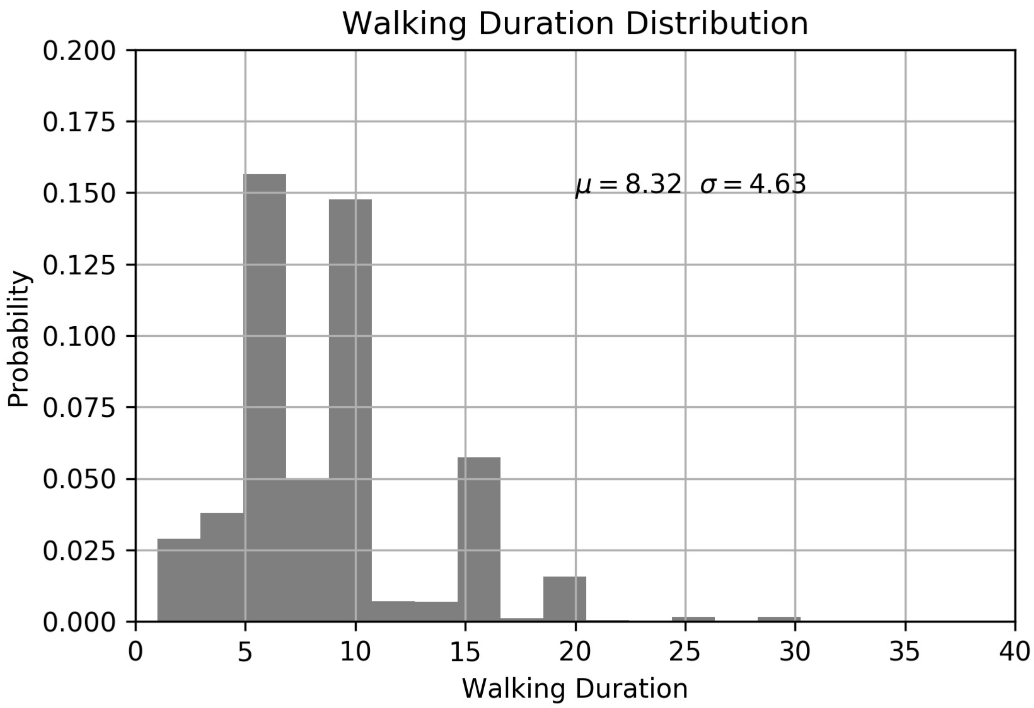


Figure 2.4: Distribution of walking duration

2.4 Methods

2.4.1 Necessary hypothesis

Based on the statements outlined in the parts of introduction and review, three basic assumptions are proposed in this study.

- A1: The distribution of departures and destinations of people with different individual characteristics in Fukuoka, Japan, is random in space.
- A2: The acceptant walking duration of people with same individual characteristics should subject to the normal distribution.
- A3: The respondents are viewed as they can accept the walking duration that they answered.

Based on A1 and A2, when examining the threshold of walking duration t, set the proportion of k group in the whole surveyed sample as r_k, the proportion of people whose walking durations are under the given threshold of t is marked as $r_k^{<t}$, the proportion of people whose walking durations are over the given threshold of t minutes is marked as $r_k^{>t}$. If individual characteristics have no significant correlation with acceptant walking durations, there should be no significant differences among r_k, $r_k^{<t}$, and $r_k^{>t}$; otherwise, the three proportion should show significant differences, and the differences will show regularities at different threshold t.

Based on H3, if someone gives the answer t minutes, it means this respondent can accept the walking duration of t minutes and any walking duration that less than t minutes. Indeed, this respondent perhaps can accept the walking duration over he answered, but this study just concern about if he can accept the given threshold of the walking duration other than how long he can accept.

2.4.2 Random forest decision model

In this study, the variable of walking duration to the rail transit station is a continuous one, while the variables of individual characteristics are multi-categorical, for which the multiple linear regression is considered not applicable. Although there were some studies using the multiple linear regression to estimate the walking duration, the result showed no clear relationship (Krygsman et al., 2004). Since what we concern with is the feature distribution of individual characteristics over or under the thresholds of walking durations, there is no need to find the linear relationship between the individual characteristics and walking durations.

According to the assumptions proposed before, when giving a specific threshold of walking duration, passengers walking longer or less than this threshold should have different feature distributions of individual characteristics. That is, if the walking duration of a passenger is more than a given threshold of walking duration, it can be considered that this given threshold of walking duration is an acceptant one, and this passenger should have a relatively high probability of choosing to use rail transit at any threshold of walking duration that less than that given one. This willingness of using rail transit can be reflected by the distribution of walking duration in terms of passengers' individual characteristics, based on which this study examines the relation of passengers' individual characteristics and this distribution. The abstract model describing this relation is expressed as follow, of which Equation 2.1 gives the value of Y_i^T, and Equation 2.2 is the probability that $Y_i^T = 1$ at given vector of X_i.

$$\begin{cases} Y_i^T = 1, & (t_i > T) \\ Y_i^T = 0, & (t_i < T) \end{cases} \tag{2.1}$$

$$P(Y_i^T = 1 \mid X_i) = F(X_i) \tag{2.2}$$

Where:

t_i is the walking duration answered by passenger i (in minutes).

T is the threshold of walking duration that to be examined (in minutes).

Y_i^T is a binary variable. According to assumption 3, $Y_i^T = 1$ means passenger i can accept the threshold of T; while $Y_i^T = 0$ indicates that it is uncertain whether passenger i can accept this threshold.

X_i is the vector whose component is the individual characteristics of passengers.

The essence of such issue can be viewed as a binary classification problem, for which the models of decision tree, Bayesian, support vector machine (SVM), logistic regression, and neural network are widely adopted. In this study, limited by the volume of sample and features, also the unknown internal relationship among the features, the decision tree model should be a good choice because of the good generalization for different forms of data. Furthermore, to avoid the structure of the tree being too complicated, and to improve the robustness of the model, this study will use an improved model of decision tree, the random forest decision model, to train and test the samples. Random forest decision is an ensemble learning method mainly for classification and prediction. The random forest decision is an extension and improvement for the decision tree model, it is operated by constructing a multitude of decision trees and randomly selecting the features at training time (Ho, 1995, 1998). To improve the efficiency and accuracy of the model, and reduce the number of invalid branches in the model, it is necessary to filter the invalid features before estimating the model.

Briefly, the general process of this study can be summarized as follows. Firstly, data is extracted on the Fukuoka city-wide, including 4,254 samples, and 6 main categories of features. Secondly, the features with importance in explaining the willingness of walking duration are selected by using analysis of variance (ANOVA). Thirdly, random forest decision for each threshold of walking duration is estimated, the result obtained from the random forest decision is the probability of walking longer than the given threshold of walking duration for each passenger in terms of individual characteristics. Finally, the accuracy of this result is evaluated using the method of simple moving average by descending order of the predicted probabilities.

2.5 Results

The procedure of feature selection includes two steps, the first is the disposal of features with low variance, and the second is the univariate feature selection. The first one is used for filtering the features accounting for very little in the total samples, because a too small quantity in the sample cannot present significant influence on the result. In this study, the multi-categorical features in the original dataset are reclassified into larger groups, thus dealing with the features with low variance. The second step is to select the valid features that indeed have relationship with the independent variable, this step is based on statistical tests. In this study, the Analysis of Variance (ANOVA) is used for estimating the importance of each feature.

2.5.1 Reclassifying features

The feature of trip purpose has 15 subcategories in the original dataset, this study reclassifies them into 5 categories including commuting to work, commuting to school, official business, private purpose (such as shopping, entertainment), and going home. The feature of occupation is reclassified from 14 subcategories into 5 categories as well, they are service, technology, administration, student, and the other. Table 2.2 reports some of the statistical description for each feature. Overall, the average walking duration to rail transit stations is 8.32 minutes; more than 75% of the passengers walk less than 10 minutes; most passengers walk to stations costing 5-10 minutes. In detail, there are some significant differences in the statistical description for each feature, the differences are summarized as follows.

- People with the purpose of commuting account for the major of rail transit users.
- Most people do not accept a more than 15 minutes walking duration to transit stations.
- The walking duration in peak hour is longer than off-peak hour.
- Young and old people are inclined to spend less time on walking to rail transit stations.
- Passengers with the trip purposes of official business and going home tend to take shorter time in walking to rail transit stations.
- Passengers with the trip purpose of commuting to work are willing to accept a longer walking duration than that with other purposes significantly.

2.5.2 Univariate feature selection

According to the achievements from previous studies, the walking distance to rail transit stations ranged generally from 400 meters to 1000 meters in terms of different city types, travel habits, also the needs of research purpose (Guerra et al., 2012; Murray, Davis, Stimson, & Ferreira, 1998; O'Sullivan & Morrall, 1996; Keijer & Rietveld, 2000; F. Zhao, Chow, Li, Ubaka, & Gan, 2003; Alshalalfah & Shalaby, 2007). If converting this walking distance into walking duration by using the walking speed of 4.8 km/h, the walking duration would range from 5 minutes to 13 minutes (Bohannon, 1997).

Table 2.2: Statistical description of walking duration for each feature

Features	Categories	Count	Percentage	Mean	Std	10th	25th	50th	75th	90th
	Total	4254	-	8.32	4.63	3	5	8	10	15
Sex	Male	2257	53.1%	8.41	4.63	3	5	8	10	15
	Female	1996	46.90%	8.22	4.63	3	5	7	10	15
Peak hour	Peak	2976	70.00%	8.51	4.66	3	5	8	10	15
	Off peak	1277	30.00%	7.88	4.52	3	5	7	10	15
Age	5-24	543	12.80%	8.18	4.7	3	5	7	10	15
	25-44	1992	46.80%	8.36	4.42	3	5	8	10	15
	45-64	1407	33.10%	8.43	4.83	3	5	8	10	15
	65-	311	7.30%	7.78	4.81	3	5	7	10	15
Occupation	Service	1256	29.50%	8.32	4.57	3	5	8	10	15
	Tech	721	17.00%	8.58	4.89	3	5	8	10	15
	Office	1076	25.30%	8.44	4.48	4	5	8	10	15
	Student	325	7.60%	8.12	4.73	3	5	7	10	15
	Null	875	20.60%	8.03	4.62	3	5	7	10	15
Purpose	Commuting to work	2697	63.40%	8.69	4.51	4	5	9	10	15
	Commuting to school	287	6.70%	8.19	4.60	3	5	8	10	15
	Official business	153	3.60%	7.10	5.05	2	3	5	10	15
	Private purpose	789	18.60%	7.82	4.78	3	5	7	10	15
	Going home	327	7.70%	7.17	4.67	2	5	5	10	15

In this study, more than 40% of the samples walk less than 5 minutes, and the walking duration within 13 minutes covers about 85% of the total. The 5 minutes walking duration can be accepted by most passengers, while a walking duration longer than 13 minutes cannot be accepted by most passengers. In addition, as the average walking duration in this study is about 8 minutes, the three representative thresholds of 5, 8, 13 minutes are picked as the typical threshold for estimating the relationship between individual characteristics and walking durations.

The valid features are identified using ANOVA. The features with p-value less than 0.05, which means this feature relevant with the dependent variable at the confidence level of 95%, are picked out and listed in Table 2.3. In the Effect column of this table, L means the individual with this feature tend to walk longer than the given threshold of walking duration, while M is the opposite meaning.

As shown in 2.3, at the threshold of 5 minutes, the features of trip purposes and peak hour play the most important role in determining the willingness of walking duration. Situations are changed a little at the threshold of 8 minutes, the importance of age and gender raised in some extent, while the features of trip purposes changed little from that of 5 minutes. The feature importance shows a big difference from 5 and 8 minutes at the threshold of 13 minutes. This result of feature selection is also consistent with common sense and partly confirmed by the previous research. Most of the walking duration is distributed around the average walking duration of 8 minutes, it can be considered that walking duration between 5 and 13 minutes is sensitive to individual characteristics. The walking duration more than 13 minutes is not accepted by most people even if they have different individual characteristics, and the threshold less than 5 minutes is also not sensitive to individual characteristics since the 5 minutes threshold is generally accepted by most passengers.

2.5.3 Estimation of features

The valid features (Table 2.3) are used in the random forest decision to estimate the probability that walking longer than the given threshold of walking duration. As to the estimation, the dataset is divided into two parts, 50% of the sample are used for fitting the model thus obtaining the coefficients, the rest 50% are used for testing the ability of prediction. The prediction process is presented in Figure 2.5. For the

Table 2.3: Valid features and the effect at each threshold

5 min threshold		8 min threshold		13 min threshold	
Features	Effect*	Features	Effect*	Features	Effect*
Age over 65	L	Female	M	Age 45-64	M
Peak hour	M	Age 25-44	M	Peak hour	M
O_Null	L	Peak hour	M	P_commuting to work	M
P_commuting to work	M	P_commuting to work	M	P_ private purpose	L
P_official business	L	P_official business	L		
P_private purpose	L	P_ private purpose	L		
P_going home	L	P_ going home	L		

***Note:** L means the individual with this feature tend to walk longer than the given threshold of walking duration, while M is the opposite meaning.

random forest decision, the prediction is not obtained from the only one decision tree but the multitude of decision trees constructed by random selection of features and samples. The dependent variable at the given threshold of walking duration to rail transit stations is calculated by equation 2.2 based on the mean prediction.

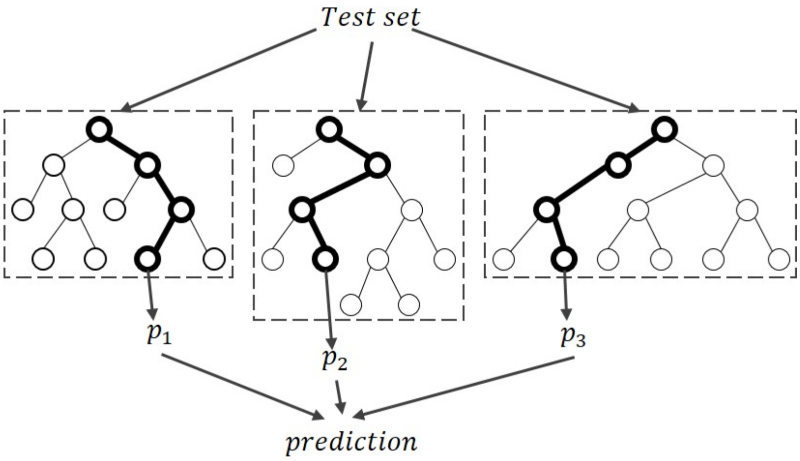


Figure 2.5: Prediction process in random forests decision

The accuracy of results are evaluated by using the method of simple moving average. Figure 2.6 is the trend line of probabilities that walking less than the given threshold. The trend line of the surveyed values based on the test set is calculated by the mean probability of a group people who have close predicted values. The trend line is drawn by a descending order of predicted values. From the comparison of predicted values and surveyed values, it can be known that the prediction at the threshold of 5 minutes has the best fitness, and the prediction for the threshold of 13 minutes is also slightly good, while it is not so good in the case of 8 minutes. In fact, if checking the prediction for individuals, the accuracy of this model is still not enough to explain the individual behavior. But the trend line in Figure 2.6 infers that this result can reflect the behavior of people with specific individual characteristics at a given threshold of walking duration.

2.6 Discussion

This study described and analyzed 4254 records of rail transit trip. Three thresholds of walking duration are examined. From the result in Table 2.3, trip purpose is the most important factor in determining the walking durations at all the 3 thresholds. The feature of peak hour is also significant in explaining the walking duration.

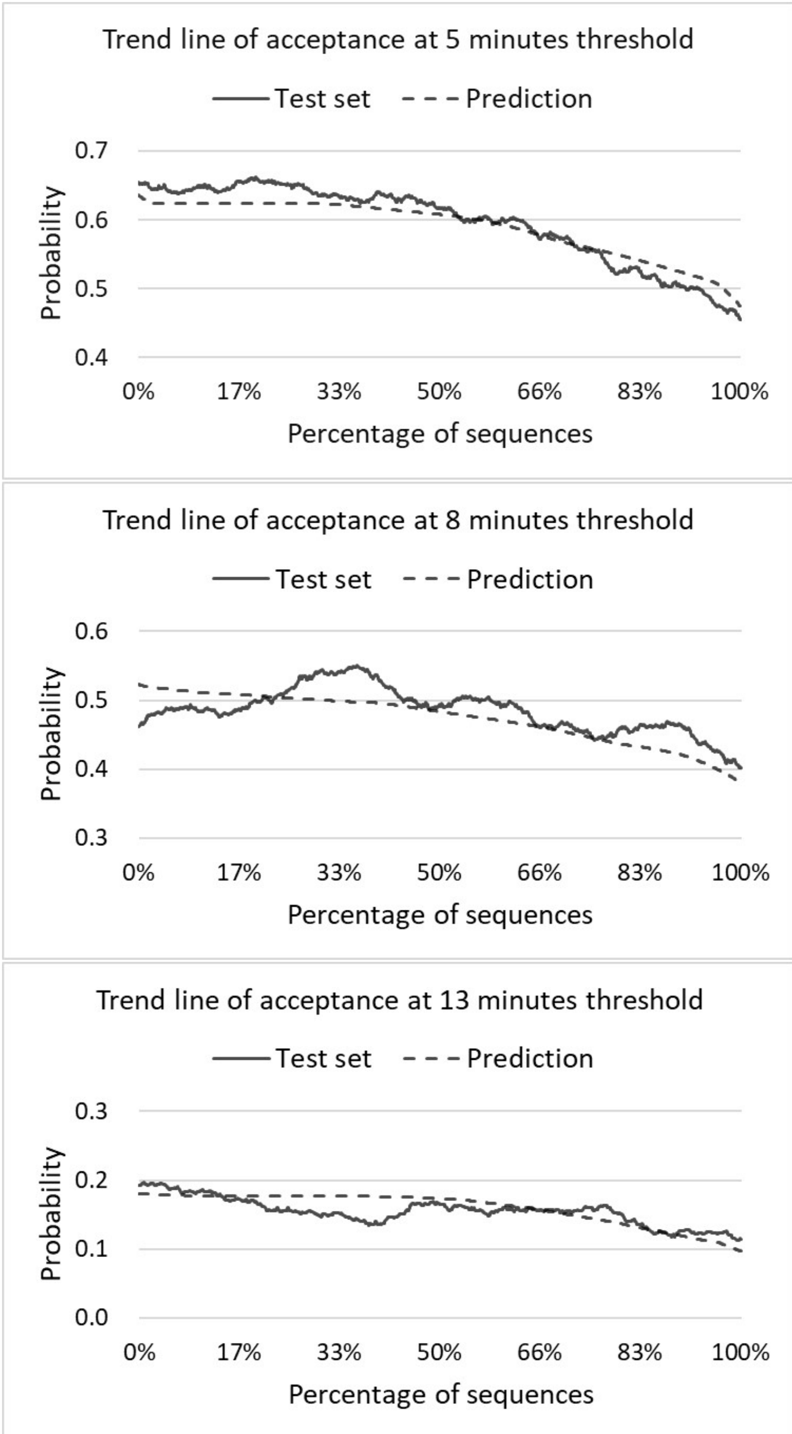


Figure 2.6: Trend line of prediction and test set

People tend to walk a longer time to stations at peak hours, while people with private purpose or on the way going home are not willing to choose a stations far away. For the details of each threshold, the unemployed people and the elderly people tend to walk less than 5 minutes to stations; people whose age is between 25 and 44 are unwilling to walk more than 8 minutes; people aged from 45 to 64 can accept walking more than 13 minutes to stations more easily.

