Potential Environmental Factors to Increase the Public Transportation Users by AI Applications Using Traffic Big Data

交通ビッグデータと人工知能を活用した公共交通利用者増加の潜在的環境要因に関する研究

February, 2023

Hsiang Chuan CHANG
張　翔筌

Potential Environmental Factors to Increase the Public Transportation Users by AI Applications Using Traffic Big Data

交通ビッグデータと人工知能を活用した公共交通利用者増加の潜在的環境要因に関する研究

February, 2023

Waseda University Graduate School of Creative Science and Engineering

Department of Civil and Environmental Engineering, Research on Transportation Studies and Planning

Hsiang Chuan CHANG
張　翔筌

Acknowledgment

This dissertation would not have been possible without the presence of many and I am sincerely thankful for those who helped this research and thesis with me. Please pardon me for not acknowledging each and every single of you, but all of you have played an important role in my life.

Firstly, I would like to express my great gratitude towards Professor Akinori Morimoto for constantly overlooking my work and providing research guidance at the weekly scheduled seminar, and most importantly taking me into as one of the doctoral students at the Transportation Engineering and Urban Planning laboratory. His work ethic has also inspired me deeply and in a way that made me realize the importance of dedications towards everything one does. I have nothing but great respect for him and it has been an absolute honor to have worked under his supervision.

Secondly, a word of appreciation must be given to my fellow laboratory members, who embraced me with open arms the first moment we met. Despite the language barriers, my 4 years here in Japan would have been tedious and less enjoyable if we had not spent time together. These memories are important parts of my journey. I genuinely hope these friendships and connections do not end here, but can be something I treasure for the rest of my life.

Last but not the least, I am grateful for the ongoing supports given by my family, especially my mother. She has sacrificed immeasurable amount of time and money to provide me with the opportunities to study overseas in America and Japan and these opportunities truly broadened my horizons. Moreover, she is always given valuable life lessons and have helped me through adversities on many occasions in the past 32 years. I am forever indebted to her for what she has done.

Morimoto Laboratory, Waseda University
Tokyo, January 2023
Hsiang Chuan CHANG

Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Background

Accompanying the rapid economic development around the world, there raise some challenges we never face before. Especially, in developed countries, many cities have met some problems such as budget problems, vacant homes due to the reduction of population, aging society. On the other hand, environmental concern is a big issue such as the global warning. Public transportation systems are considered as important method to reduce environmental problem by reducing the usage rate of private vehicles. As a result, increasing usage rate of the public transportation systems is significant topic. Especially, the new transportation systems such as BRT (Bus Rapid Transit) and LRT (Light Rail Transit) are put into more emphasis nowadays. The BRT and LRT are customer-oriented transportation systems which have the merits of delivering fast and comfort. BRT is a new solution to improve the efficiency of traditional bus systems. LRT was first introduced in North America to describe the new concept of tram transportation (Thompson, 2003). A well-established LRT system can assist city become a TOD (Transit-Oriented Development) City or Compact City through adapting new system, and then the city can be more attractive for citizens (Takami *et.al*, 2003). As shown in Fig. 1.1, the new LRT systems increased significantly from year 1985 to year 2015.

It is critical to identify potential public transportation passengers and enhancing their usage rate. Nonetheless, it is difficult to predict the passengers of new transportation systems since it is lack of enough information and existing data. Also, because of the improvement of the software and the hardware techniques of computer science, new analysis based on AI (Artificial Intelligence) methods are put on emphasis again these days. Furthermore, due to the widespread of the use of cell phone and the availability of cell phone data from some company, researches that are related to the application of this kind of dataset become more popular recently. Because the numbers of cell phone data are huge, AI analysis is often conducted to find useful information.

Therefore, the research of conducting AI analysis by utilizing cell phone data in transportation field also seems to be put into emphasis more and more. Image of background of this research is shown in Fig. 1.1.

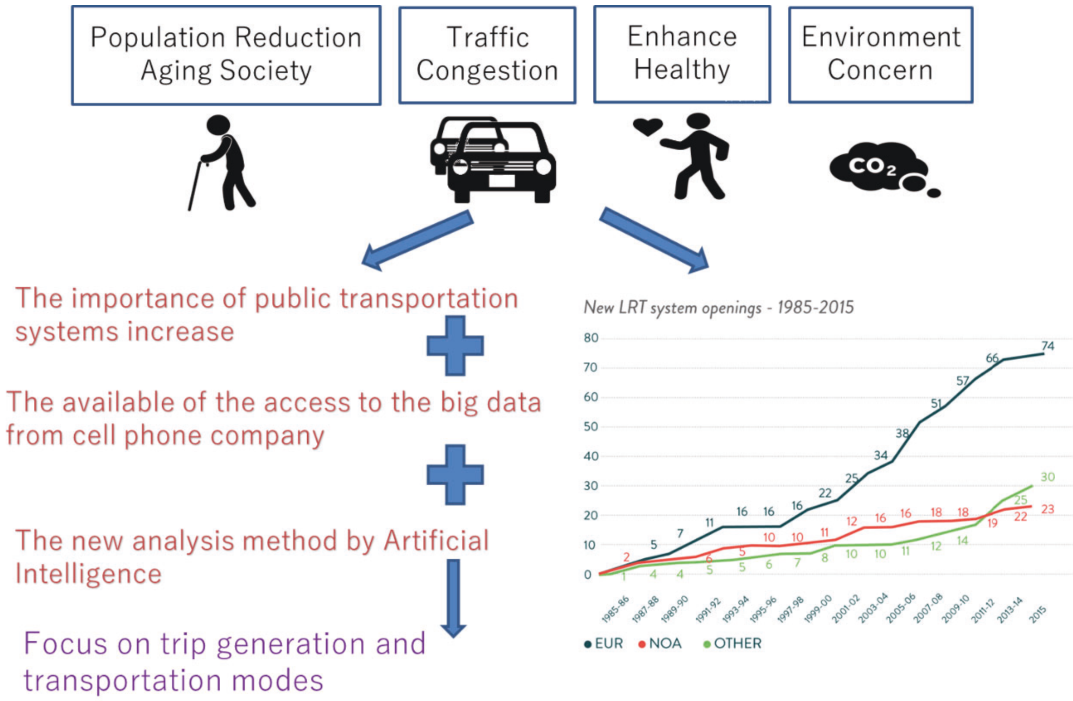


Fig. 1.1 Background of this research

1.2 Objective

The main objective of this research is to understand the factors that can make contribution to increase the users for new transportation systems from different point of views. Firstly, supply perspective from government or business sectors and demand perspective from passengers were considered. Hereafter, three categories of focus parts were divided in order to figure out related factors in each part. The conceptual diagram of above-mentioned ideas was shown in Fig. 1.2. Considering the limitation of acquiring a big data set via using questionnaire survey, the cell phone data can be more competitive for decision making. Nevertheless, traditional survey data can also be applied to Artificial Intelligence (AI) methods to make the calculation faster or obtain more important information. Thanks to the improvement of computer science, powerful machine learning methods based on AI technologies are emphasized again nowadays. The purpose of this research is using the machine learning method of Association Analysis to identify the key factors of trip generation based on the distance to infrastructures in cities. Also, another goal is applying the machine learning method of Inverse Reinforcement Learning (IRL) to figure out the tendency of selecting the transportation modes for different trip purposes. The results can be a reference for new policy planning, including re-planning the exiting routes of bus systems or integrating different public transportation systems, by the local government. Furthermore, I expect the proposed methods outperform traditional analysis methods to improve the precision accuracy.

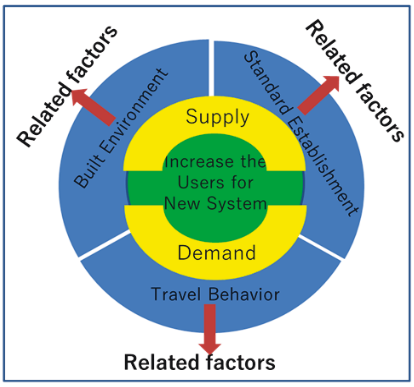


Fig. 1.2 Conceptual diagram of this research

1.3 Organization of the Dissertation

This dissertation is structured into eight chapters. Firstly, Chapter 1 is a brief introduction. Chapter 2 reviews the literature on main dimensions related to the main part of this research. Chapter 3 introduces the research design, including the data processing, case study, and the flowchart of the research. Chapter 4 is an independent one, showing the LRT indicators that can be used for evaluating the performance of LRT systems. Chapter 5 to Chapter 7 is the main core of this thesis. Chapter 5 focuses on the analysis area along the LRT route in the Utsunomiya City. Chapter 6 emphasis on the difference before and after the COVID-19 pandemic. Chapter 7 takes all area in Utsunomiya City into account and broadens the analysis by also using the questionnaire data. Finally, Chapter 8 demonstrates the conclusion, followed by summaries of the limitations and future implementation.

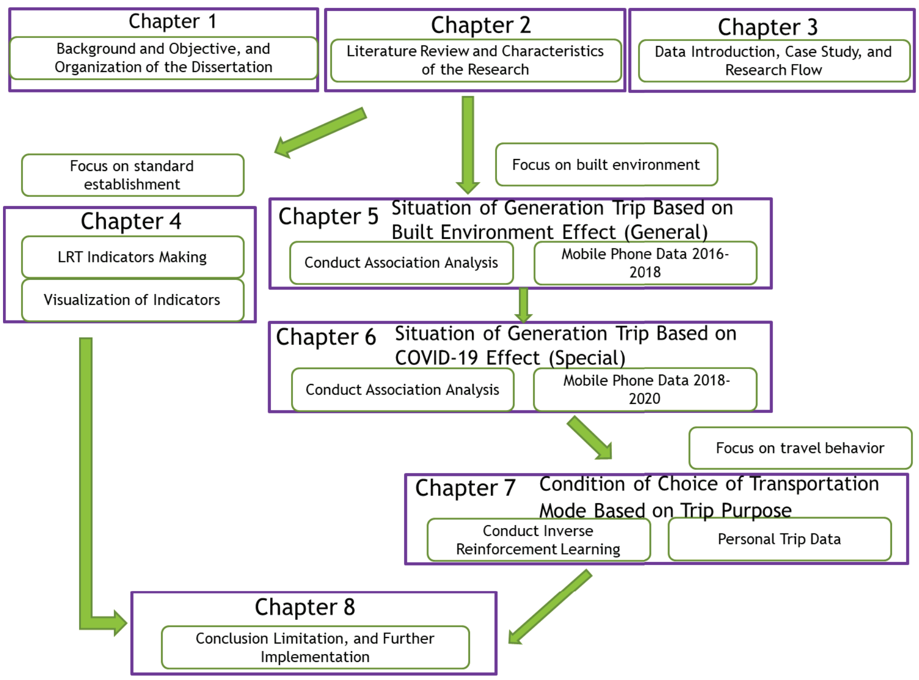


Fig. 1.3 Organization of doctoral dissertation

Chapter 2

LITERATURE REVIEW

2.1 Literature Review

The literatures review was conducted for the following three themes. (1) Public transportation system and land use (2) Mobile phone big data, and (3) Machine learning methods related to this study.

Numerous studies can be found in literature for analyzing the data sets from new transportation systems. Some recent studies, for example, Zhao *et al*. (2019) quantified the impacts of Urban Rail Transit system on land use change and tied them into the future land maps. Zhang *et al*. (2020) identified several factors that influence the estimates of BRT impacts based on 23 empirical studies. Pacheco-Raguz (2010) studied the impacts of the development of LRT in Manila on land price, land usage, and population size for the cities along the route. Using correlation and regression methods, the aforementioned variables are analyzed for an accessibility index and network distances obtained from a model built within a Geographic Information System (GIS). Sakamoto *et al*. (2015) examined the change in urban population size before and after introduction of LRT in 27 cities in Europe. They also analyzed the changes from the population size or areas along the LRT route over time for four cities in Europe. Kriss *et al*. (2020) conducted a study for Toyama City and pointed out that the LRT project in Toyama City got significant success based on the increasing ridership of the LRT. Sato *et al*. (2018) predicted the population distribution in Utsunomiya City, Japan until 2050 by supposing different integration pattern of the LRT and feeder bus systems. Takasugi *et al*. (2018) studied the differences between the BRT and LRT systems and the influence on urban population distribution for Maebashi City. Most of aforementioned researches are mainly concerned about the impacts on the cities after the introduction of public transportation systems.

It is worth to note that the findings of aforementioned researches were highly dependent on using questionnaire survey method. Questionnaire survey is time consuming and asks high cost for acquiring data. Today, cell phone is popular and becomes one of the required devices for people. Cell phone data are more competitive and complete than questionnaire survey data. Using cell phone data for studying the impacts of public transportation systems on related areas have earned more attentions in the near past decade. Widhalm *et al*. (2015) proposed a method to reveal activity patterns that emerge from cell phone data by analyzing relational signatures of activity time, duration, and land use. Also, Wang *et al*. (2018) provided a review of existing travel behavior studies by using mobile phone data. They discussed the potential of mobile phone data in advancing travel behavior research and raised some challenges that needed to be dealt with in the planning of traveling process.

Compared with traditional statistical modeling methods, machine learning methods in the AI area can be more powerful for prediction but could be weaker to explain the relationship between the response variable and explanatory variables due to an implicit functional form of hidden layers structure. Moreover, machine learning methods has the potential to automatically learn and update the predicted results through a well design system from updated data sets. Haenlein *et al*. (2019) takes a first insight into AI applications by summarizing seven articles published in this special issue that present a wide variety of perspectives on AI. When the relationships among features in data are not clear, the Association Analysis method is one of most competitive machine learning methods to identify the potential relationship between different features with less subjective assumptions. Tan *et al*. (2005) proposed a general idea to set up models by different algorithms. By showing some examples with programming, one can easily understand how to implement Association Analysis method.

The IRL is another machine learning method which is applied in this research. The IRL is about studying from humans. In practice, IRL can be used to study an agent's objectives, rewards and values with the aid of using insights of its behavior. Arora *et al*. (2021) mentioned about advantages and challenges about using IRL. You *et al*. (2019) showed a great example of mobile application via using the IRL for data analysis. Kitani *et al*. (2012) conducted analysis and prediction of walking behavior using IRL model.

2.2 Characteristics of the Research

Many existing researches related to the trips condition were based only on the personal trips survey but lack of abundant trial. Here, cell phone data was utilized and conducted the analysis as an estimated trip frequency. A new idea by integrating static and dynamic cell phone datasets from different sources with several analysis steps are proposed in this study for providing to understand the trip generation in the target city. Hereafter, personal trips data was also applied by machine learning methods to make insight analysis and comparison of the results. Therefore, simultaneous utilization of the static data and dynamic data can be a significant contribution of this study. The image of some examples in both the static and the dynamic data sources is reported in Fig. 2.1. The static-type public facilities and boundary data and dynamic-type personal trip survey and cell phone signal data are used in this study. The machine learning methods of Association Analysis and IRL are applied for data analysis to obtain valuable information for making new policy planning by the government of the target city.

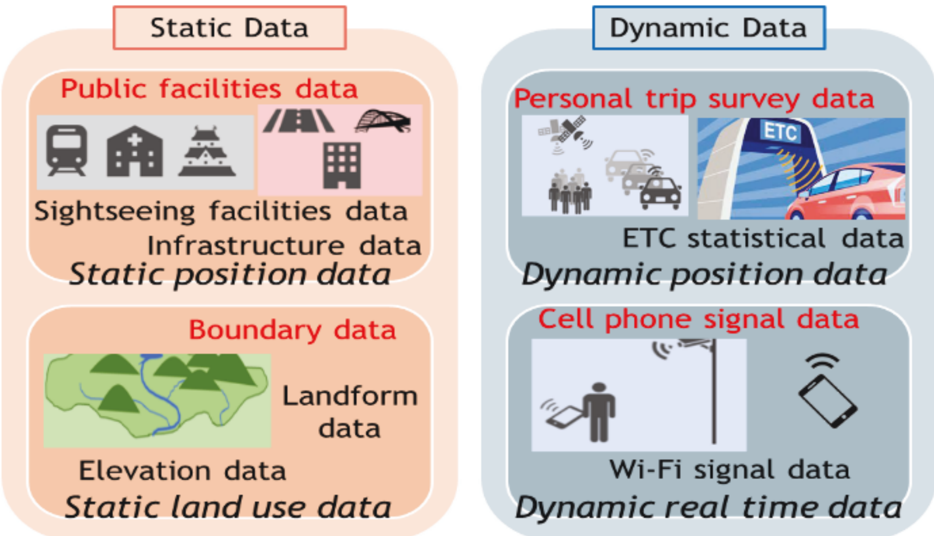


Fig. 2.1 Image of data sources

Chapter 3

RESEARCH DESIGN

3.1 Introduction of Data

3.1.1. Data and software utilized

There are variety of data can be used for analysis. Here, due to the data availability, the below three data were utilized in this research. Fig. 3.1 is the summary of source data from KDDI company and National Land Numerical Information download service. One is the cell phone data of Utsunomiya City from KDDI company. In the KDDI cell phone data set, the data was acquired in June from 2016-2020. Here, de-identification trip data extracted from the GPS logs of cell phones with permission contributes to the first data set. The second data set is about the geographic information from NLNI (National Land Numerical Information). At the same time, personal trip survey data was also prepared for the purpose of broadening the research content.

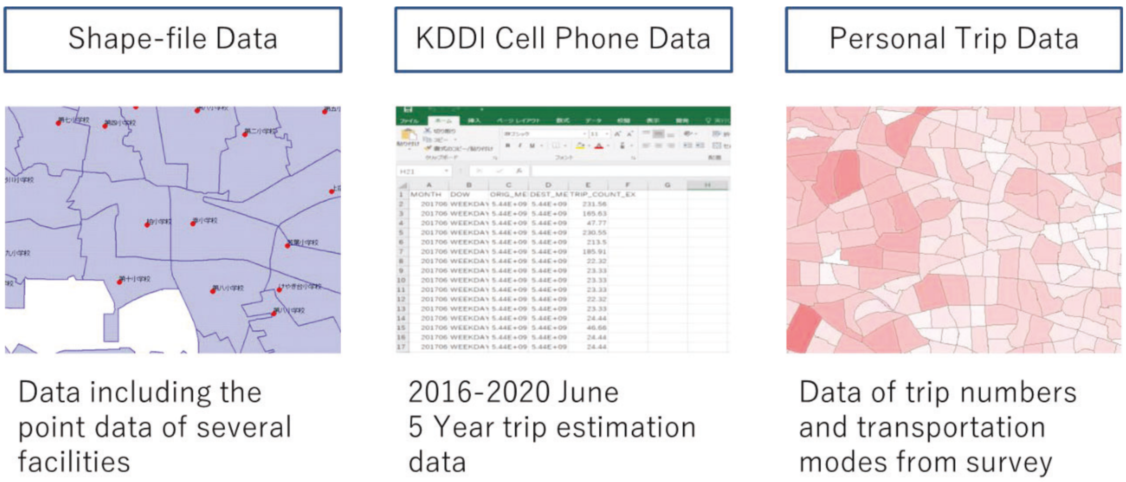


Fig. 3.1 Data collected for this research

The software used here is ArcGIS. ArcGIS is one of the GIS software provided by ESRI, similar to QGIS which can read data, create maps, and output. Another software used in this research is the software related to Python programming. Python is a language that emphasizes code readability. Python has gained a lot of support in the field of deep learning because it is simple code, easy to read, and has abundant libraries that can be used for calculation and statistical processing.

3.1.2. Data arrangement

The estimated trips count from cell phone data was organized in order to know how many trips start or end at one specific mesh. The way of getting the estimated trips count is presented in Fig. 3.2. Here, the estimated trips count was received from KDDI company and the calculation method was conducted by multiplying the expanding factor. For example, if the trips found is 20 and the percentage of users of this cell phone company is 10 percent. The estimated trips count would be 20/0.1. As a result, 200 will be the trips count been used in this situation. Although the judgement based on different distance and time may affect the results, 300m and 15 minutes were set here by the data processing process by KDDI company. Hereafter, since the data was available in 2016-2020, the frequency of trip increasing/decreasing was evaluated. I am interested in evaluate the trip frequency is increased or decreased in 2016-20 and identifying the causes for the resulting trip frequencies.

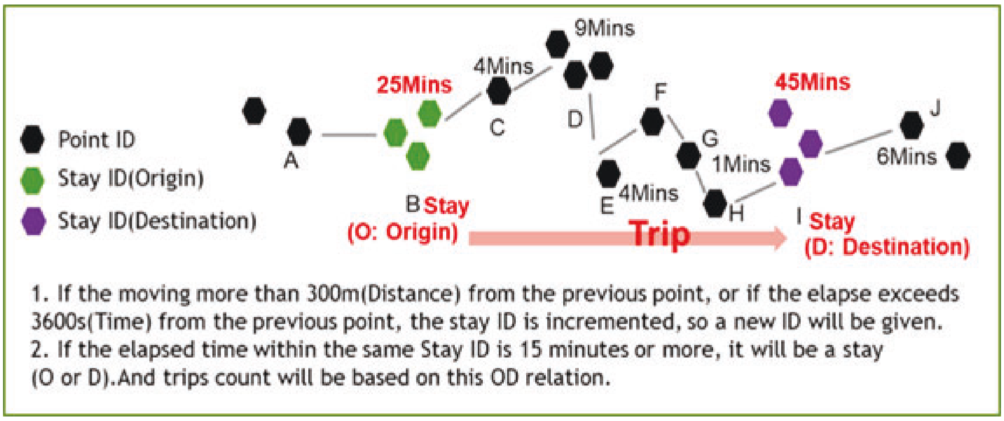


Fig. 3.2 Data arrangement for this research

3.2 Case Study

3.2.1. Overview of Utsunomiya City

Utsunomiya City, located in the central part of Japan, is in Tochigi prefecture within the Kanto Region. The prefecture government is located in Utsunomiya City and is located approximately one hour away from Tokyo station. Utsunomiya City is an industrial city and the population size is 26^{th} in Japan within 16 sub districts. Also, it is famous for the good environment for bicycle users. For instance, Japan Cup Cycle Road Race, which is held every October in Utsunomiya City is a well-known event. Nonetheless, by assimilating new challenges, Utsunomiya City is one of the leading suburban cities tackling declining birthrate and aging population (Nishiyama, 2019). The basic geography information of Utsunomiya City is shown in Fig. 3.3.

