Research on Determinants of Rail Transit Ridership: Taking Fukuoka, Japan as a Study Case

陳, 琦

出版情報：Kyushu University, 2018, 博士（工学）, 課程博士
バージョン：
権利関係：

Research on Determinants of Rail Transit Ridership

– Taking Fukuoka, Japan as a Study Case –

Graduate School of Human-Environment Studies

Kyushu University

Author: Chen Qi

Supervisor: Prof. Zhao Shichen

Last Edited: January 23, 2019

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

Contents

List of Figures

List of Tables

Acknowledgments

The journey of Ph.D. has both pleasure and pain. But I deeply trust that what I went through today, both the good and the bad, will be the valuable life experience in the future. If mentioning the most memorable thing during my Ph.D., 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 pursuing Ph.D. by 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 have to say. I want to express my gratitude one by one in the acknowledgments, but words and sentences are always not enough to express it. Anyway, I will have a try.

First and foremost, I would like to thank Prof. Zhao Shichen, my supervisor, that the guidance and instruction he gave to me not only helped me in my studies but more in work and life attitude. The time of drinking with Prof. Zhao is also joyful and memorable. The drinking party 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 for my co-supervisors, Prof. Sakai Takeru and Prof. Divigalpitiya Prasanna for giving me so many comments on my dissertation. They also gave me the 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 of the future.

I am 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 advice on my research.

I would like to thank my friends 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 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. 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 Ph.D. period, how to spend the boring time and adjust the pace of work is always a serious problem puzzling Ph.D. students. I would like to thank my ex-roommate at Kyushu University International House and his current-roommates. They provide us the party place, where we hold hotpot party and even play mahjong all night. I would like to thank my mahjong partners, the time spent with them was so happy. That is also the key that I can deal with my research with full power.

The most important thing always come last, all the expense during the 4 years Ph.D. 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 so much.

Abstract

Urban transport is an important part of urban planning and management. Urban traffic problems such as traffic congestion, traffic accident, and air pollution, always appear along with the development of the city, while they also restrict the development of the city simultaneously. At present, Japan is facing serious problems of aging society and low birthrate. Reflected in the local central cities of Japan, these problems are showing the trends of the migration to the urban edge and the spread of urban areas. What followed is the traffic congestion and the continuous increase in the traffic mode share of private cars. From the perspective of environmental protection and the full use of social public resources, how to improve the use of public transportation has become an important issue in all Japanese cities. To promote the utilization of public transit, it is necessary to explore and understand the determinants of transit ridership, based on which the specific improvement plan can be proposed regarding land use, buildings and facilities.

With such background and demand, this study aims to explore and quantitatively analyze the determinants affecting urban rail transit by using the rail transit system of Fukuoka, which is a typical local central city in Japan, as the study case, so as to provide references to future planning. Around the purpose of this research, this dissertation mainly explores and discusses the determinants of Fukuoka rail transit ridership from 3 aspects: passenger attributes, environmental factors around the station, and passenger transfer between station and station. According to the research purpose proposed, this dissertation is organized into six chapters, and the contents of each chapter are as follows.

In chapter 1, the research background is elaborated. Firstly, the overall research purpose is put forward on the basis of the practical problems existing in most cities around the world (particularly in the local central cities of Japan). Then through

the collation and summary of the previous studies, some specific research questions area extracted in terms of the details in analyzing the determinants of rail transit ridership. Finally, the rail transit system of Fukuoka City is selected as the research object, the content of this dissertation is organized around the research purpose and the specific research points.

Chapter 2 uses on passengers' attributes, including socio-economic characteristics and trip chain information, to explain the variation in walking access among individuals. The purpose is to grasp the utilization trend of rail transit. It has been confirmed in many studies that there is no linear relationship between the dependent variable of walking access and the explanatory variables. To make an intensive exploration of the relationship between walking access and passengers' attributes, this study examines the attribute characteristics of passengers that walked the given walking duration or more, thus trying to explain the preference for walking access. The research object is the probability that passengers walk the given threshold of walking duration or more. For the analysis method, the ANOVA is used to extract the effective passenger attribute at each given walking duration. Then the extracted effective attributes are used to predict the probability that the walking access is greater than the given threshold. The accuracy of the prediction is verified from the perspective of both group and individual. Three thresholds of walking duration 5, 8, 13-minute are selected to examine in this chapter. From the results of feature extraction and probability prediction, passengers with the attributes of more than 65 years of age, unemployment, and business, private or going home travel purposes are more likely to walk for more than 5-minute; passengers who travel for business, private or going home purposes are more inclined to accept a walking duration longer than 8-minute; only the passengers with private travel purposes are more likely to accept a more than 13-minute walking duration. According to the evaluation of the prediction results, the passenger attributes extracted from the 5-minute walking duration have relatively strong explanatory power in explaining the walking duration preference. The explanatory power at 13-minute walking duration is slightly weaker than that at the 5-minute threshold. The prediction at 8-minute walking duration has the lowest accuracy.

Chapter 3 examines all the subway stations in Fukuoka City regarding the characteristics of rail transit ridership, also the characteristics of land use around the station. The subway stations are classified into 5 types (low-density residence, high-density residence, downtown commerce, office, and education respectively) according to the land use characteristics, then the characteristics of transit ridership are summarized in terms of the 5 types. On the basis of fully grasping the characteristics of both transit ridership and land use, the relationship between land use and passenger flow is estimated by Quantification I Method. It is concluded that the building area of commerce and office is important on affecting transit ridership, while population density plays an important role in the growth of transit ridership.

Chapter 4 explores and estimates the influencing factors of transit ridership at station level. Aiming at a small sample case of Fukuoka City, this chapter proposes a method to extract effective influencing factors of the regression model, and finally, 9 effective indicators are identified. Also, to tackle the problem resulted from the spatial autocorrelation of explanatory variables, Moran index is introduced to describe this spatial autocorrelation thus discriminating the global and local indicators. Then the Mixed Geographically Weighted Regression model is used to estimate the impact on transit ridership. The results show that all the 9 selected indicators are statistically significant. Among them, raising the indicators of official office area, transportation facility area, land use aggregation, transfer station, bicycle parking, and bus accessibility can lead to an increase in rail transit ridership. According to the comparison of the residuals and the spatial distribution of residuals, it is clear that the results from MGWR have a significant improvement than that from OLR.

Chapter 5 examines the influence of land use types on passengers' choice of destination from the perspective of OD transit ridership. This chapter proposes that the probability of choosing destination station will be affected by the type of land use around the departure stations. Based on this argument, this research question is converted to a binary choice issue and is described by the logistic regression. Using the classification of stations obtained from chapter 2, several typical stations selected from each type are used as the research objects, they are used as the destination station to analyze if the land use types around the departure station affect the probability of choosing them as the destination. According to the results of logistic regression, the land use type around the departure station has significant influences on the probability of choosing the destination, based on which the argument proposed in this chapter can be verified. The results show that the probability of choosing a destination belongs to low-density residence type has no tend of raising regarding to the variation on land use types of the departure station. For the destination station of any type, the education land use of the departure station contributes to an increase in the probability of being chosen as the destination. The probability of destination station being chosen tend to decrease if the land use type is similar between the departure and destination stations.

Chapter 6 summarizes the findings and conclusions of each chapter in this dissertation. In the recommendation part, the feasibility of integrating passenger attributes, environmental factors around the station, and ridership transfer between OD station to predict rail transit ridership is put forward, which is expected to be used to improve the accuracy of transit ridership forecasting.

Keywords: Urban planning; Rail transit ridership; Walking duration; Land use; Passenger attributes; Station level; Station-to-station level; Catchment area

Chapter 1

Introduction

1.1 Background

1.1.1 The demand 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 of the world lives in urban areas by 2014, this population shift will continue 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 which are now experiencing a rapid change caused 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. Unlike developing countries, developed countries mainly have the problems of aging population and low birthrate. With a rapidly increasing proportion of aged people in the population, governments are forced to increase expenditure of 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, both 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 tackling the economic slowdown which is caused by 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 its 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 of 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 of travel demands, taking rail transit station as the core of a newly planned area is becoming more and more common. 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

What explains rail transit ridership? How to increase the rail transit ridership? As interpreted before, these questions are placed in front of us. The answers seem to be both obvious and complex. The elements existing within the catchment area of stations should associate with rail transit ridership, like population, road network, parking, income, transit network, building density etc. But the relative importance of these various factors and the internal relationships 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 an area that the rail transit station can serve. 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 on the utilization of rail transit***. Centering on this theme, three subtopics are proposed to be discussed in this dissertation.

1. Exploring the factors influencing walking duration to transit stations.
2. Exploring the factors influencing transit ridership at station-level.
3. Exploring the factors influencing transit ridership at station-to-station level.

1.3 Literature review

Based on the research purpose and the collation of previous studies, rail transit ridership is commonly thought to be determined by the catchment area of stations and the environmental factors around the stations, and the accuracy of estimation on the factors' influence mainly depends on the choice of model. Therefore, the literature review is organized from the catchment area, method, influencing factor and 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

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			•						

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). However, 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 real road network. According to the existing studies, the PCA generally ranges 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-meter (half-mile) walking distance, which has been verified to be suitable for most cases, has been widely accepted as a principal reference to determine the catchment area (Kuby, Barranda, & Upchurch, 2004; Gutiérrez, Cardozo, & García-Palomares, 2011; Cardozo, García-Palomares, & Gutiérrez, 2012; J. Zhao, Deng, Song, & Zhu, 2013). This 800-meter, however, is just a general range of the catchment area for rail transit stations but not considering 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 is passenger 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 determine the catchment areas of different stations even in different cities and countries. Till now, restricted by analytical method and data, there is no better option than this general 800-meter walking distance in determining rail transit catchment area.

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, extensive 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 put forward in four-step models, range from the scale of blocks 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 a neighborhood-scale taking transit stations as cores, not to mention the estimation of influencing factors of the transit ridership (Cervero, 2006; Chu, 2004; Duduta, 2013). Additionally, the trip generation stage in the four-step approach is usually conducted on the base of empirical model by using socioeconomic variables like population, employment, auto ownership. However, the impact of traffic indicators will not be exactly the same in different cases (Jones & Nichols, 1983). It is also unclear that whether these selected indicators significantly relate to traffic volume. linear relationship between traffic volume and these selected indicators.

Direct models estimate 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 model has gained popularity in estimating travel demand at neighborhood-scale. The advantages of direct models firstly stem from the ease of estimation. The only critical requirements are GIS, statistics for built-environments around stations, and the corresponding transit ridership (Guerra et al., 2012). The accuracy of prediction is also one of the advantages. It is easier and faster than other travel demand models, meanwhile, it can also provide strong predictive power (Lane, DiCarlantonio, & Usvyat, 2006). Compared with the disadvantage mentioned in four-step models, since the estimation of direct models is based on the historical statistics, it can justify the importance and validity of the indicators expected to have influence on the 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 places for human activity, and it has been widely proved to have a strong relationship to ridership. Also, land-use diversity has a significant effect on ridership since it reflects the balance between traffic demands and supply in 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 to passengers going to take public transit. Better accessibility is thought to be attractive to passengers that are living far from the stations. The factors of 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 thought to be more attractive. Because a terminal station 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, if the car ownership is higher in the catchment area of a rail transit station, people living there are more likely to choose private car than public transit (Chiou, Jou, & Yang, 2015; F. Zhao, Chow, Li, & Liu, 2005). A higher percentage of low-income household can lead to an increase in the utilization of 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 fields 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 systematic 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. For the estimation model, little progress has been made in the direct model so far. Most existing studies are conducted based on the 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 to select valid factors of specific cases are still insufficient.

On the base of fully understanding the existing literature, this dissertation will be expanded upon 4 specific research points corresponding to the main purpose.

1. Exploring the influencing factors of the walking duration to transit stations in terms of passengers' personal attributes.
2. Summarizing and analyzing the characteristics of rail transit ridership, land-use, population etc. using the case of Fukuoka city.
3. Exploring the factors influencing transit ridership at station-level with the small sample study case of Fukuoka.
4. Exploring the factors influencing transit ridership at station-to-station level.

1.5 Dissertation organization

The dissertation uses the rail transit system of Fukuoka City, which is the fifth 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 passengers' personal attributes and surveyed walking duration to rail transit station.

- ***Chapter 3*** clarifies the characteristics of transit ridership and land-use, and explores the overall relationship between them.

- ***Chapter 4*** explores the factors influencing transit ridership at station-level

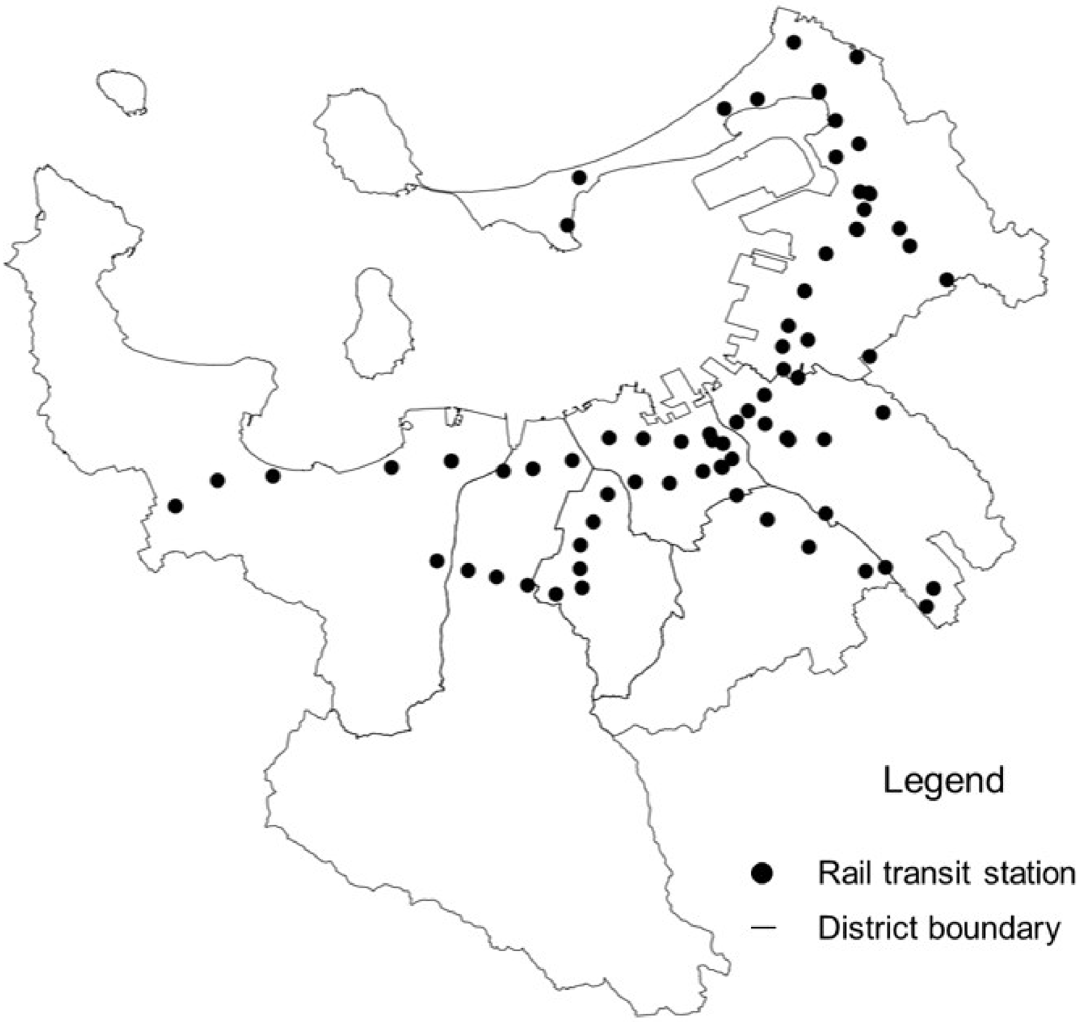


Figure 1.1: Research Area

focusing on some key questions in the small sample case.

