博士論文

Mobile Phone Big Data Mining for Potential Demand Detection of Urban Shared Transportation

（シェア交通サービスの潜在需要に着目した携帯電話のビッグデータの発掘）

張 浩然

Abstract

Mobility-as-a-Service (MaaS) describes a shift away from personally-owned modes of transportation and mobility solutions consumed as a service. The one significant application of MaaS is the shared transportation system. Shared transportation systems are enabled by combining transportation services from person-to-person transportation providers through a unified gateway that creates and manages the trip. The key concept behind the shared transportation system is to offer travellers mobility solutions based on their travel needs. For local authorities and policymakers, the potential to use shared transportation systems to source data on travel movements could open the door to more efficient use of capacity and new transport management tools.

The objective of the thesis is to utilize the mobile phone GPS data to analyze the market potential and its environmental performance of emerging shared transportation modes. In this thesis, the study takes three shared transportation modes, which are bicycle-sharing, ride-sharing, bus-sharing, as the study cases.

Firstly, the study proposes a market-oriented methodology for non-dock bicycle-sharing system planning. The potential demand for bicycle-sharing use is identified using mobile phone data. Then, a link network is constructed to represent the spatial distribution of potential markets. The algorithm of community detection is applied to the link network to discover the densely connected nodes in the area and define the division of sub-service areas. Finally, the potential patterns of bicycle-sharing use, the spatial distribution of the potential demand, and the potential emission reduction are analyzed based on the resulting division of sub-service areas. The proposed method is tested using a data set of approximately 34 million GPS trajectories obtained in Tokyo. The area of Tokyo is subdivided into 21 sub-service areas. Suggestions for bicycle management, infrastructure development, and bicycle-sharing system planning are given regarding sub-service areas with different properties. The potential emission reduction for the introduction of a bicycle-sharing system is calculated in the area of Tokyo.

Then, a model is proposed for analyzing the potential reduction in emissions associated with the adoption of a bicycle-sharing system. Methods are presented for extracting human travel modes from mobile phone GPS trajectories, together with a geometry-based probability model, to support particle swarm optimization. A comparison study is implemented to analyze the model's computational efficiency. Based on the resulting optimal layout for the network of bicycle docking stations, a multi-scenario integer linear programming model is proposed to optimize rebalancing procedures (i.e., moving bicycles between docking stations according to demand). Mobile phone GPS trajectories from approximately 3.7 million local mobilities are used to construct a case study for Setagaya Ward, Tokyo. The results show that, compared with the previous methods, the optimal layout solved by the proposed method could reduce emissions by

a further 6.4% and 4.4%. With an increase from 30 to 90 bicycle stations, the adoption of bicycle-sharing can reduce CO2 emissions by approximately 3.1–3.8 thousand tonnes. However, emission reduction will maximally decrease by 21.26% after an offset by bicycle production and rebalancing-generated emission.

For ride-sharing, this study proposes an analysis framework to bridge the gap in city-level adoption potential analysis. This study chooses the case study of the Tokyo area with over 1 million GPS travel records and trained a deep learning model to find out this potential. On average, from the computation result, nearly 26.97% of travel distance could be saved by ride-sharing, which told us that there is more similarity in people's travel patterns in Tokyo. There is considerable potential for ride-sharing. Moreover, the exhaust emission of CO, NMHC, and NOx can be reduced maximally by 15.55 tons, 0.63 tons, and 0.63 tons, respectively. Ride-sharing can improve the air quality of these center business districts and alleviate some city problems like traffic congestion. This chapter proposed a new framework to analyze the ride-sharing potential for ride-sharing service providers and decision-makers.

Finally, the thesis introduces a method to generate planning suggestions for sharing-bus lines and stops massive demand data. From the demand input, a link network is constructed to represent the sharing route of the demand. Community detection is applied to the link network to segment the link network into network communities with similar travel routes. By examining the core-peripheral structure and match the core part of communities into the road network, customized bus lines are generated. By analyzing the potential demand, boarding and alighting hotspots are identified as the suggestion for customized bus stops. A case study is conducted using mobile phone data in Tokyo, in which 29 bus lines are extracted. With inputting one-day sample data, our algorithm can generate the result in approximately 1 minute. According to the shape and spatial location of the bus lines, three types of bus lines serving different travel patterns are classified, including radiation type lines, ring-type lines, and suburban lines. By analyzing the emission reduction potential of the bus lines extracted, the bus lines generated by the proposed method can reduce emission pressure on urban expressways and mitigate approximately 13% of road traffic emissions.