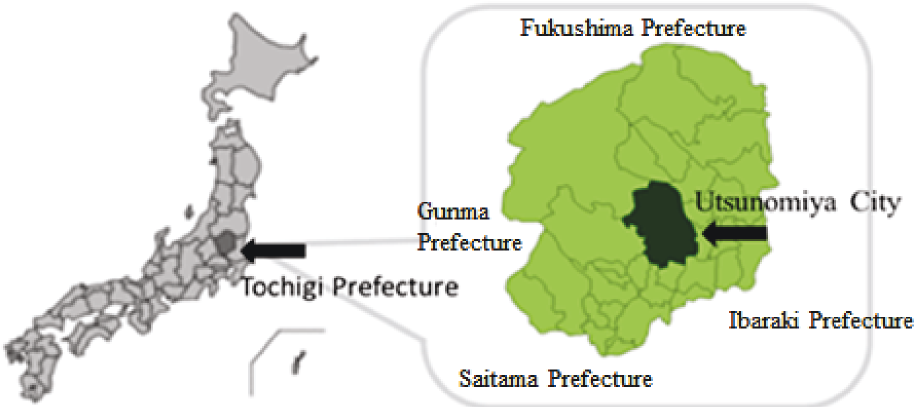
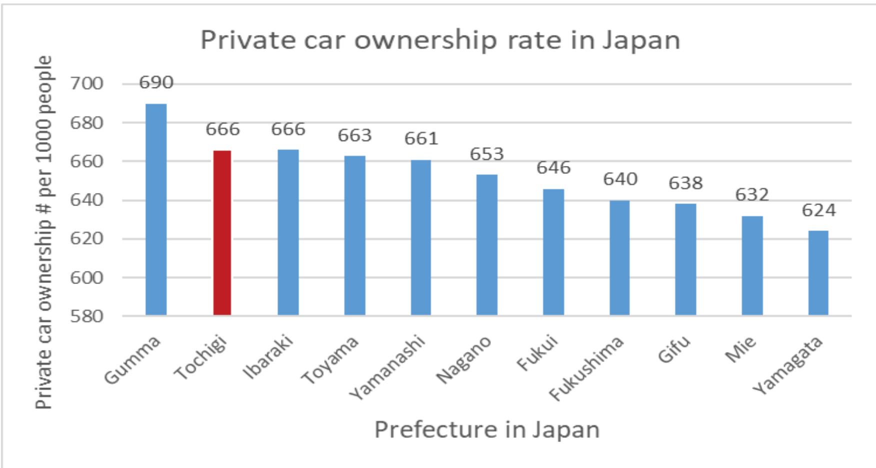


Fig. 3.3 Geography information of Utsunomiya City

3.2.2. Overview of transportation situation in Utsunomiya City

The Utsunomiya City is located in Tochigi Prefecture. Tochigi Prefecture has the second-high private vehicle ownership rate in Japan in 2018. One may find this result based on Fig. 3.4. Most citizens in Utsunomiya city still preferred to use private car (Ohmori, 2017). The new LRT system in Utsunomiya City is under construction, and is scheduled for operation in 2023. Since there are existing public transportation systems in Utsunomiya City, for example, bus systems and railway systems, the new LRT system in Utsunomiya City will surely serve as a good plan to integrate the public transportation systems. As a result, the potential to attract more citizens or tourists to select public transportation modes seems to have high possibility.



*Data used: year 2018

Fig. 3.4 Private car ownership in Tochigi Prefecture

3.3 Research Flow

Due to the availability of cell phone data from KDDI company, Utsunomiya City in Japan was selected as the target city in this study. Moreover, some open-source datasets were also downloaded by showing on the map on ArcGIS software. To make the research results reliable, I implemented a data cleaning process to ensure data is correct and consistent. Two steps in Fig. 3.5 are used for data analysis through using Association Analysis in Step 1. In order to make further understanding based on the findings of the Step 1, that is, to confirm the effect of the distance of different infrastructures, IRL was applied to traditional personal trip data in Step 2 to figure out the relation between trip purposes and selection for transportation modes. Finally, results were then shown and concluded. The data preparation and main calculation parts of the research flow are shown in below Fig. 3.5.

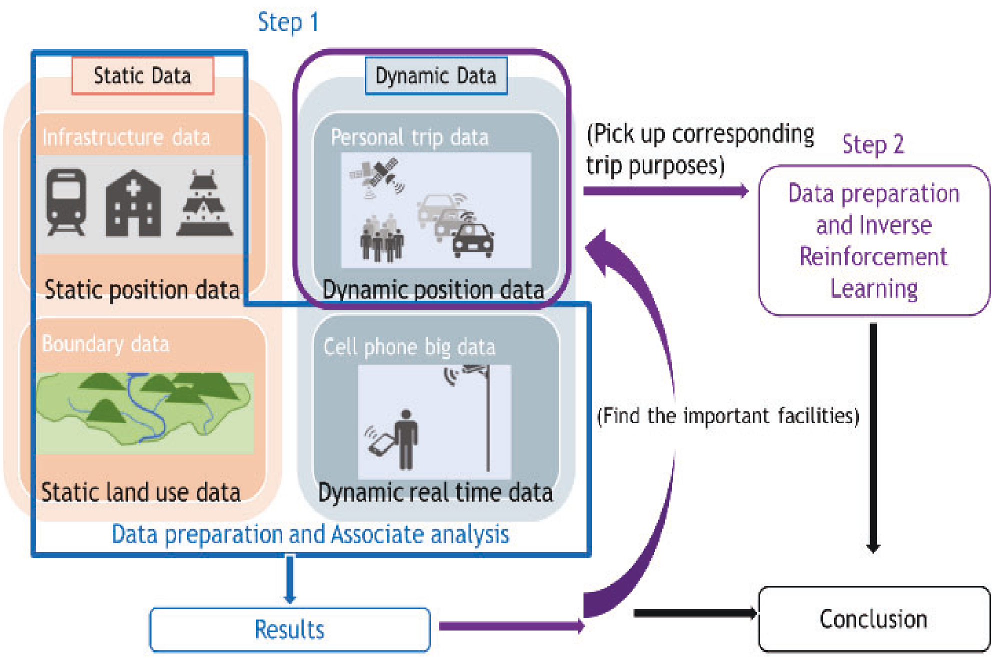


Fig. 3.5 Flowchart of the research

Chapter 4

LRT PERFORMANCE THROUGH MAKING STANDARD[1]

[1]The materials of the current chapter were accepted to:
International Conference of Asian-Pacific Planning Societies 2018(Independent Chapter)

Many countries face the challenges such as population deduction, local income reduction. The LRT system is considered a good method to improve the situation of these problems by making cities become TOD (Transit-Oriented Development) City or Compact City through adapting LRT system. The rapid growth of LRT systems around the world is an obvious fact for the recent years. Especially, there are numerous new projects in China for the past ten years. Nevertheless, there is a lack of clear definition of LRT. This research aims to make a good standard through several process for future reference. Although the fast growth of LRT development in cities in China, there is lack of research related to city development and LRT. Hence, the comparison studies in Society, Economic, and Environment dimensions between with and without LRT in cities in China was conducted. At the same time, case studies for Suzhou and Nanjing were conducted by this standard.

After the process of standard making and case studies, mainly two results were found. Firstly, cities in China with relatively higher increasing rate in economy perspective are more willing to develop LRT system in public transportation network. Nonetheless, the purpose of developing LRT in China is different to developed countries since China still expanding the city scale while developed cities using LRT to solve the problem of shrinking society by making cities into Compact City. Secondly, a fair standard was released to easily make a primary approach to understand and evaluate LRT system. From case studies, one can find the LRT in Suzhou and Nanjing perform well in this standard. Nevertheless, the design part is the weakest part of these two systems, representing more specialties should be considered when making a new LRT system.

4.1. Introduction

4.1.1. Background

The term LRT, referring to Light Rail Transit was first introduced in 1972 in North America to describe the new concept of rail transportation (Thompson, 2003). Although once regarded as not efficient since the capacity is quite low compared to other heavy rail transportation system; recently, emphasis on using LRT become more and more obvious because many merits can be found in LRT system. For instance, the construction cost is lower than other rail systems. Also, the stability and quiet when the operation is an advantage for passengers and residents.

More than 169 cities in 57 countries around the world have already embraced the usage of LRT system, at the end of 2016 (Williams *et al*., 2018). Furthermore, there are many cities are under construction of new LRT system. Not to mention that there is much planning for future construction. Many Asia cities are constructing new LRT projects. Especially China has about 5 cities been under construction and 12 cities are under planning of new LRT systems, at the end of 2014 (Xia, 2017).

Obviously, the significance of LRT definitely will be more and more huge because of numerous newly established project. In developed countries, many cities have met some problems such as budget, vacant homes due to the reduction of population, aging society. LRT would become a good method to improve the situation of these problems by making cities become TOD (Transit-Oriented Development) City or Compact City through adapting LRT system (Takami *et al*., 2008).

4.1.2. Objective

Each city has the significant difference in several aspects such as geography, culture, and history, leading to the different preference of transportation mode. Nevertheless, when a new LRT system was established, the service that LRT can provide to all passengers would be similar no matter where the system was constructed. What elements should be comprised of an LRT system is quite important. Hence, a standard can be taking referenced when setting up a new LRT in the future was released in this research.

4.1.3. Literature review

i. BRT standard:

The Theses of BRT standard were first reviewed to clarify the concept what elements should be composed of a standard. The paper BRT and LRT-Model selection of public transportation, mentioned about the result of BRT scores based on the standard by ITDP around the world. Gold, Silver, Bronze or Basic BRT were then classified depending on different scores the BRT systems acquired (Henke, 2012).

ii. Present LRT situation in China:

Also, review of research article related to the current condition of LRT development in China was conducted. The impact of urban rail transit on commercial property value: New evidence from Wuhan, China, mentioned about condition in China LRT. China government is planning to develop LRT with a significant increase rate in order to fulfill the balanced development not only in well-known and big cities but also in other cities. At the same time, traffic congestion was terrible due to fast increase in the numbers of private car ownership. LRT projects seems to be a suitable project to increase the public transportation usage rate (Xu *et al*., 2016).

iii. LRT impact to cities:

In addition, in order to understand the impact an LRT can cause to the city, the research about cities development was reviewed. For example, the paper Policy process management in the introduction of a new LRT system: A case study in Toyama, Japan showed the process of adapting LRT in Toyama City, Japan (Kato *et al*., 2008). Other analysis in Economy viewpoint also shows the price of the houses along the LRT route was risen, while other places were a deduction (Shibata *et al*., 2012). The Toyama City is a very successful case of LRT operation.

4.1.4. Characteristics of this part

Many researches were conducted to make emphasis on the social effect caused by LRT system. Also, some elements of LRT were discussed in many papers. Nonetheless, none of a good and complete standard was set up to make the concept of LRT easy-understanding and clear for future reference. This research rearranges some indicators into five categories and the goal of LRT system, as shown in Fig. 4.1, and showed a relatively complete standard.

In a real LRT project, there are numerous issues should be concerned from different perspectives. For example, from business point of view, how to get as much profit as they can will be the key issue. Local government will care more about whether the LRT can become a symbol of city. Since these concerns require more professional opinions and knowledge and is very complicated issue, here this standard will be focus on only the viewpoint from LRT passengers. Thus, convenient, safety, and price…and so on will be the key concerns when established this standard, as shown in Fig. 4.2. Of course, some indicators would be suitable for different point of views.

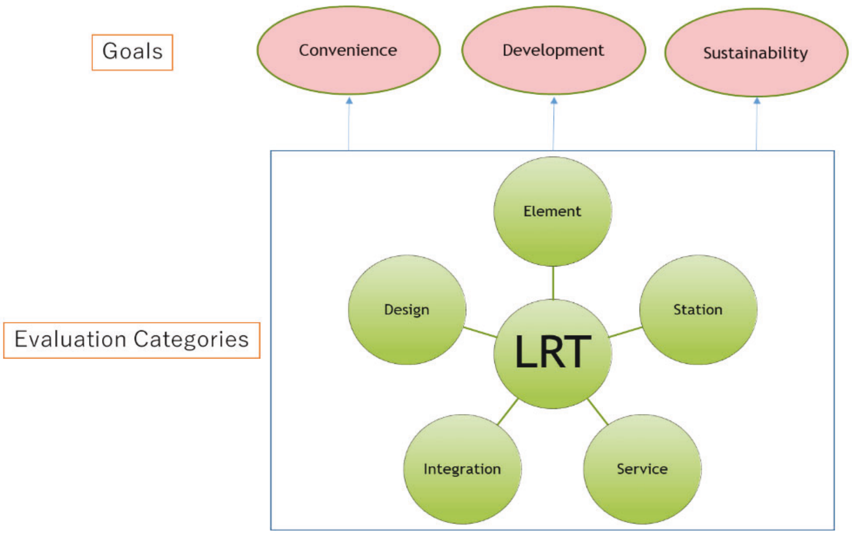


Fig. 4.1 Goals of LRT project and important parts of LRT system

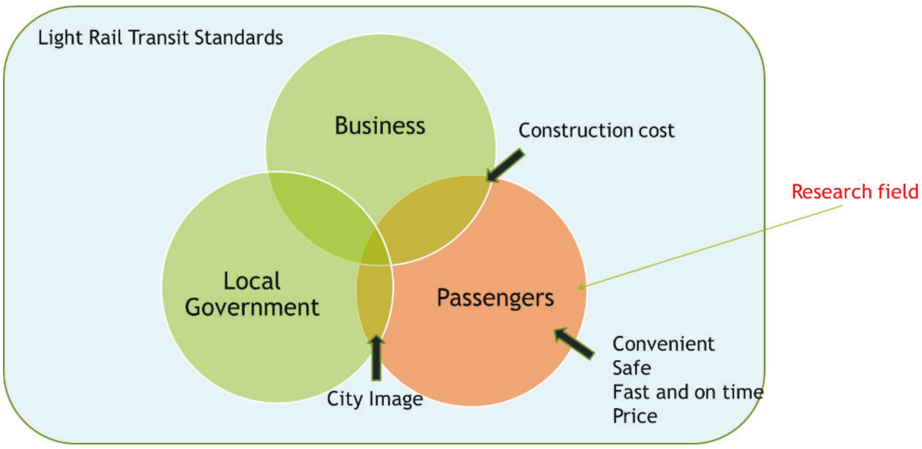


Fig. 4.2 Characteristics of this part

4.1.5. Flowchart

Firstly, literature reviews of the thesis relevant to BRT standard, LRT development in China and LRT influence were reviewed. In order to create a relatively fair standard for LRT system, five categories of indicators in LRT standards were established based on numerous thesis, books, and websites materials review. At the same time, target cities for case studies were selected here based on cities' background and cities' data in Society, Environment, and Economy viewpoint.

After the setup of standard and the choice of ideal sites-Suzhou and Nanjing, evaluation of the two systems were conducted here through scores assigned. The result of the two LRT systems in target places was made some comparison and reflection here. Also, comparison between these two LRT systems and old tram system in China was conducted. The research flow was shown in Fig. 4.3.

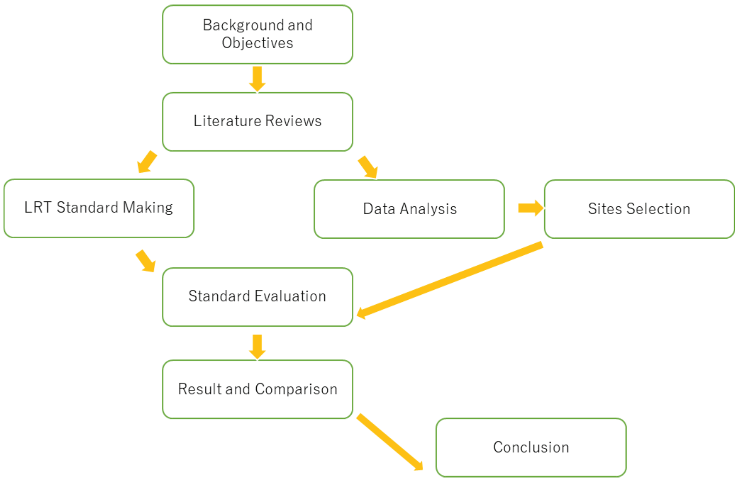


Fig. 4.3 Flowchart

4.2. Standards Assessment

4.2.1. Assessment standards for different modes

Transportation modes changed as the time changed. Nonetheless, public transportation plays an important role in our daily life is without doubt. Some public transportation methods have already developed over a long period. Therefore, in a relatively maturity stage. Based on the much experience, these transportation systems have already had many standards can be taking reference from (Ramos-Santiago *et al*., 2016).

However, some public transportation methods are relatively new in this era such as BRT, LRT. As a result, establishing a standard for the system would be preferable so that new project will have reference to make a brand-new system in the future. It can also be applied to the existing system so that the system can be improved by evaluation through the standards. Of course, the standards can also be made some adjustment in the future to make that probable for the instant changes in the new era.

4.2.2. Overview of existing BRT standards

The ITDP (Institute for Transportation and Development Policy) in the U.S.A. published an official version of BRT standard in order to evaluate the existing BRT system since there are more and more systems all around the world. There are thirty indicators in five categories of positive points with twelve indicators representing negative points. In the standards, some basic elements were put special emphasis on because they were the requirements so that the system can be called a BRT. Other are some characterizations that can make the system more willing to be used by passengers. Finally, after all the target systems were applied to all the indicators in standards. The systems around the world were rewarded with Gold, Silver, Bronze or basic BRT (Hidalgo *et al*., 2013).

4.2.3. Adaptation of BRT standard for LRT

The reason is remained unclear why there is a BRT standard but no LRT standard. The possible reason is because the BRT is an idea come from Latin America so U.S.A. want to learn from other countries' experience to understand BRT more clearly. The BRT standard released by ITDP is great and designed by a group of expertise in a big company. As a result, this LRT standard will be just a primary and basic approach to understand and evaluate LRT systems. The indicators listed here remain changeable if further adjustment is needed. Also, sometimes special considerations or adjustment should be concerned in some special culture or situations.

From the standards for BRT, part of standards for LRT can be set up since BRT and LRT shares some similarities. Nonetheless, differences between two systems should be clarified because they still have big difference while LRT is a kind of rail transportation, but BRT systems operate on the road. Therefore, the details of LRT standard needed to be determined by reconsidering the difference between both systems. The weighting of each indicator should be made some slight change, also.

4.3. Construction of LRT Standard

4.3.1. Review of existing LRT

In order to making the standard in a relatively fair and objective point of view, the LRT systems around the world were under review. The characteristics of the reviewed LRT systems are listed in Table 4.1. For the reason to make this Table precisely, discussion with several professors and specialists was conducted here. The LRT for reference here includes all the countries in different states. For example, France, England, America, and Japan. Since the Europe started to develop this idea earlier than other countries, Europe countries are many here. What elements and characteristics these LRT has can tell us what we should concern when making a new LRT system. The standard took much reference from these cases.

Table 4.1 Reviewing of LRT cities worldwide

Country	City	Year	Signal	Greening*	Integration	Catenery less	City Image	Dedicated way	Universal Access	Information provided	Parking space	Pedstrian Space
Canada	Calgary	1981			O			O		O	O	
U.S.A.	Portland	1986	O		O			O		O	O	
	Houston	2004						O	O	O		
	Tacoma	2003			O			O		O	O	
	Minneapolis	2004			O			O		O		
	Trenton	2004			O	O		O	O	O	O	
	Jersey City	2000			O			O	O	O	O	
Turkey	Istanbul	1989						O	O	O		
France	Strasbourg	1994	O	O(T+G)	O			O	O	O	O	O
	Lyon	2001		O(G)	O			O	O	O	O	
	Marseille	1994		O(Tree)			O	O	O	O		
	Nice	2007		O(Tree)		O		O	O	O	O	O
	Bordeaux	2003		O(T+G)	O	O		O	O	O	O	O
	Le Mans	2007		O(G)	O			O	O	O	O	O
	Nantes	1985	O		O			O	O	O	O	O
	Montpellier	2000		O(G)	O			O	O	O	O	O
	Grenoble	1987	O	O(T+G)	O			O	O	O	O	O
Spain	Alicante	2003		O(G)	O			O	O	O		O
	Bilbao	2002		O(G)	O			O	O	O		O
	Madrid	2007			O			O	O	O		
England	Mancester	1992		O(G)	O			O	O	O	O	O
	London	2000			O			O		O	O	
	Nottingham	2004			O			O	O	O	O	O
Portugal	Porto	2000		O(G)	O			O	O	O		O
	Lisbon	2006	O					O	O	O		O
Hong Kong	Hong Kong	1988	O		O			O		O		
Japan	Toyama	2006			O			O	O	O		
Taiwan	Kaohsiung	2015			O	O	O	O	O	O		
	*In greening part, T stands for Tree growing along the routes while G stands for Grass railway											

4.3.2. Unique indicators of LRT

Several aspects should be considered indifference between bus system and rail system. Here, more than 20 LRT systems around the world, including Europe, U.S.A., Asia were under review. What was conducted here is comparison between the characteristic of bus system and rail system by directly observation plus materials review. Also, looking into details of characteristic of LRT plays a significant role in creating new indicators for LRT indicators. The below are some indicators were established.

First, the air pollution will not cause by any LRT system, so the indicators related to pollution has been deleted. Furthermore, while bus operates on the road, the LRT operates on a rail. Therefore, LRT can be found more characteristic about environment-friendly. For instance, the greening can be made through grass ground establishment along the rail route. Some cities also grew trees at the same time LRT was made in order to make better living environment.

Also, when LRT project was decided by a city, several other rail systems may have already existed in the city for a long time. Integration the original system with newly constructed system plays an important role in the success of a new system. For example, the station integration and transportation card integration. Compare to BRT, LRT has some advantages in integration since all rail system can share some similarities, such as rail station.

Last but not least, the LRT has several operating styles because of different choice of cable lines construction. At the same time, standing up as an identity of city images, LRT can become fascinating to users not only citizens but also tourists. Cable lines quantity will have a significant impact on whether the system can become more beautiful and thus become more charming to all passengers. Hence, an indicator of cable lines selection was set up.

4.3.3. Adjustment of indicators weighting

After the discussion from the above parts, the title of all the indicators, classified into 5 categories of the standards and the weighting of each category are determined. The 5 categories and weighting of each category are shown below:

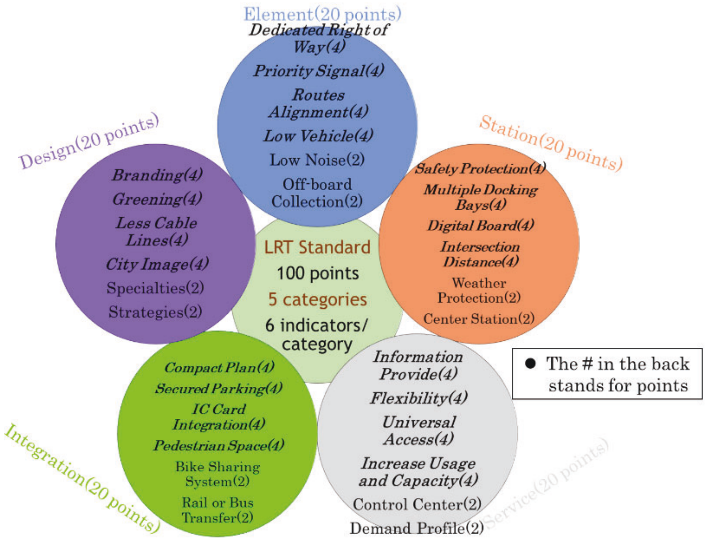


Fig. 4.4 LRT standard (General)

From Fig. 4.4, one can realize that there are 6 indicators in each points category. The total scores of all good indicators equal to 100 points. For the reason not to cause trouble by the completely weighting scores of different indicators, all the indicators here were readjusted to only two different patterns. If the indicator is relatively more important, it can stand for four points. On the other hand, some indicators only stand for two points if not many LRT systems share this characteristic. To understand clearly, details of the full version of the indicators and scoring system are shown in Table 4.2.