- ***Chapter 5*** explains the influence of land-use patterns on transit ridership at station-to-station level.
- ***Chapter 6*** makes a conclusion of the three research point, and prospects the next stage of this research.

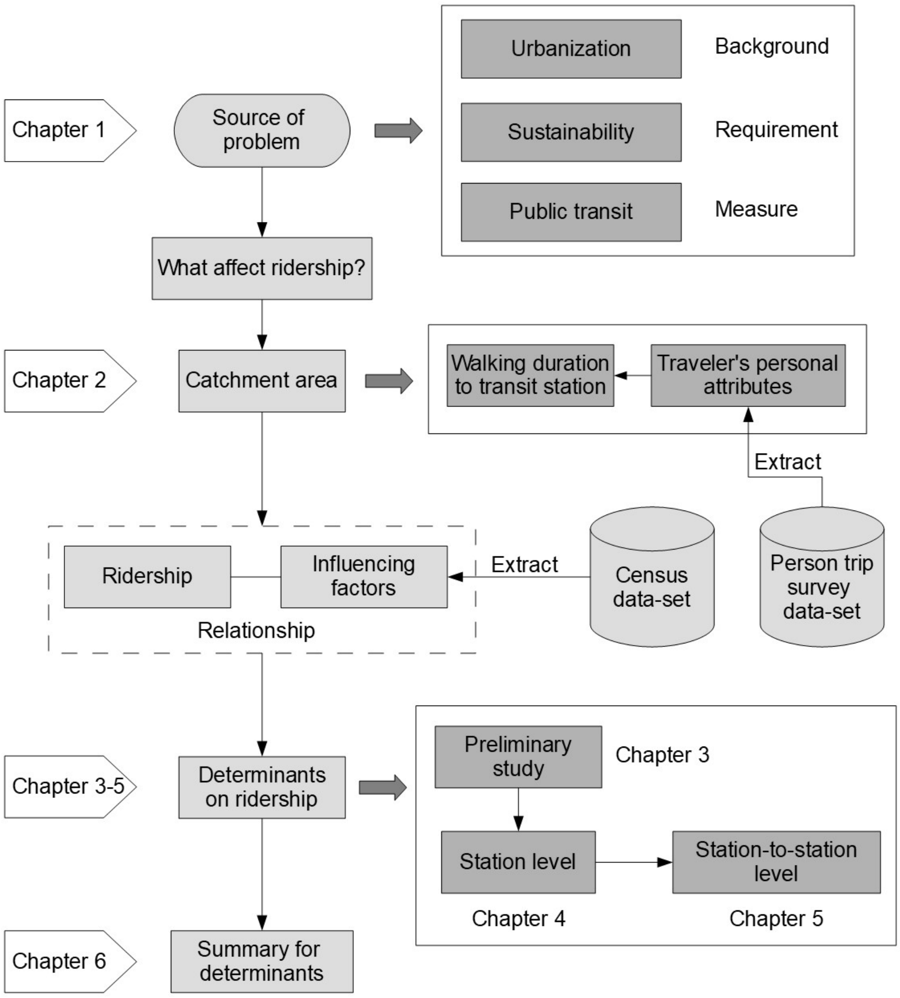


Figure 1.2: Flowchart of the Dissertation

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

Exploring the preference of walking duration to rail transit stations by individual attributes

2.1 Introduction

2.1.1 Background

Walking is the most important way to reach a transit station, people choose to take transit only when the walking access is within the acceptable range. People's willingness to walk to transit has largely defined the catchment area of a station. This catchment area determined the population served by that station, for which it is also widely accepted as a key system performance measure of the transit station (Fielding, Glauthier, & Lave, 1978; El-Geneidy, Grimsrud, Wasfi, Tétreault, & Surprenant-Legault, 2014; Chia, Lee, & Kamruzzaman, 2016).

Walking access, either distance or duration, has been paid extensive attention to determine the catchment area of stations. However, delineating this area is a complex issue as the catchment area varies due to the different transit types and people's walking preferences. At present, the 800 meters (half-mile) transit catchment area, whether radial-based or road-network-based, has been widely accepted as a principal reference to the catchment area around rail transit stations (Kuby, Barranda, & Upchurch, 2004; Gutiérrez, Cardozo, & García-Palomares, 2011; Cardozo, García-Palomares, & Gutiérrez, 2012; J. Zhao, Deng, Song, & Zhu, 2013), and a 400 meters (quarter-mile) catchment area is commonly defined around bus stops (O'Neill, Ramsey, & Chou, 1992; F. Zhao, Chow, Li, Ubaka, & Gan, 2003). On the other hand, researchers also used some other catchment areas due to different research purposes and research contents. For example, a relatively small area of 250-meter buffer was adopted to analyzing the quality of access to the BRT stop (Estupiñán & Rodríguez, 2008); a 500-meter buffer area was used in finding the factors generating boarding at metro stations in Seoul. Korea (Sohn & Shim, 2010). Jun et al. segmented the catchment area into the core area, primary area, and secondary area, corresponding to 300 meters, 600 meters, and 900 meters pedestrian area respectively (Jun, Choi, Jeong, Kwon, & Kim, 2015).

2.1.2 Statement of problem

The most widely used 400-meter or 800-meter catchment area is commonly obtained from the sampling survey by loosely asking how far or how long 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. These simplified catchment areas were applied to all the transit stations ignoring the differences in stations and the passengers, obviously, this "one size fits all" solution to catchment area should not be effective (F. Zhao et al., 2003; Chia et al., 2016). As to passengers, the acceptable walking access should not only relate to the station types, but also affected by the walking preference of passengers. Walking preferences, however, vary by destination, trip purpose, gender, age, land-use, safety, weather, and the price and availability of parking.

Various factors belong to the individual attributes pertaining to walking access have been attempted and confirmed in the literature, such as trip purpose, income level, occupation etc. (Alshalalfah & Shalaby, 2007; Weinstein Agrawal, Schlossberg, & Irvin, 2008; Chia et al., 2016). However, little is known about the quantitative relationship between walking access and individual attributes, studies based on statistical description take account the majority. Even though some studies attempted regression approach to estimate the influence of individual attributes on the walking access, very little obtained a significant result (Jiang, Zegras, & Mehndiratta, 2012; Daniels & Mulley, 2013). It seems that the linear relationship between waking access and individual attributes does not exist.

2.1.3 Research purpose and contents

This study aims to explore the quantitative relationship between walking access and individual attributes and try to predict the walking preference of accessing transit stations using individual attributes. The study case used in this study is the rail transit passengers of Fukuoka City, Japan. The contents of this study are organized as follows: section two reviews the previous research on the walking access to transit station, based on which the research questions are put forward; section three interprets the methodology; section four gives the descriptions of data; section five shows the process and results of analysis; section six discusses the results and draws the conclusion.

2.2 Literature review

In the studies of walking access, 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 acceptable range of walking access, but just the walking duration/distance between departures and stations. For this problem, some studies chose a different perspective to explain the walking access 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.

Another point of this issue is whether using the walking distance or walking duration as the research object. To date, there are a lot of studies working on the relationship between walking access and individual attributes, some of them argued that an individual has a limited amount of time spending on traveling in one day. People with faster walking speed tend to accept a further walking distance

(Marchetti, 1994; Larsen, El-Geneidy, & Yasmin, 2010). With this reasoning it is easy to think of that passengers with the same individual attributes may have the similar willingness of walking duration, but they generally have different willingness of walking distance due to different travel speed.

Among the influencing factors of walking access, individual attributes are generally thought to be the key that can affect walking distance (Besser & Dannenberg, 2005; Weinstein Agrawal et al., 2008; Krygsman, Dijst, & Arentze, 2004; Yang & Diez-Roux, 2012; Daniels & Mulley, 2013; Guerra, Cervero, & Tischler, 2012), whereas, little literature has clearly verified the quantitative relationship between individual attributes and walking access, even there is a study suggesting the walking distance should not be viewed as a function of individual attributes (Krygsman et al., 2004). Several studies have confirmed the role of travel purposes in determining walking distance, the commute trips showed particularity from the other purposes. People who have the travel purpose of commute tend to accept a longer walking access 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 access can be influenced by some specific kinds of factors, such as socio-demographic characteristics, trip purposes, and built-environment, but the problem is how the walking access is influenced by those factors.

As interpreted above, the walking access is a just direct reflection of the distance between departures and stations, which should not have a direct relationship to individual attributes. Some studies attempting regression model to deal with this problem also supported this argument that no significant quantitative relationship was found between walking access and individual attributes. However, the results from extensive studies working on the qualitative description of the walking distance distribution have clearly shown that individual attributes have an impact on the walking access to rail transit stations. It is supposed that it may due to some inappropriate understanding of the surveyed walking access, and inappropriate assumption of the relationship form between individual attributes and walking access. Focusing on the probable inadequacies summarized from the previous studies, the flow of the research process is organized as Figure 2.1.

2.3 Methodology

2.3.1 Description of realistic problem

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 to be transacted at the price acceptable to consumers. This accepted transaction price means that any price cheaper than that can be accepted by the consumers. As a result, consumers with various consumption-ability will have different characteristics distribution of any specific transaction price. Obviously, the same for rail transit users. For the operator of rail transit, it is important to know how much the cost that consumers are willing to pay, based on which thereby making rail transit more attractive. But unlike the general service and commodities, the cost refers more to the convenience but not the fare for passengers. Because the cheap enough fare is already not an important element for passengers to decide whether using rail transit, while the walking access to stations has become an important determinant.

For the issue of riding rail transit, before a potential passenger makes the decision on walking to a station, the walking access, which is the ”price” for this potential passenger, is just the reflection of the distance between departures and stations. In case this walking access is accepted by this potential passenger, this trip will happen and it can be surveyed. Based on the analogy given above, it’s easy to think of that passengers with different individual attributes should have different willingness towards walking access to stations, and the surveyed walking access can be viewed as the transaction price which has been accepted by passengers. It follows that a less walking access will lead to an increase in the willingness of using rail transit, which reflected on the survey data will be that the records of shorter walking duration are more than that of longer walking duration.

2.3.2 Conversion to mathematical problem

The study aims to find the relationship between surveyed waking duration and passenger attributes. Based on the above comparison between the behaviors of

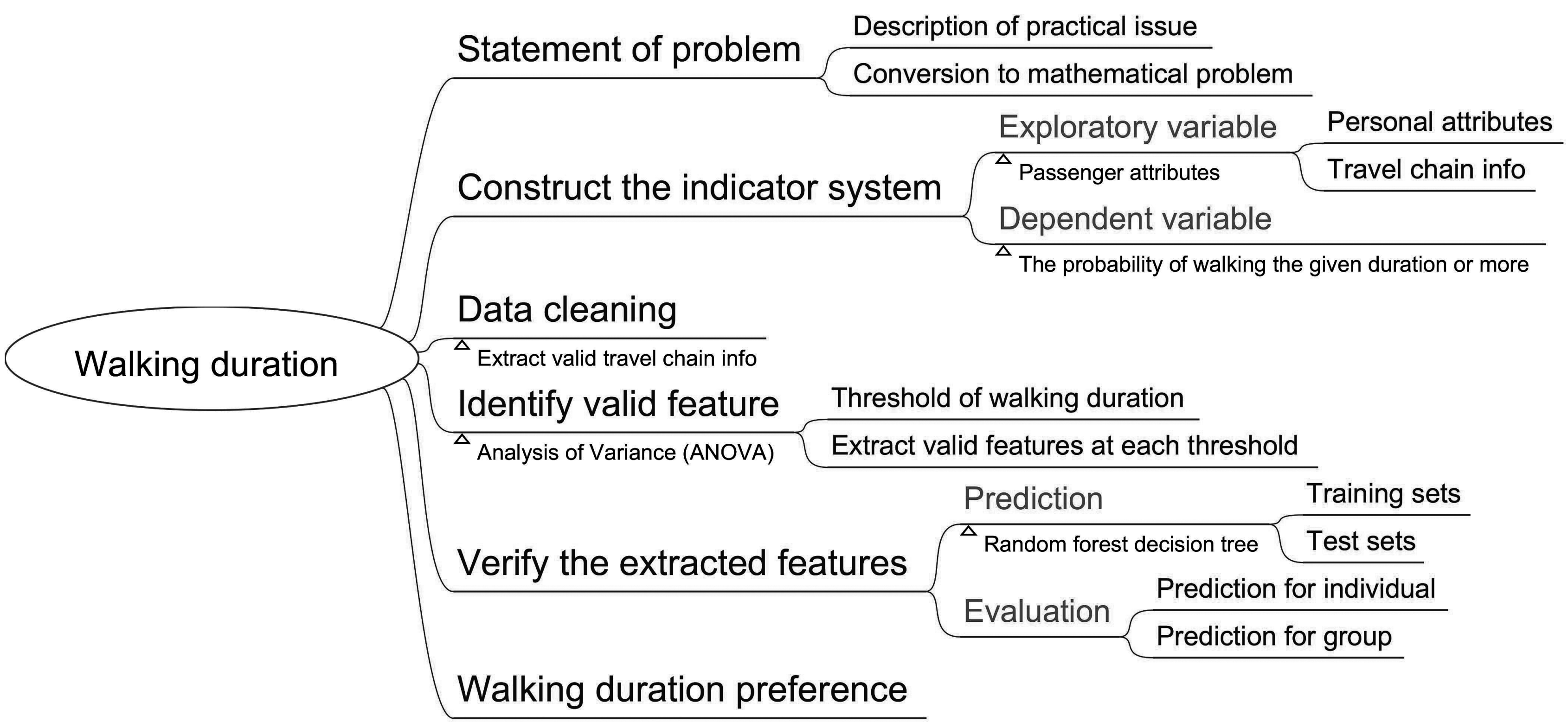


Figure 2.1: Research flowchart

using transit and consumption, it can be concluded that passenger attributes have an impact on the willingness of walking duration, but not on the surveyed walking duration. The abstract functional relationship between walking duration and passenger attributes is given in 2.1. This abstract function indicates that we need to find the expression of walking duration to describe the passenger willingness, also to find out the functional relationship between the walking duration expression and passenger attributes.

$$Expression(Walking\ duration) = Function(factor_1, factor_2, factor_3, \ldots) \tag{2.1}$$

From the perspective of consumption-ability interpreted before, if giving a specific threshold of walking duration, the probability of accepting this threshold will vary due to different passenger attributes. Based on the interpretation, this mathematical problem can be converted into that to examine the impact of passenger attributes on the probability of accepting a given threshold of walking duration. The model can be expressed as follows: Equation 2.2 gives the expression of walking duration describing the passenger willingness by Y_i^T, and Equation 2.3 is the function of this willingness and passenger attributes that the probability of walking the threshold T or more in terms of the passenger attributes of X_i.

$$\begin{cases} Y_i^T = 1, & (t_i > T) \\ Y_i^T = 0, & (t_i < T) \end{cases} \tag{2.2}$$

$$P(Y_i^T = 1 \mid X_i) = F(X_i) \tag{2.3}$$

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. $Y_i^T = 1$ means passenger i walked a longer walking duration than the threshold of T; while $Y_i^T = 0$ means walked less than the threshold.

X_i is the vector whose component is the individual attributes passengers.