However, summarized above is only the description for the distribution of walking duration in terms of each feature which is hard to apply. In fact, the data in this study is obtained from a factual investigation but not a willingness survey. Once people have made a decision of walking to rail transit stations, the walking duration is just representation of the location of departures and the walking speed. It is hard to say whether the individual characteristics affected the walking durations, maybe due to this reasons few existing studies can explain the relation between walking duration and individual characteristics quantitatively correctly. The perspective changed a little in this study, under the assumptions proposed before, the distribution of walking duration can be viewed as a reflection of the acceptability of walking duration, which should have relationship with people's individual characteristics. According to assumption 3, the behavior that a passenger chose to walk to a station means this passenger can accept the walking duration from the departure to that station. Based on the assumption A1 and A2, if the people who accepted the given threshold of walking duration shows significant differences in individual characteristics, it means they have a different acceptability of walking duration with the others.

As explained above, this study examined the differences in individual characteristics of which people who accepted the given thresholds of walking duration. As the results, people with different individual characteristics shows different acceptability at each threshold of walking duration. According to the evaluation from the method of simple moving average, the model of 5 minutes' threshold has a better explanatory ability, the model of 13 minutes' threshold is a little weaker, and the model of 8 minutes' threshold is not good. Here are some possible reasons for explaining the results. The selected thresholds of walking duration 5, 8, 13 minutes represent the lower boundary, mean value, and upper boundary of the main distribution of walking durations respectively. The commonly acceptable walking durations range from 5 to 13 minutes, therefore, it can be inferred that people who prefer walking less than 5 minutes and who can accept walking longer than 13 minutes may have significant features. However, as the mean value of walking duration, it can be thought that most of the walking durations are distributed around 8 minutes, for passengers there may be some randomness in making the decision of whether walking to rail transit stations. For the other threshold values near the mean value of 8 minutes, it also can be inferred that people may have some ambiguity in choosing whether to walk to stations or not. This explanation can also be confirmed by the result of valid features selection. There are 5 identical features in both thresholds of 5 and 8 minutes, which means people with those features are not sensitive to the threshold from 5 to 8 minutes.

2.7 Conclusion

The results of the random forest decision showed good predictive ability for groups of people, especially at the threshold of 5 minutes. However, it did not give ideal results on individual predictions, which means the application of this study needs a certain number of sample scale, there is still a lot of work should be done on improving the accuracy of prediction. In the next stage of this research, we plan to apply the results on predicting the willingness at a specific threshold of walking duration for a group of people. By using this prediction, for example, if knowing the individual characteristics of residents in a particular area locating from the rail transit station T minutes, the general acceptability of walking to the station for the residents in this area should be predictive. Therefore, this prediction of willingness is expected to be used in planning the catchment area of rail transit stations or estimating the catchment area of existing stations.

References

Alshalalfah, B., & Shalaby, A. S. (2007). Case study: Relationship of walk access distance to transit with service, travel, and personal characteristics. *Journal of urban planning and development*, *133*(2), 114–118.

Besser, L. M., & Dannenberg, A. L. (2005). Walking to public transit: steps to help meet physical activity recommendations. *American journal of preventive medicine*, *29*(4), 273–280.

Bohannon, R. W. (1997). Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. *Age and ageing*, *26*(1), 15–19.

Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, *34*, 548–558.

Daniels, R., & Mulley, C. (2013). Explaining walking distance to public transport: The dominance of public transport supply. *Journal of Transport and Land Use*, *6*(2), 5–20.

Guerra, E., Cervero, R., & Tischler, D. (2012). Half-mile circle: Does it best represent transit station catchments? *Transportation Research Record: Journal of the Transportation Research Board*(2276), 101–109.

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography*, *19*(6), 1081–1092.

Ho, T. K. (1995). Random decision forests. In *Document analysis and recognition, 1995., proceedings of the third international conference on* (Vol. 1, pp. 278–282).

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, *20*(8), 832–844.

Keijer, M., & Rietveld, P. (2000). How do people get to the railway station? the dutch experience. *Transportation Planning and Technology*, *23*(3), 215–235.

Krygsman, S., Dijst, M., & Arentze, T. (2004). Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio. *Transport Policy*, *11*(3), 265–275.

Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the united states. *Transportation Research Part A: Policy and Practice*, *38*(3), 223–247.

Larsen, J., El-Geneidy, A., & Yasmin, F. (2010). Beyond the quarter mile: re-examining travel distances by active transportation. *Canadian Journal of Urban Research*, *19*(1), 70–88.

Marchetti, C. (1994). Anthropological invariants in travel behavior. *Technological forecasting and social change*, *47*(1), 75–88.

McCormack, G. R., Cerin, E., Leslie, E., Du Toit, L., & Owen, N. (2008). Objective versus perceived walking distances to destinations: correspondence and predictive validity. *Environment and behavior*, *40*(3), 401–425.

Murray, A. T., Davis, R., Stimson, R. J., & Ferreira, L. (1998). Public transportation access. *Transportation Research Part D: Transport and Environment*, *3*(5), 319–328.

O'Sullivan, S., & Morrall, J. (1996). Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*(1538), 19–26.

Weinstein Agrawal, A., Schlossberg, M., & Irvin, K. (2008). How far, by which route and why? a spatial analysis of pedestrian preference. *Journal of urban design*, *13*(1), 81–98.

Yang, Y., & Diez-Roux, A. V. (2012). Walking distance by trip purpose and population subgroups. *American journal of preventive medicine*, *43*(1), 11–19.

Zhao, F., Chow, L.-F., Li, M.-T., Ubaka, I., & Gan, A. (2003). Forecasting transit walk accessibility: Regression model alternative to buffer method. *Transportation Research Record: Journal of the Transportation Research Board*(1835), 34–41.

Zhao, J., Deng, W., Song, Y., & Zhu, Y. (2013). What influences metro station ridership in china? insights from nanjing. *Cities*, *35*, 114–124.

Chapter 3

Analysis on the characteristics of transit ridership and land use

3.1 Introduction

3.1.1 Background

In recent years, the problem of aging population has occurred in many developed countries, which also always accompanied with decline in population. As a local central city, Fukuoka now is still in the population growth period, the population has reached 1.5 million, nevertheless, the proportion of aging population is continuously increasing as well. According to the census data, it is expected that the population will reach the peak in 15 years and shift to the population decline period, moreover, the aging population will break through one quarter of the total after 10 years (as shown in Figure 3.1). On the other hand, the data from Kitakyushu Person Trip Survey shows an inclination of that the private car share rate will keep on increasing while the rail transit share rate will turn to decrease in the future (refer to Figure 3.2), As a result, this trend of the shift in population structure and traffic share rate will lead to a decline operating income and increase in financial pressure for rail transit operator of Fukuoka, the same problem will also occur in most of the local central cities similar with Fukuoka. Addition to the financial problem for rail transit operators, traffic congestion is also becoming the problem for all the resident. Figure 3.3, which is quoted from Road Traffic Census (2010), shows the average travel speed during crowded time in major cities of Japan. As we can know from this figure, the problem of traffic congestion is becoming more and more serious in the downtown area of Fukuoka.

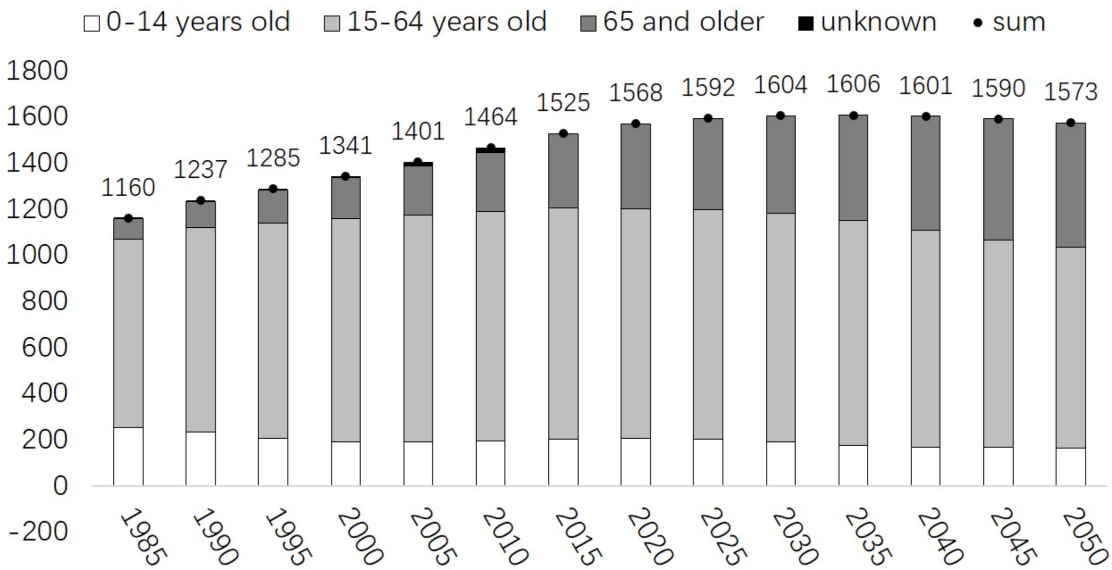


Figure 3.1: Trends of population change

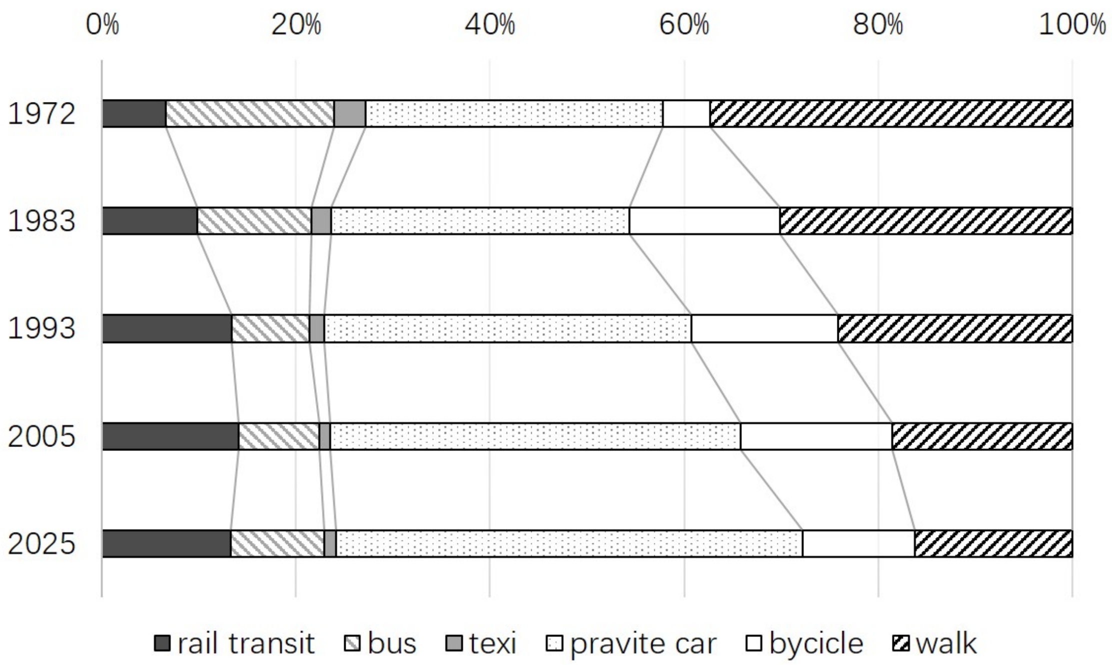


Figure 3.2: Traffic mode share rate

According to the situation stated above, obviously, the issue put in front of the local central cities like Fukuoka is how to promote the use of rail transit, thus reversing the financial dilemma of rail transit operator and making a better living environment for the resident. To achieve this goal, it is necessary to make clear

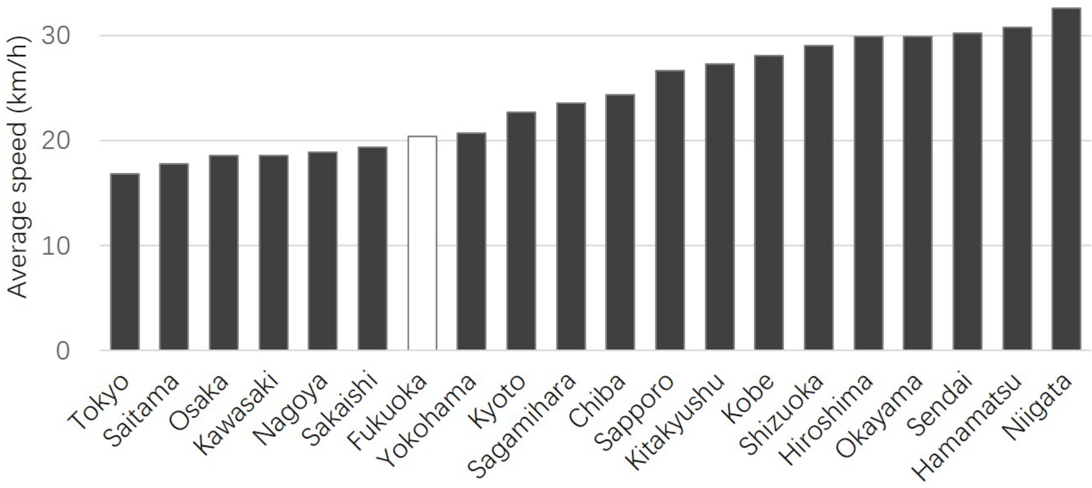


Figure 3.3: Private car travel speed ranking

what factors can influence the rail transit ridership, based on which to make new policy helping improve the role of rail transit.

3.1.2 Previous studies

Many works have been done on the topic of rail transit ridership and the environment around transit stations, while most of them focused on the trend of variation in rail transit ridership or land use, concentration on the studies of the relationship between the transit ridership and land use is still inadequate (Matsumoto & Ubaura, 2013; Nakamura, 2015b).

Depending on the research purpose, the research scale is also different. For example, the study on the changing trends in transit ridership at the scale of Shinkansen mainly aims at making clear the role of each city from the view of entire country (Matsumoto & Ubaura, 2013). At a relative small research scale, for example, some studies focused on the urban rail transit within metropolitan area to make clear of the changing trends in urban structure (Song & Deguchi, 2013; Baba, 2012). This study further narrows the research scale to the urban rail transit within cities, mainly focusing on making clear of the changing trends in rail transit ridership itself, thus providing reference for making policies of urban planning and management (Nakamura, 2015b; Yano, 2008).

Land use around transit stations is always thought as the key factor influencing the transit ridership, while on the other hand, land use is also though to be affected by the transit station. The changes in distribution and types of shop around transit stations are generally considered as good ways to reflect the influence on land use impacted by the stations, the conclusion from existing studies also supported this argument for that the changes of shops around transit stations showed clear characteristics (Sui & Zhao, 2013; Zhao & Sakamoto, 2012; Kitayama et al., 2008). The other kinds of facilities belong to different land use is also relate to rail transit ridership, such as clinic, school, and some other public facilities (Lee, Kashihara, Yoshimura, Takashi, & Sakata, 1995; Lee, Kashihara, Yoshimura, & Yokota, 1994).

Some studies also analyzed the influence on transit ridership from the perspective of land use. A study on the changing trend of transit ridership gave the main conclusion that mixed land use around rail transit stations has a constant effect on increasing transit ridership (Nakamura, 2015b). Quantitative analysis on the relationship between transit ridership and land use usually conducted using regression model (refer to Table 1.1), which is also applied on the case of Japanese cities. A study using the case of rail transit stations within Tokyo metropolitan showed a result that the land use of residence, office, and education plays the most important role in affecting the transit ridership (Nakamura, 2015a).

Land use is widely accepted as one of the most important factors influencing transit ridership, nevertheless, the problem is how to find the specific factors of land use to estimate transit ridership, and how to evaluate this effect quantitatively and precisely. Besides, the influencing factors are not only land use, some other factors such as road network, floor area ratio, transfer structure etc. also have an interactive relationship with rail transit ridership (Kondo, Oosawa, & Kishii, 2010; Inohae, Nagaie, & Hokao, 2009). There are also many factors, even though which has not been fully confirmed yet, maybe also play the important role in determining transit ridership.

3.1.3 Research purpose

With the goal of promoting the use of rail transit, this chapter focuses on making clear of the characteristics of annual changing in both rail transit ridership and land use around the station, based on which to explore the relationship between transit ridership and land use.

Specifically, the research has two main purposes:

1. To describe the characteristics of transit stations in terms of both transit ridership and land use.
2. To explore the relationship between transit ridership and land use on the base of intensive description on the characteristics of transit stations.

3.1.4 Research objective

The research objective in this study is the 35 subway stations of Fukuoka, and several reasons are given here for doing this:

1. The catchment area of subway stations cover all the downtown area of Fukuoka, and most of the urban area.
2. Subway system undertakes the major rail transit traffic within the urban area of Fukuoka.
3. Since subway system is not serving the traffic of intercity, the influencing factors on transit ridership can be easily confined within the area around transit stations, which is conductive to the analysis on the relationship between transit ridership and land use.

This study works on the transit ridership and the land use around transit stations, but how to define the area of "around"? Since this study is not aiming at predicting transit ridership using land use factors, the main purpose is to understand the trend of variation in transit ridership and land use and to explore what kind of land use factors can influence the transit

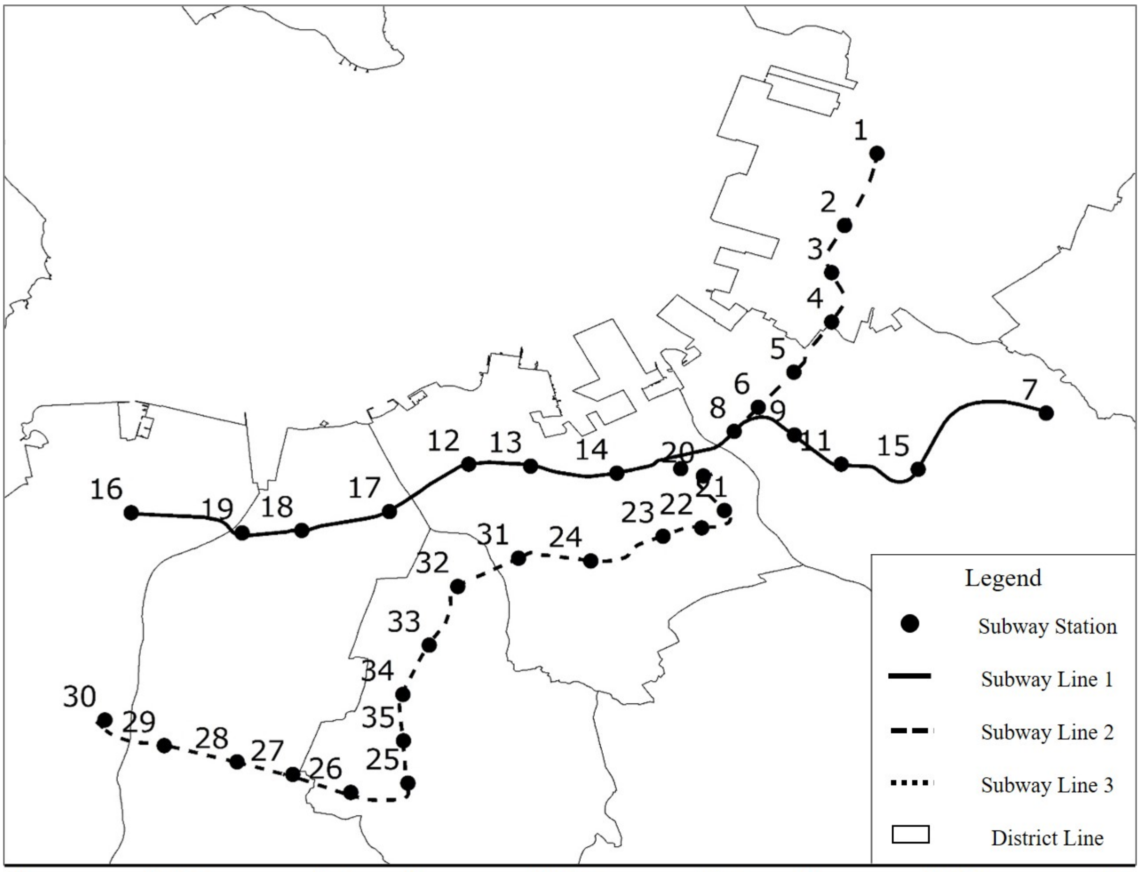
3.2 Data

3.2.1 Study case introduction

The study case in this study is the subway system of Fukuoka, some details is shown in Table 3.1. It consists of three subway lines, the Kukou, or Airport Line (Line 1), the Hakozaki Line (Line 2) and the Nanakuma Line (Line 3). The three lines are operated by the Fukuoka City Transportation Bureau, this subway system is not a large-scale one, which only has 35 stations in total. The distribution and name of the 35 stations is shown in Figure 3.4.