Originality

Due to the lack of works that summarized the development and frontiers of shared transportation potential and environmental impact analyses, it is hard to get comprehensive and high-dimensional information about the shared transportation system. This research aims to focus on the following questions: a) how to define and reinvent data-driven mobility models by studying urban dynamics, urban mobility, transportation behavior, and sharing potential. b) within a city-level urban mobility framework, how can we characterize the nature of data-enabled shared transportation services among different modes, and what are the similarities and differences. c) the existing positive and successful shared transportation systems that can be identified in the studied domain and how they can be best applied for practical success. To answer these questions, we take three shared transportation modes: bicycle-sharing, ride-sharing, bus-sharing, as the study cases to discuss shared transportation systems under the framework of MaaS and developed a series of methods. The contributions of this study include:

(1) The study maps the high-dimensional city people flow data into graph-structure forms based on the sharing characteristics. (Chapters 4 & 7)

(2) The study utilizes the community detection method to decompose the spatial complexity of the mobility sharing potential estimation. (Chapters 4 & 7)

(3) The study proposes an advanced PSO algorithm for reducing the probability of converging to locally optimal solutions of the high-dimensional model. (Chapter 5)

(4) The study extracts the key features of mobility sharing and trains the model with prior knowledge. (Chapter 6)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Mobility as a Service (MaaS) was first introduced at the European Union ITS Conference in Helsinki, Finland, in 2014. At the 2015 World ITS Conference in Bordeaux, France, MaaS began to become a hot topic in the global intelligent transportation field. The connotation of MaaS is to understand the travel needs of the public deeply. By integrating all modes of transportation into a unified service system and platform, MaaS can make full use of big data to make decisions, allocate optimal resources, and meet the needs of the travel community. Moreover, to provide external services with a unified APP. MaaS has four characteristics as followings: (1) Sharing: MaaS pays attention to the provision of transportation services rather than the ownership of vehicles; in addition, as a passenger, he is not only the enjoyment of transportation services, but also the provider and sharer of traffic data for optimizing the entire travel service; (2) Integration: The so-called integration is the integration of various models and then the integration of payment systems; (3) People-oriented: Its main goal is to provide better travel services, seamless connection, safe, convenient and comfortable travel; (4) Low carbon: let everyone increase the proportion of green travel, and reduce the usage of private cars.

The one significant application of MaaS is the shared transportation system (STS). Currently, many research groups and companies are focusing on developing STS under the framework of MaaS. The research fundamentals to achieve this target are how to assess the urban mobility characteristics, shared transportation adoption potential, and their matching performance as conditions and constraints in a transport system that is undergoing automation and is highly dependent on software, navigation systems, and connectivity. Further, another core issue is how to design MaaS platforms for STS that adapt to the evolving mobility environment, new types of transportation, and users based on an integrated solution that utilizes the sensing and communication capabilities to tackle the significant challenges that the STS industry face.

Big data-based TST development is an emerging topic both in academic and industrial aspects. By the end of 2020, there will be over 50 billion connected devices globally, collecting over 2.3 zettabytes of data each year. STS leverages on this opportunity and is an example of a business model that is supported by the growth in smartphone use. STS is a digital, data-driven service that uses several technological capabilities associated with intelligent mobility innovation. It relies on building an ecosystem of stakeholders that agree to manage the supply and demand of the services that travelers

want.

However, currently, all studies about big data in STS are fragmented. Few works have summarized the systemic knowledge on this field. Specifically, few studies focus on introducing how to screen and process the potential value from the "deluge" of unverified, noisy, and sometimes incomplete information for STS development. Also, few works were designed for high-resolution travel data (e.g., GPS data) mining for accurate shared transportation demand analysis, such as bicycle-sharing, ride-sharing, bus-sharing. However, the above knowledge is significant for stakeholders, such as researchers, engineers, operators, company administrators, and policymakers on related fields, to comprehensively understand the infra-knowledge structure and limitations of current technologies.