2.3.3 Basic hypothesis

A large number of existing studies have shown that passenger attributes can affect walking access to rail transit station. Combining with the statement of problem conversion, this study proposes the first basic hypothesis that, passengers with different attributes have distinct walking behavior to access rail transit. In addition, since walking access is the direct reflection of the distance between transit stations and departures, the distribution of passengers' departures also affects the walking access to transit. The second hypothesis aims to simplify this problem by ignoring the distribution of passengers' departures, thus focusing on the influence of passenger attributes to the walking access. Moreover, since the dependent variable in this study is converted in to the probability of accepting the given threshold of walking duration, the third hypothesis is proposed to ensure the answered walking duration is an accepted one. The specific definitions of the hypothesis are as follows.

- H1: The willingness of walking duration of people with the same attributes should subject to the normal distribution.
- H2: The departures and destinations of people with different attributes in Fukuoka are randomly distributed.
- H3: The respondents are viewed as they can accept the walking duration that they answered.

According to the H1 and H2, if considering the walking duration T, set the proportion of k group in the whole surveyed sample as r_k. The proportion of passengers walking less than the given threshold T is marked as $r_k^{<T}$, while the proportion of passengers walking more than the given threshold T is marked as $r_k^{>T}$. If passenger attributes have no significant correlation to the accepted walking duration, then the three index r_k, $r_k^{<T}$ and $r_k^{>T}$ will not show significant differences. Otherwise, the three proportion should show significant differences, and the differences will show regularities at different threshold T. Figure 2.2 shows the expected differences in the distribution of investigated walking duration in terms of passengers with different attributes.

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 longer than that he answered, but this study just concerns about if he can accept the given threshold of the walking duration other than how long he can accept.

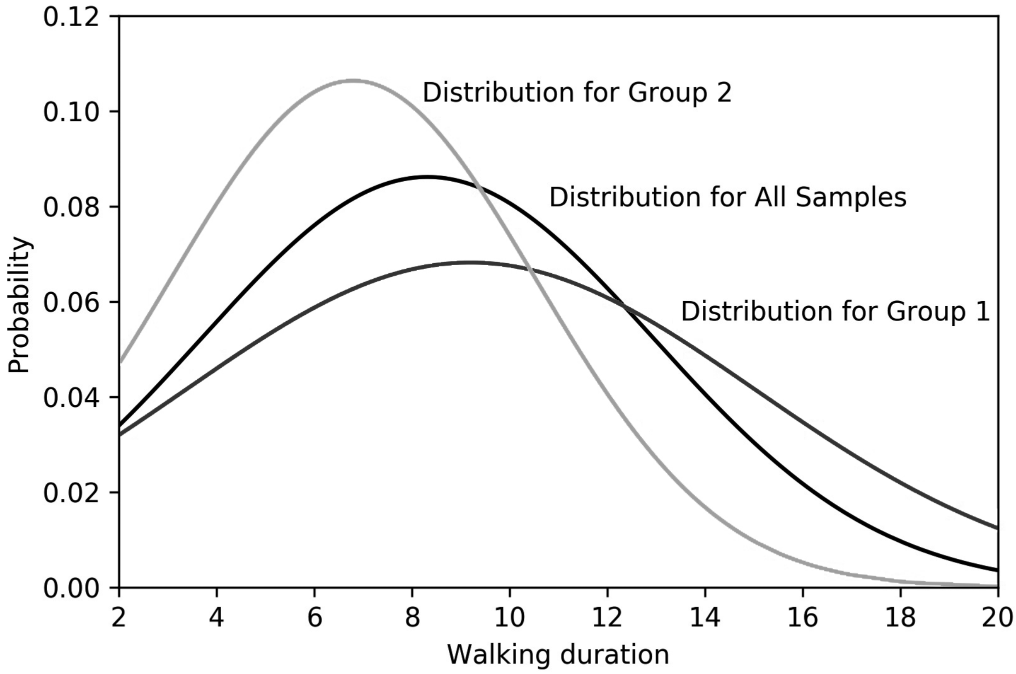


Figure 2.2: Hypothesis of walking duration distribution

2.4 Data

2.4.1 Case introduction

The research case of this study is Fukuoka, Japan. All the rail transit stations and rail transit passengers within the urban area of Fukuoka are investigated as the research objects. The data set is extracted from the Northern Kyushu Area Person Trip Survey, which is conducted about every 12 year, 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 2.3 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.

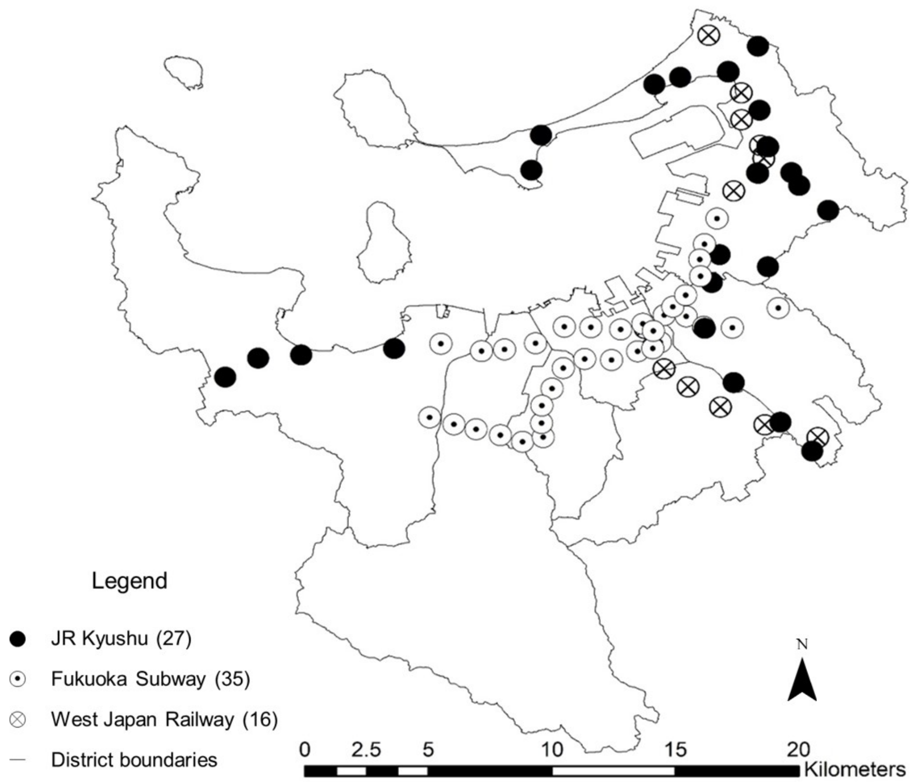


Figure 2.3: Distribution of transit stations

2.4.2 Data preparation

The original data covered the range of all the main cities in Northern Kyushu Area, which has 483,556 records of trip chaining behavior. The available data onto this study mainly includes trip chaining behavior and socio-demographic characteristics, as shown in the Table 2.1.

To analyze the walking duration from departures to 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 483,556 records. 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 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 2.4, at last the valid data set is reduced to a size of 4,254 trips.

Figure 2.5 shows the age distribution for walking trips to rail transit stations based on the finally valid data set. The passengers aged from 25 to 55 account for the majority of the whole passengers, while schoolchildren aged under 15 rarely

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

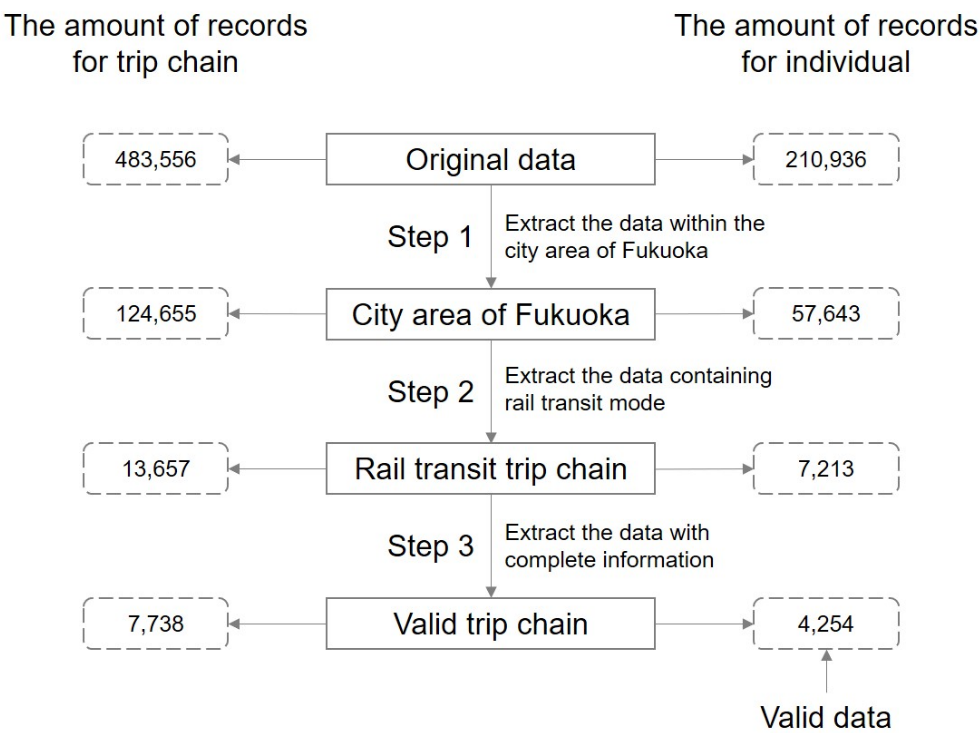


Figure 2.4: Process of data cleaning

take rail transits. The distribution graph does not show significant peak values at any specific age group. Figure 2.6 shows the distribution of real walking duration to stations, it has a mean value of 8.32, and the standard deviation is 4.63. Notably, there are several peak values at multiples of five. 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 investigation result. Despite the bias between survey and reality, as the hypothesis H3 proposed before, passengers are viewed that they can accept the walking duration what they answered. Therefore, the peak values are considered to have no significant influence on the analysis.

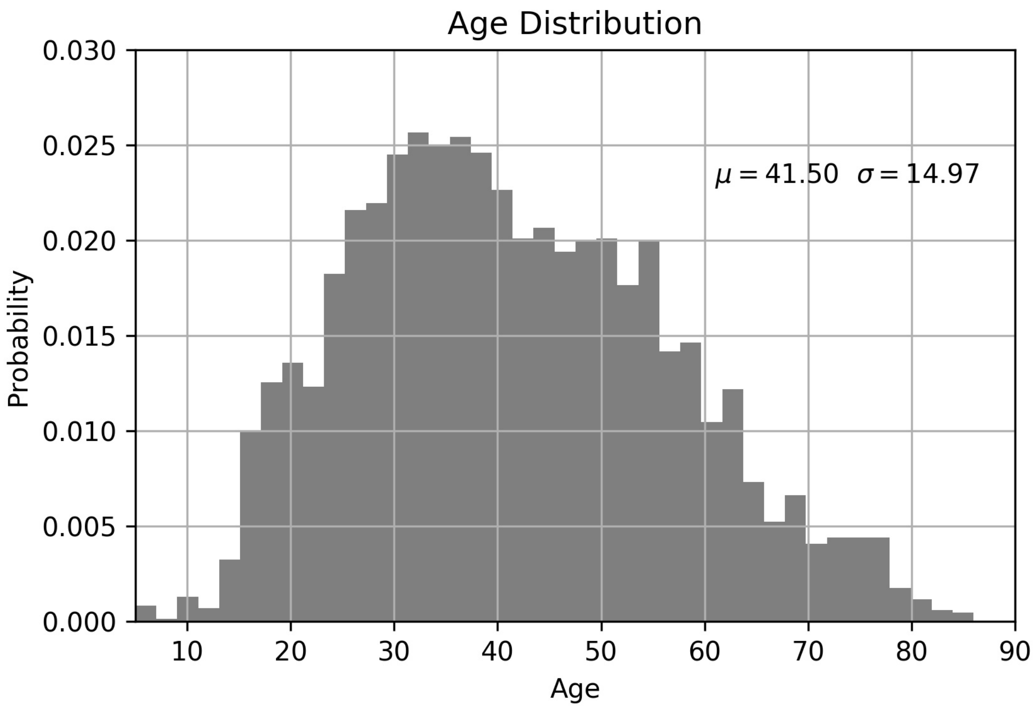


Figure 2.5: Age distribution of passengers

2.4.3 Data description

The feature of trip purpose has 15 subcategories in the original data set, this study reclassifies them into 5 categories including commuting to work, commuting to school, official business, private purpose (such as shopping, entertainment), and

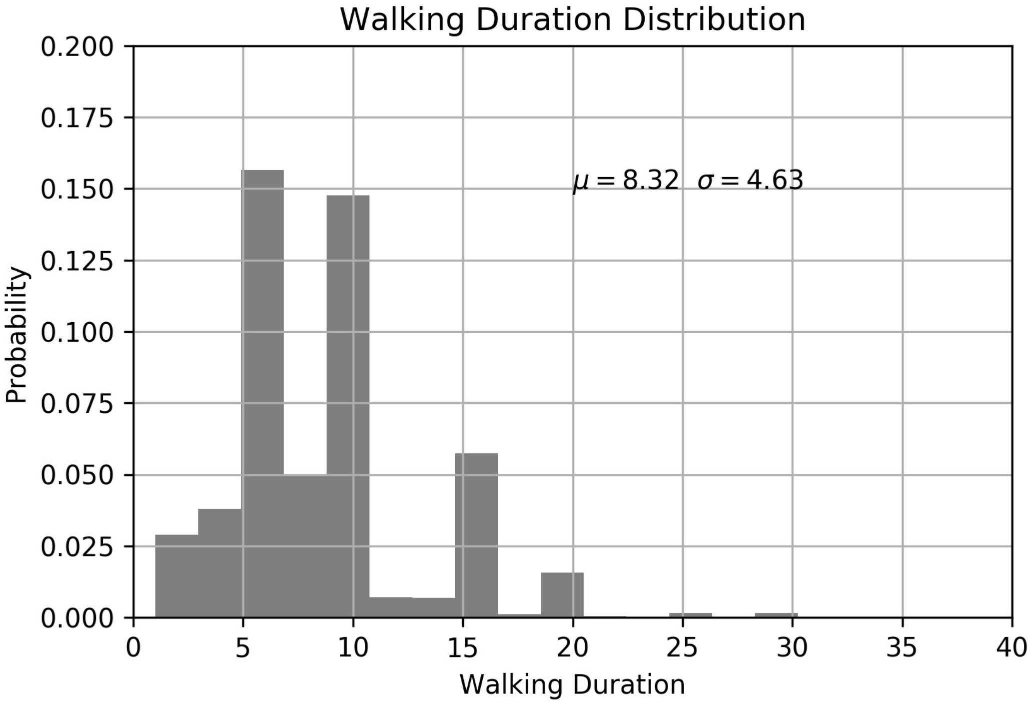


Figure 2.6: Distribution of walking duration

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 of each feature, the differences are summarized as follows.

- People with the purpose of commuting account for the majority in the rail transit users.
- Most people do not accept a walking duration to transit stations more than 15 minutes.
- The walking duration during 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 for 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.

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

2.5 Analysis

The independent variable in this study is the probability of walking the given walking duration or more. Based on the above description of the data set, several typical thresholds of walking duration are selected to explore the passenger walking preference. According to the literature, walking access to rail transit stations generally ranged 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 et al., 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). In the case of Fukuoka City, more than 40% of the samples walk less than 5 minutes, and the walking duration less than 13 minutes covers about 85% of the total samples. It can be inferred that the 5-minute walking duration is easy to accept, while a walking duration longer than 13 minutes is thought to be hard to accept. In addition, the average walking duration of this study is about 8 minutes, it can be considered that this median value is ambiguous to be accepted. As a result, the three representative thresholds of 5, 8, 13 minutes are picked as the typical threshold for estimating the relationship between individual attributes and walking duration.