Table 3.1: Information of Fukuoka subway stations

Line	Name	First section opened	Last extended	Length	Stations	Gauge
1	Kukou Line	1981	1993	13.1 km	13	1067 mm
2	Hakozaki Line	1982	1986	4.7 km	7	1067 mm
3	Nanukuma Line	2005	-	12.0 km	16	1435 mm
	Total	-	-	29.8 km	35	-



ID	Station	Line	ID	Station	Line	ID	Station	Line
1	Kaidzuka	2	13	Ohorikoen	1	25	Fukudaimae	3
2	Hakozakikyudaimae	2	14	Akasaka	1	26	Umebayashi	3
3	Hakozakimiyamae	2	15	Higashihie	1	27	Noke	3
4	Maidashikyudaibyuinmae	2	16	Meinohama	1	28	Kamo	3
5	Chiyokenchou	2	17	Nishijin	1	29	Jiromaru	3
6	Gofukumachi	2	18	Fujisaki	1	30	Hashimoto	3
7	Fukuokakuoko	1	19	Muromi	1	31	Ropponmatsu	3
8	Nakasukawabata	1	20	Tenjinminami	3	32	Befu	3
9	Gion	1	21	Watanabedori	3	33	Chayama	3
10	Tennjin	1	22	Yakuin	3	34	Kanayama	3
11	Hakata	1	23	Yakuinodori	3	35	Nanakuma	3
12	Tojinmachi	1	24	Sakurasaka	3			

Figure 3.4: Distribution of the subway stations

From the spatial distribution, the subway stations covered most of the core area of Fukuoka, Line 1 and Line 2 are connected while Line 3 is separated from the other two lines. The catchment area of stations with the number of 8, 9, 10, 11, 20 covers the downtown area, where has the higher density in both population and building. The No. 7 station is in Fukuoka airport, which mainly serves for the transfer of airport passengers. The No. 11 station, Hakata Station, is the comprehensive railway transportation hub of Fukuoka integrating Shinkansen, JR, subway, and bus terminal. The No. 10 station, Tenjin Station, is another central transportation hub, including West Japan Railway, subway, and bus terminal. The two transportation hubs undertake the role as the passenger distribution center not only within Fukuoka urban area but also extending to the cities around Fukuoka city. The endpoint stations with No. 1 and No. 16 connect to other railway lines extending to other cities.

3.2.2 Data collection

All the dataset used in this study comes from the official statistics, details for the dataset is listed in Table 3.2. The data of subway transit and population is annual statistics, while the data of urban planning basic survey is provided every 5 years. Since the subway line 3 is opened in 2005, the research period is set to the 10 years after the subway line 3 is operated. The statistics of population and land use is on the accuracy of town-chome, which is the smallest region in Japanese administrative division.

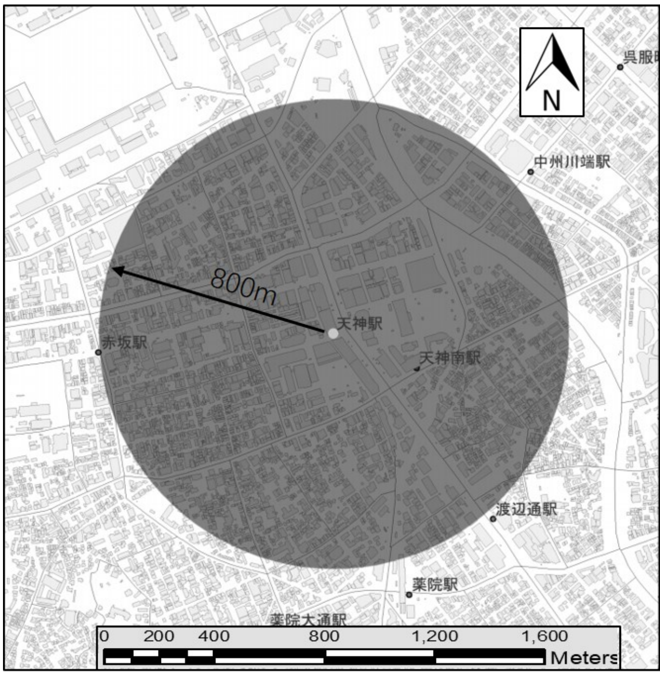
Table 3.2: Data source

Item	Source	Data accuracy	Time point
Subway transit ridership	Fukuoka Traffic Bureau	Station	2005-2014
Population	Resident Basic Account	Town-chome	2005-2014
Land use	Urban Planning Basic Survey	Building	2003, 2008, 2012

3.2.3 Data preprocessing

Since the data accuracy is not matched with population and land use, a preprocess is necessary before conducting the analysis. The data preprocess can be separated into 3 aspects.

1. Matching the region of data. Due to the data accuracy of land use is higher than that of population, the common data accuracy is set to town-chome in this study.
2. Matching the time points of data. To match the time point among different data sources, this study investigates the relationship between transit ridership and land use using the data of 2008, for which is included in all kinds of data.
3. Extracting the data of population and land use within the catchment area of subway stations. This operation is conducted by using ArcGIS, as shown in Figure 3.5, an 800-meter circle area taking the subway station as the center is drawn at first, then the statistics within this area is extracted.



Year	2008
Station No.	11
Transit ridership	108429
Population	28797
Total building area	6854646
Office building area	2100025
Government building area	197787
...	...

Figure 3.5: Extraction of data

3.3 Characteristics of transit ridership and land use

3.3.1 Characteristics of annual change in transit ridership

The subway line 3 was put into operation from 2004, the data of transit ridership was fully recorded from the year of 2005. Figure 3.6 gives the trend of transit ridership during 2005-2014. The total transit ridership has exceeded 0.6 million per day, of which the transit ridership of line 1 accounts for the most reaching about 0.5 million per day.

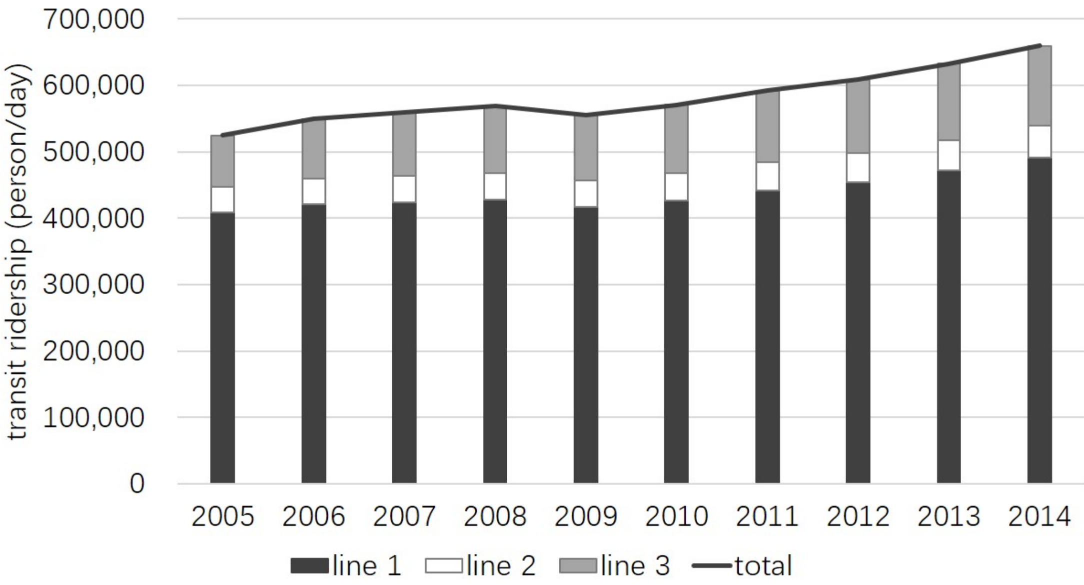


Figure 3.6: Variation in the transit ridership of Fukuoka subway 2005-2014

From the annual change rate of transit ridership in terms of subway lines, as shown in Figure 3.7, notably, at the first three years after the line 3 was opened, the growth rate of line 3 is much higher than that of line 1 and line 2. After the year of 2009, the growth rates of three lines tend to be stable and consistent with each other.

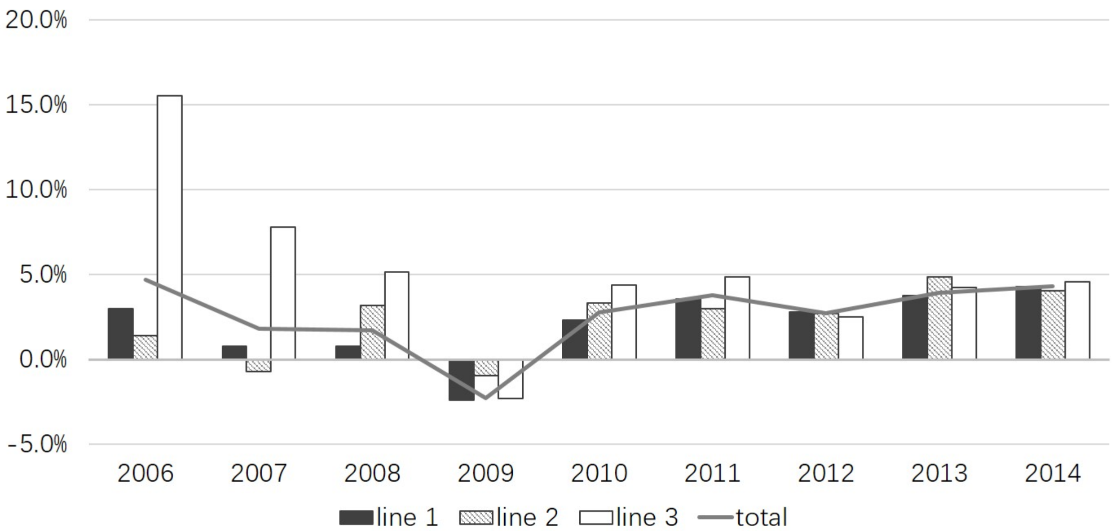


Figure 3.7: Annual growth rate of transit ridership by subway lines

The characteristics of transit ridership are investigated from two aspects, the growth rate variation and transit ridership variation. Different types of station have different characteristics, specific to each subway station, the transit ridership and growth rate is as shown in Figure 3.8, which is sorted by ascending order on transit ridership. The transit ridership is the daily average on the year of 2010, the growth rate is the average of 10 years from 2005 to 2014.

Station No.	Transit ridership on 2010	Growth rate during 2005-2014	Line	Station No.	Transit ridership on 2010	Growth rate during 2005-2014	Line
10	107642	1.50%	1	4	7116	3.50%	2
11	90109	2.77%	1	32	6667	6.32%	3
Hub-scale				2	5655	1.26%	2
16	36664	0.69%	1	5	5609	1.94%	2
7	34578	1.60%	1	35	5324	7.57%	3
17	33055	1.62%	1	31	5292	3.58%	3
20	30874	4.16%	3	3	5231	4.03%	2
Large-scale				27	4771	4.10%	3
8	22652	4.31%	1	6	4617	2.12%	2
14	22090	1.61%	1	21	4117	6.64%	3
18	16815	1.32%	1	34	4052	3.62%	3
15	14464	4.49%	1	23	3790	5.35%	3
12	14247	3.12%	1	30	3717	8.04%	3
1	13119	1.63%	2	29	3471	4.89%	3
13	12888	3.51%	1	28	3231	5.47%	3
19	11419	0.90%	1	33	2998	5.52%	3
22	11289	4.91%	3	24	2327	6.58%	3
9	10285	2.25%	1	26	1619	5.50%	3
25	9269	6.08%	3	Small-scale			
Medium-scale							

Transit ridership	Total	Line 1	Line 2	Line 3
Hub	2	2	0	0
Large	4	3	0	1
Medium	11	8	1	2
Small	18	0	4	4

Growth rate	Total	Line 1	Line 2	Line 3
0-2.5	12	8	4	0
2.5-4.5	11	5	2	4
4.5-6.0	6	0	0	6
6.0-	6	0	0	6

Figure 3.8: Variation in transit ridership of each station

Figure 3.9 plotted the information of each station on the map, the stations in the newly opened line 3 generally have lower transit ridership but higher growth rate. The stations with the higher growth rate in line 1 and line 2 are mainly located close to the downtown area. From the spatial distribution of both transit ridership and growth rate, it shows a clear central agglomeration effect, which means the large-scale stations can attract more passengers.

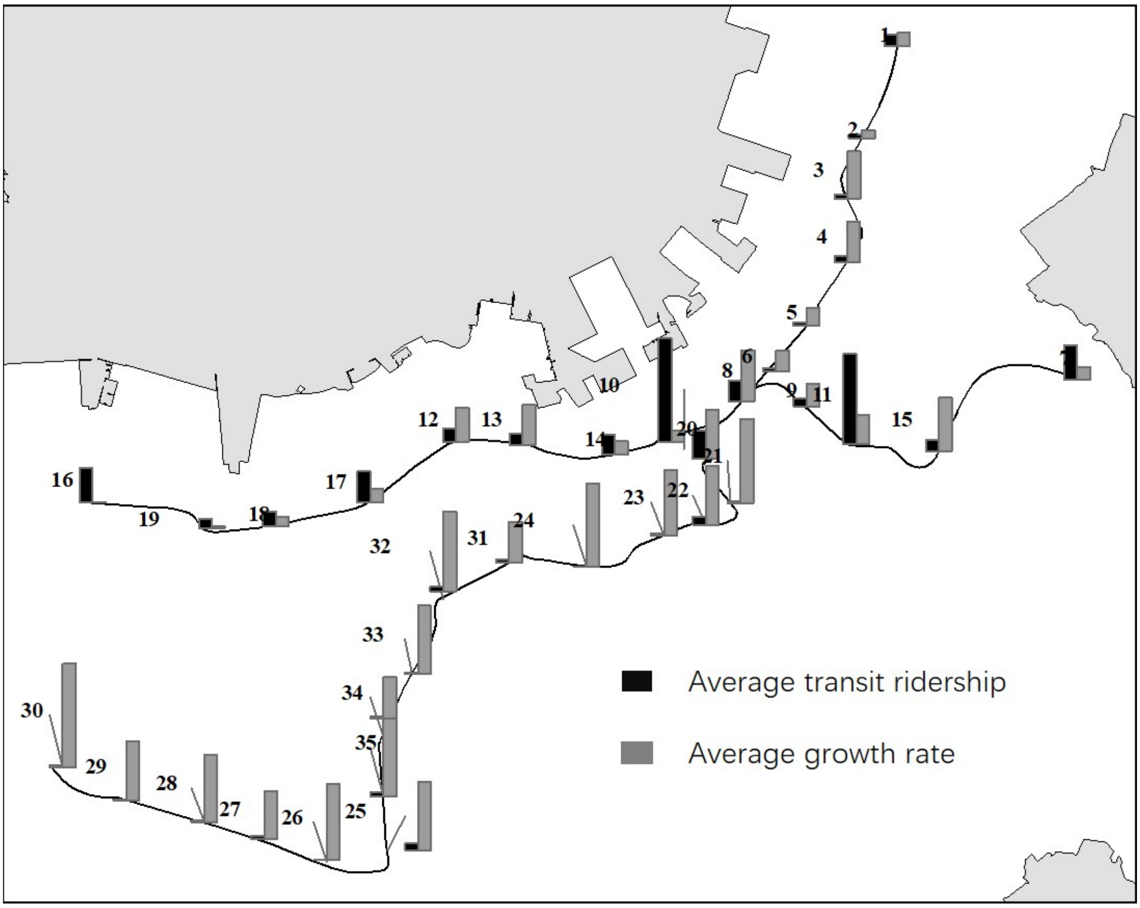


Figure 3.9: Spatial distribution of transit ridership

3.3.2 Analysis on the factors of land use around subway stations

Data extraction

As stated before, the data of land use is extracted within the 800-meter circle region taking subway station as the center. As shown in the Figure 3.10, 35 buffer areas with a 800 meters radius are created. This process of drawing the buffer area and extracting data is conducted using the tool of "Buffer" and "Tabulate Intersection" respectively in ArcGIS.

Figure 3.10: 800-meter Circular buffer areas for 35 subway stations

The land use data from Urban Planning Basic Survey is available every 5 years, to match the data period of transit ridership, the land use data on 2008 is used for analyzing characteristics of land use. The dataset of land use contains 23 types, which, however, is too many for analyzing the characteristics of land use in the case with only 35 stations. Moreover, most types of land use account for very little of the total building area, for which they are thought to have restricted influence on transit ridership. As a result, 9 types of land use, which have relatively larger building areas and commonly exist in all the catchment area of subway stations, are selected to analyze the characteristics of land use, the reserved items are listed in Figure 3.11.

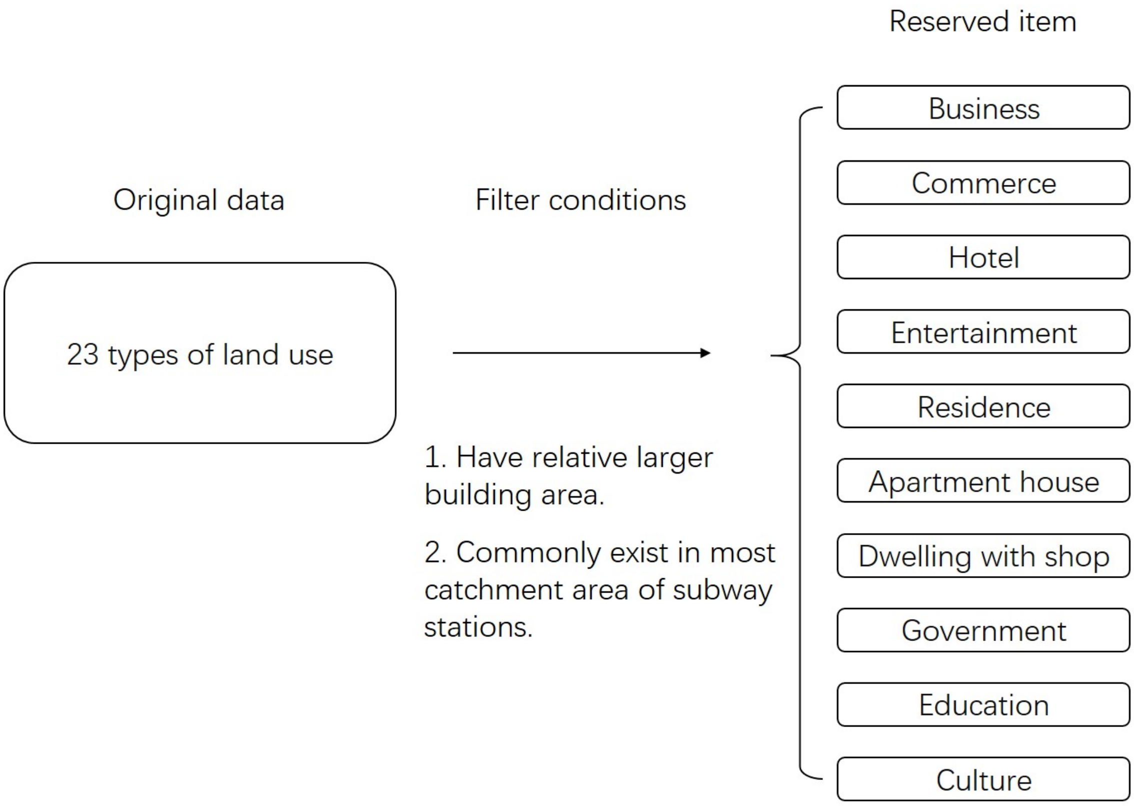


Figure 3.11: Flow of selecting indicators

Through the initial screening, there are 10 indicators being selected. Nevertheless, restricted by the sample size of 35 stations, the 10 indicators are still too many to quantitatively examine the characteristics of land use. To further explore what these indicators are expressing, thus extracting valuable information, the correlation analysis is put forward to make clear of the relationship between indicators. The result of correlation analysis is shown in the Table 3.3. There are strong correlations between the 4 types of land use which are office, commerce, hotel, and entertainment. The type of dwelling with shop also has a strong correlation with that of office, commerce, and hotel, while the culture type of land use closely relates to the type of apartment and dwelling with shop.