Therefore, in this thesis, I took three shared transportation modes, which are bicycle-sharing, ride-sharing, bus-sharing, as the study cases and try to discuss the issue of STS under the framework of MaaS via answering the following question: How to efficiently extract and utilize key feature information of high-dimensional city people flow data and assess mobility sharing potential at an urban scale?

1.2 Shared Transportation Systems

1.2.1 Bicycle-Sharing System

Bicycle-sharing is a mobility strategy that could provide a healthy and eco-friendly alternative to motorized public transportation for short-distance trips in urban areas. Meanwhile, bicycle-sharing in reducing emissions as promoting cycling has been put forth as one of the major strategies to mitigate GHGs emissions and to slow down their catastrophic consequences. As a form of urban public transportation mode, promoting bicycle-sharing systems in the urban areas has several advantages: (1) Bicycle is an environmentally friendly travel mode, it produces zero emissions. It provides a comfortable and livable environment for the city [1,2]. (2) The specific benefits of cycling on health outcomes include reductions in mortality and weight gain. The public health of the city with more people choosing bicycle can be largely improved [3]. (3) Meanwhile, a bicycle-sharing system can be a supplement of the public transportation system, as an efficient way to solve the "last mile" problem -- connecting people from a public transit network to their final destination, which is often the least efficient and least cost-effective part of travel [4,5].

Bicycle-sharing is not a new thing. In the early stage, bicycle-sharing systems were operated with bicycle docking stations and were mainly promoted by the government. Users have to reach the bicycle docking stations to get access to the bicycle, which leads to another "last mile" problem. Since 2014, Internet-supported bicycle-sharing mode without bicycle docking stations had emerged and gradually become warmly

applauded by the public. Through the mobile phone application, users can see the location of nearby bicycles and reserve one. Data released by China E-commerce Research Center in June 2017 shows that China's share economy scaled up to 3.945 trillion RMB in 2016, increased by 76.4%, serving roughly 60 million people and creating 5.85 million jobs, and the total size is still expanding [6].

However, in recent years, there are some problems revealed in the existing bicycle-sharing programs. For example, bicycles in some areas do not earn the acceptance of the public, but in some areas, the demand is unsatisfied. Bicycles without enough parking lots also block the streets or roads and cause traffic congestion[7]. These problems are mainly resulting from the inappropriate layout of bicycle-sharing systems and the lack of an accurate grasp of the bicycle-sharing market. The layout design of the bicycle-sharing system will have a significant influence on the system's implementation performance [8]. A well-designed system that matches the actual bicycle-sharing demand will be readily accepted by the public and can minimize bicycle rebalancing and reduce the emission [9]. The optimization of bicycle-sharing systems generally starts with the definition of coverage areas. Then, optimizing the bicycle dock stations or rationing of bicycles are adapted to satisfy the demand. It is also essential to propose compatible management in areas with different demand patterns and different rent and return behaviors [10].

Nevertheless, the definition of the coverage area and the division of sub-service areas are highly essential and fundamental. However, it usually gets less attention in the bicycle-sharing system planning. In most large cities, bicycle-sharing programs are usually operated based on the administrative division or limited to the city center and gradually extending it to reach the peripheral areas. This kind of area consideration has several disadvantages. Firstly, the actual travel demand is failed to be considered. The frequent cross-area bicycle-sharing behaviors will lead to considerable effort on bicycle rationing and rebalancing. Secondly, without dividing areas into small market segments, operators of bicycle-sharing services cannot adapt compatible management strategies for areas with different demand patterns. In order to determine the coverage area, a more reasonable way should be a market-oriented sub-service area division based on the bicycle-sharing pattern.

1.2.2 Ride-Sharing System

As one significant component of shared transportation, Ride-sharing is an emerging transportation mode under a new concept of MaaS [11]. The essential connotation of this concept is to understand the travel demand of the public sincerely[12]. By integrating travel information into a unified service system and platform, ride-sharing can make full use of big data to make decisions, allocate optimal resources, and meet the travel demand of communities[13]. Ride-sharing provides external services and collects massive user data with a unified APP[14], which has three characteristics: (1) Integrated: Ride-sharing pays attention to the provision of transportation services rather than the