At each threshold of walking duration, the exploration of the walking preference in terms of passenger attributes is extended to three steps. The first step is to identify the effective features using the Analysis of Variance (ANOVA). Secondly, training the model to obtain the functional relationship between the dependent variables and the independent variable. With this trained model, inputting a new set of passenger attributes can obtain a probability of walking the given walking duration or more. Finally, inputting the test data set into this trained model to evaluate the model description ability.

2.5.1 Effective attributes selection

The effective features are identified using ANOVA. The feature which has the p-value less than 0.05 (this feature relevant with the dependent variable at the confidence level of 95%) is picked out and listed in Table 2.3. In the column of "Effect" in this table, the ”M” means the passengers with this feature tend to walk more than the given threshold of walking duration, while the ”L” is opposite that the passenger with this feature tends to walk less than the given threshold.

As shown in 2.3, at the threshold of 5-minute, the features of trip purposes and peak hour play the most important role in determining the willingness of walking duration. Situations are changed at the threshold of 8-minute, the importance of age and gender raised in some extent, while the features of trip purposes are not changed from that of 5-minute. The features of 13-minute show obvious differences from that of 5-minute and 8-minute. 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 ranging from 5 to 13 minutes are sensitive to individual attributes. The walking duration more than 13 minutes are not accepted by most people even if they have different individual attributes. The threshold less than 5 minutes is also not sensitive to individual attributes since the 5-minute threshold is generally accepted by most passengers.

2.5.2 Feature estimation

This issue has been converted into a binary choice problem as stated before. The models of the decision tree, Bayesian, support vector machine (SVM), logistic regression, and neural network are widely adopted to estimate this type of problem. With the limitation of the volume of sample and features, also the unknown feature distribution, the decision tree model is considered to 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, also to reduce the possibility of over-fitting of data which is easily occurred in the decision tree model, this study introduces an improved model of the decision tree, the random forest decision model, to tackle this problem. 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 when training the model (Ho, 1995, 1998).

The effective features identified by the ANOVA (Table 2.3) are used in 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	M	Female	L	Age 45-64	L
Peak hour	L	Age 25-44	L	Peak hour	L
O_Null	M	Peak hour	L	P_commuting to work	L
P_commuting to work	L	P_commuting to work	L	P_ private purpose	M
P_official business	M	P_official business	M		
P_private purpose	M	P_ private purpose	M		
P_going home	M	P_ going home	M		

***Note:** The ”M” means passengers with this feature tend to walk more than the given threshold of walking duration, while the ”L” is opposite.

random forest decision to estimate the probability of walking the given walking duration or more. As to the estimation, the data set is divided into two parts, 50% of the sample are used for fitting the model thus obtaining the functional relationship, the rest 50% are used for testing the ability of prediction. The prediction process is presented in Figure 2.7. For the 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 2.3 based on the mean prediction.

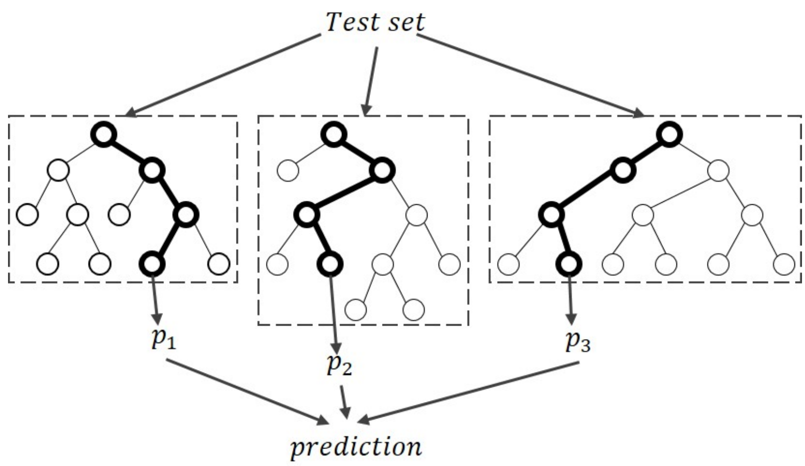


Figure 2.7: Prediction process in random forests decision

Note: The final prediction is obtained by the vote of results obtained from the same set of features in different trees

2.5.3 Prediction and evaluation

The prediction is obtained from the forest random decision model, the flow is shown in Figure 2.8. The accuracy of results is evaluated from the perspective of both individual and group. The summary of the prediction results is shown in Table 2.4.

Figure 2.8: The flow of getting prediction from a trained model

The prediction for individuals is evaluated by Cohen's kappa coefficient (K). Cohen's kappa coefficient is a statistic which measures inter-rater agreement for

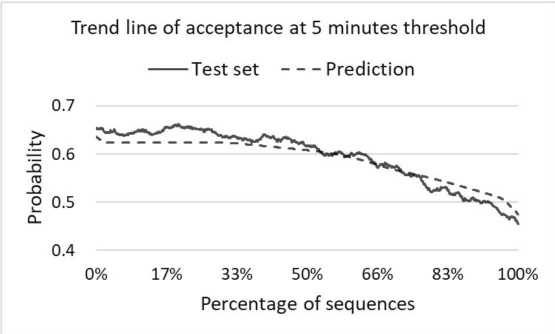
Table 2.4: Confusion matrix of the prediction at each threshold

		Prediction		
		0	1	Total
5-minute	0	394	494	888
test data set	1	377	862	1239
	Total	771	1356	2127
		0	1	Total
8-minute	0	699	442	1141
test data set	1	588	398	986
	Total	1287	840	2127
		0	1	Total
13-minute	0	1404	399	1803
test data set	1	229	95	324
	Total	1633	494	2127

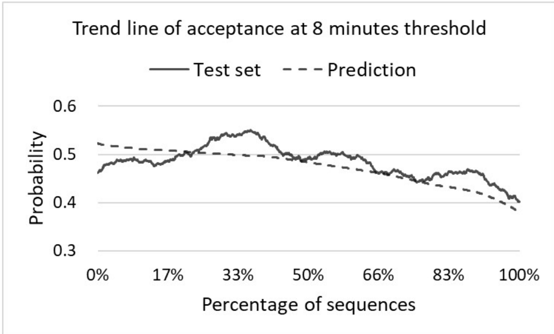
categorical items. It is generally thought to be a more robust measure than simple percent agreement calculation, as K takes into account the possibility of the agreement occurring by chance. Based on the confusion matrix of the prediction result, the Cohen's kappa coefficients of each the walking duration threshold 5, 8, 13 minutes are calculated as shown in Table 2.5. Magnitude guidelines are also put forward in the literature. Perhaps the first was Landis and Koch (Landis & Koch, 1977), who characterized values < 0 as indicating no agreement and 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1as almost perfect agreement.

The prediction for groups is expressed by using using the method of simple moving average, and is evaluated by the coefficient of determination (R^2). Figure 2.9 is the comparison between the trend lines of test data set and prediction. The trend line of the surveyed values based on the test data set is calculated by the mean probability of a group people that have close predicted values. The trend line is drawn by a descending order for predicted values. From the comparison of predicted values and surveyed values, the coefficients of determinations at each threshold are obtained. As is shown in Table 2.5 the prediction at the threshold of 5 minutes is better than that at the threshold of 8 and 13 minutes, which has a R^2 of 0.843. The prediction for the 13-minute is slightly good for a R^2 of 0.426, while it is not so good for the case of 8 minutes (R^2 is 0.221).

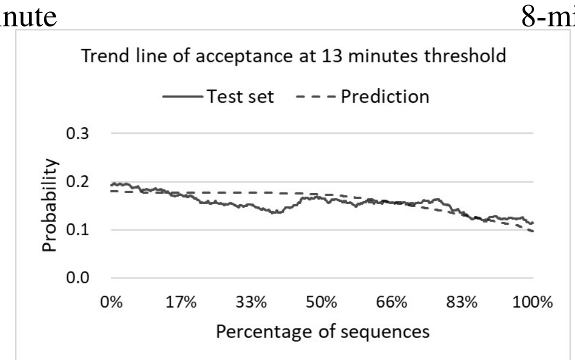
According to the evaluation, if checking the prediction for individuals, the accuracy of this model is still not enough to explain the individual behavior. But the trend line of group prediction infers that this result can reflect the behavior of people with specific individual attributes at a given threshold of walking duration.



5-minute



8-minute



13-minute

Figure 2.9: Trend line of prediction and test set

Table 2.5: Evaluation of the prediction

	5-minute	8-minute	13-minute
Kappa index	0.142	0.016	0.060
Coefficient of determination	0.843	0.221	0.426

2.6 Discussion and conclusion

This study described and analyzed 4254 records of rail transit trip. Three thresholds of walking duration are examined. From the result shown in Table 2.3, trip purpose is the most important factor of determining the walking duration at all the 3 thresholds. The feature of peak hour is also significant in explaining the walking duration. People tend to walk a longer time to stations at peak hours, while people with private purpose or on the way going home is 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 of the distribution of walking duration in terms of each feature which is hard to apply. In fact, the data onto this study is obtained from a factual investigation but not a willingness survey. Once people have made a decision on 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 attributes affected the walking duration, maybe due to this reasons few existing studies can explain the relation between walking duration and individual attributes quantitatively correctly. Under the hypothesis proposed before, the distribution of walking duration can be viewed as a reflection of the acceptability of walking duration, which should have relationship to people's individual attributes. According to hypothesis 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 hypothesis H1 and H2, if the people that accepted the given threshold of walking duration shows significant differences in individual attributes, it means they have a different acceptability of walking duration to the others.

As explained above, this study examined the differences in individual attributes of passengers that accepted the given thresholds of walking duration. As the results, people with different individual attributes shows different acceptability at each threshold of walking duration. According to the evaluation from the method of simple moving average, the model of 5-minute threshold has a better explanatory ability, the model of 13-minute threshold is a little weaker, and the model of 8-minute 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, the mean value, and the upper boundary of the main distribution of walking duration respectively. Since the commonly acceptable walking duration ranges from 5 to 13 minutes, it can be inferred that the passengers having different preference to walking duration may have significant features. However, as the mean value of walking duration, it can be thought that most of the walking duration 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 of both 5-minute and 8-minute threshold, which means people with those features are not sensitive to the threshold from 5 to 8 minutes.

The results of the random forest decision showed some predictive power for groups of people, especially at the threshold of 5 minutes. However, it did not give available prediction on individual predictions, which means the application of this study needs a certain number of sample scale, there is still a lot of work need to be done on improving the accuracy of prediction. In the next stage of this research, we plan to apply the results to predict the willingness of a group of people at a specific walking duration. 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.

Chia, J., Lee, J., & Kamruzzaman, M. (2016). Walking to public transit: Exploring variations by socioeconomic status. *International journal of sustainable transportation*, *10*(9), 805–814.

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.

El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., & Surprenant-Legault, J. (2014). New evidence on walking distances to transit stops: identifying redundancies and gaps using variable service areas. *Transportation*, *41*(1), 193–210.

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.

Fielding, G. J., Glauthier, R. E., & Lave, C. A. (1978). Performance indicators for transit management. *Transportation*, *7*(4), 365–379.

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.

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.

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.

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.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.

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'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.

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

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 be more than one quarter of the total after 10 years (as shown in Figure 3.1). On the other hand, the data of 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 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 decrease in the income and an increase in financial pressure in the case of Fukuoka. The same problem will also occur in most of the local central cities similar with Fukuoka. Addition to the financial problem, traffic congestion is also becoming the problem to all the resident. Figure 3.3, which is quoted from Road Traffic Census (2010), shows the average travel speed during crowd time for the major cities of Japan. As we can know from this figure, the problem of traffic congestion is becoming more and more serious for the downtown area of Fukuoka.

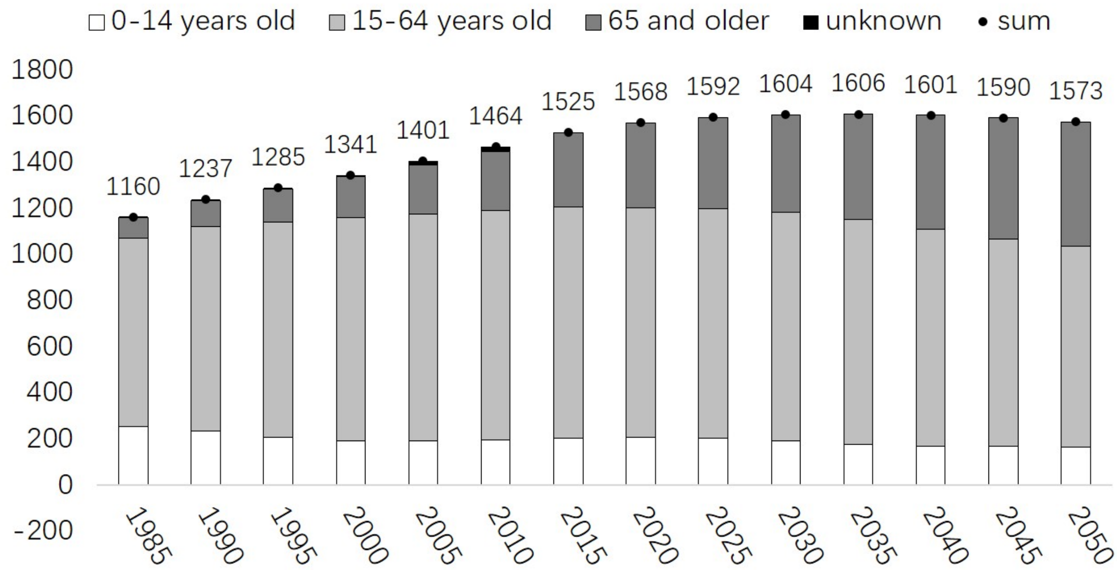


Figure 3.1: Trends of population change

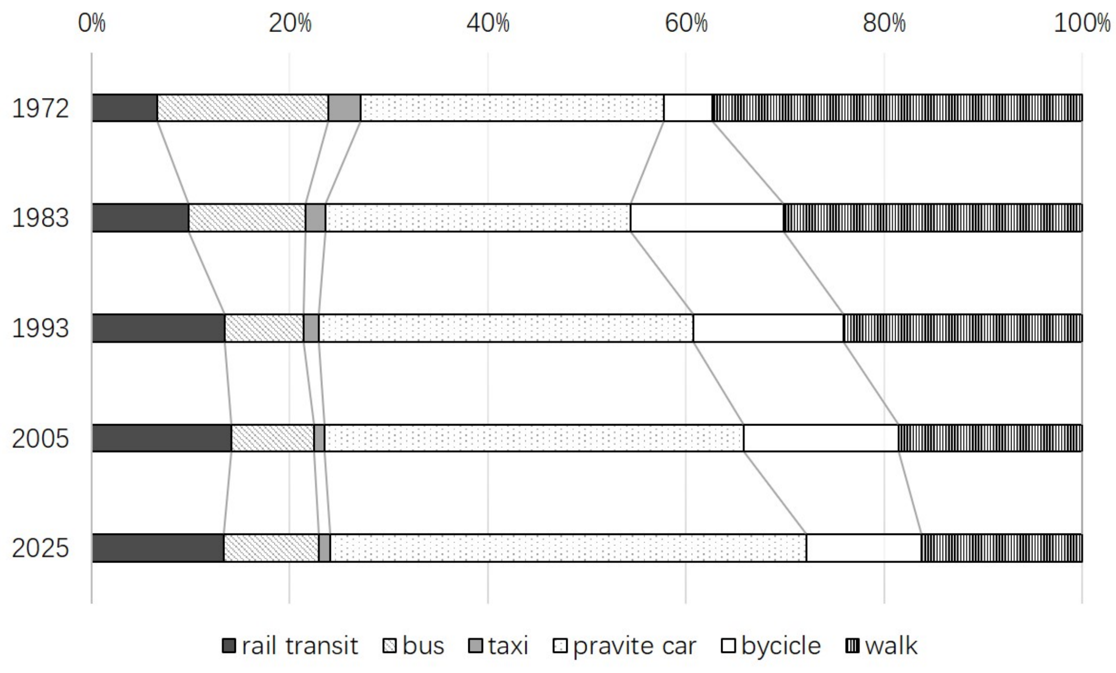


Figure 3.2: Traffic mode share rate

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 to reverse 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