Table 3.3: Correlation analysis for indicators

Indicator	1	2	3	4	5	6	7	8	9	10
1. Business	1.000	0.848	0.987	0.775	-0.385	0.378	0.848	0.615	0.006	0.656
2. Commerce	0.848	1.000	0.828	0.765	-0.293	0.345	0.734	0.645	0.065	0.530
3. Hotel	0.987	0.828	1.000	0.716	-0.361	0.382	0.815	0.585	0.000	0.619
4. Entertainment	0.775	0.765	0.716	1.000	-0.322	0.281	0.663	0.439	-0.018	0.552
5. Residence	-0.385	-0.293	-0.361	-0.322	1.000	0.205	-0.287	-0.164	-0.175	-0.063
6. Apartment house	0.378	0.345	0.382	0.281	0.205	1.000	0.649	0.408	0.087	0.729
7. Dwelling with shop	0.848	0.734	0.815	0.663	-0.287	0.649	1.000	0.641	0.038	0.808
8. Government	0.615	0.645	0.585	0.439	-0.164	0.408	0.641	1.000	0.395	0.549
9. Education	0.006	0.065	0.000	-0.018	-0.175	0.087	0.038	0.395	1.000	-0.012
10. Culture	0.656	0.530	0.619	0.552	-.063	0.729	0.808	0.549	-0.012	1.000

Factor analysis

To deal with the strong collinearity among indicators, also to further make clear of the internal relationship, the exploratory factor analysis is introduced to reduce the dimension of the indicators, thus exploring the meanings that the indicators express. The factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables, for which the factor analysis is usually used for dealing with the indicator set with strong collinearity. The procedure of factor analysis is as below.

1. KMO and Bartlett's Test

The KMO measure of sampling adequacy reflect the correlation among all the indicators, referring to Table 3.4. The test result of 0.756, which is greater than the suggested value of 0.7, means this indicator set has enough collinearity to conduct the factor analysis. As to the Bartlett's Test, if the variables are independent to each other, the common factor cannot be extracted from it, and factor analysis cannot be applied. The Bartlett sphere test judges that if the correlation matrix is a unit matrix, the factor analysis method of each variable is invalid. The test results show that $Sig. < 0.05$, which means that each variable has correlation and factor analysis is effective.

Table 3.4: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.756
Bartlett's Test of Sphericity	Chi-Square	346.086
	df	45
	Sig.	0

2. Fact extraction

The factors are extracted using principal component method, Table 3.5 lists the communalities of the indicators. As shown in this table, most information contained in the indicators can be explained by the extracted factors. Less information loss during the process of extracting factors means a good effect.

As a result, there are four factors with the eigenvalue greater than 1.00 being extracted, these four factors account for 82.90% of all the variance. It means the

Table 3.5: Communalities of each indicator

Indicator	Initial	Extraction
Office	1.000	0.943
Commerce	1.000	0.799
Hotel	1.000	0.890
Entertainment	1.000	0.725
Residence	1.000	0.710
Apartment house	1.000	0.849
Dwelling with shop	1.000	0.885
Government	1.000	0.756
Education	1.000	0.923
Culture	1.000	0.811

three factors can explain most part of the original 10 indicators. The Table 3.6 shows the total variance explained.

Table 3.6: Total variance explained

Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	5.068	50.676	50.676
2	1.851	18.510	69.187
3	1.372	13.716	82.902

3. Factor naming and interpretation

The rotated component matrix (Table 3.7) shows the attribution of each indicators on the newly extracted factors. The three factors can be named and explained as follow, they are Office & commerce, mixed-residence, education.

- **Factor 1: Office & commerce**
  This factor represents the land use mainly including office area, large commercial area, and some commercial supporting facilities.

- **Factor 2: Mixed residence**
  The indicators of apartment, residence, and culture mainly attribute to this factor, which can reflect the attribute of residence with supporting facilities.

- **Factor 3: Education**
  Except for the indicator of education, the indicator of government also at-

tributes a large part of this factor. It means that the land use of education and government have a relatively strong correlation.

Table 3.7: Rotated Component Matrix

Indicator	Component		
	1	2	3
Office	0.962	0.122	0.060
Hotel	0.934	0.123	0.050
Commerce	0.878	0.105	0.131
Entertainment	0.85	0.044	-0.030
Dwelling with shop	0.834	0.417	0.125
Apartment house	0.301	0.86	0.134
Residence	-0.521	0.62	-0.234
Education	-0.066	-0.039	0.958
Culture	0.627	0.644	0.057
Government	0.570	0.305	0.582

3.3.3 Classification of stations

To further explore the characteristics of land use in terms of each station, the cluster analysis is used to classify all the 35 subway stations, and the characteristics are summarized respect to the classifications.

The indicators used for classifying the subway stations are selected on the base of correlation analysis and factor analysis conducted before. Although the indicator of office and commerce have a strong collinearity, the definitions and usage of them are quite different, therefore, both of the two indicators are used in cluster analysis. Also with the consideration of the internal correlations shown in Table 3.3, the indicators with relatively high independence are selected. At last, there are 4 indicators of land use being selected for classification, they are office, commerce, residence, education, adding to one more indicator of population density is selected to describe the characteristic of density.

As to the procedure of cluster analysis, the cluster method is Ward Method, the measurement of interval uses the Squared Euclidean Distance, the data is standardized into 0.00-1.00 range. The 34 of all the 35 subway stations fall into 5 types, while the No.7 station located in the Fukuoka airport differs from the others. The

5 types are: low-density residence (9 stations), high-density residence (9 stations), education (6 stations), downtown commerce (5 stations), and office (5 stations) respectively. The result of classification is in the Figure 3.12, the right part of this Table is the pie chart of land use which shows the proportion of land use. The description of each type of stations is given below.

Type	Station	Population-density	Office	Commerce	Residence	Education
Low-density residence	30	33	1.70%	41.50%	44.95%	0.18%
	26	54	2.81%	1.23%	58.26%	2.85%
	29	78	3.02%	5.40%	75.86%	6.94%
	27	79	4.30%	4.24%	75.87%	0.77%
	28	81	3.55%	3.42%	73.46%	6.24%
	3	91	4.01%	8.58%	59.97%	6.25%
	33	101	1.74%	2.10%	81.71%	7.84%
	24	103	2.14%	2.17%	66.82%	8.30%
	34	109	2.38%	3.28%	86.12%	0.86%
	Average	81	2.85%	7.99%	69.23%	4.47%
High-density residence	19	121	2.17%	3.15%	76.63%	0.12%
	13	125	11.11%	6.48%	46.08%	4.51%
	16	142	4.59%	5.03%	72.75%	2.68%
	18	142	2.27%	1.96%	64.58%	6.80%
	31	143	4.14%	2.26%	54.17%	18.28%
	17	144	6.02%	7.28%	55.45%	10.60%
	12	145	4.16%	1.77%	66.22%	4.39%
	32	170	1.17%	1.28%	73.54%	9.93%
	23	193	10.00%	6.18%	44.52%	2.56%
	Average	151	5.43%	4.03%	59.66%	7.47%
Education	25	35	1.38%	2.31%	30.04%	39.83%
	1	56	8.70%	1.91%	48.01%	17.89%
	35	58	5.63%	3.78%	55.18%	21.37%
	4	67	4.93%	3.49%	33.52%	14.63%
	2	77	4.03%	11.75%	44.03%	23.24%
	5	89	11.53%	7.26%	30.51%	7.30%
	Average	64	6.03%	5.08%	40.22%	20.71%
Downtown commerce	9	64	44.97%	20.55%	10.62%	0.70%
	20	67	34.60%	36.74%	11.34%	1.26%
	8	67	42.97%	27.86%	12.23%	0.59%
	11	79	37.12%	17.38%	18.47%	1.89%
	10	92	30.76%	26.26%	15.72%	1.74%
	Average	74	38.08%	25.76%	13.68%	1.24%
Office	15	64	31.63%	7.80%	25.80%	5.73%
	6	96	34.56%	17.29%	21.61%	1.75%
	14	105	26.97%	17.27%	17.61%	2.80%
	21	136	22.86%	16.12%	33.66%	2.70%
	22	167	21.38%	12.53%	36.79%	2.50%
	Average	114	27.48%	14.20%	27.09%	3.10%
Airport	7	19	3.55%	1.45%	33.45%	1.25%

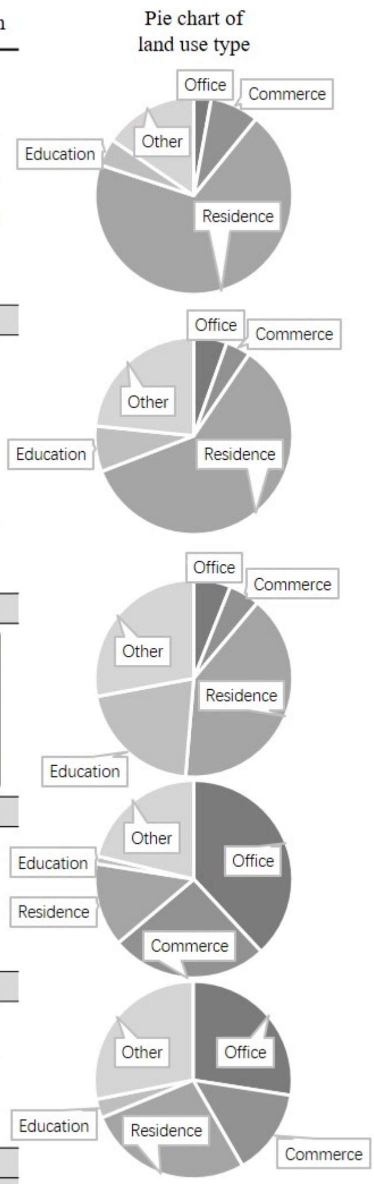


Figure 3.12: Classification of stations in terms of land use

- **Characteristics of low-density residence type**

  This factor represents the land use mainly including office area, large commercial area, and some commercial supporting facilities.

- **Characteristics of High-density residence type**

The indicators of apartment, residence, and culture mainly attribute to this factor, which can reflect the attribute of residence with supporting facilities.

- **Characteristics of education type**
  Except for the indicator of education, the indicator of government also attributes a large part of this factor. It means that the land use of education and government have a relatively strong correlation.

- **Characteristics of downtown commerce type**
  The type of downtown commerce includes 5 stations, which constituted the CBD area of Fukuoka. These stations have the highest proportion of office and commerce, while proportion of residence is the lowest and population density is relatively lower.

- **Characteristics of office type**
  This type includes 5 stations, of which the main land use is office. These stations mainly distribute around the CBD area, mixed land use is one of the main characteristics.

The No.7 Kukou station, which locate at Fukuoka airport, is different from the other 5 types of stations, the main component of land use is transportation facilities. From the view of land use, reflecting into the result of classification, the No.7 station does not belong to any type of stations.

3.4 Influencing factors of transit ridership

3.4.1 Relationship between transit ridership and land use

The data of land use is conducted every 5 years, while the data of subway transit ridership is annual. The 2008 data included in both data is used to analyze the relationship between subway transit ridership and land use.

According to the classification result (refer to 3.12), adding to the average transit ridership of each type of station, as shown in the Table 3.8, the subway transit ridership varies a lot in different types of land use. The station of low-density residence type has an average transit ridership of only 3416, while that of high-density residence type has 15633 on average. The downtown commerce type has the highest average transit ridership of 52300, while that in office type is only 11038, even though similar with the type of downtown commerce the office type has a high proportion of land use in office and commerce. The proportions of office and commerce in education type are similar with that in the two residence types, but from which the average transit ridership is quite different.

Table 3.8: Classification Characteristics

Type	Office	Commerce	Residence	Education	Transit ridership
Low-density residence	2.85%	7.99%	69.23%	4.47%	3416
High-density residence	5.43%	4.03%	59.66%	7.47%	15633
Education	6.03%	5.08%	40.22%	20.71%	7391
Downtown commerce	38.08%	25.76%	13.68%	1.24%	52300
Office	27.48%	14.20%	27.09%	3.10%	11038

According to the classification result of both subway transit ridership and land use around the stations, the cross table is shown in the Figure 3.9 below. All the stations belonging to low-density residence type have small-scale of transit ridership, which is under 10000 per day. Both the two hub stations belonging to the downtown commerce type located in the CBD area of Fukuoka. Overall, it can be inferred from this table that the area with lower density either in population or building has a lower demand in using rail transit; besides the high density in population and building, hub stations should also have particularities in location and function.

Table 3.9: Cross tabulation of land use and transit ridership

Type	Hub	Large-scale	Medium-scale	Small-scale
Low-density residence	0	0	0	9
High-density residence	0	2	4	3
Education	0	0	1	5
Downtown commerce	2	1	2	0
Office	0	0	3	2

Through the characteristics analysis of subway transit ridership and land use, the general trend of the relationship between them can be understood. However, the conclusion obtained from the trend analysis cannot be fully trusted due to the lack of quantitative analysis, also, how the factor land use can influence transit ridership is still not clear.

3.4.2 Influencing factors on transit ridership

The quantification method I is introduced to further explore the influence of each factor on the transit ridership. Different from regression model, the quantification method I converts the continuous independent variables into categorical variables thus investigating the correlation between categorical independent variables and the continuous dependent variable. This method is thought suitable to do the exploratory analysis on the data for the first time due to the procedure of discretizing the continuous independent variables. This discretization can partly reduce the deviation caused by the uneven distribution of the sample in the whole.

The quantification method I can be viewed as an improvement of regression model which is used for dealing with the exploratory analysis at the beginning. Therefore, in the process of selecting explaining indicators, the indicators with strong collinearity should be excluded. Referring to the Table 3.3, the indicators of office, residence, education, and government have less collinearity. Notably, the indicators of office and commerce have strong collinearity, but either of the two indicators accounts for a large part in total, for which they should not be easily ignored. In order to reserve more information, a new indicator named commerce & office which represents the sum of commerce and office building area is proposed. Coupled with the indicator of population density, there are 5 influencing indicators selected as the independent variables. The division of continuous variables are determined based on the Squared Euclidean distance between groups. Both the transit ridership and growth rate of transit ridership are estimated using the 5 indicators of population density, commercial & office, residence, education, and government. The result is shown as Table 3.10 and Table 3.11.

As the results, the coefficients of determination with the value of 0.513 and 0.537 in both models respectively are not satisfactory. However, the results are not for predicting the transit ridership in the future but for exploring the influence of land use on the transit ridership. In view of this, the results are thought to have a certain referential value.

As to the result of influence on the transit ridership, the commerce & office contributes most of the variation in the transit ridership, which means that the commerce & office plays important role in explaining the transit ridership. The building area of education has the weakest influence, while even building area of government is much smaller than that of education, the indicator of government contributes more to the variation in transit ridership.

Table 3.10: Results of quantification method I on transit ridership

Factor category	Category	Number	Score	Range
population density (person/ha)	0-40	3	4644	13004 13.84%
	40-80	13	2283	
	80-120	8	164	
	120-160	8	-2481	
	160-	3	-8360	
Commerce & Office (m^2)	0-100,000	16	-6249	43288 46.07%
	100,000-400,000	9	-6617	
	400,000-1,000,000	5	-4765	
	1,000,000-	5	36671	
Residence (m^2)	0-300,000	4	-1450	16240 17.28%
	300,000-800,000	20	-5575	
	800,000-	11	10664	
Government (m^2)	0-1,000	13	-2957	12267 13.06%
	1,000-10,000	9	-5501	
	10,000-	13	6766	
Education (m^2)	0-10,000	5	4842	9159 9.75%
	10,000-50,000	11	993	
	50,000-100,000	11	-4317	
	100,000-	8	1544	
Independent variable		Sample size		35
Transit ridership		Coefficient of determination		0.513

The same indicator set is also used for estimating the influence on the growth rate of transit ridership (Table 3.11), as a result, the coefficient of determination is a little higher than that of transit ridership. The indicator of commerce & office explains only 10% of the total variation in growth rate, while it accounts for almost 50% in the case of transit ridership. This result shows that the driving force of transit ridership and of transit ridership growth rate is different. The factor of population and residents play the key roles in promoting the use of subway.

Table 3.11: Results of quantification method I on growth rate of transit ridership

Factor category	Category	Number	Score	Range
population density (person/ha)	0-40	3	0.0155	0.0265 28.44%
	40-80	13	0.0020	
	80-120	8	-0.0001	
	120-160	8	-0.0110	
	160-	3	0.0056	
Commerce & Office (m^2)	0-100,000	16	-0.0016	0.0097 10.44%
	100,000-400,000	9	0.0006	
	400,000-1,000,000	5	0.0069	
	1,000,000-	5	-0.0028	
Residence (m^2)	0-300,000	4	-0.0183	0.0226 24.31%
	300,000-800,000	20	0.0043	
	800,000-	11	-0.0012	
Government (m^2)	0-1,000	13	0.0111	0.0228 24.54%
	1,000-10,000	9	0.0008	
	10,000-	13	-0.0117	
Education (m^2)	0-10,000	5	-0.0078	0.0114 12.27%
	10,000-50,000	11	-0.0025	
	50,000-100,000	11	0.0034	
	100,000-	8	0.0036	
Independent variable			Sample size	35
Growth rate of transit ridership			Coefficient of determination	0.537

3.5 Conclusion

This study investigated the variation of subway transit passengers from 2005 to 2014, and then analyzed the types of land use around the stations. On the base of understanding the characteristics of transit ridership and land use, the relationship between them was also estimated. The 35 subway stations were classified into 5 types with typical characteristics in terms of land use. The transit ridership of each type of stations showed significant differences. The results from quantification method I showed the quantitative relationship between transit ridership and land use. Even though the accuracy of results was not enough to make a prediction, it provided references for selecting more valid indicator to make a prediction in the future research.

The major finding of this study can be summarized as follow.

- From the comprehensive description of the study case and the investigation on the transit ridership, it can be known that the subway transit ridership is still increasing at present, but it probably turns to decrease within the near future.

- The subway line 3 has a greater potential for growth in transit ridership. Even though the transit ridership also the population and building density are still lower at present, the stations in subway line 3 are under rapid developing.

- The spatial variation in transit ridership shows the characteristics of central aggregation. The hub stations with higher transit ridership near to the downtown area have a higher growth rate in transit ridership.

- According to the classification of stations in terms of land use, different types of stations have quite different scales on the transit ridership.

- The same indicators of land use have different effects on transit ridership and the growth rate of transit ridership.

This study is the first step to explain the influencing factors on rail transit ridership, which aims to give a comprehensive description of the research objectives also provide references for further explaining transit ridership. Based on the understanding of the insufficiencies in this study, some recommendations are given for the next research.

- The determination of catchment area needs more investigation. Since the 800-meter circle catchment does not consider the differences in the form of the road network, the same 800 meters catchment area may represent different walking distance reflecting on the real road network.

- The selection of indicators explaining the transit ridership needs more exploration. The other categories of indicators about such as facilities, socio-economic, and urban design should also have influence on the transit ridership.

- The selection and usage of estimation model need more investigation. The issue of transit ridership is not only a simple regression problem, but it also relates to the location of stations. A model which is suitable for spatial analysis may be better than an ordinary regression model.

- The approach of dealing with small sample case should be considered. Statistical analysis needs a certain sample size, otherwise, it cannot say the estimation result is credible. The problem led by small sample also reflects on the procedure of selecting the valid explanatory variables.