ownership of vehicles; in addition, as a passenger, he is not only the enjoyment of transportation services but also the provider and sharer of traffic data for optimizing the complete travel service; (2) People-oriented: Its main goal is to provide better travel services, seamless connection, and convenient travel; (3) Sustainable: Ride-sharing enables people to share vehicles and trips, thereby reducing energy consumption and air pollution, and relieving the traffic jam during rush hour[15]. The whole process of ride-sharing is shown in Figure 1. 1. From the ride-sharing concept, people with additional resources could be organized through the Internet and MaaS APP. After intelligent matching, the integral system efficiency will be significantly increased[16]. Therefore, developing ride-sharing is a fashionable trend for smart and green urban planning[17]. According to the existing cases, ride-sharing is generally introduced and operated by some emerging companies as a commercial project, rather than governmental public welfare[18-20]. Hence, the estimated production value of ride-sharing sector is the determinant for whether the investment would be determined and the pre-estimation for the potential of ride-sharing is necessary[21].

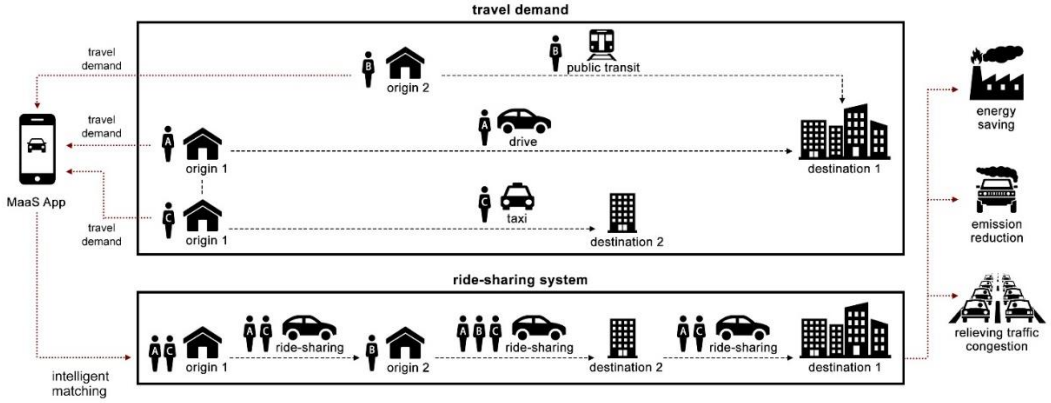


Figure 1. 1 Ride-sharing system

Nevertheless, the potential analysis is a highly complex issue[22]. Specific knowledge gaps need to be further solved. (1) Demand assessment (the scale of driver): Generally, the previous studies for the regional potential of ride-sharing are mainly based on the urban population distribution, which is rough and imprecise[23]. (2) The scenario analysis on the different extent of participation in ride-sharing. (3) Matching simulation: With the uncertainty of mobility trajectory and fuzziness of matching requirement, it is hard to predict the matching feasibility for each driver and rider[24]. As shown in Figure 1. 2, the commuting route for a driver is not unique, especially on an urban scale. There will be multiple routes for the same origin-destination (OD). Meanwhile, choosing which route for a driver will depend on the current traffic condition, driver's habit, and even governmental traffic control scheme. For a real-world case, it is hard to observe that information from the individual level. (4) Interaction among the above factors: Those inaccuracies will introduce many uncertainties into the issue, which will significantly impact the matching result. Besides, the determinant for whether a driver and a rider would be matched or not is only the similarity of their trip's OD but also that of their expected time scheme (departure and arrival time). Here, the driving time asynchronism

also should be taken into consideration. For example, in Figure 1. 2, the matching condition of the driver and rider #2 is that the picked-up time of rider #2 should be similar to its expected departure time, and so do the arrival time. However, how to measure those similarities of not only distance but also the duration is a fuzzy question. Since the uncertainty of mobile trajectory and fuzziness of matching requirement, it is incredibly complicated to make the matching work even for a small scale of users. When the range expands to an urban level, the complexity of the issue will be increased exponentially. (5) The existing optimization methods cannot support the big data-based potential and emission reduction analyses due to the computational complexity.