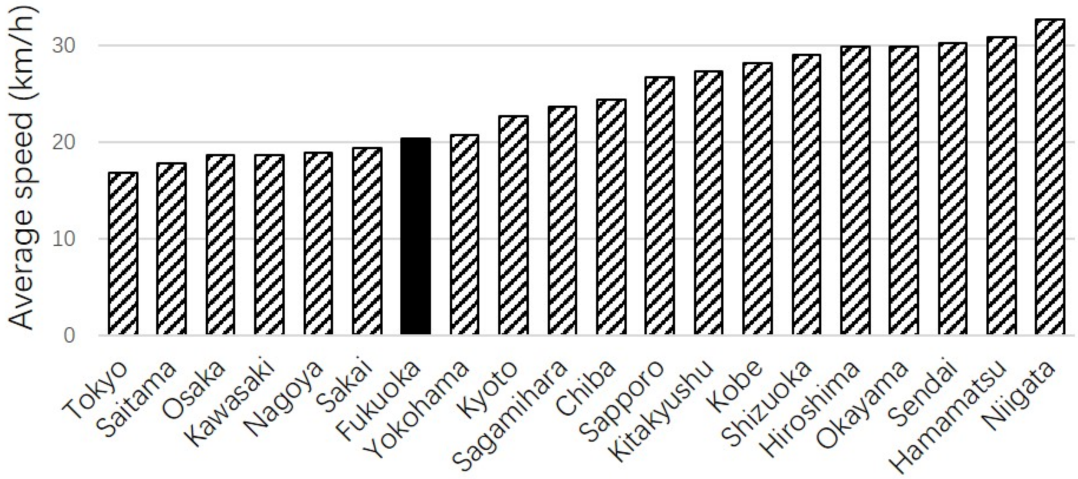


Figure 3.3: Private car travel speed ranking

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 about 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. It mainly focuses on making clear about the changing trends in rail transit ridership itself, thus providing reference to 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 thought 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 drawn 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 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 to the case of Japanese cities. A study using the case of rail transit stations within Tokyo metropolitan indicated that the land-use factors of residence, office, and education have significant influences on rail 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 to 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 about the characteristics of annual change in rail transit ridership and the land-use around the transit stations. Then Based on the characteristics 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 objects

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

1. The catchment area of subway stations is covering 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 of transit ridership can be easily confined to the area around transit stations. It 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 characteristics of transit ridership and land-use around the stations. On the base of understanding the characteristics, then to explore the relationship between land-use factors and rail transit ridership.

3.2 Data

3.2.1 Study case introduction

The study case in this study is the subway system of Fukuoka, some details are shown in Table 3.1. It consists of three subway lines, the 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 names of the 35 stations are shown in Figure 3.4.

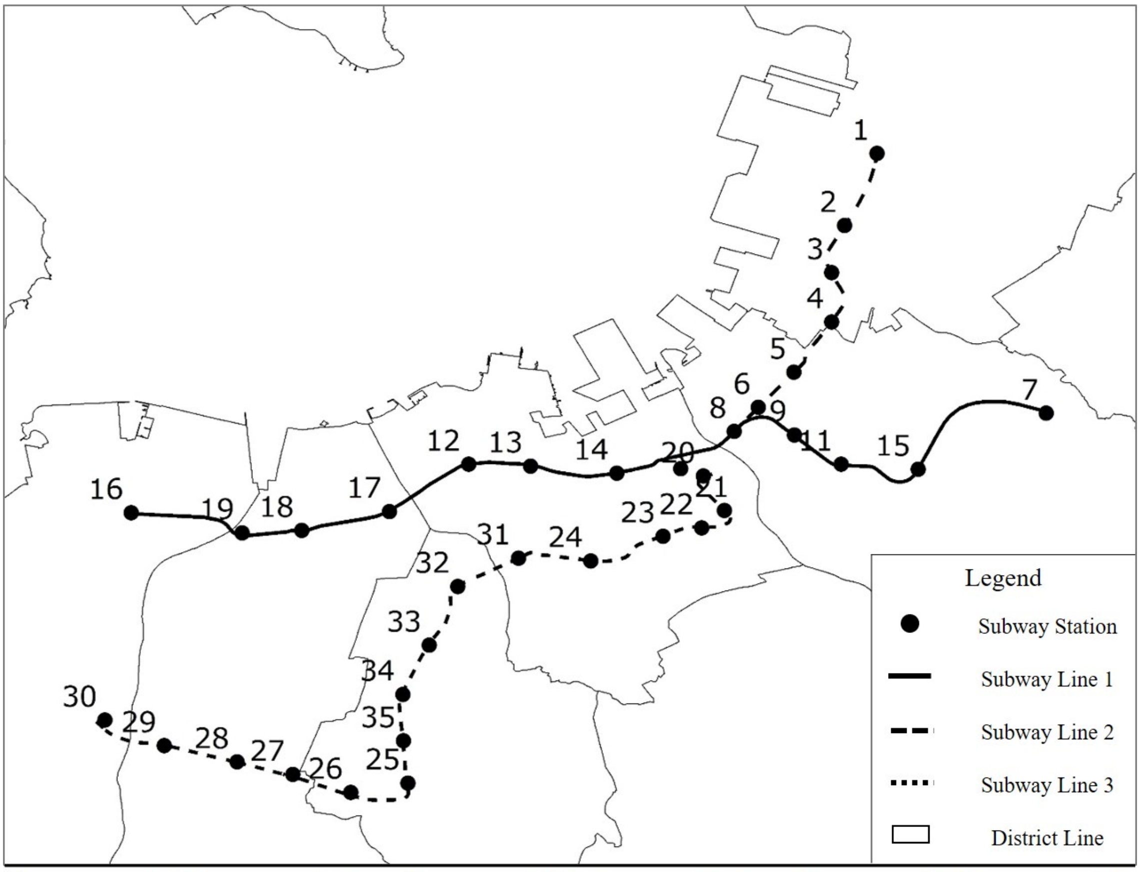
Table 3.1: Information of Fukuoka subway stations

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	Nanakuma Line	2005	-	12.0 km	16	1435 mm
	Total	-	-	29.8 km	35	-

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 a higher density of both population and building. Station 7 locates in Fukuoka airport, which mainly serves the airport. Station 11, Hakata Station, is a comprehensive railway transportation hub of Fukuoka integrating Shinkansen, JR, subway, and bus terminal. Station 10, 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 other cities around Fukuoka city. The endpoint stations of 11 and 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 onto subway transit and population is annual statistics, while the data onto urban planning basic survey is provided every five years. Since the subway line 3 is opened by 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.



ID	Station	Line	ID	Station	Line	ID	Station	Line
1	Kaizuka	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	Maidashikyudaibyoinmae	2	16	Meinohama	1	28	Kamo	3
5	Chiyokenchouguchi	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	Sakurazaka	3			

Figure 3.4: Distribution of the subway stations

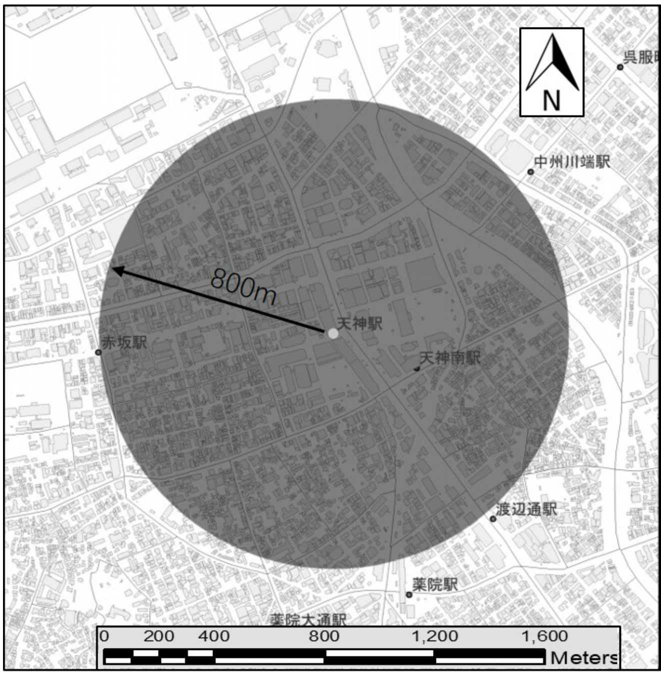
Table 3.2: Data source

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, the preprocess are necessary before conducting the analysis. The data preprocess can be separated to 3 aspects.

1. Matching the range of data unit. The minimum data unit of census and land-use is unified into block.
2. Matching the time points of data. To match the time point in 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 onto 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 onto 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.

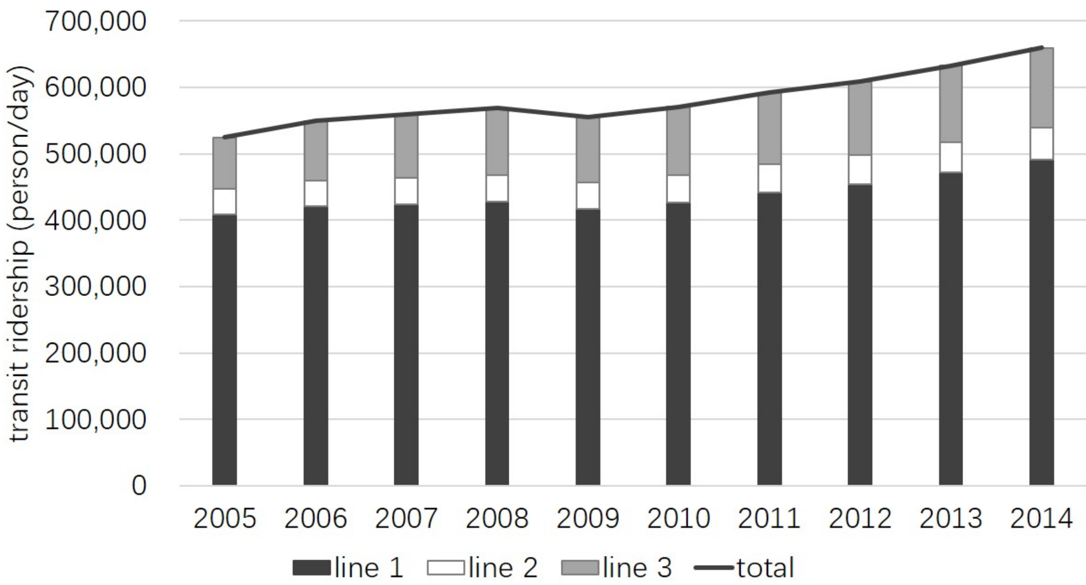


Figure 3.6: Variation in the transit ridership of Fukuoka subway 2005-2014

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 more 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.

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 shown in Figure 3.8. It is sorted by ascending order for 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.

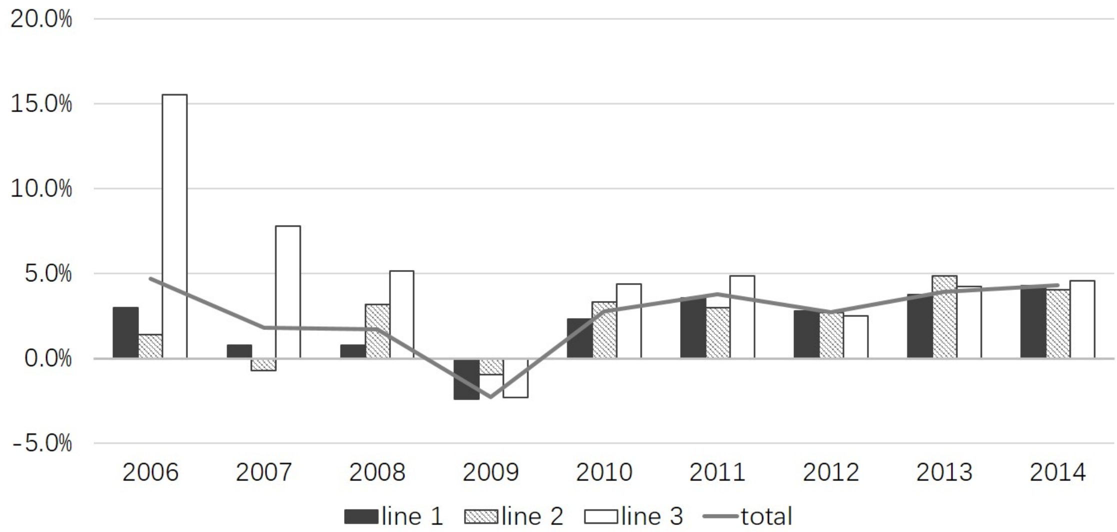


Figure 3.7: Annual growth rate of transit ridership by subway lines

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.

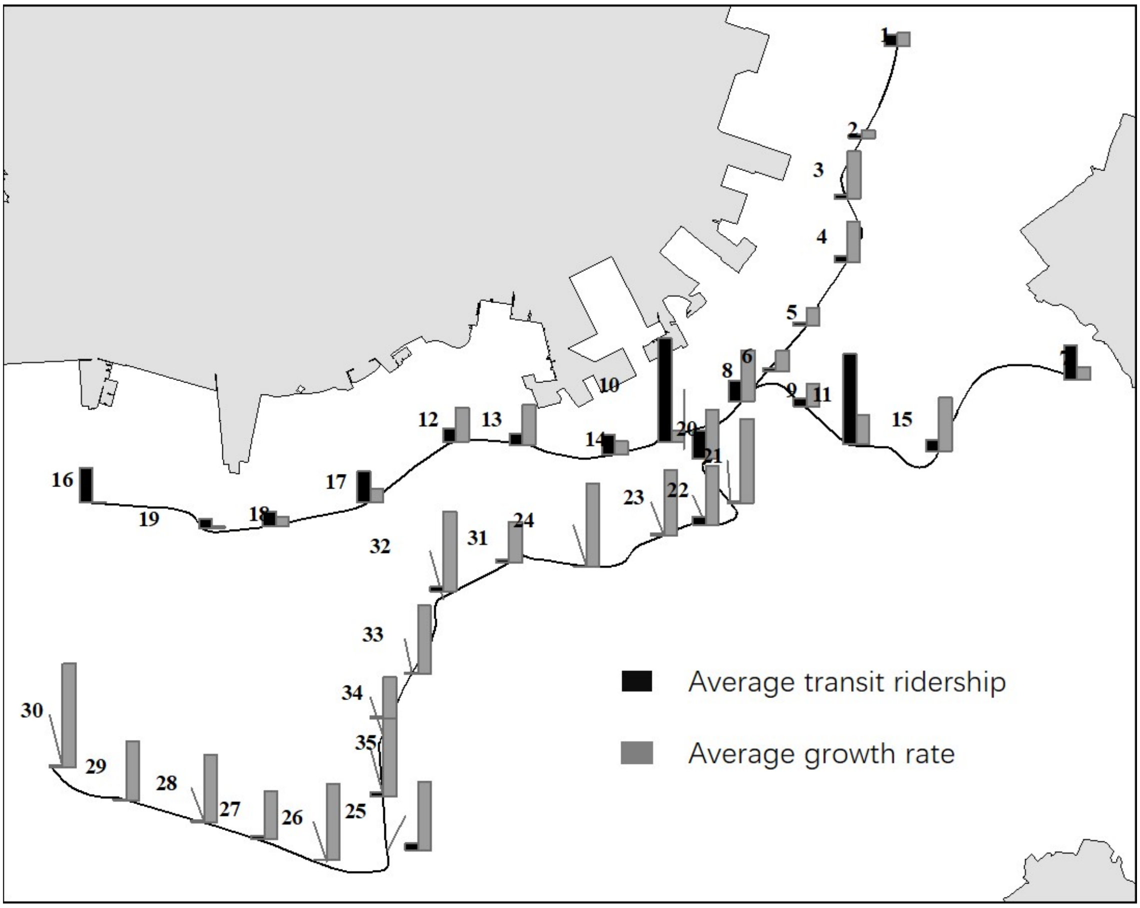


Figure 3.9: Spatial distribution of transit ridership

3.3.2 Analysis on the factors of land-use around subway stations

Data extraction

As stated before, the data onto land-use is extracted from 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. Drawing the buffer area is conducted using the tool of “Buffer” in ArcGIS, extracting data is using the tool of “Tabulate Intersection”.