References

Baba, Y. (2012). Change of personnel and passenger railway in the vicinity of osaka and study on indicators related to it.

Inohae, T., Nagaie, T., & Hokao, K. (2009). A study on formation process of road network and its unity by considering station as a core. *Journal of the City Planning Institute of Japan*, *44*(3), 541–546.

Kitayama, T., Yamada, Y., Kawashima, K., Kojima, K., Negami, A., & Uozaki, K. (2008). Study on the actual condition of urban development accompanies metro improvement on shanghai, china. *Journal of the City Planning Institute of Japan*, *43*(3), 1–6.

Kondo, A., Oosawa, M., & Kishii, T. (2010). On the railway transfer station in tokyo suburban area with focusing the passenger volume, the transfer system and the floor area ratio designated around the station. *Journal of the City Planning Institute of Japan*, *45*(3), 703–708.

Lee, M., Kashihara, S., Yoshimura, H., Takashi, Y., & Sakata, K. (1995). On predicting generation of community facilities around railway station. *Journal of Architecture, Planning and Environmental Engineering, AIJ*, *477*, 71–79.

Lee, M., Kashihara, S., Yoshimura, H., & Yokota, T. (1994). On the temporal change of distribution patterns of community facilities around railway stations. *Journal of Architecture, Planning and Environmental Engineering, AIJ*, *455*, 77–86.

Matsumoto, E., & Ubaura, M. (2013). A study on changes and problems of land use around non-central shinkansen stations in the suburbs. *Journal of the City Planning Institute of Japan*, *48*(3), 627–632.

Nakamura, T. (2015a). An empirical analysis of situations around station as a factor affecting the number of station users. *Regional Policy Research*, *17*(3), 15–26.

Nakamura, T. (2015b). A study on the relationship between land use around railway stations and the railway station passengers. *Journal of the City Planning Institute of Japan*, *50*(3), 1324–1329.

Song, J., & Deguchi, A. (2013). Evaluation and typology of railway station areas in a 30 km circumference surrounding central tokyo from view of transit-oriented development. *Journal of Architecture and Planning*, *78*(684), 413–420.

Sui, H., & Zhao, S. (2013). Research on trend of location and distribution of commercial type about small shops around railway station. *Journal of Architecture and Planning*, *78*(683), 141–148.

Yano, H. (2008). *Study on change of surrounding land use accompanying urban development centering on railway station* (Unpublished doctoral dissertation). Kyushu University.

Zhao, H., & Sakamoto, I. (2012). A study on the changes of shopping malls nearby railway station in tokyo. *Journal of the City Planning Institute of Japan*, *47*(3), 823–828.

Chapter 4

Influencing Factors on Transit Ridership at Station Level

4.1 Introduction

4.1.1 Background

Traffic demand forecasting is always an important part of urban planning also urban management. To date, most of the studies on predicting rail transit ridership are conducted using the samples of large-scale cities where commonly have a large rail transit system with hundreds of stations. As to medium-sized cities, however, there is also the necessity for traffic demand forecasting.

4.1.2 Research purpose

In many developed countries like Japan, as the increasingly serious problems of weakness in population growth, adding to the tendency of using private transport in the local central city, operators of public transport, especially urban rail transit, are now facing financial pressure due to the huge operating costs. How to increase the use of public transportation has become an important issue for the government of the local central city in Japan. With the main purpose of turning around the bad financial situation, some efforts already have been made to attract public transit use (Nakamura, 2015b), nevertheless, still far from adequate, particularly for the medium-sized cities. A further and clearer understanding of how various factors affect the ridership is necessary for the policymaker.

4.1.3 Research contents

Using the case of Fukuoka, which is a typical medium-sized local central city in Japan, as the research object, this study can be viewed as an extension of existing station-level ridership model relative to small sample cases. Different from the case with hundreds of samples, a small sample case with dozens of stations has a higher risk of both type I and type II errors in statistic when identifying the valid variables that should be put into the regression model. For this problem, the aim of this study can be stated as: exploring and explaining the factors influencing subway ridership using the small sample case. The main work includes 3 aspects: a. to build the index framework based on prior study; b. to identify the valid indexes that do affect subway ridership; c. to explain the relationship among variables in generating subway ridership quantitatively.

4.2 Review

To estimate the relevance of various factors and ridership, the model of linear regression is a widely adopted one (Cervero & Kockelman, 1997; Chakraborty & Mishra, 2013). However, due to the insufficient consideration of heteroscedasticity and spatial autocorrelation, the result of regression often leads to large standard errors or low level of significance. Thus, linear regression model is not available to all the stations with different characteristics. To improve the generality of the model, the extension such as Geographically Weighted Regression (abbreviated as GWR), Weighted Least Squares Regression (abbreviated as WLS) and Poisson Regression etc. have been successively introduced into the issue of direct ridership model.

Chu and Choi et al. estimated the ridership of bus using the model of Poisson Regression (Choi, Lee, Kim, & Sohn, 2012; Chu, 2004). The problems occurred in ordinary linear regression such as the contravention between fact and estimated coefficients, and the low level of significance was well addressed. Therefore, both generality and explanatory ability in the regression model were enhanced. Some studies have achieved a high coefficient of determination at the first stage with the ordinary least square model (abbreviated as OLS), however, the result showed a significant heteroscedasticity and a non-random distribution of estimated residual (Kuby, Barranda, & Upchurch, 2004). To deal with this deviation at the second stage, the WLS model was brought in to eliminate heteroscedasticity, in which the data points were weighted using the standard error. The result showed that WLS was effective in eliminating heteroscedasticity and improving the explanatory ability of the model.

The data points in OLS are regarded to be independent of each other, however, each data point has different geographical location in the issue of direct ridership model, the observed values are not considered to be independent of each other in terms of the fact that they are distributed continuously in space. For one data point in regression, the observed value is related to the data point nearby in geographical location, and the regression parameters in different geographical locations usually have different performances in their characteristics (Brunsdon, Fotheringham, & Charlton, 1996). For the problems of spatial autocorrelation and spatial heterogeneity, Cardozo made a comparison of OLS and GWR with the same regression parameters, the result showed that the coefficient of determination had a significant improvement and the standard errors turned to be less in GWR (Cardozo, García-Palomares, & Gutiérrez, 2012). On the basis of common GWR, Jun and Zhao introduced Mixed Geographically Weighted Regression (abbreviated as MGWR) to this issue in the consideration of that some regression parameters did not have special autocorrelation (Jun, Choi, Jeong, Kwon, & Kim, 2015; F. Zhao, Chow, Li, & Liu, 2005). They set part of the parameters as global independent variables, and the others as spatially autocorrelated variables, to make model closer to fact.

Although existing studies have done a lot on the issue of direct ridership model, there is still some insufficiency in each study due to the limitation of study case and data source, especially for small sample case. For the selection and construction of factors, the simple and direct indicators such as population, employment etc. are roughly the same with those in the existing studies. But the definition of indicators obtained by secondary calculating such as land use diversity, bus service etc. are not the same, and the effects of such factors have been neither well verified nor widely accepted. For the model, most of the coefficient of determination in OLS were not very ideal (less than 0.7), there was still more than 30% of the change in ridership not being explained by the model (Gutiérrez, Cardozo, & García-Palomares, 2011; Jun et al., 2015). Additionally, even though some of the studies have obtained the high coefficient of determination exactly, there was still another problem that one factor had too strong effect while the rest of the factors had very little influence on the ridership.

This study will focus on the small sample case and try to address some of the shortcomings stated above. The main work of this study can be divided into 3 parts. 1) Build the index framework based on the previous studies, and proposes new indicators to help describe the variation in the subway ridership of Fukuoka. 2) Optimize the procedure of identifying a valid explanatory variable, thereby making it applicable for small sample cases. 3) Improve the accuracy in the estimation of the regression model in the terms of small sample cases.

4.3 Data

4.3.1 Case introduction

This study focuses on 35 subway stations in Fukuoka City of Japan, which is a typical local central city in Japan having not so large population and not so complex rail transit system. A city like Fukuoka commonly has the demand in developing rail transit, but usually, the data support is insufficient due to the lack of station samples. Figure 4.1 is the research area and the distribution of subway stations. Until now Fukuoka has three operating subway lines (The first line was operated since 1981, the second line was operated in 1993 and the latest third line began in 2005), and has a total of 29.8 kilometers operating mileage, some details of the three lines are shown in Table 4.1. The transport system carries a daily average of more than 0.6 million passengers by 2015 that accounting for more than 20% in total motorized travel(Ministry of Land, Infrastructure n.d.). Although the subway system of Fukuoka is not a large-scale one, it plays a crucial role in public transportation in terms of the city scale and population.

Most of the data used in this study is open source, which can be freely downloaded or bought from the government official website. The data of 2010 year is used as a reference. All the resource of data used in the study is listed below.

- Subway Ridership

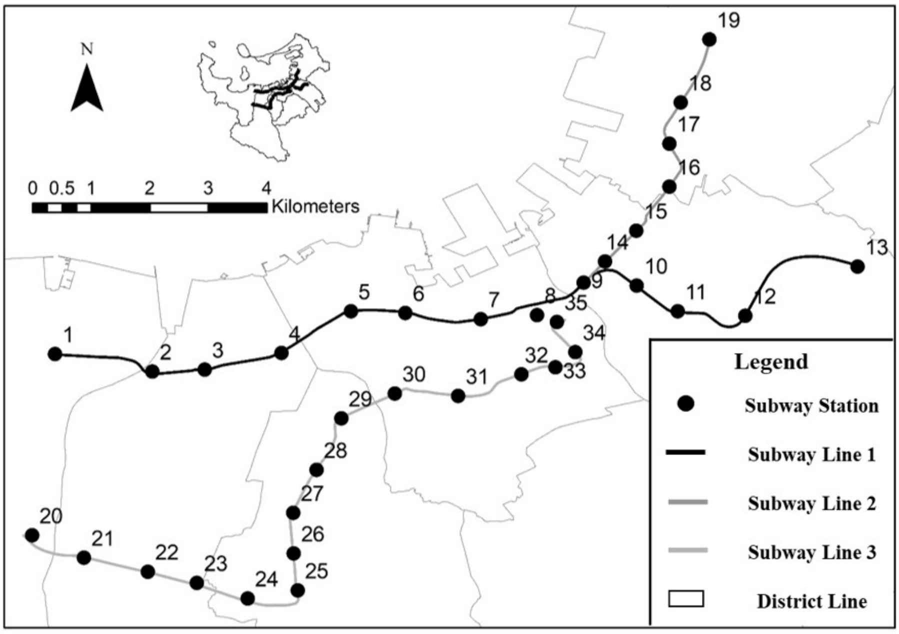


Figure 4.1: Research area and the distribution of subway stations

Table 4.1: Details for the three subway lines in Fukuoka

Item \ Lines	Line 1	Line 2	Line 3
Total Stations	13	7	16
Operating Distance	13.1km	4.7km	12.0km
Transfer Station	5	2	2

- Basic Survey of City Planning
- Census
- Digital Map (Basic Geospatial Information)
- National Land Numerical Information

4.3.2 Catchment area of stations

The scale of the catchment area is an important prerequisite for this kind of issue. At present, in the United States, a half-mile (800 meters) distance has become the practical standard for rail-transit catchment areas based on TOD (Guerra & Cervero, 2013). The distance of 800 meters corresponds to the distance people can walk in 10 minutes at the speed of 4.8 km/h. A Japanese case study also supported this 800 meters catchment area for TOD by using the survey data of <2010 big city traffic census metropolitan area report> (Nakamura, 2015a). Based on the report of basic survey along the subway third line in Fukuoka (refer to 4.2), more than 70% of the passengers choose to walk to the station, about 16% choose bicycles, the total percentage of non-motorized trip accessing to station accounted for about 90%. It can be considered that an 800 meters catchment area covered most of the subway passengers in Fukuoka. On the base of existing literature and the current situation of Fukuoka, the distance threshold of 800 meters is adopted in this study. All the data based on geographical information will be covered by the 800 meters catchment area using the areal interpolation method.

4.3.3 Index framework

In this study, the average daily ridership of each subway station is used as the dependent variable. There are 16 variables from the three categories being adopted as candidate independent factors (as shown in Table 4.3). Some of these variables have already been estimated many times in previous studies, however, there are still some indicators, which may help explain the variation of ridership, ignored in the previous studies. Besides, there is another problem that should be dealt with, the more comprehensive the index framework is, the problem of multicollinearity will occur more easily. For this problem, the candidate variables should be identified and selected by interpret ability in the next stage.

Table 4.2: Description for accessing to transit stations

Station No.	Pedestrian		Bicycle	
	Duration	Proportion	Duration	Proportion
30	8.9	46.5%	8.0	27.5%
29	8.1	84.5%	10.6	9.1%
28	8.9	53.5%	8.1	29.2%
27	9.0	46.2%	8.5	29.5%
26	7.2	91.3%	10.0	8.8%
25	13.1	44.3%	11.6	16.8%
35	6.7	85.2%	8.5	14.8%
34	7.2	97.2%	6.0	2.8%
33	6.0	90.8%	5.0	9.2%
32	7.7	77.3%	10.1	15.9%
31	8.3	60.5%	7.9	27.9%
24	7.3	97.5%	0.0	0.0%
Average	7.7	72.9%	8.4	16.0%

Land use factors

The land and buildings provide people space for living, working and recreating, and people tend to have different travel preferences with different trip purposes. Therefore, the floor area is commonly thought to be the primary factor that can affect ridership. Another compound indicator, land use diversity, is also thought to be an important factor that can influence transit ridership, because people living in the catchment area with higher land use diversity can do most of their daily activities at different types of buildings without taking public transport.

The same with previous studies, this study choose several types of land use with higher proportion to assess the indicator of land use diversity, including residential, office, commercial, education. The four main types of land use account for about 90% of all the floor area in Fukuoka, especially in subway catchment area, reaching more than 95%. However, different from the general definition of land use diversity, the proportion of land use is not evenly distributed in the case of Fukuoka City. As shown in Table 4.4, obviously, it is significantly different from average distribution, and the proportion varies with the variation of the catchment area. To describe the land use diversity closer to the facts, the referenced balance proportion of land use types is decided by the average proportion of all subway

Table 4.3: Statistical description of candidate independent variables

Category	Variable	Unit	Expected sign	Min Value	Max Value	Average
Land use factors	Commerce	ha	+	0.29	81.13	11.44
	Office	ha	+	0.26	84.00	16.71
	Residence	ha	+	11.08	106.75	52.85
	Education	ha	+	0.03	30.56	5.97
	Government	ha	+	-	12.85	2.09
	Transportation Facility	ha	+	0.02	13.28	2.12
	Land use Aggregation	-	+	0.09	0.75	0.31
Transit-related factors	Transfer Station	Dummy	+	1.00	4.00	1.34
	Bicycle Parking	100	Unknown	0.64	43.75	7.78
	Bus Capacity	-	Unknown	3.00	260.00	58.48
	Bus Accessibility	-	Unknown	4.00	455.0	89.71
	Road Density	km/km2	-	19.10	47.90	29.90
Demographic and socioeconomic environment factors	Population	Person	+	1.91	19.39	9.81
	House Member	Person	-	1.86	2.79	2.18
	Population Job Balance	-	Unknown	1.27	2.61	1.80
	Ratio of Rental House	%	-	0.15	0.65	0.43

station catchment area (800 meters pedestrian distance) in Fukuoka.

Table 4.4: Proportion of land use in terms of the range of PCA

Range of PCA	Residence	Office	Commerce	Education	Total
600	51.80%	19.30%	13.80%	5.50%	90.50%
800	55.30%	17.50%	12.00%	6.20%	90.90%
1000	57.60%	16.50%	10.80%	6.00%	90.90%
1200	59.20%	15.50%	10.20%	5.90%	90.80%
Ave	62.80%	10.10%	7.50%	5.30%	85.70%

Note: The value marked with gray is adopted in this study.

In addition, to make indicator more intuitive and simpler, the index of land use diversity is redefined as the aggregation of land use. The Euclidean Metric is used for evaluating the deviation of land use aggregation in each subway station with respect to a reference value. The value of this indicator is arranged from 0 to $\sqrt{2}$, in which the lower value represents a higher level of mixing in land use function, while the higher value means the land use function is more monotonous. This indicator of land use aggregation is defined as Equation 4.1, it is speculated to have a negative impact on ridership.

$$A = \sqrt{\sum (S_i - P_i)^2} \tag{4.1}$$

Where:

i represents the type of land use (respectively government, commercial, residence and education).

A is the indicator for aggregation of land use functions.

P_i is the average proportion of the land use with type i.

L_i is the floor area of land use with type i within the PCA.

Transit-related factors

The transit-related factors are thought to affect subway ridership in two opposite ways. Higher accessibility for accessing into the catchment area of subway station means people will have more transportation options to get into the area instead of the subway, such as private cars, bicycle, bus or walking etc. On the other hand, higher accessibility also allows people to arrive station more easily, and then attract passengers to use the subway. Considering different travel modes, accessibility is separated into 3 parts to be interpreted in this study. First is for users of the road network; second is for passengers transferring from the bus; third is for rail transit interchange.

Road network is commonly used in previous studies. Users of the road network are not only the vehicle but also non-motorized travelers including bicycle riders and pedestrians. As is known, the higher coverage of road network is thought to have better accessibility for both motor travel and non-motorized travel. Additional, bicycle parking may also be an important factor influencing the accessibility of catchment area for non-motorized travelers. Therefore, the accessibility index for road network is set into the number of bicycle parking and road density in this study.

Another accessibility for passengers transferring from the bus is also considered from both positive and negative way. Bus service in catchment area can reflect the connectivity with other regions, by which the potential subway ridership can be shared by bus to some extent. This sharing ridership can be represented by the service capacity of all the bus operating in the same catchment area of the subway station, and the service capacity BC is expressed by Equation 4.2 as follow. In contrast, the bus station close to subway can be used as transfer station which can increase the accessibility of subway for people who want to take the subway but living far away from subway station. In this case, the bus station will play a positive effect on ridership of subway. The indicator should represent the accessibility between subway and bus, thus it is defined as the number of bus lines within the catchment area of subway station. The indicator BA representing accessibility of bus is defined as Equation 4.3 in the below.

$$BC = \sum_{k}^{K} \sum_{r}^{R} f_r^k \tag{4.2}$$

$$BA = \sum_{k}^{K} R_k \tag{4.3}$$

Where:

K is the numbers of bus stations within PCA.

R is the number of lines at one bus station.

f_r^k is the frequency of the rth line at the kth station

R_k is the number of bus lines passing through the kth bus station.

Because the existing rail transit network is relatively stable and unchanging, the convenience of railway interchange is roughly determined by the characteristic of the subway station. The accessibilities are different in diverse types of station: intermediate, terminal, interchange, intermodal. Terminal station is generally thought to have a larger catchment area since the terminal station is probably the only choice for the people living far away from the station when they want to take the subway. And people can accept spending more time on getting into the terminal station (O'Sullivan & Morrall, 1996). Interchange station and intermodal station are attractive for passengers since it can connect to other lines of railway transit or other modes of transportation (Kuby et al., 2004). To distinguish the difference between different types of stations, the dummy variable for describing the number of railway lines passing through each station is introduced into this study.

Demographic and socioeconomic environment factors

For this group of factors, except for the population and employment factors, age structure and household member are also considered to have effect on travel habit. In addition, the family with more working people tends to generate more travel, therefore the indicator of population-to-employment is also thought to be one of the determinants. The factor of tenant proportion was considered to have negative effect on subway ridership in previous studies. Because renters are commonly commute-oriented, they prefer to live close to where they are working and usually commute by walking.

All the 16 candidate indicators are shown in Table 4.3. According to the previous studies, an expected influence is shown in the column of expected sign. With the consideration of small sample case in this study, the method is divided into two phases, identifying the valid variables and estimating the coefficients.