Table 4.2 Scoring methodology

Category	#	Indicator	0(No)	2 or 1(Slight)	4 or 2(Strong)
1.Element	1.1	1.Dedicated Right of Way(4)	Below 50% routes have dedicated way	50%-80% routes have dedicated way	100% routes have dedicated way
	1.2	2.Priority Signal(4)	None use of all intersections	Part intersections use	All use of all intersections
	1.3	3.Routes Alignment(4)	Meet Tier 3 Configuration	Meet Tier 2 Configuration	Meet Tier 1 Configuration
	1.4	4.Light Rail Vehicle(4)	None of LRV used	Mix of old tram and LRVs	All vehicles are LRVs
	1.5	5.Low Noise(2)	Reported noise problems	Obvious sound but not noise	No noise is found
	1.6	6.Off-board Collection(2)	All Stations On-board Payment	Mix of both kinds	All Stations Off-board
Category	**#**	**Indicator**	**0(No)**	**2 or 1(Slight)**	**4 or 2(Strong)**
2.Station	2.1	1.Safety Protection(4)	None any safety strategy	Use safety d oor or camera	Use safety door and camera
	2.2	2.Multiple Docking Bays(4)	All station have only one	One station have	More than one station have
	2.3	3.Digital Information Board(4)	None station have	Some stations have	All stations have
	2.4	4.Intersection Distance(4)	Less than 50% meet the suggested distance	50-80% meet the suggested distance	Above 80% m eet the suggested distance
	2.5	5.Weather Protection(2)	None any strategy	Some stations have	All stations have
	2.6	6.Center Station(2)	None center station	Have but not yet serve	Have and serve for multiple routes
Category	**#**	**Indicator**	**0(No)**	**2 or 1(Slight)**	**4 or 2(Strong)**
3.Service	3.1	1.Information Provide(4)	None of clear information	Information is weak	Enough information provide
	3.2	2.Flexibility(4)	None of Flexibility	Some space for emergency	Some space and railway flexibility
	3.3	3.Universal Access(4)	None of universal access	Some disable friendly	All disable friendly(Wheelchair, deafness…)
	3.4	4.Increase Usage and Capacity(4)	None of strategy	Specialty or discount	Specialty and discount
	3.5	5.Control Center(2)	None control center	Share control center	Specific control center
	3.6	6.Demand Profile(2)	None of research	Based on demand profile	High demand routes
Category	**#**	**Indicator**	**0(No)**	**2 or 1(Slight)**	**4 or 2(Strong)**
4.Integration	4.1	1.Compact Plan(4)	None any compact idea	Compact transfer or compact land use	Compact transfer and compact land use
	4.2	2.Secured Parking(4)	None of parking space	Bicycle or car parking space	Bicycle and car parking space
	4.3	3.IC Card Integration(4)	None with existing systems	Part with existing systems	Entirely with existing systems
	4.4	4.Pedestrian Space(4)	None increase of space	Increase walking space	Increase space and become Transit mall
	4.5	5.Bike Sharing System(2)	None system have	One system along routes	More than 2 systems along routes
	4.6	6.Rail or Bus Transfer(2)	None transfer of two	One transfer of two	Both transfer of two
Category	**#**	**Indicator**	**0(No)**	**2 or 1(Slight)**	**4 or 2(Strong)**
5.Design	5.1	1.Branding(4)	None obvious name	Have name but not clear	Special name and easy remember
	5.2	2.Greening(4)	None of greening strategy	Grass railway or trees growing along the routes	Grass railway and trees growing along the routes
	5.3	3.Less Cable Line(4)	All routes have catenary	Part of routes have catenary	All routes catenary less
	5.4	4.City Image(4)	None relation with city image	Vehicle or platform have city image	Both vehicle and platform have city image
	5.5	5.Specialties(2)	None of specialty	One specialty found	Two or more specialties found
	5.6	6.Strategies(2)	None of	Special strategy	Special strategies

In order to understand the indicators easier, visualizations of most indicators were made. Take the integration part for example, as shown in Fig. 4.5, compact plan and the integration with other service such as bike sharing service or secured parking is of great importance.

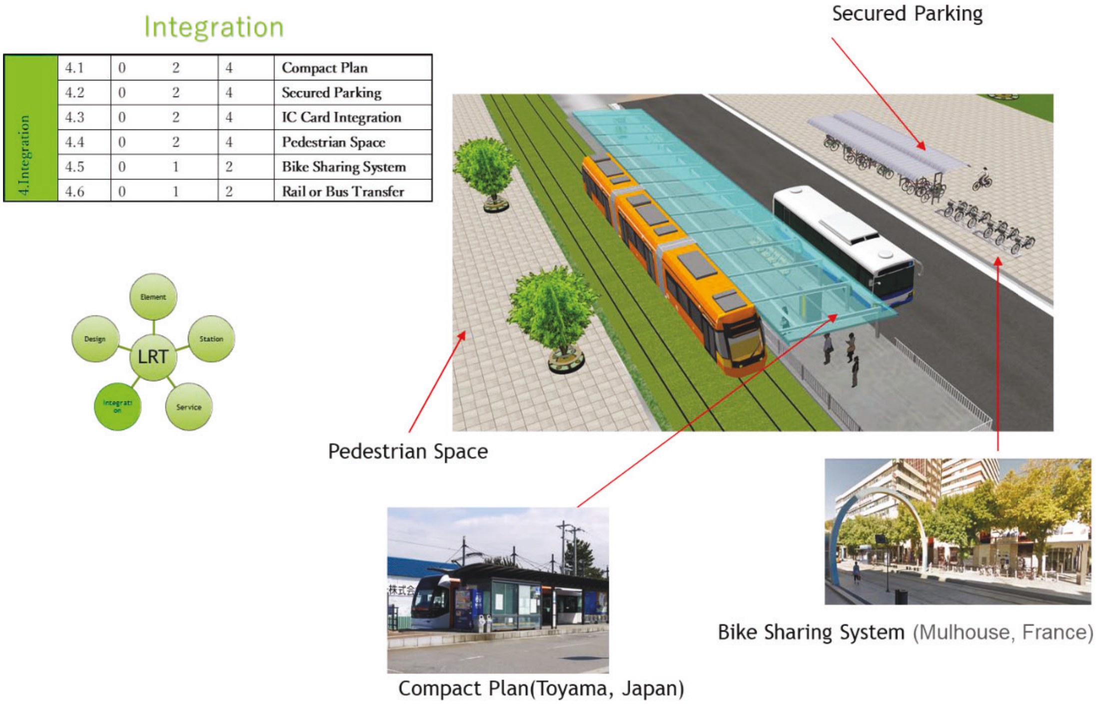


Fig. 4.5 Visualization of the indicators in integration part

4.4. Case Studies for China

4.4.1. Transportation background in China

In China now, more than 170 cities now home to more than one million residents. Because of the fast urbanization, these cities require efficient public transport systems. China's metro construction has dramatically development for the past ten years. However, Chinese cities are increasingly looking to light rail to solve their public transport issues, and in particular to serve new residential developments (Schulz *et al*., 2015).

Despite the unprecedented pace of metro construction, buses remain the main form of public transport in cities across China. Many cities in China decided to put BRT systems to increase the public transportation networks. Actually, some of BRT got big success and reputation. For instance, the BRT in Guangzhou is a good and successful case. It is rewarded Bronze title based on BRT standard set up by ITDP.

4.4.2. LRT background in China

Nevertheless, after year 2000, many governments in China decided to put LRT in their public transportation networks. Until 2015, there are already 8 LRT systems in China. In addition, there are about fifteen plans relevant to LRT projects are under review or under construction. The reasons are the capacity of LRT is higher than BRT while on the other hand, the prices of making LRT vehicle is cheaper. Besides, the air pollution is another serious problem in many cities in China. LRT have more advantage to improve this problem than bus. Also, LRT can be made more greening because LRT is operated on railway, not road. Some LRT systems even grow trees along the routes to make greening more obvious.

As for the heavy metro, there are several reasons for choosing LRT, rather than metro systems in China. As urban populations continue to expand, many cities are experiencing the development areas known as New District or New Town, which are often situated some distance from the traditional city center.

This condition therefore putting pressure on transport planners to deliver public transport networks since the traffic congestion on roads in China are already serious for many years. Nonetheless, metros are expensive and time-consuming to build the infrastructure for operating metro systems. Actually, they are also expensive to operate and maintain because the running report showed many governments are operating metro system at a loss. The construction costs of LRT are only about 20% of a metro system. As a result, particularly in small and medium-sized cities are embracing the merit of LRT now, therefore developing the LRT project in a fast growth way. At the end of year 2014, eight Chinese cities operated light rail networks, with a cumulative distance of 192.6km, and several cities are on course to open their first lines in near future. Plans are now in place to develop more than 2000km of lines by 2020 and up to 4000km of lines by 2050 (Schulz *et al*., 2015) (Briginshaw, 2014).

4.4.3. Data used and sources

Here from the book China City Development General Index 2016 (Zhou *et al*., 2016), three different dimensions of Society, Environment, and Economy ranking for 295 cities in China were shown. By reviewing this book, the basic ranking of cities in China was shown and clearly understood.

The data also used here mainly come from two resources. One is from the National Statistical Bureau official website from China. The other is from some year book published by local government or central government. For example, China City Statistical Year Book and Suzhou City Statistical Year Book. Numerous information can be found in the website version of the year book.

4.4.4. With and without LRT comparison

Here, seven different variables were selected for the analysis process due to data availability. That are, GDP, Population, Housing Prices, Travelers, Noises, PM2.5, and government budget. The number of cities selected here is 37. In these cities, 8 have LRT system, while other 29 cities without any tram system or only very old tram system.

The Discriminant Analysis Method was firstly conducted here. The only result here is that the construction of LRT is related most in GDP performances ranking. That is, the higher the GDP is, the cities are more willing to select to develop LRT projects. Other variables are relatively in low relation.

The main comparison shown in Fig. 4.6, Fig. 4.7, and Fig. 4.8 comes from the cities with or without LRT. First, top ten cities in China were compared in ten cities data, checking that how Suzhou and Nanjing performed. Here, some of the cities' environment data ranking was compared and shown in Fig. 4.6. Next, the easy comparison using the average increasing rate of several variables of all the cities from year 2011 to year 2016 was conducted here. Selecting the most important two variables here and making the Fig. 4.7, Fig. 4.8 was hereafter released. As one can find, the cities with LRT perform better in GDP and housing rates. From this result, one can say that LRT is mainly constructed in cities that is developing in a higher rate.

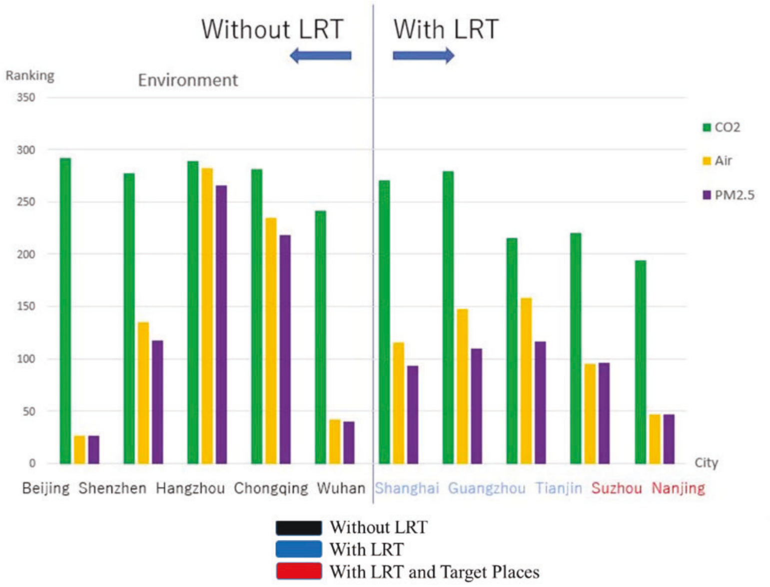


Fig. 4.6 Comparison results-Environment

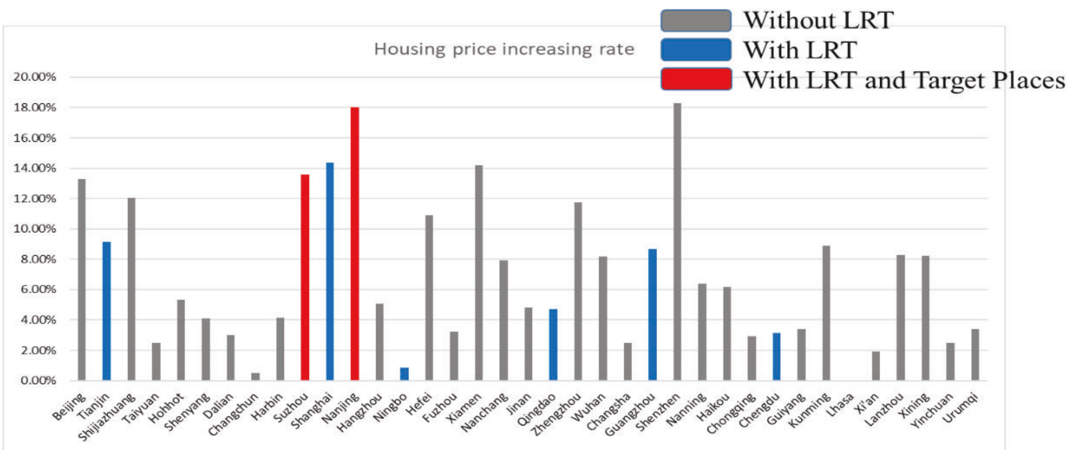


Fig. 4.7 Comparison results-Housing prices

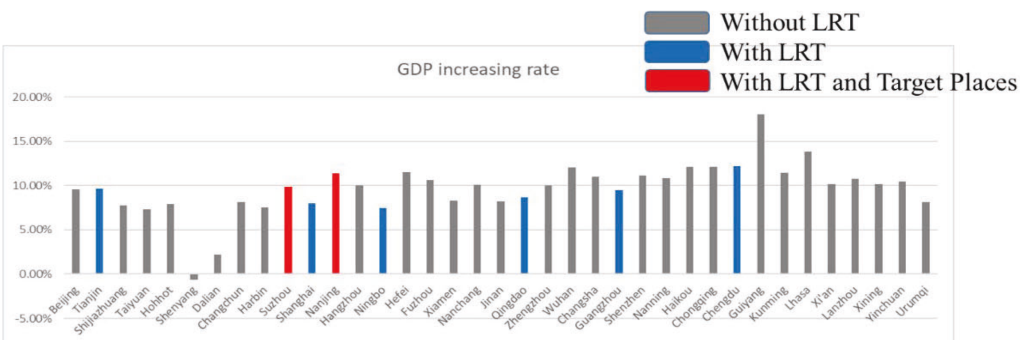


Fig. 4.8 Comparison results-GDP

4.5. Study Area

4.5.1. Sites selection methodology

Since the number of LRT in China has increased fast during the past ten years and numerous projects were under construction or under review. Case studies starting from China was determined. From the book China City General Development Index, the ranking of city performance in Society, Environment and Economy were evaluated through various data analysis. Due to the availability of information and data analysis, the two target cities-Suzhou and Nanjing were selected because both are ranked in overall top ten cities based on the above three dimensions. In addition, the two cities have new constructed LRT in the past ten years. Therefore, these two cities were the target sites for this research.

As mentioned before, the cities that has LRT in China is 8 now. Here for the reason to apply the standard to some cities to get more findings. Two cities were selected here-Suzhou and Nanjing. The reasons for selecting these two cities is because these two cities share the below similarities. Firstly, both cities operated the brand-new LRT system in year 2014. Secondly, they are both located in same province-Jiangsu province. Also, the two cities performed similar characteristics in the above with and without data comparison. For example, their environment ranking is relatively worse than other important cities in China. In the economy perspective, Suzhou and Nanjing have bigger potential than other cities since the average increasing rate of GDP and housing prices from year 2011 to year 2016 is huge comparing to other cities. This is based on the figures shown last section.

4.5.2. Overview of selected cities

i. Suzhou:

A major city located in southeastern Jiangsu Province of East China, about 100 kilometers northwest of Shanghai. It is a major economic center and focal point of trade and commerce, and the largest city in the province. The city is situated on the lower reaches of the Yangtze River and the shores of Lake Tai and belongs to the Yangtze River Delta region. The system in Suzhou, called Suzhou Trams was operated from October in year 2014, with operation distance 18.1 Kilometers. The line has seven stations, operates on dedicated tracks and is situated entirely within the Gaoxin Qu urban development area, which covers $52km^2$ and is located 5km west of Suzhou in Jiangsu province, on the banks of Taihu Lake (Wikimedia Foundation, 2018).

ii. Nanjing:

Nanjing is a city situated in the heartland of the lower Yangtze River region in China, which has long been a major center of culture, education, research, politics, economy, transport networks and tourism. It is the capital city of Jiangsu province of the People's Republic of China and the second largest city in East China region. Nanjing tram is the tram system of Nanjing city in the Jiangsu of China. There are two lines which are not connected to each other. The two lines were also started from the same year with Suzhou LRT in 2014. The LRT system were put into operation before the second Summer Youth Olympic Games between August16 and 28, 2014. Catenary is installed on only 10% of the Hexi Line route since the LRVs can be recharged from the catenary at stops and during acceleration[11)].

4.6. Calibration of Scoring System

4.6.1. Rubric construction

After the indicators were decided, detail definition of scoring system should be constructed, in order to make the clarification of scoring given. Here, the definition of each indicator and what is the way to give a proper score to this indicator was released. In all indicators, basically there are two kinds of indicators:

i. With or Without:

Literally, the point would be given once the LRT has this characteristic, while zero points would be assigned if without this characteristic. Some difference in points would be given based on a different situation. For example, the Branding, represents whether the system has a powerful name. If yes, full score will be given.

ii. Different Level:

In this part, the indicators are not only with and without. Nonetheless, more points would be given if better usage of the indicators. For instance, the Greening and the Pedestrian Walking Space.

4.6.2. Assigning scores to selected cities

The next step of the process is applying all the indicators in standards to target LRT system. The information collection has several resources. One is from existing websites of LRT. Much desired background information of basic operation and design of system were acquired here. In addition, some other information was found in some operation reports. After LRT was put into operation, there have annual or season statistical reports, dealing with operation status and some troubles LRT have met. About one-third of the indicators can be assigned based on this part information. Last, some indicators need more detail of examining the system. The best way to this part can be done by going to real sites examining. However, due to time limitation, this kind of indicators was given points based on some articles or research papers or videos on website. In their discussion, extracting the things needed is the methodology used here.

After collection the information needed, the scores are given to the two selected sites. In order to giving the scores in an objective way, several meetings and discussions were conducted here with some members in laboratory, professors, and specialties. The results are shown here in Fig. 4.9.

	Suzhou	Nanjing
Element	16	14
Station	18	16
Service	14	18
Integration	16	16
Design	10	12
	74	76

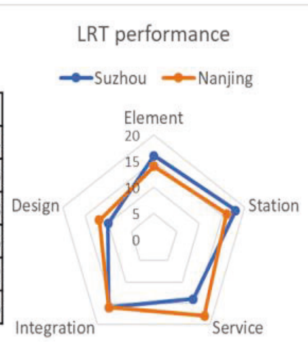


Fig. 4.9 Scores of LRT in target cities

4.6.3. Analysis of results

Here, the scores assigning for old tram system with new LRV (Light Rail Vehicle) in Dalian City in China was also conducted. The comparison between scores for the LRT system in two target places and for tram system in Dalian City was therefore compared. Obviously, in all the 5 categories the new LRT system performed better than tram system. Especially, in the integration part, the old tram system got almost no points. As a result, the new LRT system will be more suitable for this new era since it can show more characteristics of convenience and integration. At the same time, this can verify that the new established standard for LRT is fair and useful. The compared results after comparison analysis is shown here in Fig. 4.10.

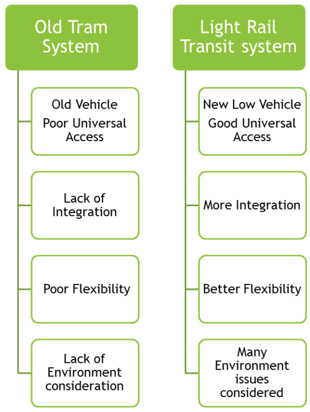


Fig. 4.10 Comparison results between tram and LRT

From the above result in Fig. 4.9, obviously, two systems perform well in Station category. Integration category also in good level, better than old tram systems in China. Nonetheless, the Design category need more improvement, representing the specialties of specific LRT system should be considered.

4.7. Summary

4.7.1. Summary of this chapter

In this research, a new standard for LRT system was established. No doubt, different cities in different countries would have an entirely different situation in public transportation, leading to the different design of LRT system. Nonetheless, the standard was set up based on relatively fair consideration, trying to be appropriate to any system. By using the existing cities data in Social, Environment and Economy dimensions and making simple comparison with and without LRT in cities in China, the condition in LRT development was shown. Next, by assigning scores to the target cities, the analysis the characteristics of both LRT systems was conducted. Last, comparison between the new LRT systems and old tram system in China was conducted, thus verify the two systems are good enough to become new LRT systems. At the same time, the standard was verified to be good and fair for the future use.

In conclusion, from this standard, one can understand the characteristic of LRT well, also standard can show good evaluation for LRT systems.

Actually, from the process of research. There is one more point should be clarified is that the main purpose of constructing LRT in China is quite different to other developed countries. While developed countries are facing shrinking scale of city, cities in China are still experiencing the fast growth of population and scale. As a result, the main reason to construct LRT in China is to expand the scale of cities whereas cities in developed countries are constructing LRT in order to decrease the scale of city into a Compact City.

4.7.2. Future implementation

The scores were assigned to the two target cities. The overall cities comparison between with and without LRT has been done. Further steps, more analysis and Comparison will be conducted through some figures making. Some indicators in standards would be validated and changed the weighting if needed.

In the long-term plan, the LRT standard should be suitable for more cities around the world. As a result, next step will start from expanding the standard to more cases in China. Then, in order to cover more countries, next research will probably expand to more cities in Asian countries.

In order to verify this standard for worldwide use, next research will be related to expanding this standard to LRT systems in other countries. For instance, Japan and Taiwan. Japan has the successful Toyama LRT that has already put into operation for more than ten years. In Taiwan Kaohsiung City has the first LRT system without catenary line along all the stations around the world. Therefore, expanding to the above two cities to make more comparable studies would be interesting and important to know more knowledge of LRT system.

In addition to the above consideration, trying to comprise other data in city development be another way to make comparable studies more meaningful. Also, trying to focus on condition changes along the LRT routes, that is, making the data scale smaller is a good and possible way.

Chapter 5

GENERAL SITUATION OF TRIP GENERATION BASED ON BUILT ENVIRONMENT EFFECT[2]

[2]The materials of the current chapter were published in:
Journal of the Eastern Asia Society for Transportation Studies

Many cities in Japan have met some problems such as budget problems, vacant homes due to the reduction of population, aging society nowadays. As a result, new transportation systems such as LRT (Light Rail Transit) are put into high emphasis recently since promoting the ridership of public transportation systems serves as a good way to solve these problems. Here, in order to understand basic condition of trip generation along the new LRT system in Utsunomiya City in Japan, cell phone data was used and calculated. By showing the results on ArcGIS map, one can know that which areas have the trend of increasing trips clearly. Furthermore, association analysis was conducted by utilizing cell phone data and shape-file data of infrastructures. Findings are the infrastructures such as hospital, parks and culture facilities are the most important elements in the area that the trends of trips are increasing in this case study.

5.1. Introduction

5.1.1. Background

Accompanying the rapid economic development around the world, there raise some challenges we never face before. Especially, in developed countries, many cities have met some problems such as budget problems, vacant homes due to the reduction of population, aging society. As a result, public transportation systems, especially the new transportation systems such as BRT (Bus Rapid Transit) and LRT (Light Rail Transit) are put into more emphasis nowadays. The term LRT was first introduced in 1972 in North America to describe the new concept of rail transportation. Good planning of LRT system in a city can make city become more attractive for citizen by making city become TOD (Transit-Oriented Development) City or Compact City. For example, the Toyama City, Japan serves as a great example to adapt LRT system. The new LRT project in Utsunomiya City, Japan is scheduled to put into operation in 2023.