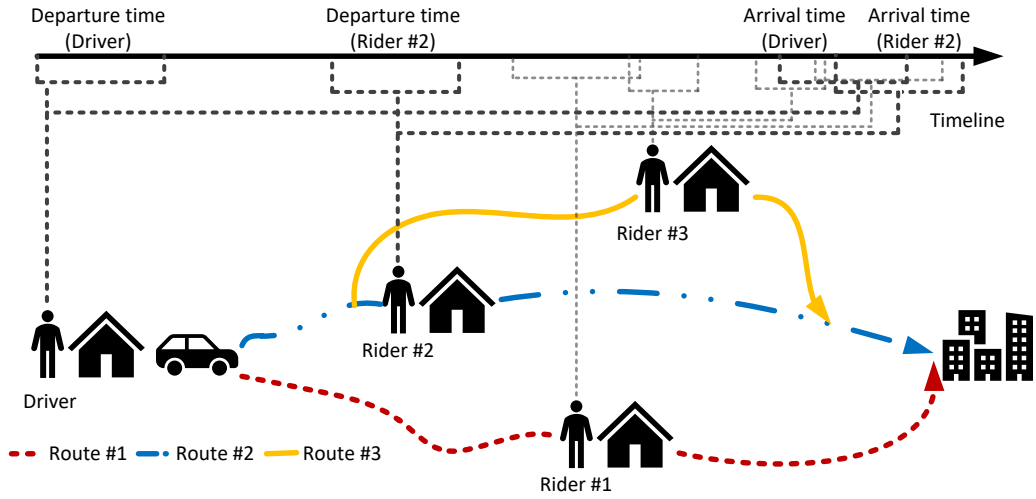


Figure 1. 2 Uncertainty of mobile trajectory and fuzziness of matching requirement

1.2.3 Bus-Sharing System

Bus-Sharing is a new type of Internet-supported public transportation mode that emerged and was warmly applauded by the public in recent years[25]. A bus-sharing system aims to provide more demand-oriented, express, and efficient transit services than traditional public transit bus lines. Meanwhile, promoting a bus-sharing system can be one of the major strategies to reduce the usage of private cars and mitigate GHG emissions from road traffic. The performance of the bus-sharing system largely depends on the system design. Currently, the bus-sharing systems in operation usually collect demand data from online surveys and manually plan bus-sharing lines, costly and inefficient and can hardly be competitive than traditional bus transit. A real-time dynamic bus-sharing line planning system based on the demand can primarily improve the performance and the public acceptance of the bus-sharing system.

Chartered bus sharing on demand is a group of people spread in a specific region, interested in going to the same place or region at approximate times. Bus-sharing systems must have the ability to offer customized routes to passengers. There have been some bus sharing platforms that have the intelligence to group demands and suggest

routes according to the customized presets (maximum Km, maximum route time, minimum vehicle occupancy, etc.), so as a suggestion, the routes will be published in a specific landing page. The passenger can buy/reserve the seat online.

However, developing such a bus line planning system faces some new challenges: (1) bus-sharing systems should have an Internet-based platform to collect passenger's daily travel demand efficiently. (2) The system should handle massive demand input and generate results in a short time of computation. (3) Due to the constant change of demand, the algorithm of generating bus-sharing lines and stops also requires to be dynamic. Dynamic bus line planning algorithms should be integrated into the system to generate bus-sharing networks according to the real-time demand input or history data.

1.3 Objective and Outline

The objective of the thesis is to utilize the mobile phone GPS data to analyze the market potential and its environmental performance of emerging shared transportation modes. In this thesis, The study took three shared transportation modes, which are bicycle-sharing, ride-sharing, bus-sharing, as the study cases and aimed to discuss the following issue: How to efficiently extract and utilize key feature information of high-dimensional city people flow data and assess mobility sharing potential at an urban scale?

The remainder of this thesis is organized as follows: In chapter 2, The study gives a literature review of the researches on shared transportation systems. In chapter 3, The study introduces the data source and preprocessing method. Chapter 4 and chapter 5 focus on the bicycle-sharing system and discuss market-oriented sub-area division and layout optimization issues. In chapter 6, the study explores the issue of city-level potential analysis for ride-sharing systems. Chapter 7 introduces the proposed method for city-level dynamic lines design for the bus-sharing system. Chapter 8 concludes this thesis and introduce some point which can improve in the future.