4.4 Methods

The exploratory regression tool in ArcGIS is introduced into this study to help to conduct the process of identifying valid variables. As is known, finding a proper OLS model is the main problem in this kind of study, especially when there are lots of candidate explanatory variables being estimated in the regression model. The Exploratory Regression tool provides a reference for choosing a valid combination of explanatory variables. It is a data mining tool that will try all possible combinations of explanatory variables to see which model can pass the necessary OLS diagnostics. This study proposes a two-step procedure to explore the final model with an optimal combination of explanatory indicators, rather than selecting the best one from all possible combinations. The first step is selecting the variables having effectiveness in explaining the subway ridership, and the second step is choosing the best combination of the valid variables in the final indicators.

GWR is a spatial regression technique, which is used for dealing with the explanatory variables with spatial dependence. The coefficients of explanatory variables are varied with the spatial location of the data point in GWR, and the closer the distance between a data point and observation point is, the greater weight the data point is. Different from general GWR, the explanatory variables in MGWR can be either spatial dependent or spatial independent. The variables with spatial dependence (called local variable) are the same with that in GWR, varied with spatial location of data points; while the variables without spatial dependence (called global variable) are the same with that in OLS, constant in all data points. Before estimating the MGWR model, it is necessary to determine whether the variable is spatial dependent or not. To prevent this small sample case from becoming data-driven, repeating test is conducted to reduce the probability of occasional mistakes. The local/global variables were determined by the spatial dependency of each exploratory variable, in which the variable with spatial dependency is treated as a local term, otherwise, is treated as a global term.

4.5 Results

4.5.1 Selection of effective variables

The first stage of selecting effective variables is conducted based on three judgment factors: a. Multicollinearity, which is expressed by the factor of VIF; b. Validity, which is expressed by the number of times that shows statistical importance; c. Stability, which is shown by the percentage of negative and positive effect to the dependent variable (Stage 1 in Table 4.5). Generally, the variable is thought to be multicollinearity if the value of VIF factor is more than 7.5, as shown in Table 4.5 there are four variables with higher VIF which are marked by dark color. For the factor of validity, three variables are filtered since they rarely show their statistical significance (less than 10%). Even though some variables have statistical significances in the regression model, they are still not credible since their performances are not stable in different models (sometimes they are positive for the independent variable but sometimes are not). As shown in the underneath of stage 1 in Table 4.5, the three variables marked with gray are shown not stable in explaining dependent variable. Therefore, as a result, 10 valid variables of 16 candidate variables are reserved in the first stage.

At the second stage, the exploratory regression is conducted again to select an optimal combination of explanatory variables based on a statistical test for regression. As shown in the stage 2 of Table 4.5, the variable of household members did not pass the statistical significance test, there are 9 valid variables entering the model at last (at 95% confidence level). The Jarque–Bera test (Abbreviated as JB test) is not significant in the final model; it indicates that there are no biased standard errors due to heteroscedasticity. The Breusch–Pagan test is not showing statistical significance, it represents that the residuals are not deviating from a normal theoretical distribution. The test of SA is not significant; it means the residuals are not spatial autocorrelated. The optimal combination with 9 explanatory variables will be evaluated by using MGWR in the next part to obtain a better result with fewer residuals.

Table 4.5: Results of the exploratory regression

Variable	Unit	Stage 1 Statistical Information				Stage 2 Test Model		
		VIF	Validity	Stability	Times	B	Sig	VIF
Government	ha	1.8	42.30%	100.00%	33	490	0.02	1.36
Transportation Facility	ha	3.0	67.90%	100.00%	53	1180	0.00	2.31
Land use Aggregation	%	2.5	50.00%	100.00%	39	124	0.03	1.42
Transfer Station	Dummy	3.0	34.60%	100.00%	27	6014.28	0.00	2.79
Bicycle Parking	100	4.5	44.90%	100.00%	35	754	0.00	2.6
Bus Capacity	-	6.6	37.20%	96.60%	29	-68.19	0.01	3.56
Bus Accessibility	-	6.8	19.20%	100.00%	15	49.37	0.00	4.71
Job-Resident Balance	%	3.0	37.20%	100.00%	35	-47.08	0.05	1.94
Tenant Proportion	%	2.2	21.80%	100.00%	29	-138.2	0.03	1.32
Household Members	%	3.0	39.70%	100.00%	27	-	-	-
Commerce	ha	8.8	32.10%	100.00%	25			
Office	ha	10.4	9.00%	71.40%	7	Residual sum of squares		337744990
Residence	ha	27.5	12.80%	70.00%	10	Adjusted R2		0.96
Education	ha	1.8	7.70%	100.00%	6	AICc		694.39
Road Density	km/km2	2.3	1.30%	100.00%	1	Jarque-Bera test (Sig)		0.61
Population	person	23.9	12.80%	60.00%	10	Koenker (BP) test (Sig)		0.85

Note: The value marked with gray means it is not valid.

4.5.2 Estimation of MGWR

The MGWR is estimated is estimated by using a fixed Gaussian kernel function, the optimal bandwidth size is determined by using a "golden-section search" method. As a result, the optimal bandwidth is obtained when its corresponding AICc value gets to the minimum. The variation of AICc at different bandwidths are shown in Figure 4.2, and the optimal bandwidth is to the lowest AICc of 5.7 km.

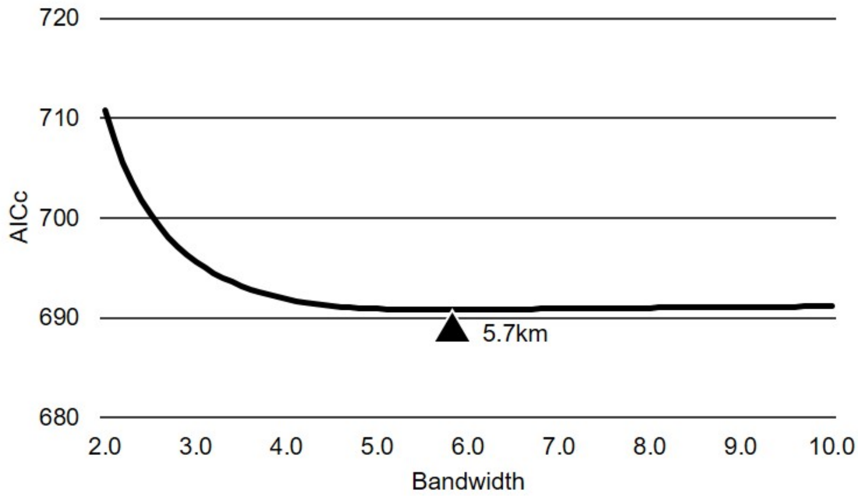


Figure 4.2: Variations of AICc at different bandwidths

The determination of local or global is processed in two steps: firstly, using Moran's index to examine if the variable is spatial autocorrelation or not; secondly, re-examining the variable with spatial dependency in terms of the indicator of "DIFF of Criterion" provided by GWR4. The test value describing the spatial relationship is summarized in the left part of Table 4.6. Five variables are found to be spatial autocorrelation; they are transportation warehousing, bus capacity, bus accessibility and the tenant proportion respectively. Thereby in the first step, the other 4 variables are considered as global variables while the 5 variables with spatial dependency are considered as local variables. Nakaya, the author of the GWR4 software suggested in the GWR4 user manual that the assessment of the spatial variability of the kth varying coefficient is conducted by comparing the fitted GWR with a model in which only the kth coefficient is fixed while the other coefficients vary spatially (Nakaya, 2014). If the original model is better than the model with the kth coefficient fixed, that coefficient can be considered as spatial autocorrelation. GWR4 also provides the indicator of model comparison which is "DIFF of Criterion". The user manual suggests that a positive value of "DIFF of

Criterion", especially greater than or equal to 2, means the local term is better to be assumed as a global one. As shown in the left part of Table 4.6, all the values of indicator "DIFF of Criterion" are no more than 2, which means that all the local terms are adapt to the model. Therefore, all the local variables passed the test in the second step.

4.5.3 Interpretation of coefficients

The right part of Table 4.6 reports the result of MGWR model. The global variables in MGWR model have statistical significance at 95% confidence level, it means most of the variation in ridership can be explained by the 9 variables in this present model. The signs of all terms for local variables are consistent with that of OLS model in Table 4.5, and the values of coefficients in MGWR maintain a high consistency with OLS model. The result shows that there are 3 variables, which are bus capacity, job-resident balance, and tenant proportion, impacting negative effect on ridership, while the others 6 variables play positive effect. The value of the explanatory variable changes by 1 unit, the subway passenger will change the value of the coefficient. It is interesting to note from the coefficient that although the variables of bus capacity and bus accessibility have a strong positive correlation with statistical significance, they perform completely opposite effect on the independent variable. Moreover, comparing with the result of OLS regression, the result of MGWR has an improvement in both adjusted R2 and AICc value, and there is a 12% decrease in residual.

4.5.4 Evaluation of results

As the result from both models, MGWR has a better performance than OLS, in which the residual of MGWR is 12.27% less than that of OLS. Also, the homogeneity of residual distribution is evaluated by using Moran's index shown in Table 4.7. As can be seen, the Moran's index in MGWR is closer to expected value than that in OLS. The MGWR model also shows less variance and greater likelihood of random distribution (with lower z-score and higher p-value) than OLS model. The spatial autocorrelation analysis presents that the residual of OLS model is more likely to be aggregative in space.

Since the MGWR is a kind of variable coefficient regression model, the coeffi-

Table 4.6: Results of MGWR model

Variable	Spatial distribution characteristics of final explanatory variables						MGWR model		
	Unit	Moran's Index	P-Value*	Pattern	Type	DIFF of Criterion	Coefficient	SE	t
Government	ha	0.04	0.51	Random	Global	-	490.00	190.00	2.59
Transportation Facility	ha	0.29	0.00	Clustered	Local	-1.95	1020.00	200.00	-
Land use Aggregation	%	-0.01	0.84	Random	Global	-	133.84	54.06	2.48
Transfer Station	Dummy	0.13	0.12	Random	Global	-	5968.65	1198.72	4.98
Bicycle Parking	100	-0.12	0.36	Random	Global	-	771.70	89.80	8.59
Bus Capacity	-	0.70	0.00	Clustered	Local	0.18	-55.16	5.94	-
Bus Accessibility	-	0.45	0.00	Clustered	Local	0.04	48.61	2.43	-
Job-Resident Balance	%	0.77	0.00	Clustered	Local	-0.17	-24.11	3.64	-
Tenant Proportion	%	0.24	0.01	Clustered	Local	1.02	-103.05	7.56	-
							Best bandwidth		5.7km
							AICc		690.6
							Residual sum of squares		296311499

Note: Confidence level 95%. If p-value is less than 0.05, the corresponding variable can be viewed as a local one.

Table 4.7: Residual spatial autocorrelation of OLS and MGWR

Index	OLS	MGWR
Moran's index	0.08	0.03
Expected index	-0.03	-0.03
Variance	0.01	0.01
z-score	1.09	0.61
p-value	0.27	0.54

cients and adjusted R^2 of each data point are different depending on the location of the data point, as shown in the Figure 4.3. For the 5 local variables in all 35 data points, the spatial distribution of significance level for the local variables are mapped in Figure 4.4. For the variables of building area of transportation, bus capacity, bus accessibility, job-resident balance and tenant proportion, there are 33, 34, 35, 24 and 32 data points having statistical significance at confidence level 95% respectively. As can be seen, the two indicators for bus proposed in this study have good performance in terms of significance and stability, while the variable of 'tenant proportion' does not have a strong stability in explanatory ability. Relatively, the other 4 variables have higher reliability in explaining the variety of ridership.

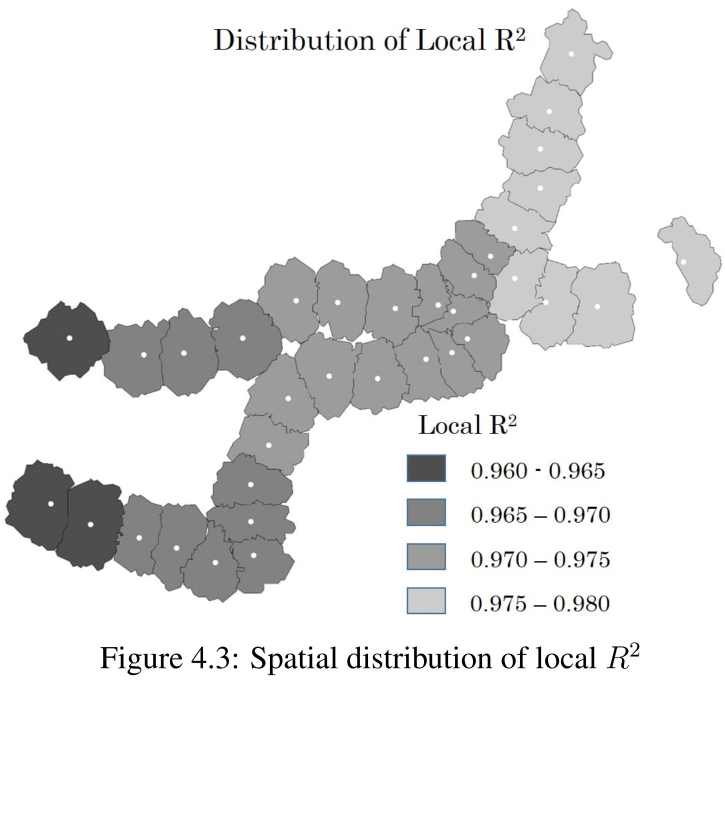


Figure 4.3: Spatial distribution of local R^2

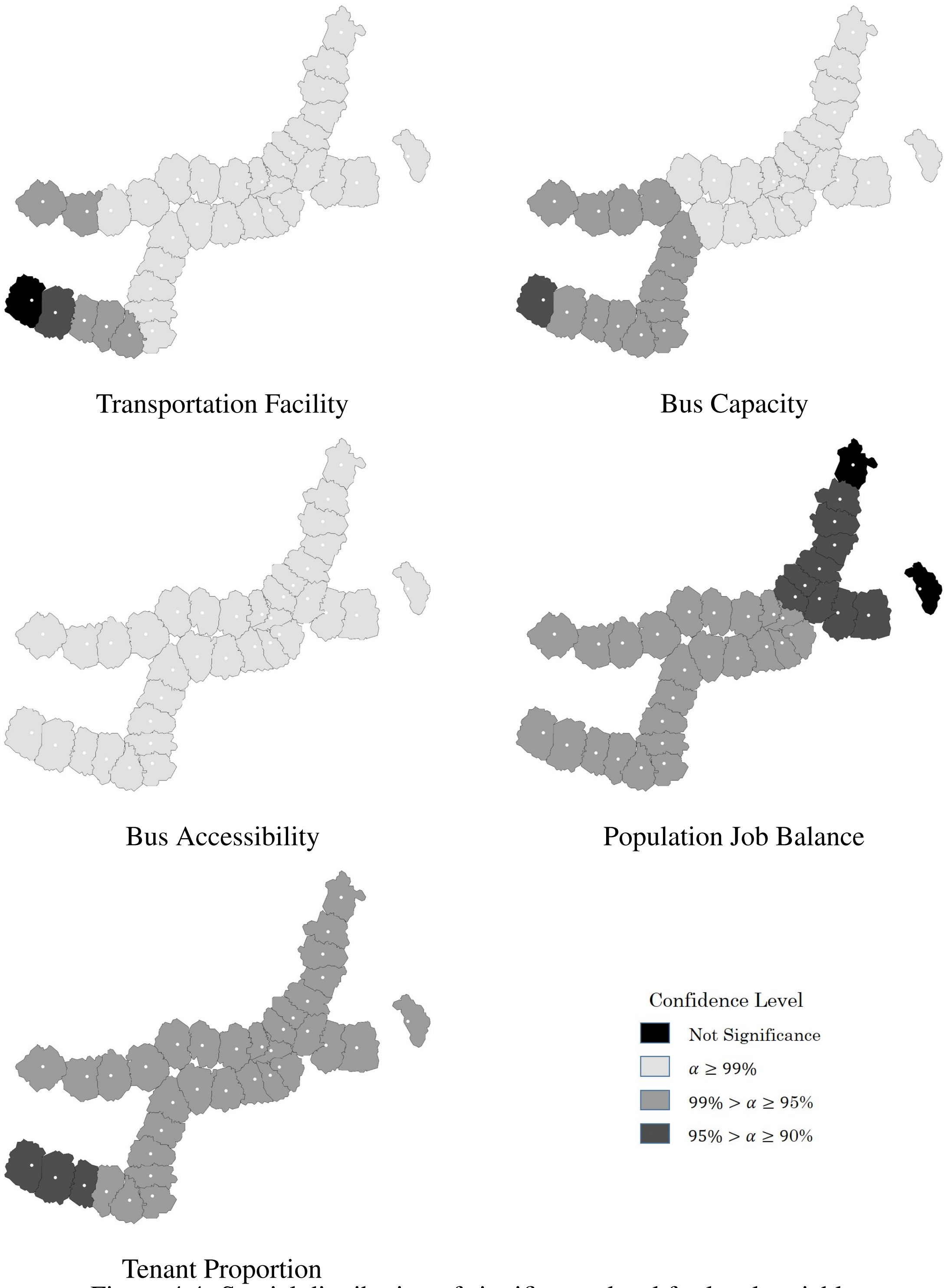


Figure 4.4: Spatial distribution of significance level for local variables

4.6 Discussion

This study focuses on small sample size, using 9 explanatory variables to describe the variety of subway ridership in a local central city with 35 subway stations, while regarding the outcomes of some existed studies in Table 1.1, most of these studies have a middle sample size (ranged from 150-450), and the number of explanatory variables is general about 10. The adjusted R^2 (0.96) in this study is a relatively higher one comparing with other studies, which means that most of the variation in ridership can be explained by the 9 variables in this present model.

However, the adjusted R^2 is not the only judgment criterion in the regression model, especially in the small sample case. The result of this study shows that there are 3 variables, which are bus capacity, job-resident balance, and tenant proportion, impacting negative effect on ridership, while the others 6 variables plays positive effect. It is interesting to note from the coefficient that although the variables of bus capacity and bus accessibility have a strong positive correlation with statistical significance, they perform fully opposite effect on the independent variable. Regarding the coefficient showed in Table 4.5 and Table 4.6 obtained from the two models and 9 variables, the coefficients in both OLS model and MGWR model have consistent signs and similar values. The rationale of estimated results will be discussed from the three categories of indicators.

Factors of land use

Land use has been considered as a critical driver for influencing transit ridership, and some results of empirical studies have supported it (Sohn & Shim, 2010; J. Zhao, Deng, Song, & Zhu, 2013; Chakraborty & Mishra, 2013). Different kinds of land use have various demand in using subway (Chakraborty & Mishra, 2013). In this study, two variables about building floor area (government area and transportation facility area) are found to be valid in explaining subway ridership (with statistical significance at 0.05 level). The coefficients showed in Table 4.6 implies that there is an increase of 5/10 passengers per additional 100 m^2 building floor area for government/transportation facility respectively.

Refer to the results of previous studies, the building floor area of office and commerce was normally considered as the crucial driver for generating transit ridership. However, travel habit and culture are not the same in all cities. The result of this study stated that in Fukuoka people working at or visiting government office are likely to use the subway. Another valid indicator is building floor area of transportation facilities, which mainly represents the scale of public transit in the urban area. Obviously, the larger station usually has a larger scale of passengers, but the causal relationship is that forecasted ridership determined the scale of the station, rather than the opposite direction. Thus, this indicator of transportation facility can be viewed as an index for posterior evaluation, to judge if the planning is consistent with the fact, but not a predictable index. In fact, the variables of office area, commerce area, and residence area are also placed on the candidate list, but they are not showing statistical significance. One possible conjecture can be given here: these three indicators represent the major category in building type, but they also contain several subcategories which cannot be expressed in the indicators. It means that each of these indicators is interpreting multiple issues, for which they cannot be consistent with statistics.