Also, because of the improvement of the software and the hardware techniques of computer science, new analysis based on AI (Artificial Intelligence) methods are put on emphasis again these days. Furthermore, due to the widespread of the use of cell phone and the availability of cell phone data from some company, researches that are related to the application of this kind of dataset become more popular recently. Because the numbers of cell phone data are huge, AI analysis is often conducted to find useful information. Therefore, the research of conducting AI analysis by utilizing cell phone data in transportation field also seems to be put into emphasis more and more.

5.1.2. Objective

The main objective of this research is to provide a primary approach of understanding the trip generation condition along the new LRT route in Utsunomiya City through available cell phone data. By extracting the trips condition near the stations of the LRT system and making them visualized on the map, some other planning and integration with other public transportation systems can be considered based on the results here. Also, trying to find the potential relations between infrastructures and trips through several steps of data arrangement and association analysis is another goal in this research.

5.2. Literature Review and Flowchart

5.2.1. Literature review

The literature reviews were conducted for the following three themes. (1) LRT system and the land use condition near the LRT system, (2) mobile phone big data, and (3) AI.

As for the LRT system and nearby land use, Pacheco-Raguz (2010) analyzed the impacts of development of LRT in Manila on land price, land use, and population along the route. Using correlations and regressions, these variables are analyzed against an accessibility index and network distances obtained from a model built within a Geographic Information System (GIS). Other study includes Sakamoto *et al*. (2015) examined change in urban population before and after introduction of LRT in 27 cities in Europe. They also analyzed change in the population or areas along the LRT route for four cities in Europe. In addition, Higgins *et al*. (2014) reviewed the previous literature on LRT and other rail rapid transit systems in North America, demonstrating that rail transit alone is not a primary driver of land use change. As for the case in Japan, Kriss *et al*. (2020) conducted a detail study of Toyama City and pointed out the LRT project in Toyama City got significant success by showing the increasing ridership of the system. Papers which analyze the effect of LRT include Mochizuki *et al*. (2007) and Mizokami *et al*. (2007). Mochizuki *et al*. (2007) conducted questionnaire-based survey and compared the visitors in Toyama City who used different transportation modes when traveling. Mizokami *et al*. (2007) demonstrated profitability of conversion of Kumamoto Electric Railway into an LRT system. Also, Sato *et al*. (2018) conducted the estimation of population distribution in Utsunomiya City, Japan until 2050 by supposing different integration pattern of LRT system and feeder bus system. The most researches related to this field were mainly concerned about the effects that are resulted after the operation of LRT system.

Since the papers mentioned above are mainly based on questionnaire survey. However, due to the difficulty of acquiring survey data, the possibility of using other data source was taken into consideration. Hereafter, the articles related to application of cell phone data were reviewed. Cell phone big data also applied to many other fields. In transportation field, Widhalm *et al*. (2015) proposed a method to reveal activity patterns that emerge from cell phone data by analyzing relational signatures of activity time, duration, and land use. Also, Bachir *et al*. (2019) made the mobile network trajectories by using mobile data in Paris region. Some papers focusing on literatures review were also reviewed here. Huang *et al*. (2019) gave huge amount of literature reviews and clear discussion of this field. This paper figured out that simple rule-based methods making use of geodata were often employed. On the other hand, Wang *et al*. (2018) provided a review of existing travel behaviour studies that have applied mobile phone data, and presents the progress that has been achieved to date, and then discusses the potential of mobile phone data in advancing travel behaviour research and raises some challenges that need to be dealt with in this process.

Because the amount of cell phone big data is always huge, AI methods are needed for conducting efficient analysis. Regarding the articles published recently in transportation field related to AI study, Haenlein *et al*. (2019) takes a first insight into AI applications by summarizing seven articles published in this special issue that present a wide variety of perspectives on AI. Also, Abduljabbar *et al*. (2019) pointed out how is a good way to make various of application in transportation field by constructing models in several phases. Finally, since the association analysis serves as a good method to find the potential relationship between different items that would be difficult to observe directly, this methodology was selected. For instance, the document written by Tan *et al*. (2005) gave us general idea of how to set up models by different algorithms.

By showing some examples with programming, one can easily understand how to set up a method to conduct association analysis.

5.2.2. Characteristics of this part

Most researches related to the LRT systems were focus on the existing systems. Nevertheless, study of the system that is under construction was lack of abundant information. This research aimed to provide a preliminary study to understand the possibility of potential passengers to use LRT system before the operation by combining the cell phone big data and open-sourced data. Usually, the data from personal trip surveys were often utilized in researches that related to this field. Nonetheless, personal trip surveys are only conducted in specific year and the frequency is not very high. As a result, due to the difficulties of acquiring the trips data by questionnaire survey, cell phone big data was chosen for estimating the trips in this research. Here, two datasets based on different sources were processed and calculated through several steps and conducted association analysis to propose a new trial to understand the basic trip generation condition along the LRT system in target city before operation.

5.2.3. Flowchart

Firstly, literature reviews of the thesis relevant to LRT system and the nearby land use, mobile phone big data, and AI were reviewed. Due to the availability of cell phone data from KDDI company, Utsunomiya City was selected as the target city. At the same time, some open-sourced datasets were downloaded. Hereafter, the two different datasets were made some arrangement and were shown on map on ArcGIS software. Hereafter, datasets were calculated and analyzed through association analysis by programming on python language. Finally, results were then shown and discussed, followed by conclusion part. The flowchart of this part is shown in the below Fig. 5.1.

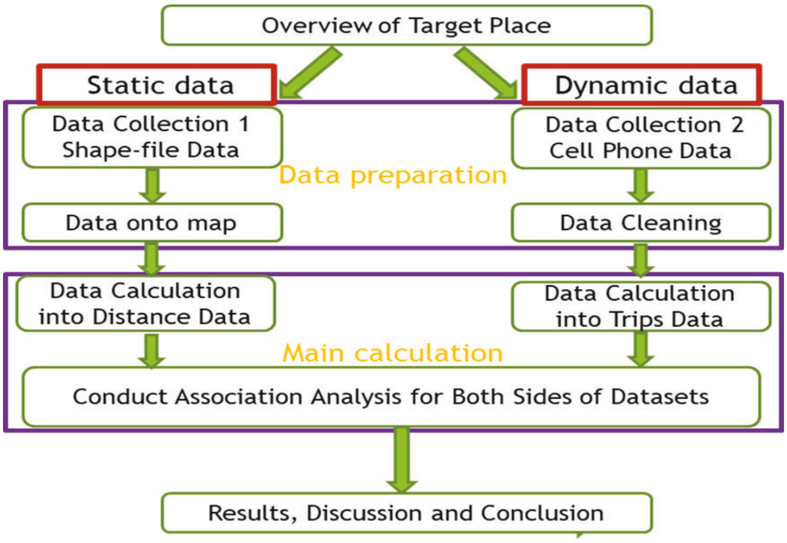


Fig. 5.1 Flowchart of this part

5.3. Case Study

5.3.1. Overview of Utsunomiya City

Utsunomiya City is a capital city in Tochigi prefecture and located about l00 km north of Tokyo in Japan. Notwithstanding the traditional face, Utsunomiya City is now called the city of gyoza, cocktail, jazz and, bicycle. The Japan Cup Cycle Road Race is held in this city every October from 1992. By assimilating new challenges, Utsunomiya City is one of the leading suburban cities tackling declining birthrate and aging population. Some basic information of Utsunomiya City is listed in Table 5.1.

Table 5.1 Summary of the demographics of Utsunomiya City

Total Population	519, 171 (Male: 259,613 Female: 259,558)
Area	416.85 km^2
Dimension	East West: 23.97 *km* North South: 29.53 *km*
Household Number	226,732
Number of Sub Districts	16

5.3.2. The overview of LRT system in Utsunomiya City

The new LRT system in Utsunomiya City is under construction, and is scheduled to put into operation in 2023. It will consist of 19 LRT stations, with total length 14.6 kilometer in the operation of first stage. Since there are existing public transportation systems in Utsunomiya City, for example, bus systems and railway systems, the new LRT system in Utsunomiya City will surely serve as a good plan to integrate the public transportation systems.

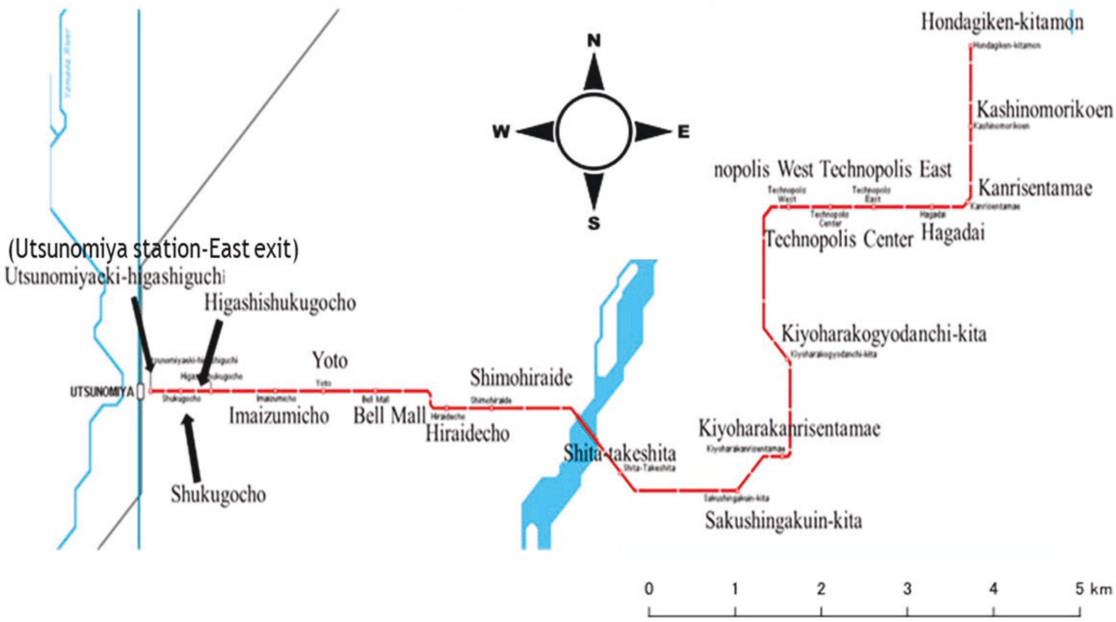


Fig. 5.2 LRT route in Utsunomiya City

5.4. Introduction of Data and Analysis Method

5.4.1. Data used and data source

There are two sets of original data. One is cell phone data of Utsunomiya City from KDDI company. Another one is geographic information from NLNI (National Land Numerical Information). Listed below in Table 5.2 is the summary of source data from KDDI company and National Land Numerical Information download service.

Table 5.2 Introduction of two datasets

	Cell phone data	Shape files data
Source	KDDI Research Inc.	National Land Numerical Information
Data Content	Estimated Trips(numbers)	The School, Culture Facilities(numbers),etc
Scale of data	Mesh (250m)	National, Local, Municipality
File format	Text file(.xlsx)	Point, Line, Polygon(.shp)

In the KDDI cell phone dataset, the data were acquired in June from 2016-2018. Here, only that users who have given permission to use the data are included in the analysis. Also, the personal information was not included in this study. The data can be acquired for three consecutive years was extracted here so there are total 3414 mesh data was available in this dataset. Signals was tested every 15 minutes here. When one signal is stayed in one mesh more than 15 minutes, this condition will consider “stay”, so that the same signal will not be counted repeatedly in order to increase the accuracy of estimation. Detail information of this dataset are listed below in Table 5.3. In this table, the data that was used in this research is the user's trip data extracted from the GPS logs of cell phones. The method of extended estimation of trips count will be explained in next chapter.

Table 5.3 Detail information of KDDI cell phone data

Data properties	Detail Explanation
Time of data	Year and month at departure (June in 2016-2018)
Day property	Type of day at departure (Weekday, Holiday)
Origin point (Location 1)	250m mesh cord of departure point (JGD2011 Coordinate)
Destination point (Location 2)	250m mesh cord of arrival point (JGD2011 Coordinate)
Estimation trips count	Number of trips after extended estimation

Data source: KDDI Research Inc.

5.4.2. Basic concept of Association Analysis

The association analysis, or often called the market basket analysis is an analytics technique often conducted by retailers and consultants to understand the purchase behavior of customers. Association Analysis or the market basket analysis is used to determine what items are frequently bought together and associate with one another. It uses this purchase information, or the transaction data, to leverage the effectiveness of sales and marketing. Moreover, Association Analysis looks for combinations of products that are frequently purchased together, and has been actively used in the industry once the immense amount of data has been available due to the creation of electronic cashier system.

5.4.3. Principles of Association Analysis

In this research, the association analysis was conducted to extract key features influencing the trips count. The purpose is to find the potential relations between the distance from mesh to infrastructures and the trips count in each mesh. Through constructing binary transaction data, it is possible to mine the hidden relationship of each item sets within the data set. Here, the association analysis provides information on how strong the association rule, or the if-then relationship, is found to be true in the dataset. In this analysis, the following equations are used to provide inferences on the cause-and-effect relationship.

$$\text{Support}(X) = |\{t \in T; X \subseteq t\}|/|T| \quad (5.1)$$

$$\text{Confidence}(X \Rightarrow Y) = \frac{\text{supp}(X \cap Y)}{\text{supp}(X)} \quad (5.2)$$

$$\text{Lift}(X \Rightarrow Y) = \frac{\text{supp}(X \cap Y)}{\text{supp}(X) \times \text{supp}(Y)} \quad (5.3)$$

$$\text{Conviction}(X \Rightarrow Y) = \frac{1 - \text{supp}(Y)}{1 - \text{conf}(X \Rightarrow Y)} \quad (5.4)$$

where,

X and Y: independent item set

T: set of transactions

t: individual transactions, and

X→Y: association rule that were under proposing

In the above equations, the support provides how frequently the item set appears in the data set and is defined as the number of transactions divided by the total number of transactions. The confidence indicates how often the association rule itself was found valid and is defined as the proportion of both item *X* and *Y* being purchased to the total purchase of *X*. The lift value provides the strength of association rule and is defined as Equation (5.3). Lastly, the conviction provides the inference on randomness, and is defined as the expected frequency that *X* occurs without *Y*. In order to conduct this analysis, Jupyter environment for Python programming was selected as the operation software here. Python has gained a lot of support in the field of deep learning because it is simple code, easy to read, and has abundant libraries that can be used for calculation and statistical processing. Jupyter is an operation system that common use for doing Python programming analysis.

5.5. Basic Arrangement of Data

5.5.1. Steps of data arrangement

First part of this step is to arrange the data from NLNI. The point data downloaded from NLNI was put into map firstly. There are totally 10 infrastructures selected here due to the data availability. The details of 10 facilities are listed in Table 5.4. At the same time, the shape files of mesh and LRT stations and route of Utsunomiya City were also put onto map. In order to conduct the data arrangement, the original data are needed to be made some calculations. The point data was arranged to find the distance between each infrastructure and the central point of each mesh. The software used here is ArcGIS. By the function in ArcGIS, this step can be calculated easily. ArcGIS is one of the GIS software provided by ESRI, similar to QGIS, it can read data, create maps, and output. Although ArcGIS need to pay for acquiring license, it has more function than the free software QGIS. As a result, ArcGIS was selected in this part. The central point of each mesh was defined by using the function called "polygon centroids". Finally, the distance from central point of each mesh to nearest infrastructure was calculated by using the function called "Near Tool" in ArcGIS.

Table 5.4 10 infrastructures chosen for this part

Elementary school (Elem)	Park (Park)
Middle school (Midd)	Attraction facilities* (Attr)
High school (High)	Newtown (New)
Culture facilities** (Cult)	University (Univ)
Police station (Poli)	Hospital (Hosp)

*Attraction facilities: Movie theater, citizen center, etc.

**Culture facilities: Libraries, art museum, etc.

Hereafter, since the data was available in three consecutive years, the trend of this data was calculated in order to know the trips trend was increase, or decrease. Finally, the meshes have the trend of increasing trips larger than 20% were selected. The steps of data arrangement are shown in Fig. 5.3.

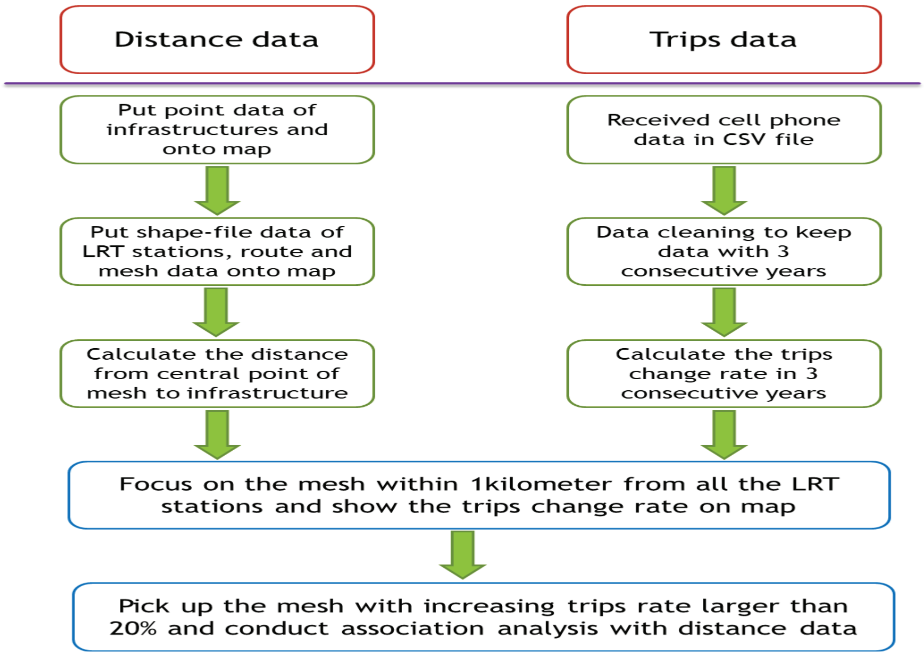


Fig. 5.3 Steps of data arrangement

5.5.2. Trip generation condition along the LRT route in Utsunomiya City

The areas within 1kilometer from all the 19 LRT stations were focused in this part as shown in Fig. 5.4. The trips count used here is estimated by the percentage of users and the calculation detail was already explained in previous section. Hereafter, the trips changing rate in each mesh in the target areas is shown in Fig. 5.5. The red triangles indicate the meshes that trips rate is increasing and larger than 20%. One can find that most meshes have the trend of increasing trips rate.

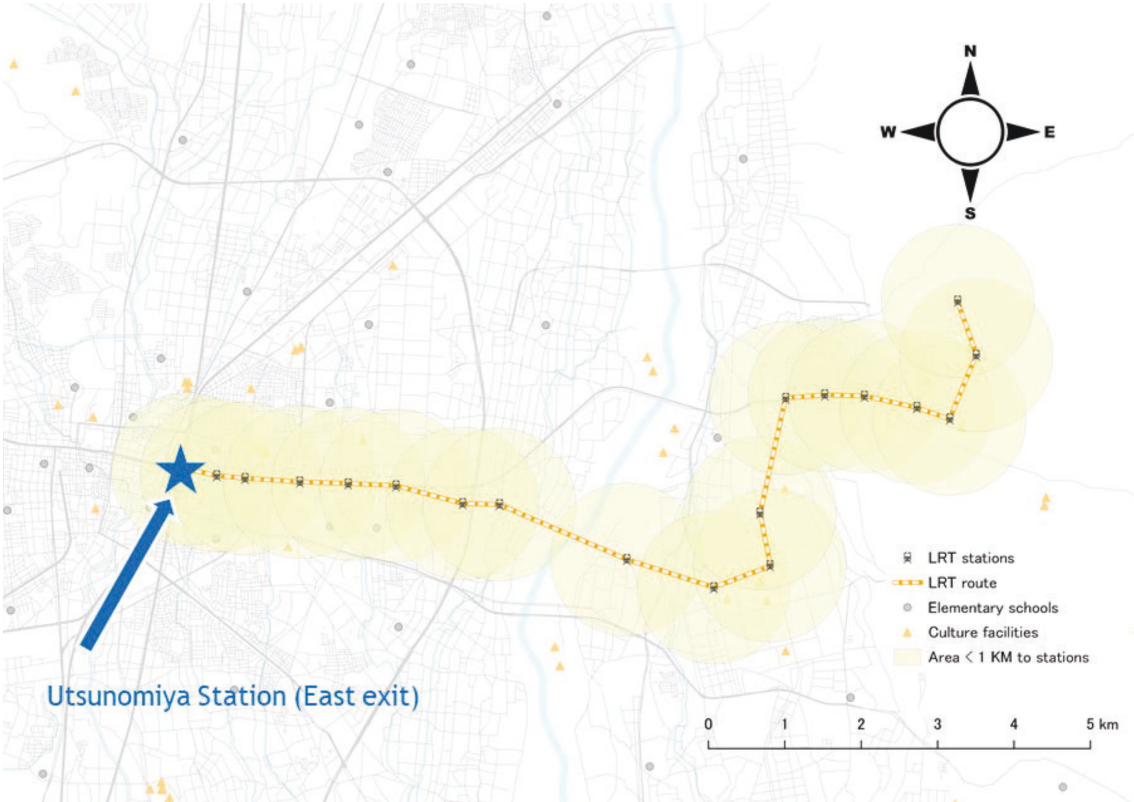


Fig. 5.4 Areas within 1kilometer from all the 19 LRT stations

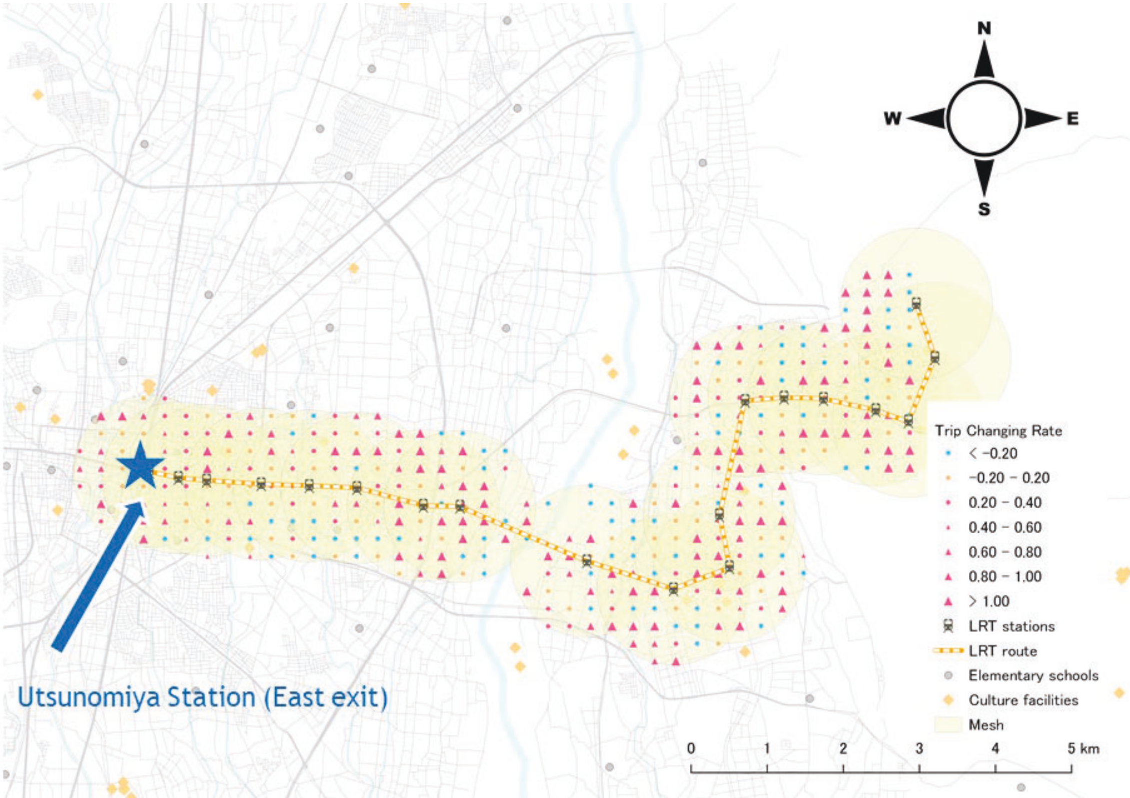


Fig. 5.5 Trips changing rate along LRT route

Hereafter, the meshes with increasing trips rate larger than 20% were classified into different categories based on different increasing trips rate. There are 236 mesh here and the result is shown in Fig. 5.6. Obviously. the numbers of mesh that the trips rate increasing in a significant rate, more than 100% is quite larger than other categories. However, some meshes with big triangles is not very close to station. As a result, some other public transportation systems such as bike sharing systems may be considered to integrate with LRT system.