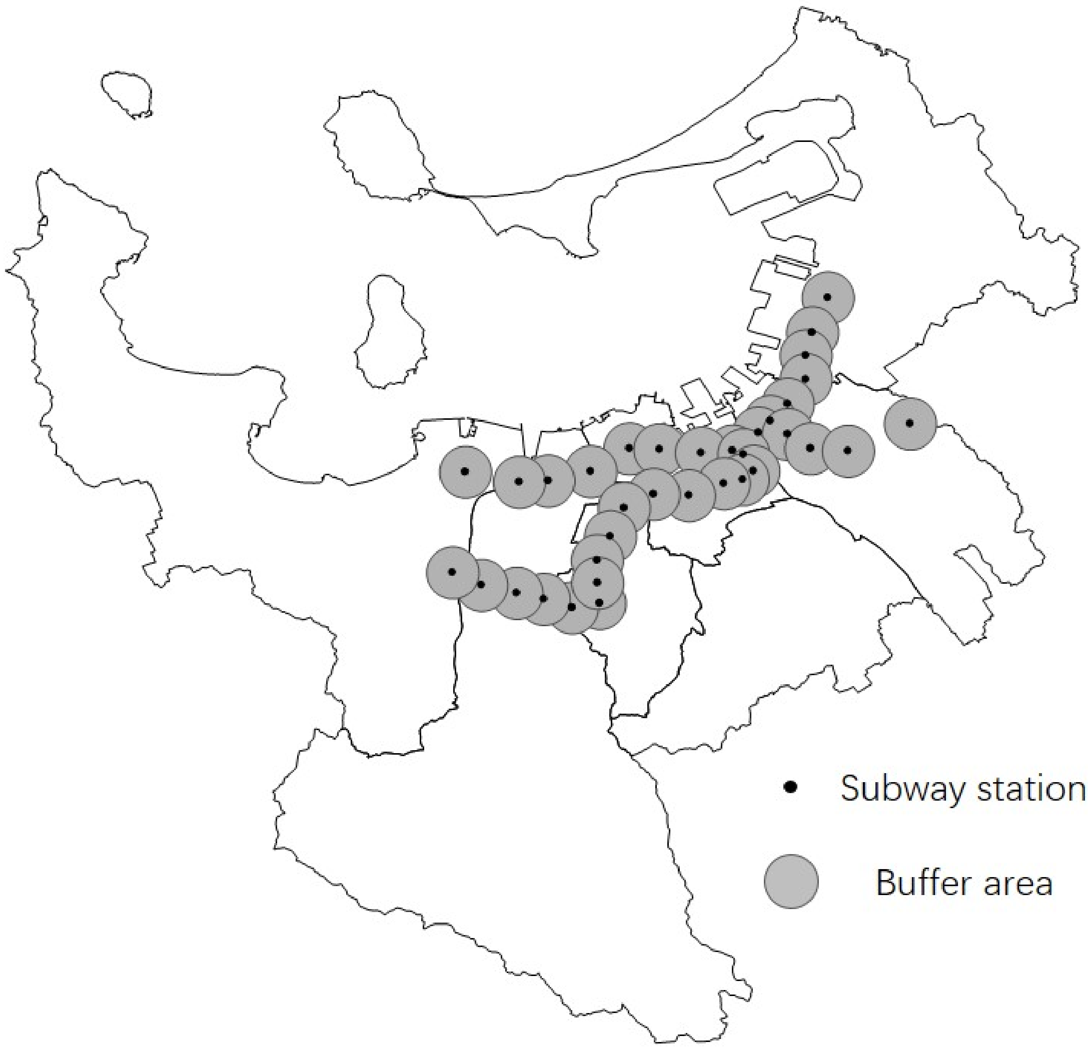


Figure 3.10: 800-meter Circular buffer areas for 35 subway stations

The land-use data onto Urban Planning Basic Survey is available per 5 years, to match the data period of transit ridership, the land-use data in the year of 2008 is used for analyzing characteristics of land-use. The dataset of land-use contains 23 types, which, however, are too many for analyzing the characteristics of land-use in the case of only 35 stations. Moreover, most types of land-use account for very little of the 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 items are listed in Figure 3.11.

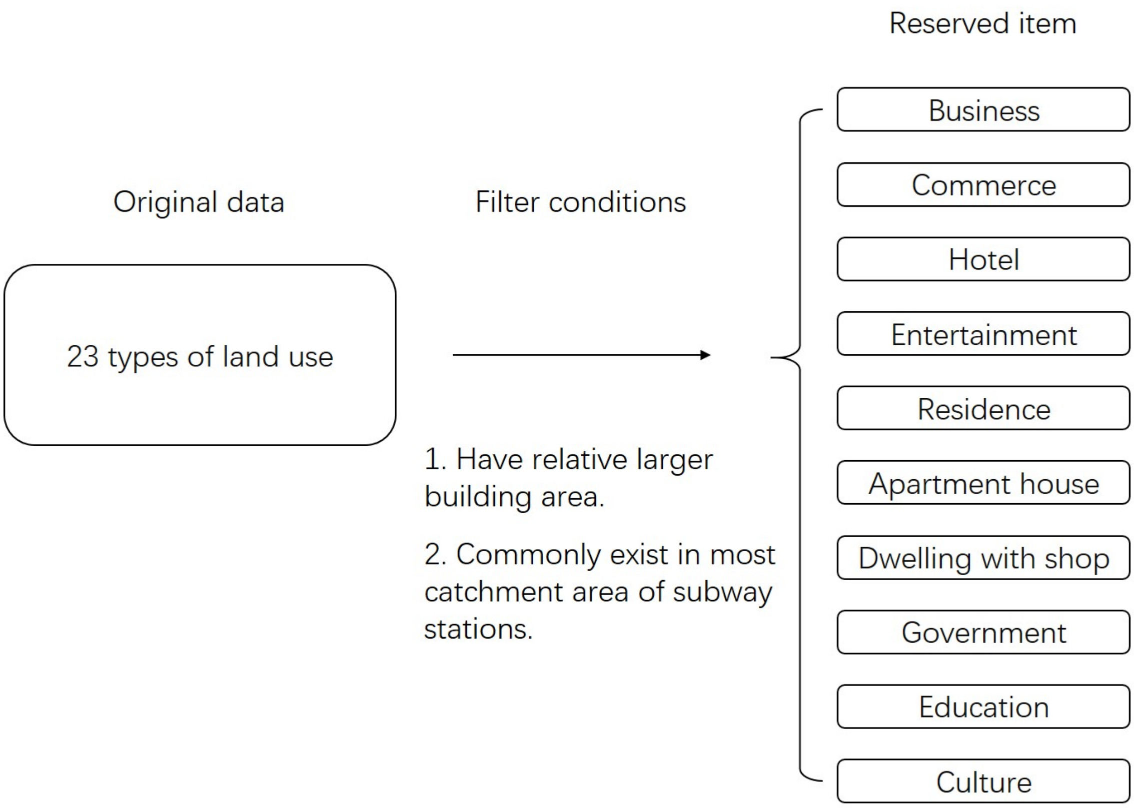


Figure 3.11: Flow of selecting indicators

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 explore what these indicators are expressing, the correlation analysis is put forward to make clear about 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 upon shop also has a strong correlation between that of office, commerce, and hotel, while the culture type of land-use closely relates to the type of apartment and dwelling upon shop.

Factor analysis

To deal with the strong collinearity among indicators, also to make clear about the internal relationship, the exploratory factor analysis is introduced to reduce the dimension of the indicators, thus to explore the meanings of the indicators. 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 reflects the correlation between 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 for conducting the factor analysis. As to the Bartlett's Test, if the variables are independent from 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 a $Sig. < 0.05$. It means each variable has correlation and factor analysis is effective.

2. Factor 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, four factors whose eigenvalue are greater than 1.00 are extracted. These four factors account for 82.90% of all the variance. It means the three factors can explain most part of the original 10 indicators. The Table 3.6 shows the total variance explained.

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

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.	0.000

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

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 indicator on the newly extracted factors. The three factors can be named and explained as follows, 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 attributes a large part of this factor. It means 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	0.850	0.044	-0.030
Dwelling with shop	0.834	0.417	0.125
Apartment house	0.301	0.860	0.134
Residence	-0.521	0.620	-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 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 station 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 indicators are put into the 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 are 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. Then the data is standardized into 0.00-1.00 range. The 34 of all the 35 subway stations fall into 5 types, while the station 7 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.

- **Characteristics of low-density residence type**
  This type has 9 stations with the characteristics of high proportion of residence, and low population-density, very few office and commerce.

- **Characteristics of High-density residence type**
  The proportion of each land-use in high-density residence is similar with that in low-density residence type, but population-density is much higher than other types.

- **Characteristics of education type**
  This type has 6 stations, they are located in the vicinity of colleges and universities, and the service population is mainly students and nearby residents.

- **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

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%

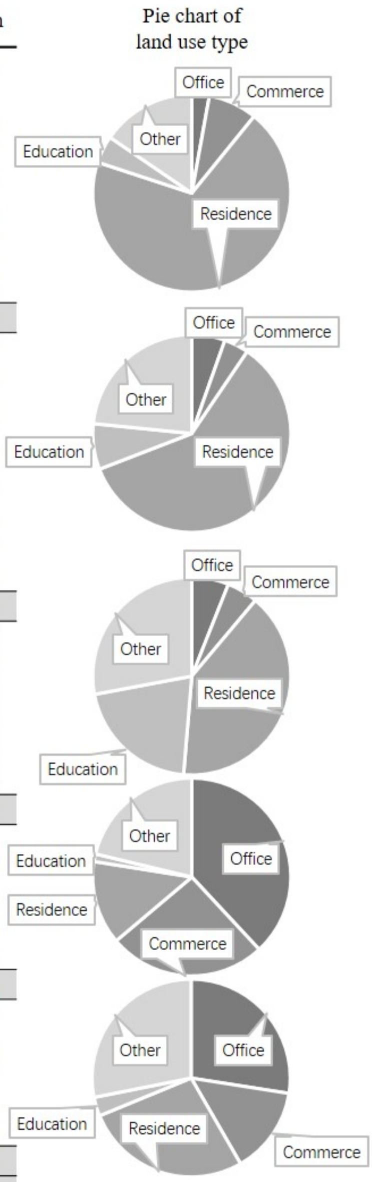


Figure 3.12: Classification of stations in terms of land-use

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 over the CBD area, of which the mixed land-use is one of the main characteristics.

The station 7, which locates 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 on the result of classification, the station 7 does not belong to any type of stations.

3.4 Exploration on the relationship between land-use and transit ridership

3.4.1 General relationship between land-use and transit ridership

The data onto land-use is updated per 5 years, while the data onto 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) and the average transit ridership of each type of station, which is shown in the Table 3.8, the subway transit ridership varies a lot due to different types of land-use. The station of low-density residence type has an average transit ridership of only 3,416, while that of high-density residence type has 15,633 on average. The downtown commerce type has the highest average transit ridership of 52,300, while that in office type is only 11,038, 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 to that in the two residence types, but from which the average transit ridership is quite different.

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 10,000 per day. Both the two hub stations belonging to the downtown commerce type located in the CBD area of Fukuoka. Overall, it

Table 3.8: Characteristic of each classification

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

can be inferred from this table that the area with lower density either in population or building has a lower demand on 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 Statistical relationship of land-use and transit ridership

The quantification method I is introduced to explore the influence that each factor impact on the transit ridership. Unlike ordinary regression models, the quantification method I divides the continuous independent variables into categorical variables. Then estimating the correlation between categorical independent variables and the continuous dependent variable. This method is thought suitable to do the exploratory analysis on the data onto 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.

The quantification method I can be viewed as an improvement on ordinary regression models which is used for dealing with the exploratory analysis at the beginning. Therefore, the indicators having strong collinearity should be excluded in the process of selecting explaining indicators. Referring to the Table 3.3, there is no significant linear correlation between these indicators of office, residence, education and government. Notably, the indicators of office and commerce have strong collinearity, however, both the indicators account 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 is 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 transit ridership, which means 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 less than that of education, the indicator of government contributes more to the variation in transit ridership.

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

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

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

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.

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 follows.

- From the comprehensive description of the study case and the investigation into the transit ridership, it can be known that the subway transit ridership is still increasing to date, but it probably turns to decrease in 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 of rail transit ridership, which aims to give a comprehensive description of the research objects also provide references to 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 models 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 to tackle the 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

Forecasting traffic demands is always an important part of urban planning and 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 forecasting traffic demands.

4.1.2 Research purpose

In many developed countries like Japan, the weakness for population growth and the tendency towards using private transport in the local central city is becoming more and more serious. The 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 deeper 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 models relative to small sample cases. Unlike the case containing hundreds of samples, a small sample case having dozens of stations has a higher risk of both type I and type II statistical errors when identifying the valid variables that should be put into the regression model. For this problem, this study mainly aims to explore and explain the factors influencing transit ridership using the small sample case. The main work include 3 aspects: 1. to build the index framework based on prior study; 2. to identify the effective indicators affecting the transit ridership; 3. to explain the relationship between variables in generating transit ridership quantitatively.

4.2 Literature review

To estimate the relationship between indicators and ridership, the model of linear regression is a widely adopted one (Cervero & Kockelman, 1997; Chakraborty & Mishra, 2013). However, the ordinary linear regression model often leads to large standard errors or low level of significance due to the insufficient consideration of heteroscedasticity and spatial autocorrelation. To improve the generality of the model, many models, for example Geographically Weighted Regression (abbreviated as GWR), Weighted Least Squares Regression (abbreviated as WLS) and Poisson Regression, have been successively introduced to the issue of forecasting transit ridership at station-level.

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 of 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 by the ordinary least square model (abbreviated as OLS), however, the result showed a significant heteroscedasticity and a non-random distribution of estimation residual (Kuby, Barranda, & Upchurch, 2004). To deal with this deviation, the WLS model was brought in to eliminate the heteroscedasticity, in which the data points were weighted using the standard error. The result showed the 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 over space. For one data point in ordinary linear 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 coefficients 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 transit ridership of Fukuoka; 2. Optimize the procedure of identifying a valid explanatory variable, thereby making it applicable to small sample cases; 3. Improve the accuracy in the estimation of the regression model in the terms of small sample cases. The flow of research process is shown in Figure 4.1.

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 on developing rail transit, but usually, the data support is insufficient due to the lack of station samples. Figure 4.2 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 onto the year of 2010 is used as a reference. All the resource of data used in the study is listed below.