Moreover, the land use mix is also widely thought to be a crucial factor that can influence subway ridership. Some researches argued that the diversity of land use has a positive relationship with ridership, which means the higher diversity of land use can attract more passengers (Gutiérrez et al., 2011; Jun et al., 2015). However, the finding of this study shows an opposite result. The index of land use mix is redefined into land use aggregation in this study, and the result shows that the more aggregated the land use is the more ridership will be generated. In another word, the high diversity of land use will lead to a decrease in subway ridership.

A possible explanation can be interpreted as following from the perspective of principles and features of TOD. A complete TOD area should have various kinds of urban function, and the main aim of TOD is to allow people to do their daily activities by walking in the TOD area, thus reducing inefficient vehicle trip. That is, if people can do most of their daily activities around the station, they will tend to reduce the use of subway. It is still not clear why this difference occurs. Unfortunately, the indicator of land use mix was not discussed in previous studies, there is little reference to explain it. A hypothesis is proposed to interpret this difference, even though there is no way to verify it, that is the proportion of each kind of land use should not be equal, the proportion in Fukuoka is as shown in Table 4.4 (Bhat & Guo, 2007). Therefore, maybe this indicator with statistical significance in the prior studies was interpreting something else related to ridership but not describing land use mix. The results may be just statistically relevant to the ridership coincidentally. And that is the reason why this study proposes the method of identifying valid explanatory variables. Especially for small sample case, repeating test can reduce the probability of contingency.

Factors of transportation accessibility

Accessibility is also thought to be a key factor influencing ridership. K. Sohn and H.Shim further divided this factor into internal and external accessibility, which was also cited and used by Zhao et al., the former represented the accessibility to station within the catchment area, and the latter reflected the connectivity to the place outside the catchment area (Sohn & Shim, 2010; J. Zhao et al., 2013). This study identified four valid variables in transportation accessibility: transfer dummy, bicycle parking, bus capacity and bus accessibility. The transfer dummy and bicycle parking can be easily classified into external accessibility and internal accessibility respectively, and both have a positive effect on ridership. It means the more easily people reach subway station, the more likely they will use the subway. This result is also consistent with prior studies (Gutiérrez et al., 2011; Cardozo et al., 2012; Kuby et al., 2004).

However, the effect of bus service on subway ridership is not so intuitive and clear as other factors. It can be speculated that bus service may have both positive and negative effects on ridership, for there are both competitive and transferring relationships between bus and subway simultaneously. Therefore, a greater transport capacity of bus service can share part of the passengers of the subway while a more accessible route network, bus service can transfer more passengers to subway from other places. And the result from this study has verified this hypothesis that the indicator of bus accessibility and bus capacity showed the totally opposite effect on subway ridership, the former have a positive effect on subway ridership while the latter is in contrary.

The indicator of bus service was intuitively considered to be related with subway ridership, and it often appears in the candidate variables list of prior studies, but is rarely estimated successfully in final model (some studies used the indicator of feeder bus rather than normal urban bus (Sohn & Shim, 2010; Cardozo et al., 2012; J. Zhao et al., 2013). Besides, the factor of trunk bus which is thought to be competitive with subway, is also considered in the previous study, but it did not show statistical significance (Sohn & Shim, 2010). It can be guessed that the influence of bus service cannot be interpreted by only one indicator since the factor of bus service contains more than one kind of information. This result provided some inspiration that every transportation mode having both competitive and transferring relationships with another transportation mode may have both positive and negative effect on the others simultaneously.

Factor of demographic and social economic environment

Regarding the factor of the demographic and socioeconomic environment, the job-resident balance and tenant proportion are found to be effective in explaining the variation of subway ridership. As shown in Table 4.6, both job-resident balance and tenant proportion are showed to have a negative impact on subway ridership. The result indicates that working people tend to use subway more than unemployed people (like children, old people, and housewife), while tenant takes subway less.

Because the travel habit is not the same between working people and unemployed people, the indicator of job-resident balance is developed to help to interpret the difference between jobs and population. It is also suggested that job-resident balance is a crucial factor that can influence house price, this indicator is thought to be related to income level, family structure, social class an so on (Song & Knaap, 2004). But whether and how it can influence subway ridership has not been verified yet in prior studies. The result from this study shows that job-resident balance can affect subway ridership, and it also can be inferred that different group of people has different travel habit. Moreover, the generally considered crucial indicator of population and employment is not showing validity in this study. One possible explanation is that these two variables have multicollinearity with other variables and they have been expressed in the combination of other variables.

In this study, tenant proportion is also verified that having a positive influence on subway ridership. But for different cities and regions, travel preferences are different. The result from the empirical case in nine US cities indicates that tenant is more likely to use subway, while the case in Seoul shows an opposite result that tenants living around station use subway less (Jun et al., 2015; Kuby et al., 2004).

It seems that renters are likely to use public transport, since they are thought to be poor, young, located in denser multifamily housing (Kuby et al., 2004). However, the discussion also suggests that the indicator of tenant percentage should be treated separated: it may have a high tenant percentage in both CBD areas and suburban apartment, but of which the travel habits may be totally different. That means even though travel preference is different in CBD area and suburban area, the indicator of tenant proportion may be almost the same.

4.7 Conclusion

This study examined the factors that may be associated with transit ridership using the case of subway stations in Fukuoka, Japan. There are 9 effective factors selected from candidate indicators to describe the variation of subway ridership in the final models. The results from both the OLS and MGWR model show that the 9 indicators performed stably. The procedure proposed in this study is verified to be effective in identifying valid explanatory indicators in terms of small sample cases. The major contribution of this study can be summarized as follows.

First, on the base of previous studies, this study reclassified and reorganized the indicator framework. Besides the indicators appeared in the previous studies, the indicator of land use diversity was redefined as an aggregation of land use to make the expression of this index more intuitive and closer to reality. Additionally, two variables representing bus accessibility and bus capacity were proposed in this study to explore how can bus service influence the subway ridership.

Second, this study proposed an approach for identifying the valid indicators from all the candidate factors. The result shows that the valid indicators identified by this approach indeed has stable effect on the subway ridership of Fukuoka. In small sample cases, this approach is expected for it can reduce the probability of statistical error by repeating the trail regression.

Third, this study proposed an approach for distinguishing the global and local variables in MGWR model by using Moran' Index. The result shows a significant improvement, even though in the case of a small sample, it performed reasonable and stable.

Finally, as a summary and prospect, direct station-level transit ridership forecasting model showed its advantages of rapid response, low cost, and high efficiency. But on the other hand, the direct model was still a part of the four step model, which could be viewed as the in-depth first step (forecasting of traffic generation) of the four step model. With the enrichment and diversification of data, the influence of environmental change in catchment area on transit ridership can be mastered more accurately with the help of GIS technology.

References

Bhat, C. R., & Guo, J. Y. (2007). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, *41*(5), 506–526.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, *28*(4), 281–298.

Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, *34*, 548–558.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219.

Chakraborty, A., & Mishra, S. (2013). Land use and transit ridership connections: Implications for state-level planning agencies. *Land Use Policy*, *30*(1), 458–469.

Choi, J., Lee, Y. J., Kim, T., & Sohn, K. (2012). An analysis of metro ridership at the station-to-station level in seoul. *Transportation*, *39*(3), 705–722.

Chu, X. (2004). *Ridership models at the stop level* (Tech. Rep.). National Center for Transit Research, University of South Florida.

Guerra, E., & Cervero, R. (2013). Is a half-mile circle the right standard for tods?

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography*, *19*(6), 1081–1092.

Jun, M.-J., Choi, K., Jeong, J.-E., Kwon, K.-H., & Kim, H.-J. (2015). Land use characteristics of subway catchment areas and their influence on subway ridership in seoul. *Journal of Transport Geography*, *48*, 30–40.

Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the united states. *Transportation Research Part A: Policy and Practice*, *38*(3), 223–247.

Nakamura, T. (2015a). An empirical analysis of situations around station as a factor affecting the number of station users. *Regional Policy Research*, *17*(3), 15–26.

Nakamura, T. (2015b). A study on the relationship between land use around railway stations and the railway station passengers. *Journal of the City Planning Institute of Japan*, *50*(3), 1324–1329.

Nakaya, T. (2014). Gwr4 user manual. *WWW Document. Available online: http://www.st-andrews. ac. uk/geoinformatics/wp-content/uploads/GWR4manual_201311. pdf (accessed on 4 November 2013).*

O'Sullivan, S., & Morrall, J. (1996). Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*(1538), 19–26.

Sohn, K., & Shim, H. (2010). Factors generating boardings at metro stations in the seoul metropolitan area. *Cities*, *27*(5), 358–368.

Song, Y., & Knaap, G.-J. (2004). Measuring the effects of mixed land uses on housing values. *Regional Science and Urban Economics*, *34*(6), 663–680.

Zhao, F., Chow, L., Li, M., & Liu, X. (2005). A transit ridership model based on geographically weighted regression and service quality variables. *Lehman Center for Transportation Research, Florida International University, Miami, Florida. http://lctr. eng. fiu. edu/re-project-link/finalDO97591_BW. pdf (accessed December 12, 2010).*

Zhao, J., Deng, W., Song, Y., & Zhu, Y. (2013). What influences metro station ridership in china? insights from nanjing. *Cities*, *35*, 114–124.

Chapter 5

Influencing Factors on Transit Ridership at Station-to-Station Level

5.1 Introduction

5.1.1 Background

Urban rail transit is a type of high-capacity public transport generally found in urban areas. Because of the rapid, punctual and environment-friendly features, the urban rail transit is becoming one of the most important travel modes in modern cities. With the popularity of the urban rail transit in modern cities and the emphasis on sustainable development, the concept of TOD (Transit Oriented Development) is put forward, intended to build the compact city (Calthorpe, 1993). Many cities around the world have adopted the policy of giving priority to the development of public transport for decreasing the share of motorized travel and increasing the willingness of using public transit. For policymakers, how to grasp the relation between land use and transit ridership has become an important issue that they must face.

Although there have been lots of research on the relationship between land use and transit ridership, most of them are conducted from the perspective of the single station. The transit ridership of a station, however, should be affected not only by the elements around that station, but also be affected by the transit ridership of other stations, since all the stations are running in the same system. Once the circumstance within catchment area changes, obviously, the ridership of this station will change as well. But the increased part of passengers will be transported to other stations, thus every station connected to that station will also have an increase in ridership.

5.1.2 Research purpose

Some researchers have discussed on the issue of inter-urban transit demand at station-to-station level (Wardman, 1997; Jones & Nichols, 1983), nevertheless, restricted by the difficulties in collecting data at station-to-station level, few studies focused on intra-urban transit demand (Choi, Lee, Kim, & Sohn, 2012). Even now, the transit ridership forecasting at station-to-station level is still difficult due to the complex interaction among stations.

Based on this realistic background, this study will focus on the relation among stations from the perspective of land use. But rather than making the prediction of transit ridership between station and station, this study tries to explain what factors of land use can influence the choice of destination station for passengers.

5.2 Review

Until now, there are many studies focusing on the relationship between various factors and transit ridership. Most of the studies are based on regression model and conducted from the view of station-level, the transit ridership is thought to be affected by the circumstance surrounding the station (Cervero & Kockelman, 1997; Taylor, Miller, Iseki, & Fink, 2003; Zhao, Chow, Li, & Liu, 2005; Estupiñán & Rodríguez, 2008; Taylor, Miller, Iseki, & Fink, 2009; Sohn & Shim, 2010; Gutiérrez, Cardozo, & García-Palomares, 2011; Jun, Choi, Jeong, Kwon, & Kim, 2015). Among them, the multiple linear regression models are the earliest and most widely used model (Cervero & Kockelman, 1997; Gutiérrez et al., 2011). However, the data point in ordinary least squares (OLS) model is treated as a single point, which is not consistent with fact since the transit node is connected to each other. To deal with the relationship among stations in the network, the approach of spatial regression is also introduced into this issue (Cardozo, García-Palomares, & Gutiérrez, 2012; Jun et al., 2015). However, this relationship among stations in spatial regression models is just the expression for the distribution relationship of stations in location, it cannot reflect the real connectivity between two station areas.

To explore the connectivity between station areas, Choi et al. conducted a station-to-station level investigation into the effect of both origins and destinations on OD metro ridership of Seoul, Korea by using the data from the automatic fare collection system (Choi et al., 2012). The factors considered to influence the OD ridership is divided into three groups: the factors of both origin and destination, and the impedance factors between stations. And where the variables of origin and destination are the same, representing the travel characteristics of O and D respectively. The influence of factors on OD ridership is estimated using multiplicative and Poisson regression, with the data of morning, evening peak hours, and midday hours. This station-to-station approach has connected stations by using the factors of both origin and destination. As the result of this empirical study, different land use functions have different travel characteristics in terms of both time and space. However, this approach still cannot reflect the connectivity between two station areas because of the aggregate processing for data.

Land use and public transit are coevolving partners in city building (Handy, 2005; Dittmar & Ohland, 2012). In the urban railway transit system, the ridership between stations is thought to be related to land use, distribution of functional regions, or travel preferences (Thompson, 1997). The catchment area of a station is the area with mixed land use of residences, business, and leisure within walking distance taking the station as the center, while the transit ridership between stations can be viewed as the connectivity of different catchment areas. The catchment area is generally referring to a compact residential district that includes mixing land use to allow people to most of their daily activities within the easy walking distance of a major transit node (Lund, Cervero, & Willson, 2004). In details, various functional buildings are the carrier for people to live, work and recreate, different functions of buildings correspond different trip purposes. When the functions of buildings within the easy walking distance of a station cannot satisfy the requirement of people's daily activities, people will choose to go to other places to conduct their business by using the transit node such as the subway. Therefore, the distribution of different functional building in a catchment area is considered to not only affect the ridership of the station where it is located but also affect the ridership of other stations connected to that station.

With the goal of explaining the variation in the ridership between stations, this study will focus on the passengers' choices for destination stations from the perspective of complementarity of building function in different catchment areas, using the case of subway network in Fukuoka, Japan. The result from this study also can provide a foundation for explaining the connectivity between different catchment areas. Factors that are expected to influence the connection of stations are stated in the next section. Then the approach and model that are adopted in this study are interpreted. Based on the approach and model, the case of the subway system in Fukuoka, Japan is investigated. The discussion for the result is also presented in the last section.

5.3 Data

5.3.1 Case introduction

This study focuses on 35 subway stations in Fukuoka City (the sixth largest city in Japan) which has the largest population in Kyushu Island of Japan (more than 1.5 million). Referring to Figure 4.1, the research area and the distribution of subway stations. Until now Fukuoka has three operating subway lines, a total of 29.8 kilometers operating mileage. The transport system carries a daily average of more than 0.6 million passengers by 2015 that accounting for more than 20% in total motorized travel (from Fukuoka City Transportation Bureau). Although the Fukuoka subway system is not a large-scale one, it plays a crucial role in public transportation in terms of the city scale and population. Moreover, one thing should be noted that the third line is not directly connected with the first and second line. The catchment area of transit stations used in this chapter is the same with that in Chapter 3 (refer to Figure 3.10). An 800-meter real walking distance on the road network is used as the pedestrian catchment area. The factors considered to have influences on the passengers' choices for destination stations are described as follow.

5.3.2 Land use factors

Land use is generally accepted as one of the determinants for transit ridership. The indicator of floor area in different building functions can be considered as a detailed expression for land use. Also, many researchers have verified the influence of floor area in different building functions on transit ridership by conducting empirical studies (Sohn & Shim, 2010; Gutiérrez et al., 2011; Chakraborty & Mishra, 2013, 2013; Jun et al., 2015). Also, all the stations in the network are connected, the change of land use within the PCA of a station will lead to the changes of transit ridership in all the stations in the same network. For the case of Fukuoka, about 90% of the trips accessing subway stations are non-motorized, in another word, the land use within the PCA of stations play a crucial role in determining the trip purposes of subway users. The accessing duration refers to Table 4.2.

The same with previous studies, this study chose several types of land use with higher proportion to assess the indicator of land use mix, including residential, office, commercial, education. The four main types of land use account for about 85% of all the floor area in Fukuoka City, especially in subway PCA, reaching more than 90%, which refers to Table 4.4. In addition to the indicator of floor area, the index of the land use mix is also assumed to be a crucial factor for explaining the connectivity of stations (Badoe & Miller, 2000; Cervero, 2004; Frank, Andresen, & Schmid, 2004). Different from the general definition of land use mix, this study redefines it into the aggregation of land use. The Euclidean Metric is used for evaluating the deviation of land use aggregation in each subway station with respect to a reference value. The value of this indicator is ranged from 0 to 1, in which the lower value represents a higher diversity in land use function, while the higher value means the land use function is less diverse. This indicator of land use aggregation is defined as for Equation 4.1, it is speculated to have a negative impact on ridership. To describe the mixture of land use closer to the facts, the referenced balance proportion of land use types is decided by the average proportion of all subway station PCA (800 meters) in Fukuoka City (refer to Table 4.4). There are five variables of land use category selected as the indicator in this study, the list of indicators also the statistical descriptions refer Table 5.1.

Table 5.1: Statistical description of data

Category	Variable	Unit	Min Value	Max Value	Average
Land use	Commerce	%	1.03	44.58	9.89
	Office	%	0.41	44.19	12.78
	Residence	%	7.15	92.26	60.34
	Education	%	0.03	48.02	7.08
	Land Use Aggregation	-	0.09	0.75	0.31
Impedance	Bus Capacity	-	58.48	2.91	259.80
	Bus Accessibility	-	89.71	4.00	455.00
	Operation Distance	km	154.50	389.30	237.58

5.3.3 Impedance factors

Another factor generally used for representing the connectivity between stations is the remoteness, which is also widely adopted as impedance in the gravity model for explaining the connectivity between traffic zones (Iwanow & Kirkpatrick, 2007; Kepaptsoglou, Karlaftis, & Tsamboulas, 2010; Nitsch, 2000). Impedance is the index that represents the connectivity and cost between origin and destination, it can affect the choice of the trip in terms of both destination and travel mode. Generally, the factor of impedance can be considered from two aspects in this issue, one is the internal impedance that representing the convenience for accessing the station (Chu, 2004; Chakraborty & Mishra, 2013); the other is external impedance that evaluating the connectivity between stations (Sohn & Shim, 2010).

For Fukuoka, there are about 90% of all the trip accessing stations by walking and bicycles (refer to Table 4.2). Because the PCA is set by the pedestrian distance, the internal road impedance has been already considered in the phase of setting the PCA. It can be thought that there is no need to create a new indicator to represent the internal impedance in this case study.

Since the focus of this study is the connectivity between stations, it can be assumed that the interconnection between stations in a subway network is an influential factor in determining the ridership between stations. Two types of external impedance are considered in this study. The operation distance of subway line between two stations is the direct indicator representing the spatial connectivity between two stations. And the impedance of competing modes is also considered in this study. Two indicators of bus accessibility and bus capacity are proposed referring to Equation 4.2 and Equation 4.3 respectively. The former one is used for representing the convenience for accessing the station which is speculated as a positive factor for ridership, while the later one is the transport capacity of the bus within the PCA of the station which is considered as a negative factor for ridership because it may share part of ridership from the subway. All the statistical description of indicators in this study are listed in Table 5.1.

5.4 Methods

5.4.1 Flow of the method

It is difficult to analyze the connectivity of all the stations simultaneously. To simplify this issue, the study will be started by one single station, and then the connectivity between this station and all the other stations connected to it will be investigated. For a passenger, the behavior of going to some places by subway can be viewed as a procedure of choice. In this selection process, the choice is decided by the trip purpose of the passenger. As stated in the introduction, the type of building functions can be mainly categorized into residence, office, education, and commerce, which represent the trip purposes of going home, business, commute, and leisure respectively. Therefore, this issue can be converted into a discrete choice model, in which the dependent variables are the building environment in the station catchment area, and the independent variable is the choice of the destination station.