1.4 Assumptions

To effectively formulate the big-data-driven shared transportation models, here we assume that Shared transportation usage demand has a spatiotemporally linear relation with Corresponding-distance total travel demand. Specifically, we assume

- all short-distance (1-3 km) walk, bicycle, and car travels in the dataset are the potential bicycle-sharing demands;
- all car travels in the dataset are the potential ride-sharing drivers' travel demands; all long-distance travels (>3km) extracted from the dataset represent the ride-sharing riders' demands;
- all car travels in the dataset are the potential bus-sharing demands.

Chapter 2

Literature review

2.1 Bicycle-Sharing System

In recent years, most of the studies of bicycle-sharing have focused on the operation and optimization of bicycle-sharing systems and system layout design [8], social and environmental benefits[26], bicycle-sharing demand [27], impacts on public health [28], and effects of built environment and weather on bicycle-sharing [29]. The success of the bicycle-sharing program largely depends on how user demand is met and how the system is designed. Much attention has been paid to these two problems.

The existing predicting methods are focusing on demand from three levels[27]: city-level, cluster-level, and station-level on the demand prediction aspect. Giot and Cherrier made demand predictions for the next 24 hours with a city-level granularity for the Capital Bike Share system[30]. Li proposed a bike-sharing demand prediction framework that introduces a Bipartite Station Clustering algorithm to individual group stations[31]. Li used a Graph Convolutional Neural Network with Data-driven Graph Filter (GCNN-DDGF) model to learn hidden different pairwise correlations between stations to predict station-level hourly demand on a large-scale bike-sharing network[27].

Research mainly focuses on the promotion policies of bicycle-sharing service and the siting of bicycle stations on the system design aspect. The existing methods can be divided into two types: mathematical programming models and GIS-based methods. Mathematical models based on developing mixed-integer linear programming (MILP) or mixed-integer nonlinear programming (MINLP) models have been frequently applied to plan a new bicycle-sharing system. Chen and Sun proposed a MILP model to formulate the layout of public bicycle stations to minimize investment budget and users' total travel time[32]. The GIS-based subjective analysis is a method for the layout design for many public infrastructures by using GIS as a support tool for assessing bicycle facility planning and proposing indicators for measuring latent demand[33-35]. Gonzalez proposed a model to optimize the bicycle stations in bicycle-sharing systems by using GIS-based analyzing tools [36]. Zhang proposed an optimization method for the layout of the network of bicycle docking stations and a model for optimal rebalancing[8]. Based on the spatial data generated from the real world, results from the GIS-based method are more objective and practical.

Most existing researches and methods are focusing on the dock-based bicycle-sharing system. As a newly-arisen business model, an Internet-supported bicycle-sharing service without bicycle docking stations is more flexible and easier to accept by the

public[37]. However, both demand prediction and the system design in the non-dock bicycle-sharing systems are different from dock-based systems.

In the previous study and existing bicycle-sharing program, the coverage area of the bicycle-sharing system gets little attention. The coverage area is usually defined based on the boundaries of the central urban district or administration division [38]. Most of the existing method to define the bicycle-sharing area is to propose models for optimizing the travel demand served by the bicycle-sharing system. Frade defined the coverage area of the bicycle-sharing system by proposing an optimization method to maximize the demand covered[39]. Zhang and Meng also consider the subarea segmentation in their research. They proposed a method based on network community division and set up a model to optimize the recourses. However, the study area was only restricted to a relatively small district[40]. The existing consideration of coverage area is usually based on a mathematical model to maximize the demand served by the bicycle-sharing system. Due to the lack of data and computing resources, these methods are usually applied to case studies in a small region.

The emergence of pervasive, geospatial data generated by individuals has recently triggered an opportunity to study individual mobility patterns[41] and spatio-temporal accessibility[42]. As a new travel survey tool, mobile phone data is more pervasive and accurate, which allows us to fully track the trip chain of the individual in both temporal and spatial dimensions. It offers a new way to anticipate the potential demand and perform a market-oriented sub-area division of the non-dock bicycle-sharing system in the city-level geospatial area.

In short, the previous methods are still limited by the dilemma between theoretically optimal sites and their real-world practicability. Moreover, the coverage area of the bicycle-sharing system gets little attention. Due to the heavy computation requirement, the existing methods have limitations in handling large-scale cases. For the related works about bicycle-sharing system layout optimization, the existing methods are usually survey (or aggregated) data-driven models which cause less accuracy in travel demand detection. Also, the existing methods are usually point (or cell)-based models which cause less flexibility in real applications. Although novel GIS-based methods have been developed to achieve a more effective balance for this challenge, further analysis, and discussion of the strength of higher-quality sample sets and corresponding solving methods are still needed.