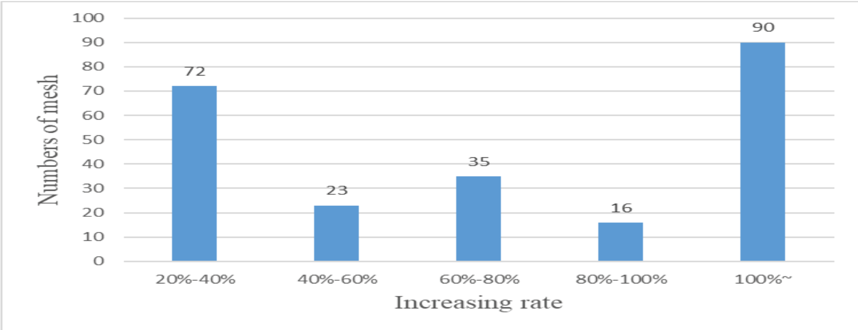


Fig. 5.6 Numbers of mesh based on different increasing trips rate

5.6. Results and Discussion

5.6.1. Transformation of dataset

In order to make the dataset available for association analysis, data transformations were made here. Based on the introduction of association analysis above, the original dataset was transformed into different format in both sides. In the part of distance data, based on different criteria, the data can be transformed into 0, 1 data. The criteria can be set in different values. Here, for reasonable calculation, 1 kilometer was chosen. In this study, since the distribution of infrastructures are the main concerned, distance was selected for setting criteria instead of the travel time for this part of analysis. In other words, distance<1KM was transformed into 1 and >1KM will stand for 0. On the other hand, the mesh with increasing rate of trips trend larger than 20% was selected here for analysis. Totally 236 mesh was chosen for this analysis.

5.6.2. Steps of Association Analysis

The final step of the process is applying the association analysis to the dataset prepared from previous section. The original results were more than 10,000 rules shown, so it was difficult to pick up useful results. Several filtered steps to extract the useful rules was therefore applied here. The steps are shown in Fig. 5.7 below.

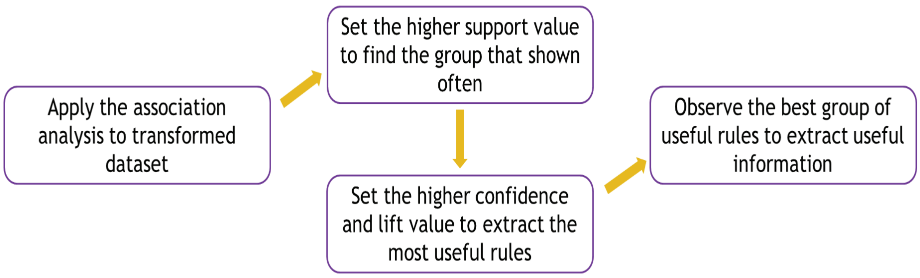


Fig. 5.7 Steps of Association Analysis

5.6.3. Results after filtered process

Finally, results from association analysis after the above filtered steps is shown below. After several trials, the final criteria that were chosen here for support value is 0.2, while the confidence value was set at 0.8 and the lift value was set at 1.2 as shown in Table 5.5. Take first 9 rules for example, the part of results can be shown in the following Table 5.6. due to limited space. There are totally 122 rules in this group.

Table 5.5 The criteria set for important parameters

Support value	0.20
Confidence value	0.80
Lift value	1.20

Table 5.6 The parts of filtered results from Association Analysis

Rule#	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
19	Poli	Cult	0.49	0.70	0.43	0.88	1.26	0.09	2.49
25	High	Cult	0.28	0.70	0.26	0.92	1.32	0.06	3.97
41	High	Midd	0.28	0.32	0.23	0.82	2.57	0.14	3.75
69	Hosp, Poli	Cult	0.48	0.70	0.42	0.88	1.25	0.09	2.45
71	Poli	Cult, Hosp	0.49	0.69	0.42	0.86	1.26	0.09	2.27
76	Midd	Cult, Hosp	0.32	0.69	0.26	0.83	1.20	0.04	1.81
84	Hosp, High	Cult	0.28	0.70	0.26	0.92	1.32	0.06	3.97
87	High	Cult, Hosp	0.28	0.69	0.26	0.92	1.35	0.07	4.14
128	Hosp, High	Midd	0.28	0.32	0.23	0.82	2.57	0.14	3.75

5.6.4. Discussion of results

I. Finding the most important elements

After the above association analysis, one may know what elements tend to be have higher relation with the mesh that have larger trip numbers. Hereafter, the numbers that how many times each infrastructure was shown in above association analysis was calculated. The result is shown in Table 5.7. below. Hospital, culture facility, police station and park performed well in this analysis. The result is shown in Fig. 5.8. However, the police station is remaining doubtful since it seems to be just fit accidently because police station is not a target people often intend to take a visit.

Table 5.7 The numbers of each infrastructure shown in above Association Analysis

	Hosp	Elem	Cult	Poli	Univ	Midd	Attr	High	Park	New
Antecedents	40	36	36	54	0	23	27	54	40	0
Consequents	42	0	77	27	0	18	0	0	42	0
Total	82	36	113	81	0	41	27	54	82	0

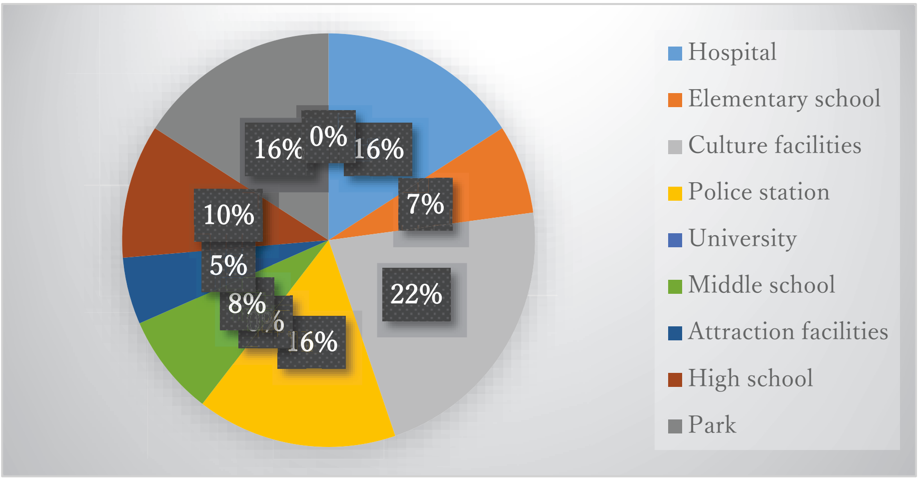


Fig. 5.8 Percentage of each infrastructure shown in above Association Analysis

II. Finding the potential groups

Also, other results were filtered by setting lower support value but higher confidence and lift value. The criteria and result are shown in Table 5.8. and Table 5.9. Table 5.9 shows the first 5 rules that were satisfied by these criteria.

Table 5.8 The criteria set for important parameters

Support value	0.05
Confidence value	0.90
Lift value	4.00

Table 5.9 Part of the results based on the above criteria

Rule#	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
2114	Midd, New	High, Elem	0.10	0.17	0.09	0.96	5.51	0.08	19.00
2689	Midd, New	Poli, High	0.10	0.21	0.09	0.96	4.51	0.07	18.13
2826	Midd, New	Attr, High	0.10	0.18	0.09	0.91	5.01	0.07	9.40
3847	Midd, Hosp, New	High, Elem	0.10	0.17	0.09	0.96	5.51	0.08	19.00
3855	Midd, New	Hosp, High, Elem	0.10	0.17	0.09	0.96	5.51	0.08	19.00

Here, the lift value was set at 4.0, a relatively high value, representing the rules that are shown in low probability but with very strong relation. Different from above section, new town is often shown in this part. As a result, some infrastructures with low density should be also taken into consideration in new planning.

5.7. Summary

5.7.1. Summary of this chapter

To conclude, this research proposed a new methodology to find the potential relations between infrastructures and trip generation. By using the cell phone data, the trips condition along the LRT route in Utsunomiya City was analyzed. Through association analysis, hospitals, parks and culture facilities are important elements that related to the trends of increasing trips rate in this case study. As a result, the future planning for integrating LRT systems with other public transportation systems such as bike sharing systems can be taken reference based on this result.

However, different cities have different transportation culture and different city scale. The results may differ in different places. Also, some relations are important indicators although the support value is not high. That is, it should be considered even the amount of this infrastructure may be not so many. On the contrast, some infrastructures may be not important factors even the results are good, such as police station. Therefore, further considerations and analysis should be taken when make planning decision based on these results.

5.7.2. Limitation and further implementation

The analysis conducted in this research serves as a primary approach to find the potential relations between infrastructures and trip generation. Nonetheless, due to the different amount and density of different infrastructures, further studies should be conducted in more detail analysis. For instance, comparing with the existing PT data or making more analysis on map would be good methods. Using some other AI methods or other different sources of dataset would also be possible research directions. For example, Point of Interest (POI) data may be used to conduct correlation analysis if the data is available.

Chapter 6

SPECIAL SITUATION OF TRIP GENERATION BASED ON COVID-19 EFFECT[3]

[3]The materials of the current chapter were accepted to:
International Conference of Asian-Pacific Planning Societies 2022

Questionnaire data is often used for the research in transportation field. Nevertheless, due to the difficulties of acquiring the big numbers of traditional survey data, using cell phone data became a new trend of research in this field. On the other hand, COVID-19 has made significant change in the travel behavior of people. Here, in order to understand the trips condition before and after the COVID-19, cell phone data in Utsunomiya City in Japan were used and calculated. One can easily found that the trend of trips was strongly affected by the COVID-19 and changed significantly in a decreasing rate. By showing the results on ArcGIS map, one can easily figure out the difference of trips condition before and after the pandemic. Furthermore, association analysis was conducted by utilizing cell phone data and shape-filed data of infrastructures. Findings are the areas that closed to hospitals and parks have the decreasing trend of trips condition. On the other hand, the infrastructures that people still need to go for necessary tasks were not strongly affected. For instance, the percentage of elementary school remains the same.

6.1. Introduction

6.1.1. Background

Questionnaire data is often used for the research in transportation field. Nevertheless, due to the difficulties of acquiring the big numbers of traditional survey data, using cell phone data became a new trend of research in this field. Wang *et al.* (2018) provided a review of existing travel behaviour studies that have applied mobile phone data, and presented the progress that has been achieved to date. Moreover, thanks to the improvement of the software and the hardware techniques of computer science, new analysis based on AI (Artificial Intelligence) methods are emphasized again these days.

On the other hand, COVID-19 has made significant change in the travel behavior of people. As shown in Fig. 6.1, the cases of coronavirus in Japan increased significantly from March 2020. Many companies asked the workers to work from home, thus reduced the trips number. Since the questionnaire data is not collected every year, cell phone data becomes a good choice to analyze the trips situation before and after the exposure of this pandemic.

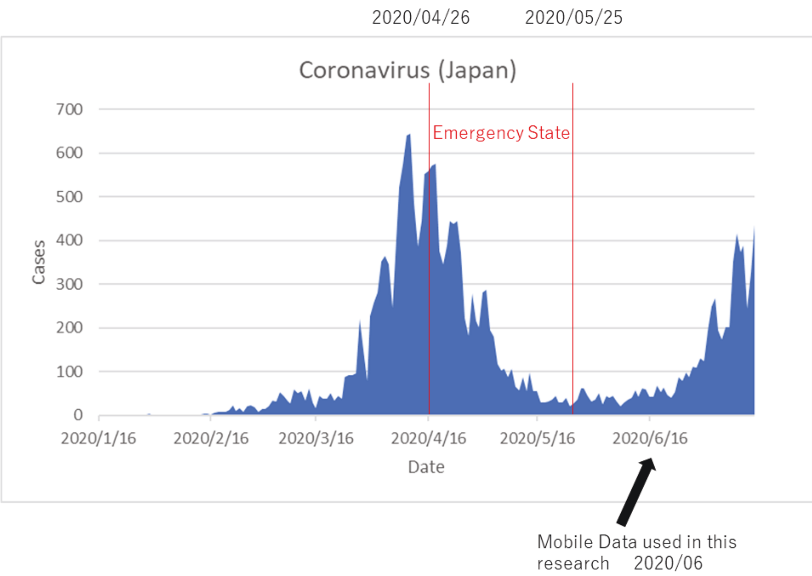


Fig. 6.1 Coronavirus cases in Japan (2020/01/16-2020/07/15)

6.1.2. Objective

The main objective of this research is to provide a new and primary approach of finding the potential relations between infrastructures distribution and trip generation in Utsunomiya City through available cell phone big data. Especially, under the COVID-19 pandemic exposure, it would be essential to understand the effect caused by this condition. By extracting the trips condition and making them visualized on the map, some other planning and integration with other public transportation systems can be considered based on the results here. Also, trying to find the potential relations between infrastructures and trips through several steps of data arrangement and association analysis is another goal in this research.

6.2. Literature Review and Flowchart

6.2.1. Literature review

The literature reviews were conducted for the following three themes. (1) Utsunomiya City and the land use of city operating the LRT system, (2) mobile phone big data in transportation field, and (3) AI and the algorithm of Association Analysis.

Since the Utsunomiya City are planned to operate the new LRT system in year 2023, some papers were reviewed related to this field. As for the LRT system and nearby land use, Pacheco-Raguz (2010) analyzed the impacts of development of LRT in Manila on land price, land use, and population along the route. Other study includes Sakamoto *et al.* (2015) examined change in urban population before and after introduction of LRT in 27 cities in Europe. In addition, Higgins *et al.* (2014) demonstrated that rail transit alone is not a primary driver of land use change. As for the case in Japan, Mochizuki *et al.* (2007) conducted questionnaire-based survey and compared the visitors in Toyama City who used different transportation modes when traveling. In addition, Kriss *et al.* (2020) conducted a detail study of Toyama City and pointed out the LRT project in Toyama City got significant success by showing the increasing ridership of the system.

Since the papers mentioned above are mainly based on questionnaire survey. However, due to the difficulty of acquiring survey data, the possibility of using other data source was taken into consideration. Hereafter, the articles related to application of cell phone data were reviewed. Widhalm *et al.* (2015) proposed a method to reveal activity patterns that emerge from cell phone data by analyzing relational signatures of activity time, duration, and land use. Also, Huang *et al.* (2019) gave huge amount of literature reviews and clear discussion of this field. In addition, Wang *et al.* (2018) presented the progress in this field that has been achieved to date, and then discussed the potential of mobile phone data and raised some challenges. On the other hand, Bachir *et al.* (2019) made the mobile network trajectories by using mobile data in Paris region.

AI methods are often relevant to big data analysis since numbers of dataset are huge. AI methods seems to have potential to deal with this problem. Regarding the articles published recently in transportation field related to AI study, Haenlein *et al.* (2019) takes a first insight into AI applications by summarizing seven articles. Also, Abduljabbar *et al.* (2019) pointed out how is a good way to make various of application in transportation field by constructing models in several phases. Finally, since the association analysis serves as a good method to find the potential relationship between different items that would be difficult to observe directly, this methodology was selected. For instance, the document written by Tan *et al.* (2005) gave us general idea of how to set up models by different algorithms. By showing some examples with programming, one can easily understand how to set up a method to conduct association analysis.

6.2.2. Characteristics of this part

Most researches related to this field were mainly based on traditional questionnaire survey data. Nonetheless, personal trip surveys are only conducted in specific year and the frequency is not very high. As a result, due to the difficulties of acquiring the trips data by questionnaire survey, cell phone big data was chosen for estimating the trips in this research. Nevertheless, study based on mobile phone big data was lack of abundant information but actually the new trend. This research aimed to provide a new and preliminary study to understand the potential relations between infrastructures distribution and trip generation in Utsunomiya City by combining the cell phone big data and open-sourced data. Here, two datasets including static data and dynamic data based on different sources were processed and calculated through several steps and conducted association analysis to propose a new trial. Therefore, the diversity of datasets is another characteristic of this research.

6.2.3. Flowchart

Firstly, literature reviews of the thesis relevant to Utsunomiya City and the city land use, mobile phone big data, and AI were reviewed. Thanks to the availability of cell phone data from KDDI company, Utsunomiya City was selected as the target city. At the same time, some open-sourced datasets were downloaded. Hereafter, the two different datasets were made some arrangement and were shown on map on ArcGIS software. Hereafter, datasets were calculated and analyzed through association analysis by programming on python language. Finally, results were then shown and discussed, followed by conclusion part. The flowchart of this part is shown in the below Fig. 6.2.

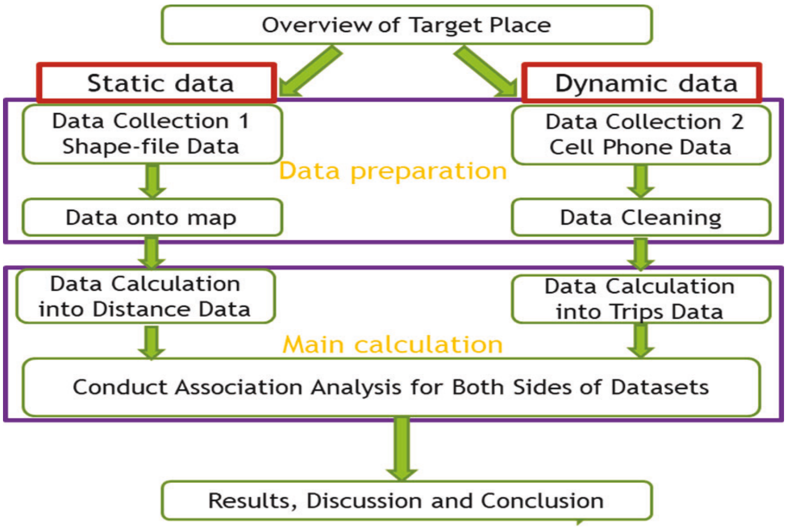


Fig. 6.2 Flowchart of this part

6.3. Introduction of Data and Methodology

6.3.1. Data used and data source

There are two sets of original data selected in this main analysis part. One is the cell phone data of Utsunomiya City from KDDI company. For the KDDI cell phone datasets, the data that was used in this research is the user's trip data extracted from the GPS logs of cell phones without the personal information. The estimated trip numbers of OD (Origin and Destination) data were extracted here. Another one is the shape-filed data downloaded from geographic information from NLNI (National Land Numerical Information). Some information of these datasets is listed in the Table 6.1.

Table 6.1 Introduction of two main datasets

	Cell phone data	Shape-filed data
Data Source	KDDI CORPORATION	National Land Numerical Info.
Data Content	OD Points, Estimated Trip Numbers	Position of Infrastructures
Scale of Data	250m Mesh(JGD2011 Coordinate)	Prefecture, City
File Format	Excel File(.csv)	Point, Line(.shp)

In the KDDI cell phone dataset, the data were acquired in June from 2018-2020. Here, only that users who have given permission to use the data are included in the analysis. The data that were used in this research is the user's trip data extracted from the GPS logs of cell phones without the personal information. The data can be acquired for three consecutive years were extracted here so there were total 4469 mesh data available in this dataset for the calculation. Signals are tested every 15 minutes here. When one signal is stayed in one mesh more than 15 minutes, this condition will consider “stay”, so that the same signal will not be counted repeatedly in order to increase the accuracy of estimation. Detail information of this dataset are listed below in Table 6.2. In this table, the data that was used in this research is the user's trip data extracted from the

GPS logs of cell phones. The method of extended estimation of trips count will be explained in next chapter.

Table 6.2 Detail information of KDDI cell phone data

Data properties	Detail Explanation
Time of data	Year and month at departure (June in 2018-2020)
Day property	Type of day at departure (Weekday, Holiday)
Origin point (Location 1)	250m mesh cord of departure point (JGD2011 Coordinate)
Destination point (Location 2)	250m mesh cord of arrival point (JGD2011 Coordinate)
Estimation trips count	Number of trips after extended estimation

*Data source: KDDI CORPORATION

6.3.2. Basic concept and principles of Association Analysis

The association analysis, or often called the market basket analysis is an analytics technique often conducted by retailers and consultants to understand the purchase behavior of customers. In this research, the association analysis was conducted to extract key features influencing the changing rate of trips count. The purpose is to find the potential relations between the distance from mesh to infrastructures and the trips count in each mesh. Through constructing binary transaction data, it is possible to mine the hidden relationship of each item sets within the data set. Here, the association analysis provides information on how strong the association rule, or the if-then relationship, is found to be true in the dataset.

6.4. Data Processing

6.4.1. Steps of data arrangement

First part of this step is to arrange the data from NLNI. The point data downloaded from NLNI was put into map firstly. There are totally 10 infrastructures selected here due to the data availability. The details of 10 facilities are listed in Table 6.3. At the same time, the shape-filed data of meshes and Utsunomiya City were also put onto map. In order to conduct the data arrangement, the original data are needed to be made some calculations. The point data was arranged to find the distance between each infrastructure and the central point of each mesh. The software used here is ArcGIS. By the function in ArcGIS, this step can be calculated easily. ArcGIS is one of the GIS software provided by ESRI, similar to QGIS, it can read data, create maps, and output. The central point of each mesh was defined by using the function called "polygon centroids". Finally, the distance from central point of each mesh to nearest infrastructure was calculated by using the function called "Near Tool" in ArcGIS.

Table 6.3 10 infrastructures chosen for this part

Elementary school (Elem)	Park (Park)
Middle school (Midd)	Attraction facilities* (Attr)
High school (High)	Newtown (New)
Culture facilities** (Cult)	University (Univ)
Police station (Poli)	Hospital (Hosp)

*Attraction facilities: Movie theater, citizen center, etc.

**Culture facilities: Libraries, art museum, etc.

Regarding the arrangement of the cell phone data, the estimated trips count from cell phone data was organized in order to know the trips condition at each specific mesh. Here, the estimated trips count was received from KDDI CORPORATION as seen in Table 6.2 and the calculation method was conducted by using the percentage of users.