5.4.2 Logistic regression

Taking one station as the destination station, it will correspond to the other departure stations in the subway system. The passengers from the all the departure stations have two options at that destination station, that is getting off and not getting off at this destination station. This probability of getting off at that destination station can be viewed as the representation of the connectivity between the departure station and the destination station. At this point, this issue has been converted into a binary choice problem that investigating the probability of getting off at one subway station. Thus, the Binary-Logistic Regression is introduced into this study to deal with this issue. Formula 5.1 shows the Binary-Logistic Regression that will be adopted in the next section.

$$p_i(y_i = 1 \mid X_i) = \frac{1}{1 + e^{-(\alpha + X_i)}} \tag{5.1}$$

Where:

p_i is the probability of getting off the subway for the ith passenger.

y is the choice of passengers.

X_i is the attribute vector of the ith passenger.

α is the residual item

For the estimation of a regression model, the model of logistic regression built in this study is a non-linear regression model, which can be directly estimated by using maximum likelihood estimation (MLE). Thus, the MLE is adopted to estimate the coefficient. In this study, the sample size is the passenger volume of all the departure stations connected to the investigated destination station, which is marked as N. The same with Equation 5.1, p_i is the probability for choosing to get off, thus, $1 - p_i$ represent the probability for choosing not to get off. Then the occurrence probability $L(\theta)$ in all observation sample can be expressed as Equation 5.2:

$$L(\theta) = \prod_{i=1}^{N} {p_i}^{y_i} (1 - p_i)^{(1 - y_i)} \tag{5.2}$$

Where:

p_i is the probability of getting off at the investigated destination station for the ith passenger.

y_i is the choice of ith passenger, of which $y_i = 1$ means getting off at the investigated destination station, $y_i = 0$ means not getting off at the investigated station.

N is the number of passengers riding rail transit at all the departure stations connected to the investigated destination station

5.5 Results

The characteristic of land use within the PCA of a station can represent the characteristic of passengers departing from that station. Since a city is the collection of different functional regions, the land use characteristics of each PCA are also different. To explore the connectivity between stations with different types of land use characteristics, this study firstly classifies all the stations in terms of the proportion of different function of land use and population density. Then the typical stations are chosen from each group of different land use type, thereby estimating the connectivity between stations by conducting the logistic regression.

5.5.1 Selection of object station

As the result, all the stations are fallen into six categories, they are medium-density residence, high-density residence, education, office, commerce, and airport (refer to Figure 3.12). Detailed statistics for each type is shown in Table 5.2, the average values of attributions in each group are represented, furthermore, the factors of bus service and land use aggregation are also put in this table to help to observe the difference among groups. As is shown, both the group of medium-density residence and the group of high-density residence have a higher proportion of residence floor area, but there is a significant difference in the population density in the two groups. It is clearly that the education type has the highest proportion of education floor area. Both the office type and the commerce type have a higher proportion in office than other types, but the commerce proportion in commerce type is also higher than other types. Additionally, the station of Fukuoka airport does not belong to any group, which mainly takes the feeder traffic from the airport to urban area. It also can be noted that the residence and office type have a relatively lower land use aggregation and higher population density.

To investigate the stations with significant differences in land use characteristic, the most typical object station is chosen from each group in terms of the guidelines as follow. 1) Avoid selecting the stations with the highest value of feature indicator. Because the high value may be caused by the particularity of its distribution of land use and road network, it does not have generality in terms of the type it is. 2) Avoid selecting the transferable stations. Because the transit ridership of this kind of

Table 5.2: Station classification

Type	Population density	Commerce	Office	Residence	Education	Land Use Aggregation	Bus Capacity	Bus Accessi-bility
Medium-density residence	98	4%	3%	83%	4%	0.34	18	28
High-density residence	173	5%	7%	76%	6%	0.26	51	80
Education	83	5%	6%	51%	22%	0.3	45	52
Office	135	8%	31%	49%	3%	0.18	83	131
Commerce	69	34%	32%	24%	1%	0.47	132	213
Airport	36	1%	3%	45%	2%	0.23	32	56

Note: the highlighted cells refer to the representative values that having much higher percentage than other land use types.

stations generally relates the other stations, which cannot reflect the distribution of land use within the PCA. 3) Avoid selecting the stations with much lower or higher population density than other stations. Because the low or high population density may lead to different travel preference due to the difference in the distribution of facilities (such as hospitals, schools, and transportation facilities etc.). Finally, six stations are selected from six groups, the factor of each station is shown in Table 5.3.

5.5.2 Estimation of logistic regression

In this study, the independent variable is the choice of getting off at the object station or not, and the influencing factors are land use and impedance between stations and stations. Notably, in the case of Fukuoka subway network, subway line 3 is not directly connected with that line 1 and line 2, thus subway line 3 is addressed separately from subway line 1 and line 2. That is, if the object station belongs to subway line 3, the connectivity will be only investigated within subway line 3. The result of estimation is shown in Table 5.3.

The estimation is as shown in Table 5.3, the values in gray cell mean that this factor showed no statistical significance at the confidence level of 0.05. The index of $Exp(B)$, which is the odds ratio, represents the extent to the effect on the probability of choice if the coefficient B changes one unit. Here giving a brief explanation of the index of $Exp(B)$, since the unit and meaning of the variables entered the models are different. For the variables of land use, the index of $Exp(B)$ represents that the 1 percent (0.01 unit) change in the variables will lead to an $Exp(B)$ multiples of increases in the probability of getting off at the object station. For the variables of impedance, it means the 1 unit change in the variables will lead to an $Exp(B)$ multiples of increases in the probability of choosing the object station as a destination.

A further explanation taking the Tenjin station as an example is given as below. From the point of land use, the increase of commerce and office floor area in the catchment area of the boarding stations can lead to a decrease in the probability of choosing the Tenjin station as the destination station; while the increase of residence and education floor area in the catchment area of boarding stations can raise the probability of choosing to go the Tenjin station. If interpreted in terms

Table 5.3: Result of Logistic Regression

Destination Station		Variables in Boarding Station								
			Land use					Impedance		
Type	Station	Statistical index	Commerce	Office	Residence	Education	Land use Aggregation	Distance	Bus Capacity	Bus Accessi-bility
Medium-density residence	Kamo	*B*	-0.05	-0.06	-0.06	-0.07	-0.03	0.07	0.01	-0.01
		P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.04
		Exp(B)	0.95	0.94	0.94	0.93	0.97	1.07	1.01	1.00
High-density residence	Fujisaki	*B*	0.01	0.00	-0.01	0.01	-0.02	-0.06	0.00	0.00
		P-value	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.00
		Exp(B)	1.01	1.00	0.99	1.01	0.98	0.95	1.00	1.00
Education	Hakozakikyudai	*B*	0.02	0.00	0.02	0.03	-0.01	-0.13	0.00	0.00
		P-value	0.00	0.53	0.00	0.00	0.00	0.00	0.04	0.00
		Exp(B)	1.02	1.00	1.02	1.03	0.99	0.88	1.00	1.00
Office	Gofukumachi	*B*	0.02	-0.02	0.02	0.05	-0.01	-0.01	0.00	0.00
		P-value	0.00	0.00	0.00	0.00	0.01	0.72	0.03	0.00
		Exp(B)	1.02	0.98	1.02	1.06	0.99	0.99	1.00	1.00
Commerce	Tenjinn	*B*	-0.01	-0.01	0.01	0.03	0.02	0.06	0.00	0.00
		P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00
		Exp(B)	1	0.99	1.01	1.03	1.02	1.06	1.00	1.00
Airport	Fukuoka Airport	*B*	-0.03	0.03	-0.01	-0.02	0.03	-0.08	0.00	0.00
		P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Exp(B)	0.97	1.03	0.99	0.99	1.03	0.92	1.00	1.00

Note: B is the coefficient
$Exp(B)$ is the odds ratio
Gray cell means the estimated value is not significant at the significance level of 0.05

of connectivity, the business type of Tenjin Station is weakly connected to stations with the same commercial type, and the stations of office type; while the stations of business type have relatively strong connectivity with the stations of residence and education type.

5.6 Discussion

From the perspective of TOD, this study investigated the connectivity between stations with different land use characteristic by examining the impact of land use and impedance on subway passengers' choice of the destination station. As shown in Table 5.2, the PCA of the station in Fukuoka shows a significant characteristic of land use distribution, and all the subway stations are categorized into six major types. The result of logistic regression shows the coefficient of factor influencing the choice of the destination station, which can be explained as the connectivity between different PCAs.

Some findings of the land use factors can be known from the result in Table 5.3. The increase in the proportion of residence area of the departure station can lead to an increase in the choice of taking the office and education type station as destination stations. This can be explained as the result of commuter traffic since there is a relatively stronger connectivity between the residential area and work area (Badoe & Miller, 2000). But the situation may change in the medium-residence type due to the different travel preferences in the area with different population density. Another interesting finding is that all kinds of land use have positive connectivity with the education type station. One speculation is that students tend to take public transit because of the low income. Overall, one kind of land use generally shows a rejection effect on the station which has a similar consist of land use type. Such as the Tenjin station located in the CBD area, the proportion of commerce area in the PCA of departure station causes a negative effect on the choice of getting off at Tenjin station. Moreover, the variable of land use aggregation shows that there is a positive connectivity between the areas with an unbalanced distribution of land use.

The factors of impedance also showed a good statistical significance, however, the results do not seem to show a certain regularity in different types of stations. Here some possible speculations of the reasons are given for helping to find the limitation of this study and explore the direction for the next study. First, the share of different transportation modes is not considered in this study. The distance between two stations may also affect the variation in the share of different transportation modes, thus it is not stable in representing the impedance. Second, the variables of bus service describe the features of one single station, but not the features that reflecting the connectivity between two stations. If considering two stations, one is in the downtown where the transportation hub locates, and one is in suburban where there are few public transportation facilities; it can be inferred that even though the bus service is rich in the downtown area, it affects very little on the station located in the suburban area.

5.7 Conclusion

This study investigated the effect of land use and impedance between stations on the OD subway ridership from the perspective of the connectivity. The effect of factors on the choice of destination station was estimated using the logistic regression model. The result showed that the influence of land use on the ridership between stations was effective, and this result could be explained consistently with fact; while the factor of impedance was still difficult to get explained.

From the results obtained in this article, in the urban rail transit, the land use within the station catchment area is an important factor in determining the destination of passengers. Correspondingly, the change in the choice of destination can be reflected to the transit ridership. Furthermore, this change in the transit ridership of each station, which is caused by the change in the choice of destination, is the reflect of the connectivity between the station and the station in terms of land use in each station catchment area. The results are clearly shown in Table 5.3.

Based on the results in this study, the next stage of this research will focus on the share ratio in different transportation modes to make a deeper exploration of subway ridership. The factor of impedance will also be rebuilt to help to describe the connectivity between stations more accurately.

References

Badoe, D. A., & Miller, E. J. (2000). Transportation–land-use interaction: empirical findings in north america, and their implications for modeling. *Transportation Research Part D: Transport and Environment*, *5*(4), 235–263.

Calthorpe, P. (1993). *The next american metropolis: Ecology, community, and the american dream*. Princeton architectural press.

Cardozo, O. D., García-Palomares, J. C., & Gutiérrez, J. (2012). Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography*, *34*, 548–558.

Cervero, R. (2004). *Transit-oriented development in the united states: Experiences, challenges, and prospects* (Vol. 102). Transportation Research Board.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219.

Chakraborty, A., & Mishra, S. (2013). Land use and transit ridership connections: Implications for state-level planning agencies. *Land Use Policy*, *30*(1), 458–469.

Choi, J., Lee, Y. J., Kim, T., & Sohn, K. (2012). An analysis of metro ridership at the station-to-station level in seoul. *Transportation*, *39*(3), 705–722.

Chu, X. (2004). *Ridership models at the stop level* (Tech. Rep.). National Center for Transit Research, University of South Florida.

Dittmar, H., & Ohland, G. (2012). *The new transit town: best practices in transit-oriented development*. Island Press.

Estupiñán, N., & Rodríguez, D. A. (2008). The relationship between urban form and station boardings for bogota's brt. *Transportation Research Part A: Policy and Practice*, *42*(2), 296–306.

Frank, L. D., Andresen, M. A., & Schmid, T. L. (2004). Obesity relationships with community design, physical activity, and time spent in cars. *American journal of preventive medicine*, *27*(2), 87–96.

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression.

Journal of Transport Geography, *19*(6), 1081–1092.

Handy, S. (2005). Smart growth and the transportation-land use connection: What does the research tell us? *International Regional Science Review*, *28*(2), 146–167.

Iwanow, T., & Kirkpatrick, C. (2007). Trade facilitation, regulatory quality and export performance. *Journal of International Development*, *19*(6), 735–753.

Jones, I. S., & Nichols, A. J. (1983). The demand for inter-city rail travel in the united kingdom: some evidence. *Journal of transport economics and policy*, 133–153.

Jun, M.-J., Choi, K., Jeong, J.-E., Kwon, K.-H., & Kim, H.-J. (2015). Land use characteristics of subway catchment areas and their influence on subway ridership in seoul. *Journal of Transport Geography*, *48*, 30–40.

Kepaptsoglou, K., Karlaftis, M. G., & Tsamboulas, D. (2010). The gravity model specification for modeling international trade flows and free trade agreement effects: a 10-year review of empirical studies. *The open economics journal*, *3*(1).

Lund, H. M., Cervero, R., & Willson, R. (2004). Travel characteristics of transit-oriented development in california. *Sacramento, CA: California Department of Transportation*.

Nitsch, V. (2000). National borders and international trade: evidence from the european union. *Canadian Journal of Economics/Revue canadienne d'économique*, *33*(4), 1091–1105.

Sohn, K., & Shim, H. (2010). Factors generating boardings at metro stations in the seoul metropolitan area. *Cities*, *27*(5), 358–368.

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2003). Analyzing the determinants of transit ridership using a two-stage least squares regression on a national sample of urbanized areas.

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? analyzing the determinants of transit ridership across us urbanized areas. *Transportation Research Part A: Policy and Practice*, *43*(1), 60–77.

Thompson, G. (1997). Achieving suburban transit potential: Sacramento revisited. *Transportation Research Record: Journal of the Transportation Research Board*(1571), 151–160.

Wardman, M. (1997). Inter-urban rail demand, elasticities and competition in great britain: evidence from direct demand models. *Transportation Research Part E: Logistics* *and Transportation Review*, *33*(1), 15–28.

Zhao, F., Chow, L., Li, M., & Liu, X. (2005). A transit ridership model based on geographically weighted regression and service quality variables. *Lehman Center for Transportation Research, Florida International University, Miami, Florida. http://lctr. eng. fiu. edu/re-project-link/finalDO97591_BW. pdf (accessed December 12, 2010).*

Chapter 6

Conclusion

6.1 Summary

In the context of promoting the use of public transit, the prediction of rail transit ridership is becoming more and more important. This research taking explaining the rail transit ridership as the overall goal analyzed the influence of various factors on ridership from station level and station-to-station level respectively. Moreover, this research also provided new explanations for the catchment area of rail transit stations. As results, this research provided an approach for selecting the valid indicators; and proposed a ridership forecasting method with the consideration of interaction among stations and stations; also, it showed a way to accurately estimate the catchment area.

Specific to each chapter, the main content and findings are:

- ***Chapter 1*** proposed the overall research purpose according to the background of real issues. By reviewing the literature relating to this field, specific research questions were proposed. Around the primary goal of explaining rail transit ridership, the description of study case and dissertation organization were given at the last of this chapter.

- ***Chapter 2*** examined the implication of surveyed walking distance/duration to transit station, and estimated the correlation between the walking duration and people's individual characteristics using random forest decision tree model. In this chapter, the distribution of surveyed walking duration was viewed as the reflection of acceptability of walking to transit stations, and this acceptability was thought to be affected by individual characteristics. As the results, the correlation between surveyed walking duration and individual characteristics was verified, which can provide a way to estimate the catchment area of rail transit stations.

- ***Chapter 3*** is the preliminary study to explore the influencing factors on rail transit ridership using the case of subway stations in Fukuoka. This study analyzed the characteristics of annual variation in subway transit ridership and land use around subway stations. Then the relationship between transit ridership and land use was explored. The results presented the trend of variation in transit ridership, the subway stations were classified into 5 types in terms of land use. Also, the estimation of the relationship between transit ridership and land use showed some correlations in statistics.

- ***Chapter 4*** analyzed the influencing factors on ridership from the perspective of station level. In this chapter, the approach for selecting the valid factors that used for explaining ridership was proposed; the index system was rearranged and the effect of newly proposed indicators was verified; the improvement on distinguishing local and global variable in MGWR model was made. As the results, the valid influencing factors on ridership were extracted from various factors, and influence of each valid factor was estimated.

- ***Chapter 5*** shifted the perspective of analyzing ridership from station level to station-to-station level. In this chapter, the type of land use and the relative location of stations were thought to be important factors influencing the ridership between station and station. The station-to-station connectivity was expressed by the probability of selecting the destination station from all the stations, and was estimated by logistic regression model. As the results, the variation in station-to-station connectivity was able to be estimated, thus the variation in ridership of one station caused by the variation in ridership of other stations within the same station network can be estimated.

- ***Chapter 6*** summarized the main content and findings from the view of both integer and each chapter. Recommendations for future work were also given for extending and improving this research field.

6.2 Contributions

This dissertation worked on several key questions in the field of explaining rail transit ridership, the main contribution can be arranged as below.

1. Reinterpreted the implication of surveyed walking distance/duration to transit stations, that it has no linear relation with people willingness but just the reflection of the distance between departures and stations. (refer to **chapter 2**)

2. Described the correlation between surveyed walking duration and people's individual characteristics from the view of probability, which can be further applied to the estimation of catchment area of rail transit stations. (refer to **chapter 2**)

3. Analyzed the trend of variations in transit ridership, and classified the subway stations of Fukuoka into 5 types in terms of land use. Confirmed the correlations between transit ridership and land use. (refer to **chapter 3**)

4. Considered the influence of bus on rail transit ridership from both the sides of bus capacity and bus accessibility, and verified that the effect of bus capacity is positive to rail transit ridership, while the effect of bus accessibility is negative. (refer to **chapter 4**)

5. Proposed the approach of screening valid indicators by introducing the exploratory regression, and confirmed its effectiveness in a small sample case. (refer to **chapter 4**)

6. Distinguished the local and global variables in MGWR model by examining the spatial distribution of each variable, and the effectiveness was confirmed in the estimation of MGWR model. (refer to **chapter 4**)

7. By examining the probability of selecting the destination station from all the stations, established a logistic regression model for describing the correlation between stations and stations. (refer to **chapter 5**)

8. Quantitatively validated the impact of land use types on the selection of destination stations. (refer to **chapter 5**)

6.3 Recommendations

This research focused on the influencing factors on rail transit ridership from both station level and station-to-station level, also analyzed the key element influencing the use of rail transit, the scale of catchment area and walking preferences. Based on the findings in this research, several recommendations for helping increase the use of rail transit are outlined.

1. People's walking preferences have significant influence on the use of rail transit. It is suggested that the design for pedestrian accessibility should also consider the individual characteristics of resident, thus helping increase the use of rail transit.

2. With the development of rail transit, if viewing from the whole transit system, the positioning of bus system is gradually inclining to be the connection between departures and rail transit. Rather than planning higher capacity for bus transport, it is suggested to plan more accessible bus routes which can help people use rail transit more convenient.

3. According to the findings, the more aggregation in land use functions the more demand in rail transit. Thus suggesting to connect more functional regions as possible when making rail transit planning.

At current stage, the obstacles of going deep into this research field are mainly in obtaining available data. Some suggestions for future work include the following:

1. Surveys of income and occupation, exploring the effect on travel behavior.

2. Integer of the whole process of estimating the catchment area of rail transit stations based on the achievements of this dissertation.

3. Joint of the estimation of catchment area and ridership forecasting, improving the accuracy.

4. Analysis of the influence on land use affected by the use of rail transit, exploring the interaction between land use and transit.

5. Analysis of balance condition among various elements in the catchment area

of transit stations, including human, resource, land use, transportation etc. achieving the goal of sustainable development.