2.2 Ride-Sharing System

As a representative of intelligent and green city development, the ride-sharing system is one of the hottest topics in transportation, public health, urban planning. Many articles pursued emerging applications, case discussion, and improving methodology for this system in the recent five years. Among these studies, many of them have focused on the history and benefits of ride-sharing systems[43], crowdsourcing

optimization[44], development policies[45], market analysis[46], framework design[47], and survey study[48]. These studies could provide some theoretical foundations and data supports for the potential analysis of ride-sharing systems. In order to make a precise classification for their contributions, Agatz, et al. [49] concluded three further broad research fields for ride-sharing systems: (1) Optimization[50], (2) Incentive[51], and (3) Choice[52]. The rest of the related work will be organized from these three aspects.

Optimization for ride-sharing systems is a traditional topic in operational research, which is also the critical task of the potential assessment. Model is the base of optimization, which could be classified into static and dynamic[53]. The former requires the complete knowledge of transportation networks for optimization; however, real-time applications usually lack it. Despite basic or dynamic, both problems need to consider how to arrange the matching pool, and different researches focused on mainly three kinds of arrangement: "Single rider, single driver[54]", "Single driver, multiple riders[55]" and "Single rider, multiple driver[56]" arrangement issues. Recently, uncertainty in ride-sharing optimization is another emerging research topic because the uncertainty in the model's parameter could destroy the performance of the optimization solution, and there are always uncertainties in real-world application. Chen et al.[57] exploited an event-driven Receding Horizon Control scheme for a Ride-Sharing System to minimize a weighted sum of passenger waiting and travel times and used Simulation of Urban Mobility as the case study evaluate their scheme. The result shows that based on an actual map and taxi travel information, the RHC performs better than the known greedy heuristic. Naoum-Sawaya, et al. [58] presented a stochastic mixed-integer programming model to optimize the allocation of shared vehicles to employees while considering the unforeseen event of vehicle unavailability. Considering uncertainties of travel time and delivery location, Li, et al. [59] proposed a methodology for ride-sharing optimization integrating an adaptive large neighborhood heuristic search and three sampling strategies for the scenario generation.

Incentive, which is another research field in ride-sharing, means assessing the degree of increasing the potential and contribution to the environment. Rising fuel costs, congestion condition[60], and climate mitigation pressure[61] may further increase the cost of private car use in the future and thus strengthen the advantages of ride-sharing and potential[62]. Studies that concerned this topic mainly focused on evaluating how far ride-sharing can be adopted and how much benefit can be brought. Santos and Xavier [63] made an economic comparison between private ride and ride-sharing travel. In their simulations, passengers paid, on average, almost 30% less than they would spend on individual journeys compared with shared travels. Based on the real-world ride-sharing trip data provided by DiDi Chuxing company, Yu, et al. [64] evaluated the direct environmental benefits of ride-sharing. The chapter showed that immediate annual energy savings are approximately 26.6 thousand tce, and annual emission reductions of CO2 and NOx are approximately 46.2 thousand tons and 253.7 tons, respectively. Yin and Liu [65] explored the CO2 mitigation potential by ride-sharing. They developed an integrated land-use transport model and applied it to the Paris Legion with some ride-

sharing scenarios in 2030. The result showed that when the vehicle occupancy rises by 50%, the CO2 emission can be reduced by 33%. Liu and Yin [66] estimated the fuel-saving of individual ride-sharing trips based on a trip-specific model and the real-world data offered by DiDi company. For each trip, the fuel saving can be 1.23 L and carbon subsideies are ¥5.38 with the strictest subsidy ceiling.

Choice, the last research topic in ride-sharing, means making a good understanding on participants' behaviors and preferences in various aspects, which is essential to participant matching[67]. Jalali, et al. proposed an assessment method for the potential of ridesharing to reduce emissions by reviewing mobility patterns of approximately 8,900 privately-owned vehicles and extracting their mobility performances [68]. The approach includes five steps: data preprocessing, trip recognition, feature vector creation, similarity measurement, and clustering. With the rising of new technology, there exist some studies that explore the potential of autonomous vehicles. Gurumurthy et al. matched AirSage's cellphone-based trip records and found significant opportunities for DRS(dynamic ride-sharing system)-enabled SAVs(shared autonomous vehicle) [69]. They showed that about 60% of single-person trips could be matched with another. The value increases to 80% if the wait time climbs to 15 to 30 minutes. Then, Pettigrew et al. interviewed 43 key stakeholders in different sectors to explore the threats and opportunities for the potential of autonomous vehicle sharing [70].