Also, since the data was available in three consecutive years, the trend of this data was separated into two parts and calculated respectively in order to know the trend of trips count was increasing, or decreasing. That is, the changing rate of trips count was calculated from year 2018 to year 2019 and from year 2019 to 2020, respectively. The results are shown in the below Fig. 6.3. One can easily found that the trend of trips was strongly affected by the COVID-19 and changed significantly in a decreasing rate. Fig. 6.4 is the overall changing rate of trips count from year 2018 to year 2020. Hereafter, these two groups are prepared for the analysis in next step.

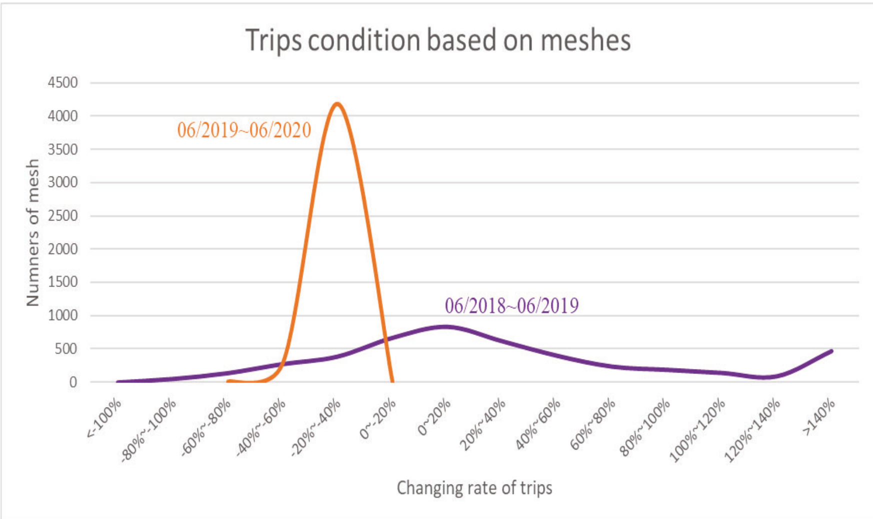


Fig. 6.3 Changing rate of trips count before and after the COVID-19 exposure

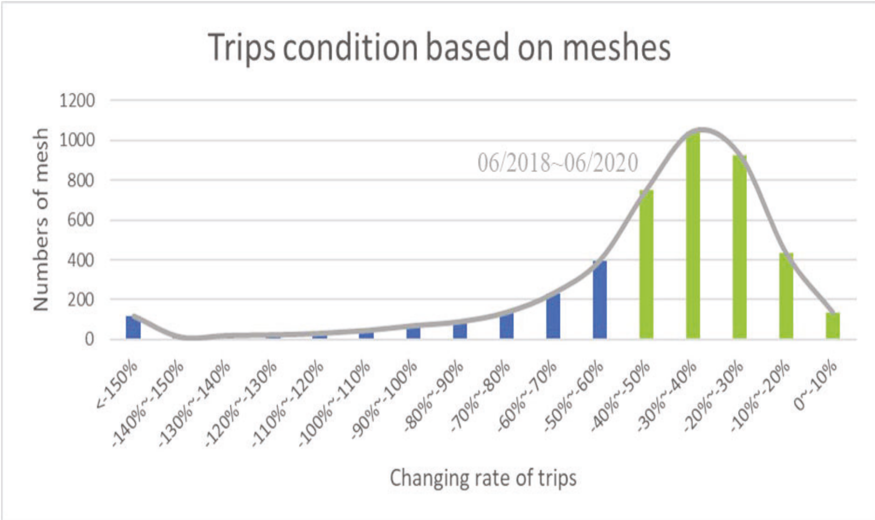


Fig. 6.4 Overall changing rate of trips count after the COVID-19 exposure

From the above two figures, one can easily notice that the trend of trips count has significantly different before and after the COVID-19 pandemic exposure. In Fig. 6.3, more meshes have the increasing rate of trips count from year 2018 to year 2019 whereas all the meshes have the decreasing rate of trips count from year 2019 to year 2020. Furthermore, the average changing rate from 2018 to 2019 is 21.66% while the average changing rate from 2019 to 2020 is -31.05%. In addition, all the meshes decreased in the range -20% to -60% from year 2019 to year 2020. In Fig. 6.4, all meshes have the decreasing rate of trips count in overall condition from year 2018 to year 2020. Also, the average changing rate from 2018 to 2010 is around -48%.

6.4.2. Visualizations of trips condition in Utsunomiya City

The meshes within Utsunomiya City were focused in this research and therefore the results from previous section were visualized as shown in Fig. 6.5 and Fig. 6.6. Totally, the 4469 meshes that have the data of 3 consecutive years are considered here. Some parts of the results that are closed to the city center, Utsunomiya station are chosen and shown below. The red ones indicate the meshes that the trips rate is increasing and the green ones represent the meshes that the trips rate is decreasing. Obviously, one can find that the changing rate of most meshes are strongly affected by the COVID-19.

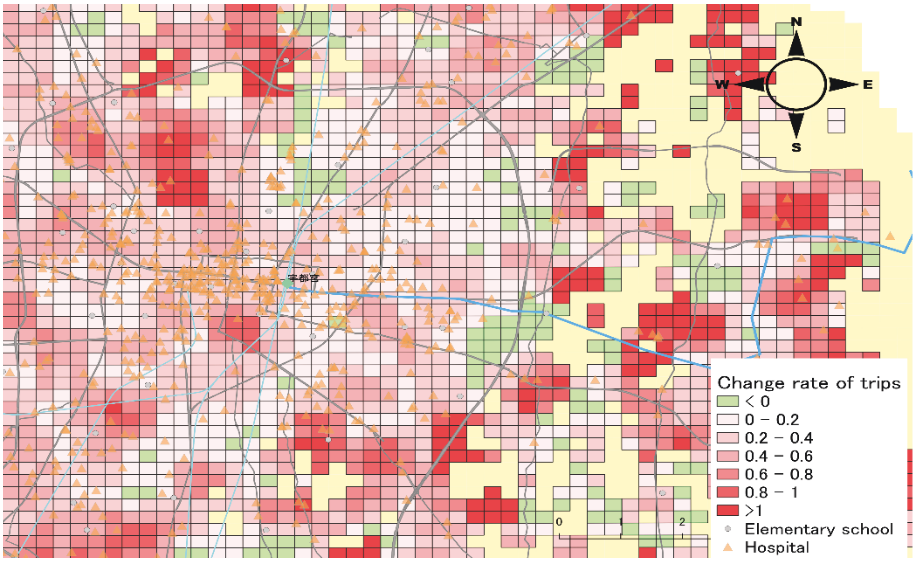


Fig. 6.5 Visualization of trips condition before the COVID-19 (2018-2019)

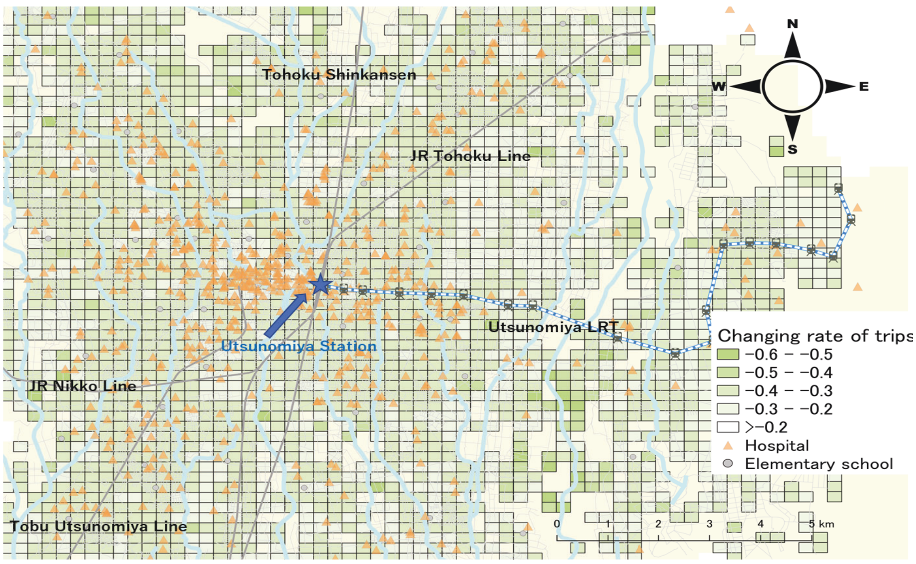


Fig. 6.6 Visualization of trips condition after the COVID-19 (2019-2020)

6.5. Results and Discussion

6.5.1. Steps of Association Analysis

In order to make the dataset available for association analysis, data transformations were made here. In the part of distance data, based on different criteria, the data can be transformed into 0, 1 data. The criteria can be set in different values. Here, for reasonable calculation, 1 kilometer was chosen. In this study, since the distribution of infrastructures are the main concerned, distance was selected for setting criteria instead of the travel time for this part of analysis. In other words, distance<1KM was transformed into 1 and >1KM will stand for 0.

On the other hand, the meshes were divided into two groups based on the difference of the changing rate of trips count. The criterion is the difference of changing rate of meshes is smaller than -50% (<-50%) or larger than -50% (>-50%). In other words, that represents the changing rate of meshes is strongly affected by COVID-19 or not. The final step of the process is applying the association analysis to the datasets prepared for both groups, respectively. Here, the distance<1KM of each infrastructure stands the variables X and Y mentioned in section 6.4.2.

6.5.2. Filtered process of Association Analysis

The original results were more than 20,000 rules, so it was difficult to pick up useful information. Several filtered steps to extract the useful rules were therefore applied here. Finally, results from association analysis after the filtered steps are shown below. After several trials, the final criteria that were chosen here for support value is 0.15, while the confidence value was set at 1.0 and the lift value was set at 0.8 as shown in Table 6.4.

Table 6.4 The criteria set for important parameters

Support value	0.15
Confidence value	0.8
Lift value	1.0

6.5.3. Results of Association Analysis

After the above association analysis, one may know the relation between what elements tend to appeal more often with the different changing rates of trips count. Here, Fig. 6.7 shows us the numbers that how many times each infrastructure was shown near the target meshes in above association analysis. The blue bar shows the results of the group that have the decreasing rate larger than 50% (>-50%). The green bar stands for the results of the group that have the decreasing rate smaller than 50% (<-50%).

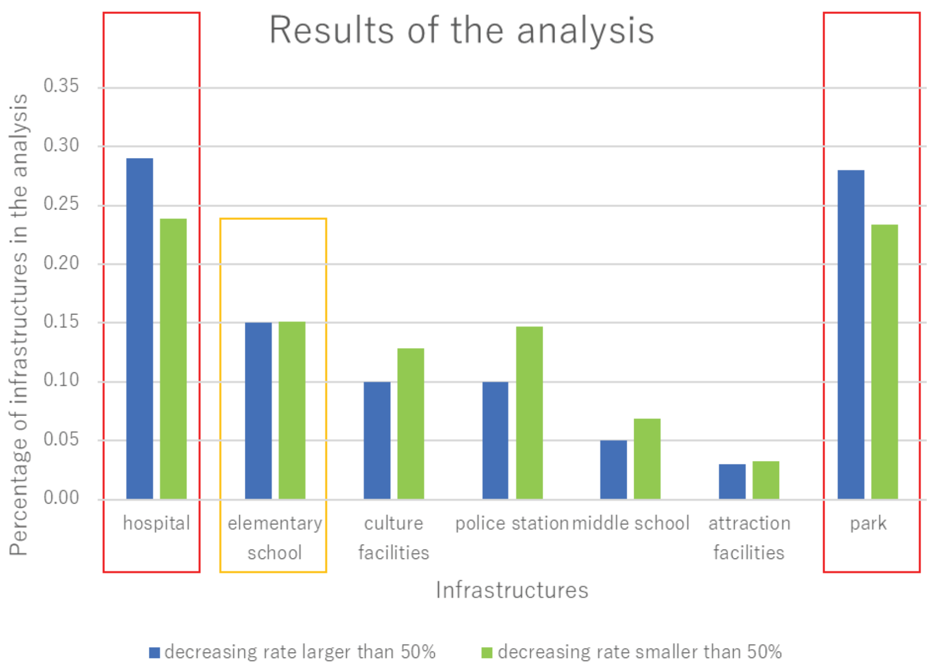


Fig. 6.7 Percentage of infrastructures in the analysis based on different trips changing rate of meshes (2018-2020)

By comparison of the results, hospitals and parks show more often in the group that have significantly decreasing rate. That is, after the COVID-19, the areas that closed to hospitals and parks have the decreasing trend of trips condition, representing that the people prefer not going for the activities that are not necessary. The infrastructures that people still need to go for necessary tasks were not strongly affected. For instance, the percentage of elementary school remains the same in both figures.

6.6. Summary

6.6.1. Summary of this chapter

To conclude, this research used the mobile phone big data and proposed a new methodology to find the potential relations between infrastructures distribution and trip generation. By using the cell phone data, the trips condition in Utsunomiya City was analyzed. One can easily found that the trend of trips was strongly affected by the COVID-19 and changed significantly in a decreasing rate. Through association analysis and comparison of the results, hospitals and parks show more often in the group that have significantly decreasing rate. That is, after the COVID-19, the areas that closed to hospitals and parks have the decreasing trend of trips condition. Possible reasons for the case of hospitals may be affected by the fear of going around the hospitals and the limitation of number of patients that were allowed to enter the hospitals. On the other hand, the infrastructures that people still need to go for necessary tasks were not strongly affected. For instance, the percentage of elementary school remains the same in both figures.

However, different cities have different transportation culture and different city scale. The results may differ in different places. Also, some relations are important indicators although the support value is not high. That is, it should be considered even the amount of this infrastructure may be not so many. On the contrast, some infrastructures may be not important factors even the results are good, such as police station. Therefore, further considerations and analysis should be taken when make planning decision based on these results.

6.6.2. Limitations and further implementations

The analysis conducted in this research serves as a new and primary approach to use the mobile phone big data and find the potential relations between infrastructures distribution and trip generation. Nonetheless, due to the different amount and density of different infrastructures, further studies should be conducted in more detail analysis. For instance, comparing with the existing PT data or making more analysis on map would be good methods. Also, due to the limitation of datasets, further analysis should be conducted in order to verify the results of the case of hospitals if more datasets related to hospitals or patients can be acquired. In addition, using some other AI methods or other different sources of dataset would also be possible research directions. For example, Point of Interest (POI) data may be used to conduct correlation analysis if the data is available.

Chapter 7

CONDITION OF CHOICE OF TRANSPORTATION MODE BASED ON TRIP PURPOSE[4]

[4]The materials of the current chapter were accepted to:
The International Journal of Information and Management Sciences

The government places emphasis on increasing the usage rate of public transportation nowadays due to public transportation having many benefits for the environment. In order to understand the key factors of trip generation and identify the key trip purposes for selecting transportation modes in a target city, the cell phone data and personal trip survey data were studied by using the machine learning methods of Association Analysis and Inverse Reinforcement Learning. Findings such as hospital, park and elementary school are the most important elements implies that the facilities for mandatory task will attract more people. Also, the middle age group has very strong tendency to use private vehicle compared to other age groups implies that attracting more young people may be a good strategy. Findings can be a reference for new policy planning, including re-planning the exiting routes of bus systems or integrating different public transportation, by the local government.

7.1. Introduction

7.1.1. Background

For developed countries around the world, problems such as budget cut, net out-migration from cities, and aging society, could block the urban development. Public transportation system stands an important role to ease the impacts from the aforenoted problems. How to increase the number of public transportation users becomes an important issue nowadays. The longitudinal data set of passengers can offer useful information for modeling to understand the growth/decade rate of the passengers by public transportation. Bus Rapid Transit (BRT) and Light Rapid Transit (LRT) are two major mass transit systems. The BRT and LRT are customer-oriented transportation systems which have the merits of delivering fast and comfort. The maintenance costs of the BRT and LRT systems are low to keep urban mobility. Moreover, BRT and LRT provide more flexibility and bus can do service on exclusive lanes along major corridors such as branch out to cover more territory as the destination approaches. The BRT and LRT can provide modern, efficient and comfortable service to public transportation users. BRT is a new solution to improve the efficiency of traditional bus systems. LRT was first introduced in North America to describe the new concept of tram transportation (Thompson, 2003). A well-established LRT system can assist city become a TOD (Transit-Oriented Development) City or Compact City through adapting new system, and then the city can be more attractive for citizens (Takami *et.al*, 2003). It is critical to identify potential public transportation passengers and enhancing their usage rate.

7.1.2. Objective

Considering the limitation of acquiring a big data set via using questionnaire survey, the cell phone data can be more competitive for decision making. Nonetheless, traditional survey data can also be applied to Artificial Intelligence (AI) methods to make the calculation faster or obtain more important information. Thanks to the improvement of the software and hardware techniques of computer science, powerful machine learning methods based on AI technologies are emphasized again nowadays. The purpose of this research is using the machine learning method of Association Analysis to identify the key factors of trip generation based on the distance to infrastructures in Utsunomiya City, Japan. Also, another goal is applying the machine learning method of Inverse Reinforcement Learning (IRL) to figure out the tendency of selecting the transportation modes for different trip purposes. The results can be a reference for new policy planning, including re-planning the exiting routes of bus systems or integrating different public transportation systems, by the local government. Furthermore, we expect the proposed methods outperform traditional analysis methods to improve the precision accuracy.

7.2. Literature Review and Research Flow

7.2.1. Literature review

The literatures review was conducted for the following three themes. (1) Public transportation system and land use (2) Mobile phone big data, and (3) Machine learning methods related to this study. Numerous studies can be found in literature for analyzing the data sets from new transportation systems. Some recent studies, for example, Zhao *et al*. (2019) quantified the impacts of Urban Rail Transit system on land use change and tied them into the future land maps. Zhang *et al*. (2020) identified several factors that influence the estimates of BRT impacts based on 23 empirical studies. Pacheco-Raguz (2010) studied the impacts of the development of LRT in Manila on land price, land usage, and population size for the cities along the route. Using correlation and regression methods, the aforementioned variables are analyzed for an accessibility index and network distances obtained from a model built within a Geographic Information System (GIS). Sakamoto *et al*. (2015) examined the change in urban population size before and after introduction of LRT in 27 cities in Europe. They also analyzed the changes from the population size or areas along the LRT route over time for four cities in Europe. Kriss *et al*. (2020) conducted a study for Toyama City and pointed out that the LRT project in Toyama City got significant success based on the increasing ridership of the LRT. Sato *et al*. (2018) predicted the population distribution in Utsunomiya City, Japan until 2050 by supposing different integration pattern of the LRT and feeder bus systems. Takasugi *et al*. (2018) studied the differences between the BRT and LRT systems and the influence on urban population distribution for Maebashi City. Most of aforementioned researches are mainly concerned about the impacts on the cities after the introduction of public transportation systems.

It is worth to note that the findings of aforementioned researches were highly dependent on using questionnaire survey method. Questionnaire survey is time consuming and asks high cost for acquiring data. Today, cell phone is popular and becomes one of the required devices for people. Cell phone data are more competitive and complete than questionnaire survey data. Using cell phone data for studying the impacts of public transportation systems on related areas have earned more attentions in the near past decade. Widhalm *et al*. (2015) proposed a method to reveal activity patterns that emerge from cell phone data by analyzing relational signatures of activity time, duration, and land use. Also, Wang *et al*. (2018) provided a review of existing travel behavior studies by using mobile phone data. They discussed the potential of mobile phone data in advancing travel behavior research and raised some challenges that needed to be dealt with in the planning of traveling process.

Compared with traditional statistical modeling methods, machine learning methods in the AI area can be more powerful for prediction but could be weaker to explain the relationship between the response variable and explanatory variables due to an implicit functional form of hidden layers structure. Moreover, machine learning methods has the potential to automatically learn and update the predicted results through a well design system from updated data sets. Haenlein *et al*. (2019) takes a first insight into AI applications by summarizing seven articles published in this special issue that present a wide variety of perspectives on AI. When the relationships among features in data are not clear, the Association Analysis method is one of most competitive machine learning methods to identify the potential relationship between different features with less subjective assumptions. Tan *et al*. (2005) proposed a general idea to set up models by different algorithms. By showing some examples with programming, one can easily understand how to implement Association Analysis method.

The IRL is another machine learning method which is applied in this research. The IRL is about studying from humans. In practice, IRL can be used to study an agent's objectives, rewards and values with the aid of using insights of its behavior. Arora *et al*. (2021) mentioned about advantages and challenges about using IRL. You *et al*. (2019) showed a great example of mobile application via using the IRL for data analysis. Kitani *et al*. (2012) conducted analysis and prediction of walking behavior using IRL model.

7.2.2. Characteristics of research

Many existing researches related to the trips condition were based only on the personal trips survey but lack of abundant trial. Here, cell phone data was utilized and conducted the analysis as an estimated trip frequency. A new idea by integrating static and dynamic cell phone datasets from different sources with several analysis steps are proposed in this study for providing to understand the trip generation in the target city. Hereafter, personal trips data was also applied by machine learning methods to make insight analysis and comparison of the results. Therefore, simultaneous utilization of the static data and dynamic data can be a significant contribution of this study. The image of some examples in both the static and the dynamic data sources is reported in Fig. 7.1. The static-type public facilities and boundary data and dynamic-type personal trip survey and cell phone signal data are used in this study. The machine learning methods of Association Analysis and IRL are applied for data analysis to obtain valuable information for making new policy planning by the government of the target city.

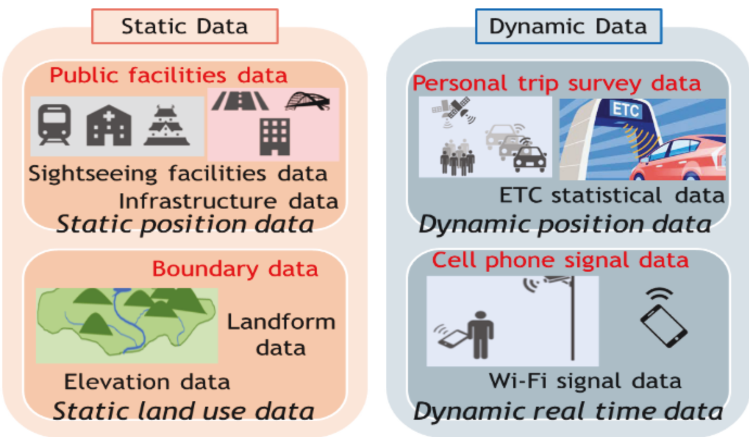


Fig. 7.1 Image of data sources

7.2.3. Research framework

Due to the availability of cell phone data from KDDI company, Utsunomiya City in Japan was selected as the target city in this study. Moreover, some open-sourced datasets were also downloaded by showing on the map on ArcGIS software. To make the research results reliable, we implemented a data cleaning process to ensure data is correct and consistent. Two steps in Fig. 7.2 are used for data analysis through using Association Analysis in Step 1. In order to make further understanding based on the findings of the Step 1, that is, to confirm the effect of the distance of different infrastructures, IRL was applied to traditional personal trip data in Step 2 to figure out the relation between trip purposes and selection for transportation modes. Finally, results were then shown and concluded. The data preparation and main calculation parts of the research flow are shown in below Fig. 7.2.