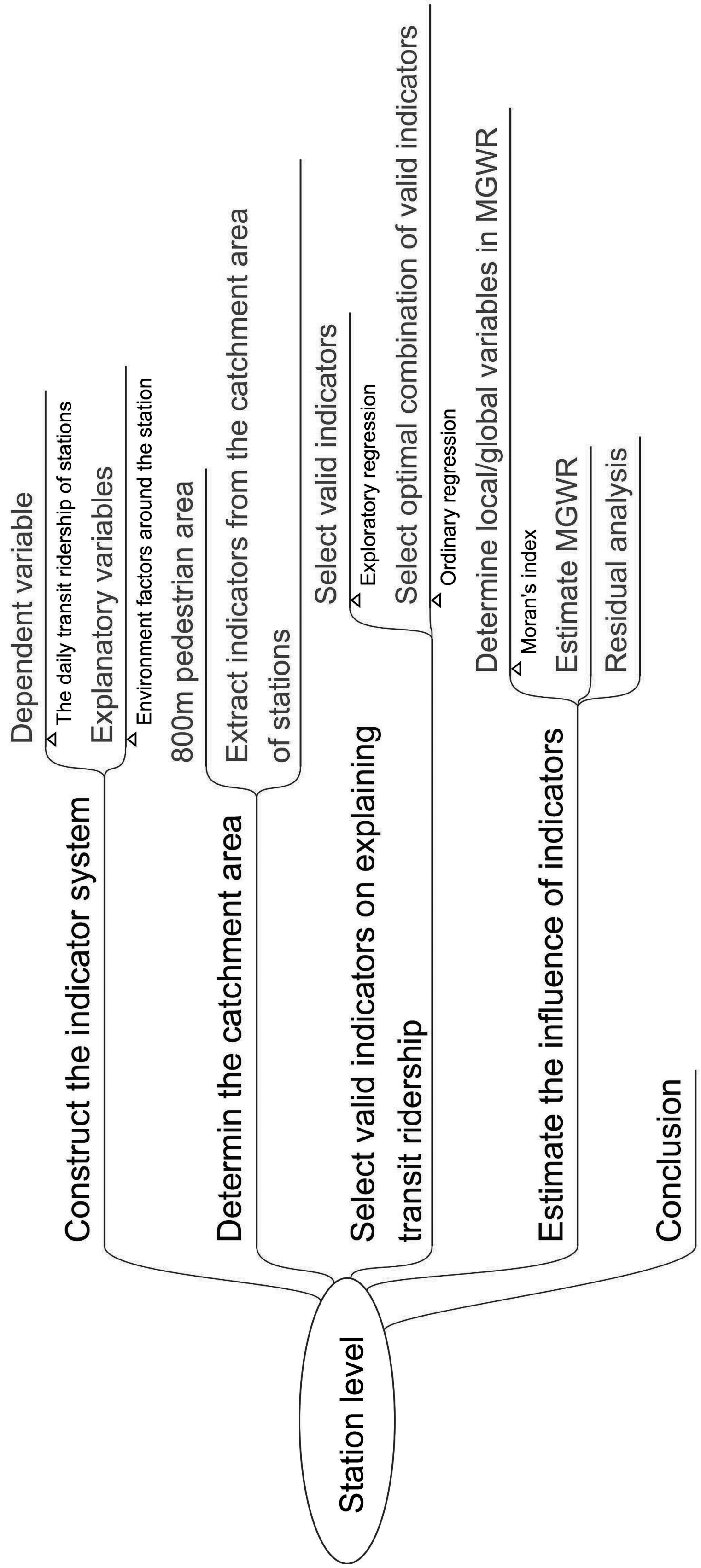


Figure 4.1: Research flowchart

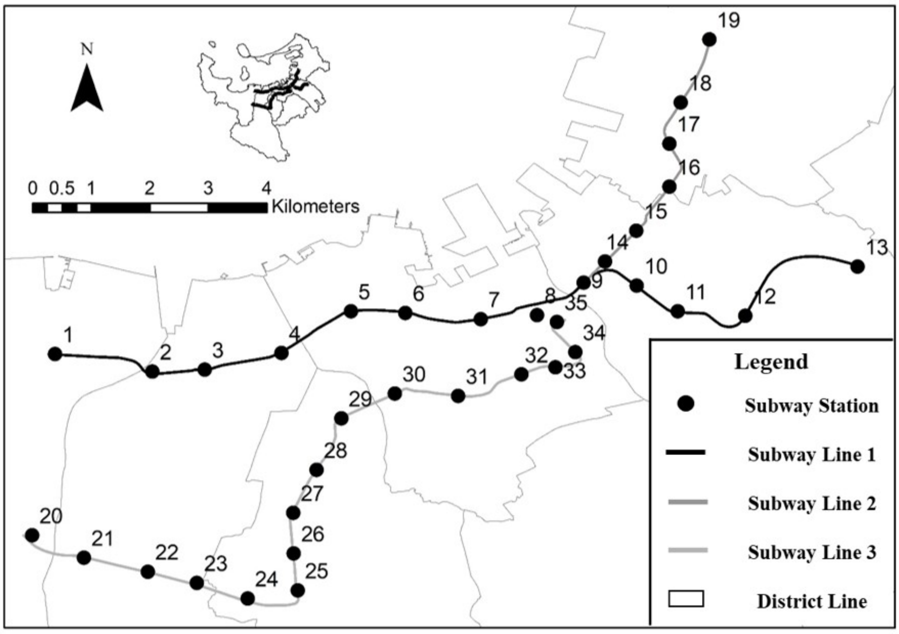


Figure 4.2: Research area and the distribution of subway stations

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

- Subway Transit Ridership
- 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. Now 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 800-meter distance 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 onto <2010 big city traffic census metropolitan area report> (Nakamura, 2015a). According to the report on basic survey along the subway third line of 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 near 90%. It can be considered that an 800 meters catchment area covered most of the subway passengers in Fukuoka. Based on the 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-meter 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). Moreover, since the more comprehensive the index framework is the problem of multicollinearity will occur more easily, the candidate variables will be identified and selected by explanatory ability in the next stage to deal with this problem.

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 affecting transit 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 chooses 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 near 90% of all the floor area in Fukuoka, that in the catchment area reaches more than 95%. The proportion of land-use in the area taking subway stations as the center is shown in Table 4.4. Obviously, it is significantly different from average distribution, and the proportion varies following the variation in the catchment area. To describe the land-use diversity closer to the facts, the balance proportion of land-use types is decided by the average proportion of all subway station catchment area (800-meter pedestrian distance) in Fukuoka.

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

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 transit ridership in two opposite ways. Higher accessibility for accessing into the catchment area of subway station means people will have more 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 divided 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 to passengers. 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 transit 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 that 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 from 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 of the people living far away from the station when they want to take the subway. And people can accept a longer access to 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

Demographic and socioeconomic environment factors also have important effect on the transit ridership. The factors of population and employment have been widely confirmed to have effect on transit ridership. The factors of 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 transit 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 of 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 at the regression model. The Exploratory Regression tool provides a reference to 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 transit 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 from 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. Unlike the ordinary 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 on 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 significance 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 of 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 of 9 explanatory variables will be evaluated by using MGWR in the next part to obtain a better result with small residual.

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 the "golden-section search" method. As a result, the optimal bandwidth is obtained when its corresponding AICc value gets to the minimum. The variations in AICc at different bandwidths are shown in Figure 4.3, and the optimal bandwidth is to the lowest AICc of 5.7 km.

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 a indicator of "DIFF of Criterion" to help justify the global and local variables. A value of more than 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 under 2. It means the distinction of global/local variables is proper.

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

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	-
Note: Confidence level 95%. If p-value is less than 0.05, the corresponding variable can be viewed as a local one.							Best bandwidth		5.7km
							AICc		690.6
							Residual sum of squares		296311499

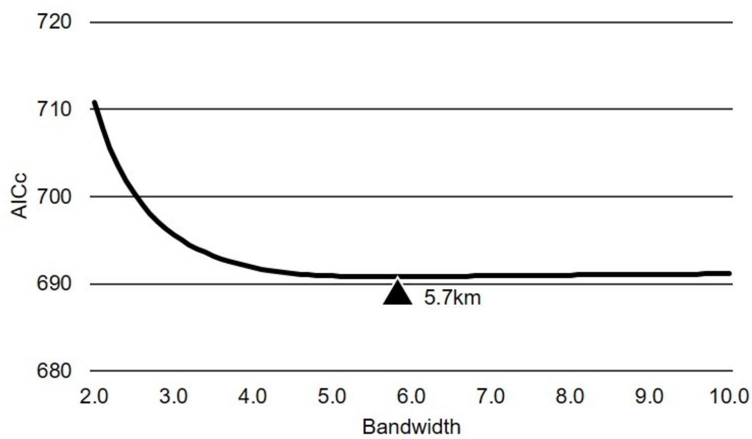


Figure 4.3: Variations of AICc at different bandwidths

present model. The signs of all the 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, 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. The result shows a 12% decrease in the 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 4.7: 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 4.4. For the 5 local variables in all 35 data points, the spatial distribution of significance level for the local variables are mapped in Figure 4.5. As can be seen, the two indicators for bus proposed to 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.

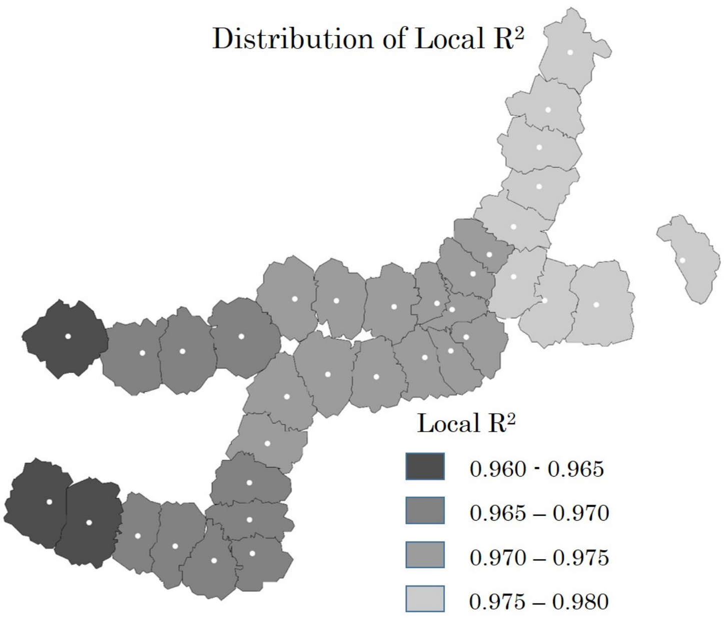


Figure 4.4: Spatial distribution of local R^2

4.6 Discussion

This study explored and estimated the influencing factors of subway transit ridership at the station level using the case of Fukuoka which has a small sample size.

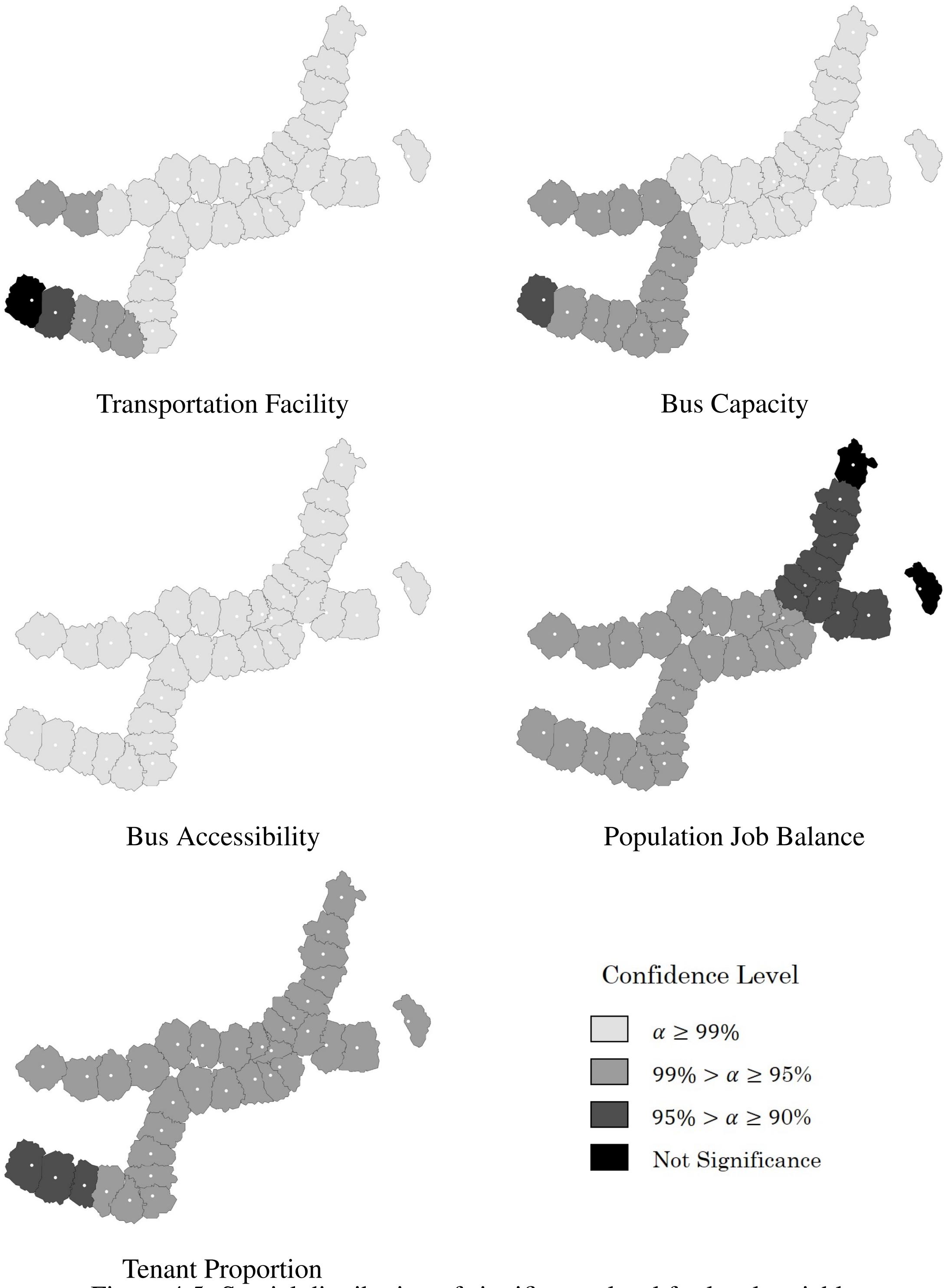


Figure 4.5: Spatial distribution of significance level for local variables

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, they perform totally opposite effect on the independent variable. Regarding the coefficient showed in Table 4.5 and Table 4.6 obtained from the two models, the coefficients in both OLS model and MGWR model have consistent signs and similar values. The rationale for 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 on 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 transit 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 is 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 transit ridership. Some researches argued that the diversity of land-use has a positive effect on transit 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 transit ridership.

A possible explanation can be interpreted as follows. 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 represents the accessibility to station within the catchment area, and the latter expresses 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 a convenient access to the subway station can attract more people to use 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 transit 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 is the indicator of bus accessibility and bus capacity showed totally opposite effects on subway transit ridership, the former have a positive effect while the latter is in contrary.

The indicator of bus service was intuitively considered to be related to subway transit ridership, and it often appears in the candidate variables list of prior studies but is rarely estimated successfully at the final model (some studies used the indicator of feeder bus but not the 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 the 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 demographic and socioeconomic environment, the factors of job-resident balance and tenant proportion are confirmed effective to explain the variation of subway transit ridership. As shown in Table 4.6, both job-resident balance and tenant proportion are showed to have a negative impact on subway transit 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.