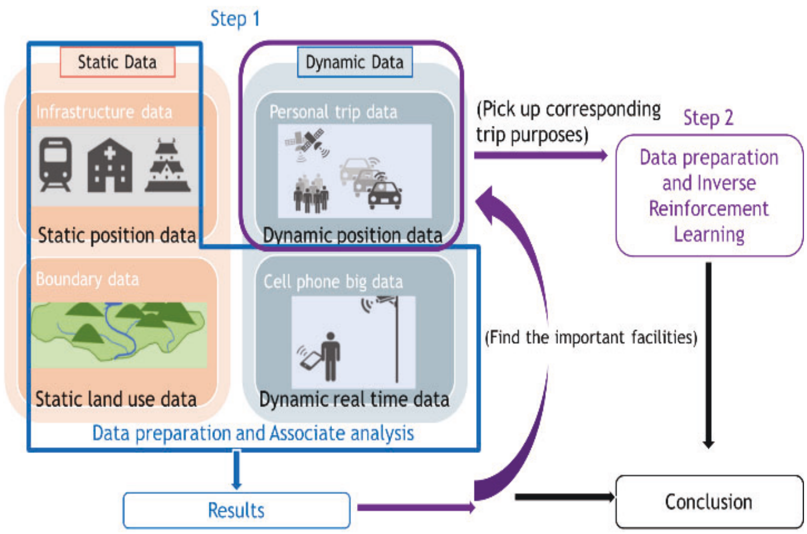


Fig. 7.2 Flowchart of the research

7.3. Introduction of Data and Data Preparation Process

7.3.1. Data and software utilized

Table 7.1 is the summary of source data from KDDI company and National Land Numerical Information download service. One is the cell phone data of Utsunomiya City from KDDI company. In the KDDI cell phone data set, the data was acquired in June from 2016-2018. Here, de-identification trip data extracted from the GPS logs of cell phones with permission contributes to the first data set. The second data set is about the geographic information from NLNI (National Land Numerical Information). At the same time, personal trip survey data was also prepared for the purpose of broadening the research content.

Table 7.1 Introduction of two main datasets

	Cell phone data	Shape-filed data
Data Source	KDDI CORPORATION	National Land Numerical Info.
Data Content	OD Points, Estimated Trip Numbers	Position of Infrastructures
Scale of Data	250m Mesh(JGD2011 Coordinate)	Prefecture, City
File Format	Excel File(.csv)	Point, Line(.shp)

In order to prepare the useful dataset, the original data is needed to be made some calculations. The point data is arranged to find the distance between each infrastructure and the central point of each mesh. The software used here is ArcGIS. ArcGIS is one of the GIS software provided by ESRI, similar to QGIS which can read data, create maps, and output. By the function in ArcGIS, this step can be calculated easily. Another software used in this research is the software related to Python programming. Python is a language that emphasizes code readability. Python has gained a lot of support in the field of deep learning because it is simple code, easy to read, and has abundant libraries that can be used for calculation and statistical processing. Jupiter Notebook is a web-based interactive computing environment that is common use for doing Python programming analysis.

7.3.2. Steps of data arrangement

After downloading the data from NLNI was put onto the map firstly. Total 10 infrastructures are selected in this study based on the consideration of data availability, see Table 7.2. The numbers of each facility in the target area are also shown in Table 7.2. At the same time, the shape files of mesh and the Utsunomiya City were also put onto the map. The central point of each mesh was defined by using the function called "polygon centroids". Finally, the distance from the central point of each mesh to the nearest 10 infrastructures was calculated by using the function of "Near Tool".

The estimated trips count from cell phone data was organized in order to know how many trips start or end at one specific mesh. Although the judgement based on different distance and time may affect the results, 300m and 15 minutes were set here by the data processing process by KDDI company. Hereafter, since the data was available in 2016-2018, the frequency of trip increasing/decreasing was evaluated. We are interested in evaluate the trip frequency is increased or decreased in 2016-2018 and identifying the causes for the resulting trip frequencies. Finally, the mesh with the symbol of trip increasing was selected, see the flowchart of data processing in Fig. 7.3.

Table 7.2 10 infrastructures chosen for this research

Elementary school (Elem) (**116*)	Park(Park) (**884*)
Middle school (Midd) (**52*)	Attraction facilities** (Attr) (**56*)
High school (High) (**24*)	Newtown (New) (**27*)
Culture facilities*** (Cult) (**254*)	University (Univ) (**9*)
Police station (Poli) (**79*)	Hospital (Hosp) (**1074*)

*The numbers of this infrastructure in target area

**Attraction facilities: Movie theater, citizen center, etc.

***Culture facilities: Libraries, art museum, etc.

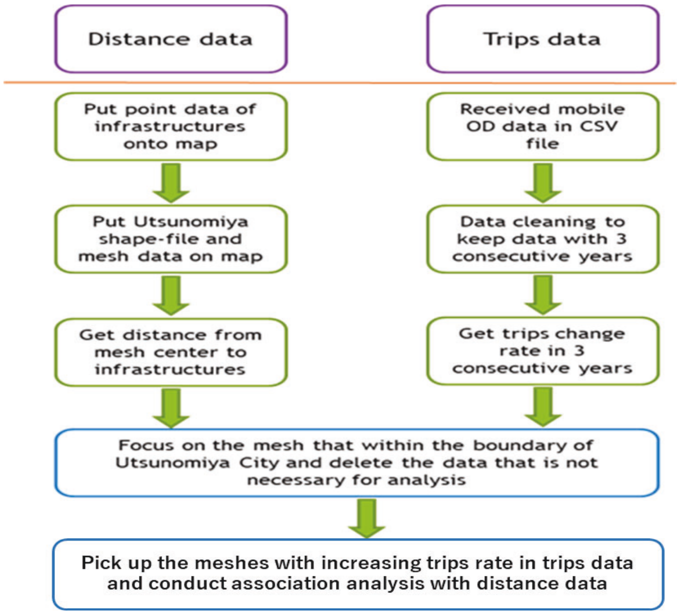


Fig. 7.3 Steps of data processing

7.3.3. Trip generation condition in Utsunomiya City

The mesh areas within Utsunomiya City was focused in the research purpose of this study. The trips changing rate in each mesh in the target areas is calculated. Total 3,328 meshes in the target area with enough data information and the average trip changing rate from year 2016 to 2018 is 0.47. The overall outlook of the trips changing conditions in the target area can be visualized on the GIS map and displays in Fig. 7.4. Although there is no obvious tendency of the trip changing condition in this map, one can find that several regions far from city center have a strong tendency of increasing trips, therefore several corresponding planning should be considered. Hereafter, the meshes were classified into different categories based on different trips changing rate. Fig. 7.5 shows most of the meshes have the increasing rate. Even though the trips count was originally estimated number by multiplying the extracted trips number and the expanding factor so some error may exist, but the trend can be observed through these data. The red highlight sets represent the target meshes that were selected to utilize for analysis in the section 7.4.

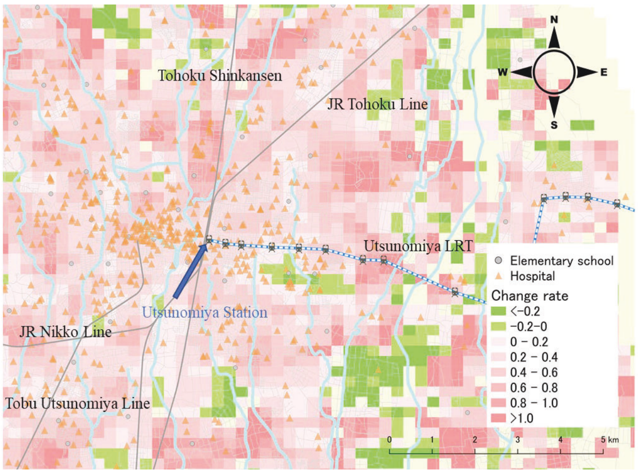


Fig. 7.4 Trips changing condition in Utsunomiya City (2016-2018)

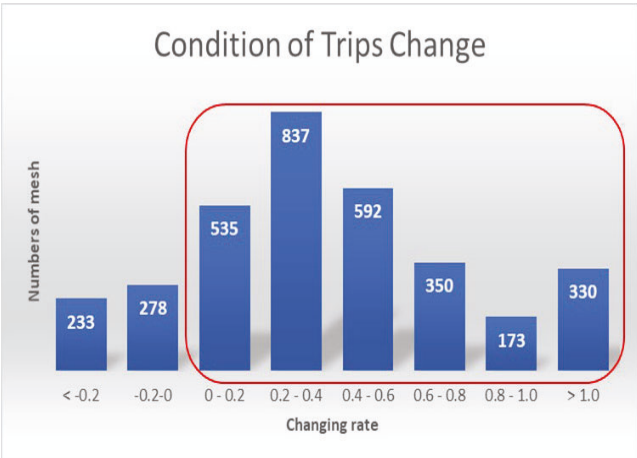


Fig. 7.5 The number of meshes based on different trip changing rates

7.4. Apply the Association Analysis

7.4.1. Concept and principles of Association Analysis

Association Analysis also named market basket analysis is an analytics technique which often conducted by retailers and consultants to understand the purchase behavior of customers. Association Analysis can be used to find interesting relationships in large datasets (Bai *et al.*, 2020). For market basket analysis, Association Analysis uses the purchase information or transaction data to leverage the effectiveness of sales and marketing. Association Analysis has been actively used in related industries once the immense amount of data has been available due to the creation of an electronic cashier system.

In this research, Association Analysis is conducted for trying to extract key features influencing the trip counts. The purpose is to find the potential relation between the distance from the mesh center to infrastructures and the trip changing condition in each mesh. More precisely, not only the combinations of the facilities with a high number of trips but the potential relations to the distribution of facilities with a high number of trips. Through constructing binary transaction data, it is possible to identify the hidden relationship of each item set within the dataset. Here, Association Analysis provides information on how strong the association rule, or the if-then relationship, is found to be true in the dataset (Sarker, 2018). The following equations are used to provide inferences on the cause-and-effect relationship.

$$\mathrm{Support}(X) = |\{t \in T; X \subseteq t\}|/|T| \tag{7.1}$$

$$\mathrm{Confidence}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \cap Y)}{\mathrm{supp}(X)} \tag{7.2}$$

$$\mathrm{Lift}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \cap Y)}{\mathrm{supp}(X) \times \mathrm{supp}(Y)} \tag{7.3}$$

$$\mathrm{Conviction}(X \Rightarrow Y) = \frac{1 - \mathrm{supp}(Y)}{1 - \mathrm{conf}(X \Rightarrow Y)} \tag{7.4}$$

where,

X and Y: independent item set

T: set of transactions

t: individual transactions, and

$X \rightarrow Y$: association rule that were under proposing

In the above equations, the support provides how frequently the item set appears in the data set and is defined as the number of transactions divided by the total number of transactions. The confidence indicates how often the association rule itself was found valid and is defined as the proportion of both item X and Y being purchased to the total purchase of X. The lift value provides the strength of association rule and is defined as Equation (7.3). Lastly, the conviction provides the inference on randomness and is defined as the expected frequency that X occurs without Y.

7.4.2. Transformation of dataset

Data transformation was made to make the dataset available for implement Association Analysis. For example, the distances are re-scaling to mapping them into 0, 1 feature. The threshold of 1 kilometer was selected, if distance<1km, it was coded as 1 and distance >1km was coded as 0. The mesh with a trip increasing rate>20% was selected in this study for analysis. Totally 2,817 meshes were chosen for this analysis as shown in the red part of Fig. 7.5.

7.4.3. Steps and results of Association Analysis

In order to reduce huge rules were generated via using Associate Analysis method for the data set, Fig. 7.6 show several filtered steps are used to extract useful rules. The method of this part is trial and error in order to avoid too many rules but keep the enough rules for discussion. Through several trials, the final criteria that was chosen here for support value 0.1, confidence value 0.8, and lift value 1.0. Pick up the first 10 rules, for example, the part of results can be shown in the following Table 7.3 for saving space. The results of antecedents and consequents are shown. Overall, there are total 165 rules are obtained in this group.

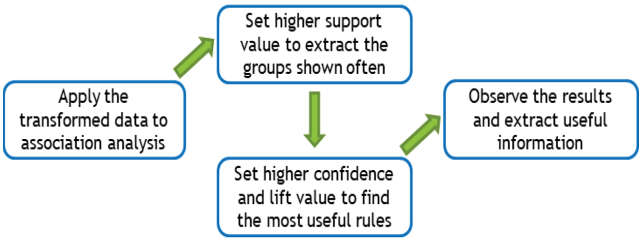


Fig. 7.6 Steps of Association Analysis

Table 7.3 The parts of filtered results from Association Analysis

Rules#	antecedents	consequents	support	confidence	lift
51	({'newt'})	({'hosp'})	0.12	0.95	1.09
54	({'cult'})	({'hosp'})	0.40	0.95	1.10
57	({'poli'})	({'hosp'})	0.38	0.95	1.10
58	({'elem', 'midd'})	({'park'})	0.21	0.94	1.14
64	({'elem', 'attr'})	({'park'})	0.16	0.88	1.07
71	({'elem', 'high'})	({'park'})	0.12	0.97	1.18
76	({'cult', 'elem'})	({'park'})	0.27	0.96	1.17
83	({'elem', 'poli'})	({'park'})	0.29	0.96	1.17
88	({'elem', 'hosp'})	({'park'})	0.48	0.92	1.12
90	({'elem', 'park'})	({'hosp'})	0.48	0.95	1.10

Through using Association Analysis, the numbers that how many times each infrastructure is shown in antecedents or consequents were known. This numbers in the contents of rules imply the trip frequency of each infrastructure. As a result, we find that hospital, elementary school, and park are the major locations where have a higher trip frequency for people, see Fig. 7.7. Short summary here might be that the facilities for mandatory task will attract more people, such as hospital and elementary school.

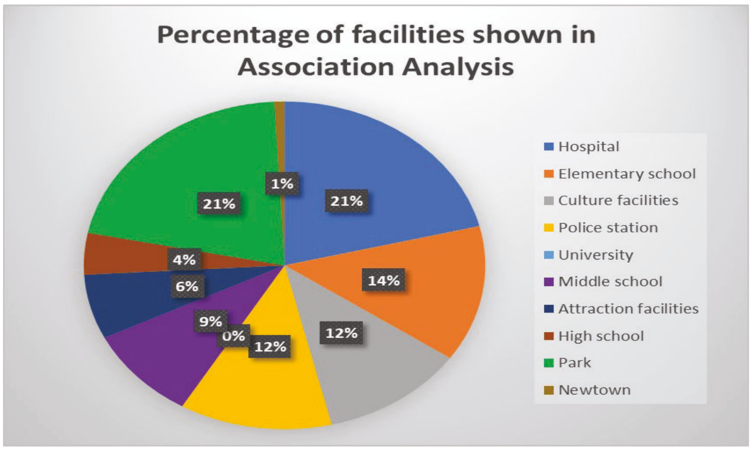


Fig. 7.7 Percentage of each infrastructure shown in above Association Analysis

7.5. Apply the Inverse Reinforcement Learning

The relations between trip generation and distance to infrastructures were obtained based on the previous analysis. In order to further understand the relationship between transportation modes and trip purposes, the machine learning method of IRL for traditional questionnaire survey data was also conducted.

7.5.1. Concept and principles of IRL

IRL method was selected as the analysis method for traditional questionnaire survey data. Reinforcement learning (RL) provides a powerful and general framework for decision making and control, but its application in practice is often hindered by the need for extensive feature and reward engineering. Deep reinforcement learning methods can remove the need for explicit engineering of policy or value features, but still require a manually specified reward function. IRL holds the promise of automatic reward acquisition (Fu, 2017). In the area of behavior analysis, it is difficult to precisely set rewards via using the RL method. The IRL method can be an alternative machine learning method to overcome the drawback of RL method. The decision-making structure of IRL can be flexibly updated. Moreover, IRL can perform iterative calculations based on the data and determine the structural pattern. Hence, IRL is particularly competitive for analyzing behavior for which the decision-making structure is unclear. As the IRL model estimates state value using continuous environment, it can consider the spillover effect. For example, if one state is often observed, IRL estimates the state value of the other states which is closely related to the observed state to be high as well as the observed state. As simple aggregation cannot capture this spillover effect, IRL can estimate more accurately than the simple aggregation. In this study, the maximum entropy (ME) method was applied to create an IRL model. It is true that various IRL method have been proposed and Bayes method and optimal margin method are often used. However, these IRL method acquire the optimal action in each state. For activity estimation like this research, it is impossible to estimate or define optimal action in each state. On the other hand, ME method can calculate the optimal reward function by inputting the action trajectory even if the optimal action in each state is unknown. Therefore, ME method was used for IRL. The calculation procedure is given in the following: The first step is to make a multidimensional space with the dimensions of the number of elements to be considered. Hereafter, the initial values of the reward function are decided. The state values and measures are therefore optimized from the initial values. Also, based on the optimal state values and measures, the corresponding action trajectory that maximizes the reward is estimated. Finally, the reward function is compared with the observed behavioral trajectory. At the same time, the reward function is updated so that the difference between the two trajectories is minimized.

Let R in Equation (7.5) denote the reward function, which is assumed from the action trajectory ζ and updates the parameter θ. Evaluating R until θ converges to a constant value.

$$\mathrm{R}(\zeta|\theta) = \theta^T \boldsymbol{f}_\zeta \tag{7.5}$$

where,

R(ζ): reward function

ζ: action trajectory

θ: parameter

$\boldsymbol{f}_\zeta$: assumed action trajectories

Numerical calculation could be needed to obtain the value of θ that maximizes the log-likelihood function L(θ) in Equation (7.6):

$$L(\theta) = -\sum_i logP(\boldsymbol{f}_{\xi_i}|\theta) \tag{7.6}$$

The function $P(\boldsymbol{f}_{\xi_i}|\theta)$ represents the probability of selecting a specific locus ξ_i. The gradient of the log-likelihood function is calculated as shown in Equation (7.7).

$$\nabla L(\theta) = \frac{\partial}{\partial\theta}\left\{\frac{1}{M}\sum_{i=1}^{M}\theta^T\boldsymbol{f}_{\xi_i} - log\sum_{i=1}^{M}exp\theta^T\boldsymbol{f}_{\xi_i}\right\} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{f}_{\xi_i} - \frac{1}{\sum_{i=1}^{M}exp\theta^T\boldsymbol{f}_{\xi_i}}\frac{\partial}{\partial\theta}\sum_{i=1}^{M}exp\,\theta^T\boldsymbol{f}_{\xi_i} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{f}_{\xi_i} - \sum_{i=1}^{M}\frac{exp\theta^T\boldsymbol{f}_{\xi_i}}{\sum_{i=1}^{M}exp\theta^T\boldsymbol{f}_{\xi_i}}\boldsymbol{f}_{\xi_i} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{f}_{\xi_i} - \sum_{i=1}^{M}P(\xi_i|\theta)\boldsymbol{f}_{\xi_i} \quad (7.7)$$

Finally, using this gradient, θ is updated by Equation (7.8) where α is the learning rate given exogenously.

$$\theta_{new} = \theta_{old} - \alpha\nabla L(\theta_{old}) \quad (7.8)$$

The steps and images of the above part is shown in Fig. 7.8.

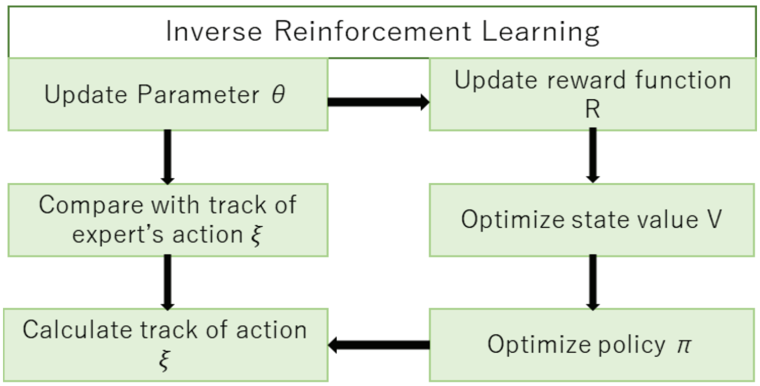


Fig. 7.8 Calculation concept of IRL

7.5.2. Data used and steps of IRL

The data of trip survey conducted in Tochigi prefecture in 2014 was used in this research. Focus on the data in Utsunomiya City, the data that origin place is within Utsunomiya City was extracted for the analysis here. The steps of IRL were conducted in the Jupyter environment through Python programming language. The variables of this part, the trip purposes and transportation modes selected for analysis are displayed in Fig. 7.9. Also, the time interval for the input was set at every 5 minutes as shown in Fig. 7.10. The number of samples of the data applied to the IRL model is almost 800,000. In this research, training data and test data are not completely divided. When training the model, data are extracted in certain sample rate. As the sample rate was 1% or 0.1% in this research, the estimated value of the accuracy is considered to be almost the same as the out-of-sample estimation.

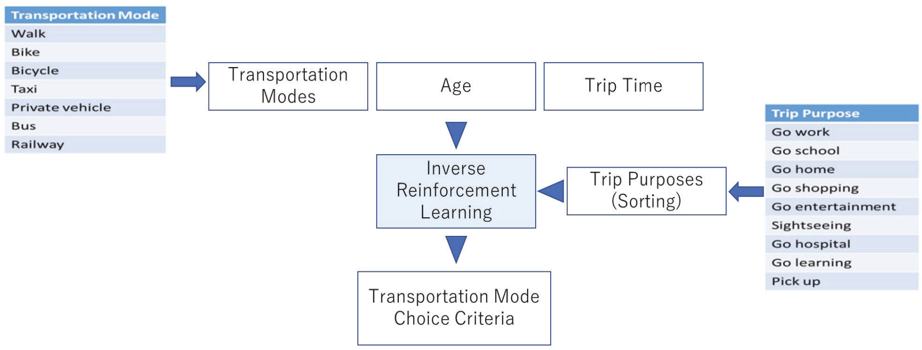


Fig. 7.9 The variables selected for this analysis

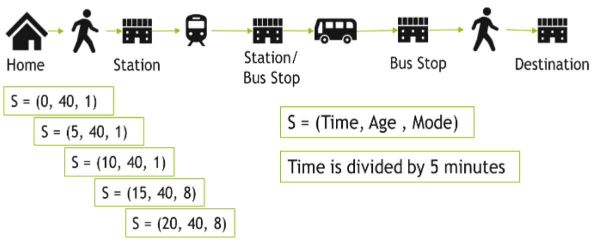


Fig. 7.10 The time interval used here is every 5 minutes

7.5.3. Results of IRL model

After analysis by using IRL method, one can get the tendency of many relationships. Fig. 7.11 to Fig. 7.14 display the pattern the transportation modes situation for the purposes of “go entertainment” and “go hospital” vs. different trip time and age, respectively. The higher state value represents that the tendency of selecting this transportation mode is higher. Valuable information can be found from these four figures. For example, Fig. 7.11 and Fig. 7.12 show that private vehicle is often selected when trip time is less than 30 minutes regardless of the purpose is to go entertainment or hospital. However, for longer trips or the trip time longer than 30 minutes, there is no dominate intension of choosing private vehicle, bus or railway to be found. From Fig. 7.13 and Fig. 7.14, we find that the middle age group have very strong intension to use private vehicle compared to other age groups. Possible reason here may be that middle age group was get used to private vehicle and it is difficult to change their habits. As a result, attracting more young people may be a possible way to increase the ridership of public transportation system.