The travel habit should be not the same between working people and unemployed people. The indicator of job-resident balance can be used to reflect how the differences in the travel habit influence transit ridership. 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 transit ridership has not been verified yet in prior studies. The result from this study shows that job-resident balance can affect subway transit 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 to have positive influences on subway transit ridership. However, this indicator performs different effects in different cases. The result from the empirical case of nine US cities indicates that tenant is more likely to use subway, while the case of 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 from 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 in subway transit ridership in the final models. The results from both the OLS and MGWR model show that the 9 indicators performed stably. The procedure proposed to this study is verified to be effective against identifying valid explanatory indicators in terms of small sample cases. The major contribution of this study can be summarized as follows.

First, 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 to this study to explore how can bus service influence the subway transit ridership.

Second, this study proposed an approach to identify 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 transit ridership of Fukuoka.

Third, this study proposed an approach to distinguish 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.

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

Impact of Land Use Types on Transit Ridership between Stations and Stations

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 putting 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 onto station-to-station level, very little studies focus on the intra-urban transit ridership (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 between station and station 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 Literature review

Until now, there are many studies focusing on the relationship between various factors and transit ridership. Most of the studies are based on the 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) models is viewed as a single point. It is not consistent with fact that the transit node is connected to each other. To deal with the relationship between stations in the same network, the spatial regression is also introduced into this issue (Cardozo, García-Palomares, &

Gutiérrez, 2012; Jun et al., 2015). Nevertheless, this relationship between stations in spatial regression models is just the expression of the distribution relationship of stations in location, which cannot reflect the real connectivity between two station areas.

To explore the connectivity between station areas, Choi worked on 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 influencing factors are divided into three groups: the factors of departure stations, the factors of destination stations, and the factors of impedance between stations. The influence of OD ridership is estimated using multiplicative and Poisson regression, at the time slot 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 a result of this empirical study, different land-use functions have different travel characteristics in terms of both time and space. But there was a problem not well tackled in this study, that although the factors of departure stations and destination stations are separately constructed, they were not distinguished in the process of estimating the model.

Land-use and public transit are coevolving partners in urban planning (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 with mixing land-use, which allows people to accomplish most of their daily activities within an easy walking distance from a major transit node (Lund, Cervero, & Willson, 2004). In detail, various types of land-use are the carrier for people to live, work and recreate, different types of land-use correspond different trip purposes. If the land-use within the easy walking distance from 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. It can be inferred that the land-use of a catchment area not only can affect the ridership of the station where it is located but also can affect the ridership of other stations connected to that station.

With the goal of explaining the variation in the ridership between stations, this study focuses on the passengers' choices of destination stations from the perspective of land-use. The study case is the subway network of Fukuoka, Japan. The contents are organized as follows: section two reviews the previous research about transit ridership; section three is the description of the data; section four interprets the methodology; section five is the result of the logistic model; section six gives the discussion on the result and draws the conclusion. The flow of research process is shown in Figure 5.1.

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). Figure 4.2, gives the research area and the distribution of subway stations. Until now Fukuoka has three operating subway lines, a total of 29.8 km 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.

The same with the catchment area used in chapter 4, it is also the road-network-based 800-meter walking distance in this chapter. Then the factors considered to have influences on the passengers' choices of destination stations are extracted from this 800-meter pedestrian catchment area. The follows give the interpretations of the indicators used in this study.

5.3.2 Land-use factors

Land-use is generally accepted as one of the determinants for transit ridership. The indicator of different types of floor area can be considered as a detailed expression for land-use. This influence of floor area is extensively confirmed in previous studies (Sohn & Shim, 2010; Gutiérrez et al., 2011; Chakraborty & Mishra, 2013, 2013;

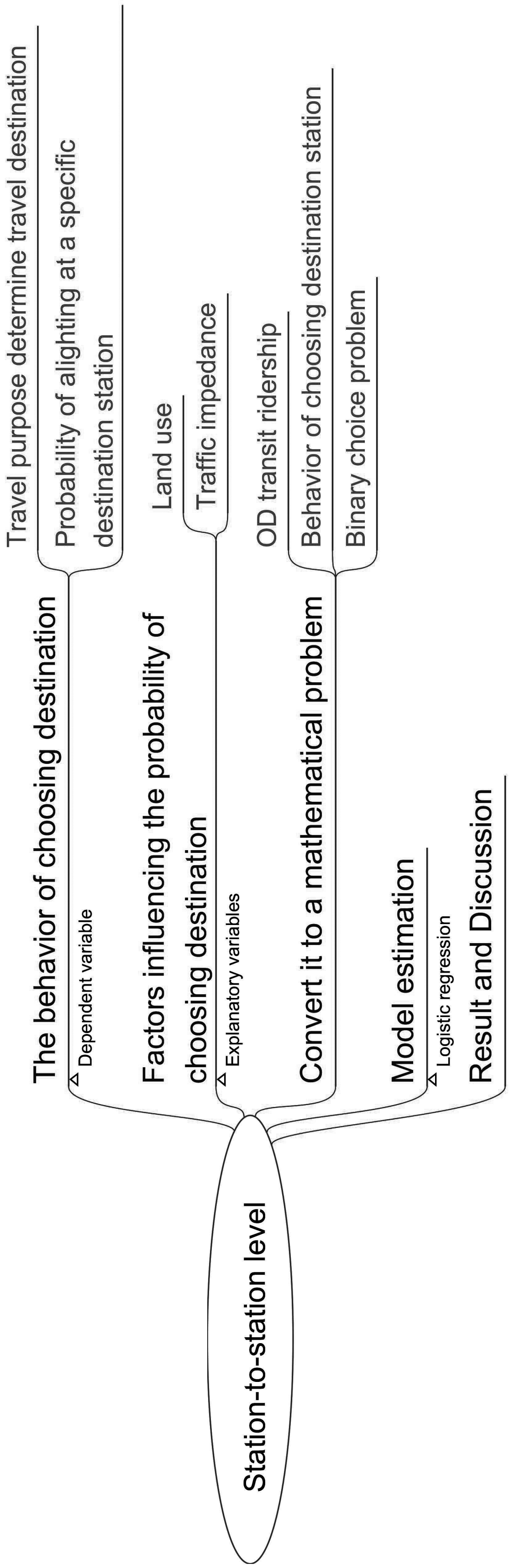


Figure 5.1: Research flowchart

Jun et al., 2015). Moreover, since 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. In Fukuoka, near 90% in the trips of accessing transit stations are non-motorized (refer to Table 4.2). It represents that the land-use around the station is playing a crucial role in determining the trip purposes of transit users.

The same with previous studies, this study chose 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 near 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 diversity is also assumed to be a crucial factor in explaining the connectivity of stations (Badoe & Miller, 2000; Cervero, 2004; Frank, Andresen, & Schmid, 2004). Unlike the general definition of land-use diversity, this study redefines it into the aggregation of land-use. The Euclidean Metric is used for evaluating the deviation between the land-use aggregation in each subway station and the reference value. The value of this indicator is ranged from 0 to 1, in which the lower value represents a higher diversity of land-use, while the higher value means the land-use 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. The 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 types of land-use being selected as the indicator in this study, the list of indicators also the statistical descriptions refer Table 5.1.

Table 5.1: Summary of the land-use proportion

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 an index which is used to represent the connectivity and cost between departure 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). Since the PCA is set by the pedestrian distance, the area is already representing the internal road impedance. The indicator of internal road impedance is not included in this study. For the external impedance, the operation distance between stations is first to be considered as the reflection of the spatial connectivity between two stations.

Additionally, 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 is used for representing the convenience for accessing the station which is speculated as a positive factor of transit ridership, while the later is the transport capacity of the bus within the PCA of the station which is considered as a negative factor of transit ridership because it may share part of ridership from the subway. Table 5.1 lists the statistical description for the indicators.

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

If investigating any one of the stations, passengers have the choices of alighting or not alighting at this object station. The probability of alighting at this object station can be viewed as the connection between the departure and the destination. In this way, this issue can be converted into a binary choice problem. Therefore the transit ridership between stations can be explained by investigating the probability of alighting at the object station. The binary-logistic regression is introduced to estimate this converted binary choice issue. Formula 5.1 shows the expression of binary-logistic regression.

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

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, $1 - p_i$ represents 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 of the two groups. It is clearly that the education type has the highest proportion of education floor area. Both office and commerce types have a higher proportion of office than other types, but the commerce proportion of 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. The station classification also shows 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 follows:

- 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.
- Avoid selecting the transferable stations. Because the transit ridership of this kind of stations generally relates the other stations, which cannot reflect the distribution of land-use within the PCA.
- 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 alighting or not alighting at the object station; the influencing factors are land-use and impedance between stations and stations. The result of estimation is shown in Table 5.3.

The estimation is shown in Table 5.3. The values of gray cell indicate no statistical significance at the confidence level of 0.05. The index of $Exp(B)$, which is

Table 5.2: Station classification

Type	Population density	Commerce	Office	Residence	Education	Land-Use Aggregation	Bus Capacity	Bus Accessibility
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.

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

the odds ratio, represents the variation in the probability of choice when the coefficient B changes one unit. Here giving a brief explanation of the index of $Exp(B)$. For the variables of land-use, the index of $Exp(B)$ represents that the 1 percent (0.01 unit) variation in the variable 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 to alight at the object station.

The following is a detailed description taking the Tenjin station as an example. From the point of land-use, the increase in 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 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 and conclusion

This study investigated the connectivity between stations with different land-use characteristic by examining the 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 shown 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 working area (Badoe & Miller, 2000). But the situation may change in the medium-residence type due to different travel preferences. 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 the alike 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 variable of bus service describes the feature of one single station, but not 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 little on the station located in the suburban area.

As the conclusions, 1. the probability of choosing a station as the destination tends to decrease if the land-use types are similar between the departure and destination stations; 2. the probability of choosing a station as the destination which belongs to low-density residence type has no tendency to raise regarding the variation of land-use in the departure station; 3. for any type of stations, the education land-use in the departure station contributes to an increase in the probability of choosing that station as the destination.

In summary, this study described the transit ridership between stations by using the choice of destination. The results indicate land-use types of the catchment area have significant influences on the choice of destination, while the influence of impedance was not clearly confirmed yet. Nevertheless, the impedance between stations is supposed to have influences on the choice of destination. Since the impedance contains various types of transportation modes, such as private cars, rail transit, bus, walking, etc., it is recommended investigating the influence of impedance from the perspective of different transportation modes.

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 estimated the influences of various factors from the perspectives of 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 to select the valid indicators; and proposed a ridership forecasting method considering the interactions 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 of exploring determinants of rail transit ridership based on the needs of sustainable urban development. By reviewing the literature relating to this field, specific research questions were proposed. Around the primary goal of exploring determinants of rail transit ridership, the description of study case and dissertation organization were given at the last of this chapter.

- ***Chapter 2*** discussed how the walking duration to rail transit station is affected by passenger attributes. Centering with this topic, 3 specific research questions were proposed according to the previous studies, they are: 1. How to understand and describe the property of walking access to rail transit; 2. how to identify the valid factors influencing the preference to walking duration; 3. how to estimate the walking duration using passenger attributes. This chapter was organized centering the 3 research questions. In the beginning, a detailed interpretation of the property and implication of walking access to transit station was given, based on which the description model of the relationship between walking duration and passenger attributes was constructed. This study argued that the probability of walking a given walking duration or more should be influenced by passenger attributes, and converted this issue into a binary choice problem. And then, the ANOVA was used to identify the feature attributes of passengers at each given walking duration threshold. With the extracted feature attributes, the random decision forest model was adopted to explore preferences to walking duration of passengers with different attributes. The probabilities of walking more than the given thresholds of walking duration were estimated using the training set, and predicted with the test set. At last, the evaluation of the prediction showed that the individual behavior of walking more than a given walking duration still cannot be predicted accurately, but the overall tendency to walking duration of a group of passengers is predictable at some extent. As the conclusion, the quantitative relationship between walking duration and passenger attributes discussed in extensive literature was verified in this study, also the possibility of predicting walking duration was provided.

- ***Chapter 3*** is a preliminary study of exploring the determinants of rail transit ridership. This chapter summarized the characteristics of rail transit ridership and land-use in the case of Fukuoka. It aimed to make a comprehensive understanding of the research object of this dissertation, thus providing reference and implication for the next research. Based on this aim, chapter 3 focused on three aspects: 1. Summarizing the characteristics of rail transit ridership of Fukuoka; 2. Summarizing the characteristics of land-use around the rail transit stations in Fukuoka; 3. Exploring the relationship between rail transit ridership and the land-use around stations. Firstly, the characteristics of transit ridership were summarized from the perspectives of total amount, growth rate and spatial distribution. Then, the internal relationships between each type of land-use around the stations were interpreted using correlation analysis and further analyzed using factor analysis, based on which the subway stations were classified into 5 types in terms of the characteristics of land-use. At last, the influence of land-use around the stations on both the amount and growth rates of transit ridership was estimated using the quantification method I. To explore and estimate the influencing factors of transit ridership, some recommendations obtained from the result of this chapter: 1. pedestrian area but not radius buffer should be considered; 2. the explanatory variables should be enriched; 3. the estimation model should be improved; 4. the approach to addressing small sample case should be considered.

- ***Chapter 4*** explored and estimated the influencing factors of rail transit ridership at the station level using the case of Fukuoka which has a small sample size. With this small sample study case, this chapter focused on 3 specific research contents: 1. Summarizing the literature and putting forwards the candidate indicators that may have the impact on the rail transit ridership; 2. improving the approach to identifying and selecting the valid indicators towards the case having a small sample size; 3. improving the estimation of Mix Geographically Weighted Regression by distinguishing the local/global variables. This chapter was organized centering the 3 contents. First, the indicator system that is considered to have an impact on rail transit ridership was constructed from three categories of built environment, traffic accessibility, and social demographic environment. Particularly, the impact of the bus system was considered to have both positive and negative effect, represented by bus accessibility and bus capacity respectively. And then, to reduce the probability of type I and type II statistical errors in a small sample case, the exploratory regression was introduced to help to identify the valid explanatory indicators among the candidate indicators. Finally, the influence of the identified indicators was estimated using MGWR, where the local and global explanatory variables in MGWR were distinguished based on the spatial autocorrelation. As the conclusion, the approach proposed in this chapter is verified to be effective against identifying valid explanatory indicators in terms of small sample cases; the impact of the bus system was verified that it has both positive and negative effects on the rail transit ridership.

- ***Chapter 5*** explained the how land-use patterns influence transit ridership at station-to-station level. The transit ridership at station-to-station level is a result of passenger transfer from station to station. With the main purpose of describing and estimating this passenger transfer, and the specific research contents were putting forward as followings. 1. Describing the passenger transfer from station to station, and convert it into a mathematical problem that can be estimated; 2. estimating the influence of land-use pattern of the catchment area on that passenger transfer. In this chapter, the passenger transfer was described using the probability of alighting at a specific transit station. This choice of destination was thought to be affected by the land-use type around the destination station. Therefore, this issue could be converted into a binary choice problem that the choice of whether alighting at a specific transit station is affected by the land-use type around that transit station. Then this binary choice problem was estimated using the logistic regression model. The results showed that the land-use type around a station has a significant influence on choosing that station as the destination. As the conclusions, 1. the probability of choosing a station as the destination tends to decrease if the land-use types are similar between the departure and destination stations; 2. the probability of choosing a station as the destination which belongs to low-density residence type has no tendency to raise regarding the variation in land-use in the departure station; 3. for any type of stations, the education land-use in the departure station contributes to an increase in the probability of choosing that station as the destination.

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

6.2 Contributions

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

1. Reinterpreted the implication of being surveyed walking access to transit stations, that it has no linear relation to 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 to screen 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 on the selection of destination stations in terms of land-use types . (refer to **chapter 5**)

6.3 Recommendations

This research focused on the influencing factors of 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 of 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 to help 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 recommended planning 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 on rail transit. It is recommended connecting 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 followings:

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.