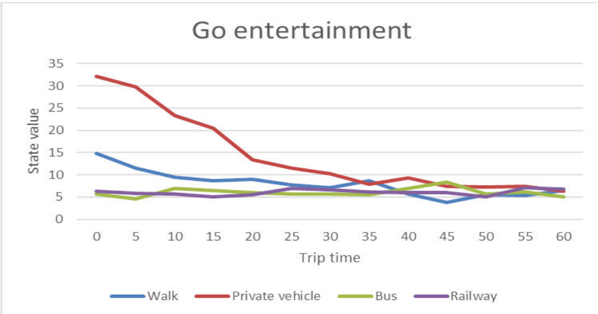


Fig. 7.11 The transportation modes situation for entertainment purpose based on different trip time

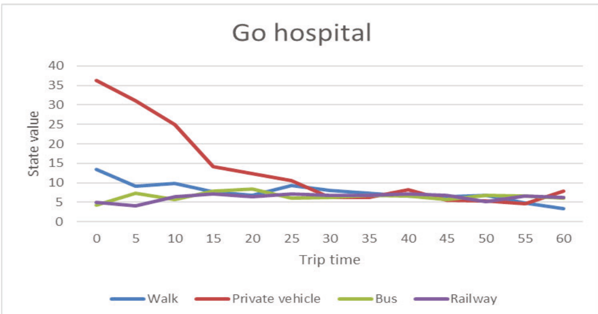


Fig. 7.12 The transportation modes situation for hospital purpose based on different trip time

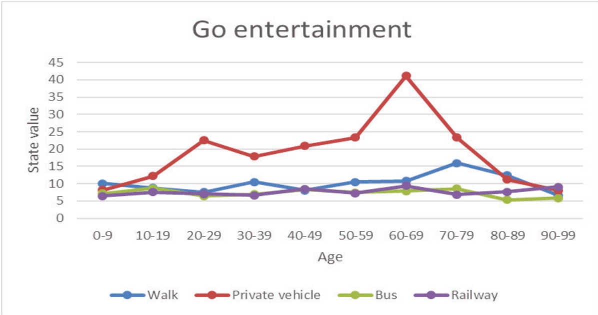


Fig. 7.13 The transportation modes situation for entertainment purpose based on different age

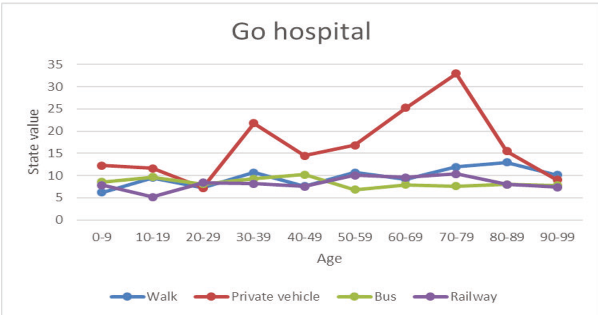


Fig. 7.14 The transportation modes situation for hospital purpose based on different age

7.5.4. Accuracy verification of IRL model

Since the application of IRL in this field is lack of enough information, the accuracy of the model should be checked after the results were acquired in order to verify the model. We compare our proposed model against the linear interpolation model. Table 7.4 shows how our proposed model outperforms the linear interpolation model. Here, the value represents the NLL (negative log-loss) index of error by the following formula (Kitani *et al*., 2012).

$$NLL(\boldsymbol{s}) = E_{\pi(a|s)}[-log \prod_t \pi(a_t|s_t)] \qquad (7.9)$$

where s_t and a_t means a sequence of states and actions, generated by the moving path for a specific configuration of the dataset. After the calculation, IRL represents the value of IRL model while Linear is the value of linear interpolation model. One can find that the precision accuracy of IRL model outperforms the linear interpolation model with a smaller value of index of error. Also, based on the below Table 7.4, the t-value test was also conducted for the reason to check the significance probability of this result. The average values of both IRL and Linear models were utilized here. Since the value of significance probability is below 0.05, the results here are good enough for taking reference.

Table 7.4 The NLL index of error of both models

	IRL	Linear
Work	0.000	9.15
School	0.705	9.35
Home	0.201	8.78
Shopping	0.000	10.50
Entertainment	3.14	6.94
Sightseeing	5.75	6.20
Hospital	0.000	13.0
Learning	0.000	7.71
Pick up	0.000	8.48
Average	*1.09*	*8.90*

7.6 Summary

7.6.1. Summary of this chapter

To conclude, this research proposed a study to utilize both static-type and dynamic-type data for two goals. The first goal is to provide an analysis procedure to find the potential relations between infrastructures and trip generation. By using the cell phone data, the trip purpose in the target city was analyzed. Through association analysis, hospital, park and elementary school are important locations in the areas to increase trip frequency in this case study in the Utsunomiya City. As a result, the future planning for transportation systems can be taken reference based on this result. The finding in this part that the facilities for mandatory task will attract more people might also be applied to the cities with similar scale. However, different cities could have different transportation cultures and different city scales. The Associate Analysis method with the proposed steps can be repeated in other target cities to find the important locations to increase trip frequency in the target city.

The second goal of this study is to apply IRL method to find that the key factors for selecting transportation modes for different trip purposes using the data obtained from a traditional personal trip survey. In summary, we can find that private vehicle is often selected when the trip time is less than 30 minutes regardless of the purpose is to go entertainment or hospital. However, for longer trips with the trip time longer than 30 minutes, the difference of frequency of choosing private vehicle, bus or railway is not clear. Moreover, the middle age group has very strong intension to select private vehicle for trips compared to other age groups. As the result, we may conclude that middle age group was get used to private vehicle and it is difficult to change their habits. Attracting more young people may be a possible way to increase the ridership of public transportation system. Also, the proposed IRL analysis procedure can also be used for other cities to find the key factors for selecting transportation modes for the target city.

7.6.2. Limitation and further implementation

The analysis conducted in this research serves as a primary approach to find some useful information behind the cell phone big data set and personal trip survey data by applying to the machine learning method of Associate Analysis and IRL. Nonetheless, due to the different amount and density of different infrastructures, further studies should be conducted in more detail analysis. Also, due to the limitation of datasets, further analysis should be conducted in order to verify the results or broaden the scope of the research if different datasets can be acquired. For example, using different machine learning methods for cell phone data and personal trip survey data can be a further study. Furthermore, because the cost for acquiring survey data is expensive, the years for obtaining the trip survey data and cell phone data are not overlap. Using overlap trip survey data and cell phone data for the proposed method can be another future study.

Chapter 8

CONCLUSION AND FURTHER IMPLEMENTATION

8.1 Conclusion

To conclude, the findings of this research were shown in the below Fig. 8.1 based on the previous conceptual diagram shown in Chap. 1. The orange part of the outer circle shows some keywords of important factors from the results. Here, since the study was focused on Utsunomiya City, the place where new LRT system will be put into operation soon, the potential factors that may be helpful to increase the passengers in the future were the core focuses in this study.

In first part of this research, I took LRT as a subject for new transportation and attempted to establish a new standard for LRT systems. No doubt, different cities in different countries would have an entirely different situation in public transportation, leading to the different design of LRT system. Nonetheless, the standard was set up based on relatively fair consideration, trying to be appropriate to any system. From this standard, one can understand the characteristic of LRT well, also standard can show good evaluation for LRT systems.

In second part, the first goal is to provide an analysis procedure to find the potential relations between infrastructures and trip generation. By using the cell phone data, the condition of generation trip in the target city was analyzed. Through association analysis, hospital, park and elementary school are important locations in the areas to increase trip frequency in this case study in the Utsunomiya City. As a result, the future planning for transportation systems can be taken reference based on this result. The finding in this part that the facilities for mandatory task will attract more people might also be applied to the cities with similar scale. However, different cities could have different transportation cultures and different city scales. The Associate Analysis method with the proposed steps can be repeated in other target cities to find the important locations to increase trip frequency in the target city. The second goal of this study is to apply IRL method to find that the potential factors for selecting transportation modes for different trip purposes using the data obtained from a traditional personal trip survey. In summary, I can find that private vehicle is often selected when the trip time is less than 30 minutes regardless of the purpose is to go entertainment or hospital. However, for longer trips with the trip time longer than 30 minutes, the difference of frequency of choosing private vehicle, bus or railway is not clear. Moreover, the middle age group has very strong intension to select private vehicle for trips compared to other age groups. As the result, I may conclude that middle age group was get used to private vehicle and it is difficult to change their habits. Attracting more young people may be a possible way to increase the ridership of public transportation system. Also, the proposed IRL analysis procedure can also be used for other cities to find the potential factors for selecting transportation modes for the target city.

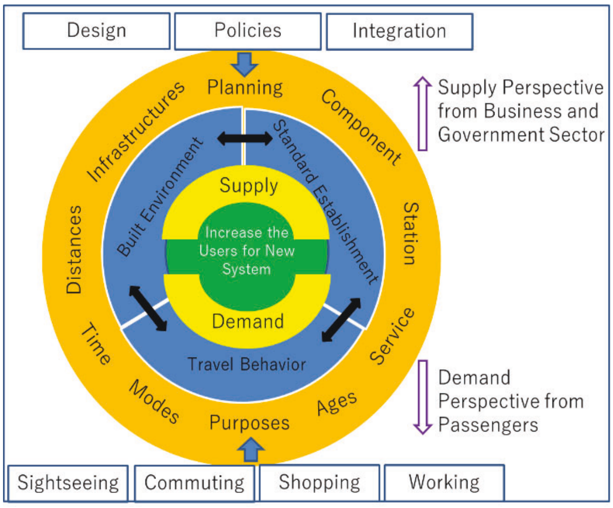


Fig. 8.1 Findings from conceptual diagram of this research

Hereafter, the findings of potential factors of each part of this study should be clarified and were shown in the below Fig. 8.2. In Chapter 4, through reviewing the existing LRT systems around the world, a standard of LRT indicators was established for the reference for future planning. The most indicators were also visualized in order to make them easier to understand. By case study and the results based on the current situation of all LRT systems, I may conclude that especially the design part should be strengthened in order to increase the primary demand.

In Chapter 5, the general situation of trip generation was focused on. The built environment data and mobile phone OD data were utilized in this part. Visualization of mobile phone data was conducted for the purpose of knowing the trip generation in general situation. By using the Association Analysis, findings are the trip generation in the areas near hospitals, parks and elementary schools have the biggest travel demand.

In Chapter 6, the special case of trip generation under the COVID-19 was focused on. The built environment data and mobile phone OD data were utilized in this part. Visualization of mobile phone data was conducted in order to know the difference of trip generation in special situation under the COVID-19 pandemic. Also, findings are the trip generation in the areas near elementary school was not strongly affected by the COVID-19. The derived demand was not decreasing obviously.

In Chapter 7, the condition of choice of transportation mode based on trip purpose was focused on. The personal trip data was utilized in this part. By using the Inverse Reinforcement Learning, findings are the private vehicle is often selected when the trip time is less than 30 minutes regardless of the trip purpose. Also, the middle and elderly age group has relatively strong tendency to use private vehicles.

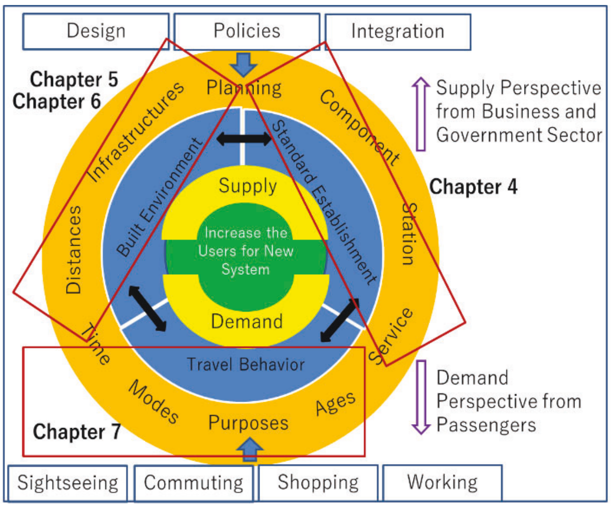


Fig. 8.2 Findings of each part of this research

8.2 Limitation

The analysis conducted in this research serves as a primary approach to find some useful information behind the cell phone big data set and personal trip survey data by applying to the machine learning method of Associate Analysis and IRL. Nonetheless, due to the different amount and density of different infrastructures, further studies should be conducted in more detail analysis. Also, different cities have different transportation culture and city scale, so the planning should be made by considering the difference between different cities. Furthermore, the numbers of trip generation by mobile phone data have the error since it is originally estimated number. It can be served as a reference of the trend of trips but not the exactly correct numbers.

8.3 Future Implementation

Due to the limitation of datasets, further analysis should be conducted in order to verify the results or broaden the scope of the research if different datasets can be acquired. For example, using different machine learning methods for cell phone data and personal trip survey data can be a further study. Furthermore, because the cost for acquiring survey data is expensive, the years for obtaining the trip survey data and cell phone data are not overlap. Using overlap trip survey data and cell phone data for the proposed method can be another future study.

References

Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability*, 11(1), 189.

Arora, S., Prashant D. A Survey of Inverse Reinforcement Learning: Challenges, Methods and Progress. *Artificial Intelligence*, 297 (2021): 103500. Print.

Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M., & Puchinger, J. (2019). Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 101, 254-275.

Bai, C., Zhou, L., Xia, M., & Feng, C. (2020). Analysis of the spatial association network structure of China's transportation carbon emissions and its driving factors. *Journal of environmental management*, 253, 109765.

Collins, G. S., & Moons, K. G. (2019). Reporting of artificial intelligence prediction models. *The Lancet*, 393(10181), 1577-1579.

Fu, J., Luo, K., & Levine, S. (2017). Learning robust rewards with adversarial inverse reinforcement learning. *arXiv preprint arXiv*:1710.11248.

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, 61(4), 5-14.

Hasugian, P. S. (2018). The data mining of cell phone most interested using apriorial algorithm. *Infokum*, 7(1, December), 7-15.

Henke, Cliff. "Bus Rapid Transit Bus Rapid Transit (BRT) and Light Rail Transit Systems Light Rail Transit (LRT) : State of Discussion." Encyclopedia of Sustainability Science and Technology, 2012, pp. 1810–1818., doi:10.1007/978-1-4419-0851-3_666.

Hidalgo, Dario. "Bus Rapid Transit Bus Rapid Transit (BRT) : Worldwide History of Development Bus Rapid Transit (BRT) Worldwide History of Development , Key Systems Bus Rapid Transit (BRT) Key Systems and Policy Issues Bus Rapid Transit (BRT) Policy Issues." Transportation Technologies for Sustainability, 2013, pp. 235–255., doi:10.1007/978-1-4614-5844-9_668.

Higgins, C., Ferguson, M., & Kanaroglou, P. (2014). Light rail and land use change: Rail transit's role in reshaping and revitalizing cities. *Journal of Public Transportation*, 17(2), 5.

History of Light Rail – LRTA, Tony Williams - LRTA, http://www.lrta.org/information.html

Huang, H., Cheng, Y., & Weibel, R. (2019). Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies*, 101, 297-312.

International Railway Journal, David Briginshaw, https://www.railjournal.com/index.php/light-rail/catenary-free-lrt-line-opens-in-nanjing.html

International Railway Journal, Dennis Schulz - Kevin Smith, http://www.railjournal.com/index.php/asia/china-embraces-light-rail.html

Kato, H., Shiroyama, H., & Fukayama, T. (2008). Process Management in Public Transit Planning: Case Study of Introduction Project of Light Rail Transit in Toyama, Japan. Transportation Research Record: Journal of the Transportation Research Board, 2042, 33-40. doi:10.3141/2042-04

Kitani, K. M., Ziebart, B. D., Bagnell, J. A., & Hebert, M. (2012, October). Activity forecasting. In *European conference on computer vision* (pp. 201-214). Springer, Berlin, Heidelberg.

Kriss, P., Miki-Imoto, H., Nishimaki, H., & Riku, T. (2020). Toyama City: Compact City Development. In *Word Bank*, Washington DC.

Landmark, A. D., Arnesen, P., Södersten, C. J., & Hjelkrem, O. A. (2021) Mobile phone data in transportation research: methods for benchmarking against other data sources. *Transportation*, 2021, 1-23.

Mizokami S., Hashiuchi J., Saito Y. (2007). Demand and cost-benefit analysis of upgrading the local railway into LRT system and its extension to the downtown. *Journal of Japan Society of Civil Engineering*, 63(1), 1-13. (in Japanese).

Mochizuki, A., Nakagawa, D., & Kasahara, T. (2007). Empirical analysis concerning the effect of Toyama Light Rail on the regional transport activities. *City planning review, Special issue, Papers on city planning*, 42(3), 949-954. (in Japanese).

Nishiyama, H. (2019). A Problem of Vacant Housing in Local Cities: Utsunomiya City, Tochigi. *The Rise in Vacant Housing in Post-Growth Japan: Housing Market, Urban Policy, and Revitalizing Aging Cities*, 123.

Ohmori, N. (2017, November). Future Perspective of Lifestyles and Mobility in Utsunomiya City. In *Proceedings of the 2nd Join Conference of Utsunomiya University andUniversitas Padjadjaran* (pp. 280-280).

Pacheco-Raguz, J. F. (2010). Assessing the impacts of Light Rail Transit on urban land in Manila. *Journal of Transport and Land Use*, 3(1), 113-138.

Ramos-Santiago, L. E., Brown, J. R., & Nixon, H. (2016). Streetcar Resurgence in the United States. Transportation Research Record: Journal of the Transportation Research Board, 2540, 30-38. doi:10.3141/2540-04

Sakamoto, S., Morimoto, A., & Daimon, H. (2015). A Study on Influence of LRT on Population Shift in European Various Cities. *Journal of the City Planning Institute of Japan*, 50(3).

Sarker, I. H., & Salim, F. D. (2018, June). Mining user behavioral rules from smartphone data through association analysis. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 450-461). Springer, Cham.

Sato, T., Sasaki, T., & Chikuma, M. (2018). Cost-benefit analysis of developing a light rail transit and feeder bus system in Utsunomiya city considering the change in population distribution. *Asian Transport Studies,* 5(1), 151-164.

Shibata, K., Fujita, K., & Horita, Y. (2012). Assessment of tram location and route navigation system in Toyama light rail transit. Paper presented at the 12^{th} International Conference on ITS Telecommunications, Taipei, Taiwan, 2012.

Takami, K., & Hatoyama, K. (2008). Sustainable regeneration of a car-dependent city: The case of Toyama toward a compact city. In *Sustainable City Regions:* (pp. 183-200). Springer, Tokyo.

Takasugi, E., SATO, T., & Chikuma, M. (2018). Development of the method of measuring the benefits of developing light rail transit and bus rapid transit considering differences of these systems and influence on urban population distribution and a case study for Maebashi city, Japan. *Journal of the City Planning Institute of Japan*, 53(3).

Tan, P. N., Steinbach, M., & Kumar, V. (2005). Association analysis: basic concepts and algorithms. In *Introduction to Data mining* (Vol. 321321367). Addison-Wesley. Boston, MA.

Thompson, G. L. (2003). *Defining an alternative future: birth of the light rail movement in North America* (No. E-C058).

Tosi, D. (2017). Cell phone big data to compute mobility scenarios for future smart cities. *International Journal of Data Science and Analytics,* 4(4), 265-284.

Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, 11, 141-155.

Widhalm, P., Yang, Y., Ulm, M., Athavale, S., & González, M. C. (2015). Discovering urban activity patterns in cell phone data. *Transportation*, 42(4), 597-623.

Xia, Haishan, et al. "Challenges in the Integration of Light Rail and Land Use: A Case Study of Yizhuang Line in Beijing." Urban Rail Transit, vol. 3, no. 1, 2017, pp. 45–53., doi:10.1007/s40864-017-0055-3.

Xu, Tao, et al. "The Impact of Urban Rail Transit on Commercial Property Value: New Evidence from Wuhan, China." Transportation Research Part A: Policy and Practice, vol. 91, 2016, pp. 223–235., doi:10.1016/j.tra.2016.06.026.

Yang, Y., Xu, Y., Rodriguez, D. A., Michael, Y., & Zhang, H. (2018). Active travel, public transportation use, and daily transport among older adults: The association of built environment. *Journal of Transport & Health*, 9, 288-298.

You, Changxi, et al. Advanced Planning for Autonomous Vehicles Using Reinforcement Learning and Deep Inverse Reinforcement Learning. *Robotics and Autonomous Systems*, vol. 114, 2019, pp. 1–18.

Zhang, M., & Yen, B. T. (2020). The impact of Bus Rapid Transit (BRT) on land and property values: A meta-analysis. *Land Use Policy*, 96, 104684.

Zhao, L., & Shen, L. (2019). The impacts of rail transit on future urban land use development: A case study in Wuhan, China. *Transport Policy*, 81, 396-405.

Zhou Mu Zhi., & Xu Lin, China City General Development Index 2016, People's Publishing House

"Suzhou Rail Transit." Wikipedia, Wikimedia Foundation, 17 June 2018, en.wikipedia.org/wiki/Suzhou_Rail_Transit.

List of Research Achievements

List of research achievements for application of Doctor of Engineering, Waseda University	
Full Name：	Hsiang Chuan Chang — seal or signature
	Date Submitted(yyyy/mm/dd): 2022/11/2
種類別 (By Type)	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む） (theme, journal name, date & year of publication, name of authors inc. yourself)
Journal Paper	(O) **Chang, H.**, Kitano, N., Morimoto, A. A Primary Study to Understand the Trends of Generation Trips along the LRT Route in Utsunomiya City by Using Cell Phone Data. Journal of the Eastern Asia Society for Transportation Studies Volume 14, Page 1568-1585, 2022.
Journal Paper	(O) **Chang, H.**, Okubo, T., Kobayashi, A., Morimoto, A. Artificial Intelligence (AI) Applications Using Big Data and Survey Data for Exploring the Existence of the Potential Users of Public Transportation System. International Journal of Information and Management Sciences, 2022. (Accepted)
Journal Paper	Takayama, K., Watanabe, Y., **Chang, H.**, Morimoto, A. Study on the Characteristics of Japanese Transit Oriented Development as Seen from Long-term Land-use Changes. Transportation Research Procedia Volume 48, Page 2313-2328, 2019.
Journal Paper	Noble, T., Kitano, N., Morimoto, A., **Chang, H.** Analyzing the Potential of the Sharing Economy in a Post-COVID World: A Comparison of Literature on Cities. Urban and Regional Planning Review Volume 9, Page 167-184, 2022.
Lecture (Int. Conf.)	(O) **Chang, H.**, Morimoto, A. A Primary Study to Evaluate LRT Performance through Making Standard-Case Studies for Suzhou and Nanjing. Asian-Pacific Planning Societies 2018 International Conference(Presentation in Ho Chi Minh City, Vietnam)
Lecture (Int. Conf.)	Takayama, K., Watanabe, Y., **Chang, H.**, Morimoto, A. Study on the Characteristics of Japanese Transit Oriented Development as Seen from Long-term Land-use Changes. 15th World Conference on Transport Research(Mumbai, India)
Lecture (Int. Conf.)	**Chang, H.**, Kitano, N., Morimoto, A. A Primary Study to Understand the Trends of Generation Trips along the LRT Route in Utsunomiya City by Using Cell Phone Data. The 14th International Conference of Eastern Asia Society for Transportation Studies 2021(Presentation in Hiroshima, Japan)(Online)
Lecture (Int. Conf.)	(O) **Chang, H.**, Minamikawa, A., Morimoto, A. A Study on the Temporal and Spatial Change of Generation Trips under COVID-19 by Using Mobile Phone Data- Case Study for Utsunomiya City. Asian-Pacific Planning Societies 2022 International Conference(Presentation in Nagasaki, Japan)