Although previous studies have made comprehensive and in-depth discussion on ride-sharing issues when dealing with the and emission reduction potential analyses, there are still some topics need to be explored further:

(1) Complexity and uncertainty of travel time and mobility behavior. Although commuting is the main factor of urban mobility behavior, other activities (such as shopping or dining outside) also account for a considerable part of ride-sharing markets but are rarely considered in previous studies. Besides, ride-sharing will attract participants not only from private vehicle travelers but also from public transit riders. Those factors have a high impact on the assessment of emission reduction. The multiple travel purposes and modes will introduce complexity and uncertainty into the issue.
(2) Various potential matching factors. Previous researches for potential analysis mainly used the similarity of two travels to decide whether they could be matched or not. However, there is a weakness in this method. With different travel modes and habits, different trajectories might vary a lot from each other despite the same OD. For example, if one participant goes to the workplace by private car, and another one goes to the same place by urban railway. Although the ODs are identical, the two participants will go through different trajectories. Using similarity methods may not be matched, but this is a good candidate for adopting ride-sharing in the real world. Furthermore, when the time feasibility issue is considered, the matching issue will be more complex.
(3) Efficient spatial analyses. City-level potential analyses are a complex issue, of which the solution is highly time-consuming. While efficient spatial analyses on the potential and emission reduction potential are significant and essential to policymakers

to make a detailed plan for ride-sharing development. However, few studies gave these results.

2.3 Bus-Sharing System

As an emerging sustainable travel mode in the public transit system, bus-sharing has earned great attention in recent researches. The current study of customized bus systems mainly focuses on the following aspects: Analysis of bus-sharing concept[25]; Bus-sharing network design[71]; Bus-sharing management strategies[72] and Passenger preference on bus-sharing[73].

The concept of bus-sharing service was introduced as subscription bus services several decades ago. Researches mainly focus on subscription bus services, guidelines for the subscription bus network, operational planning, and the pricing strategy, etc[74-79]. In recent years, Qingdao launched the first bus-sharing system in August 2013 in China. Liu et al. introduce the new customized bus concept and provide a systematic examination and analysis of the development and current state of bus-sharing practices in China [25].

Lyu et al. proposed a typical design of bus-sharing lines (Figure 2. 1). Along the bus-sharing bus line, there are two bus stops: (1) Grouped bus stops: multiple bus stops are arranged in a small region to assemble travelers. (2) Intermediate bus stops: few intermediate bus stops are arranged along the bus route to guarantee the efficiency of the bus-sharing line. In this design, the bus-sharing bus can provide more efficient and direct service than a traditional bus.

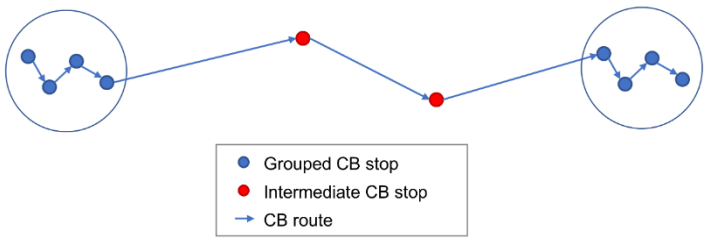


Figure 2. 1 Design of bus line and bus stops for a bus-sharing line[71,80]

Researches on bus-sharing network design mainly focus on designing bus-sharing lines' direction and stop location by using optimization models. These models are usually based on the following steps [25,81]: demand extraction; bus-sharing line and stops deployment; further optimization on scheduling, dispatching and controlling, etc. In the detailed design of bus-sharing systems, researchers have well-developed dozens of models that can consider multiple factors, including dynamic routing and timetabling, congestion detouring, etc.[71,82]

However, the previous methods of developing bus-sharing lines have their limitations:

(1) The basic concept of most bus-sharing network design research is initially to deploy the bus stop and then connect the stops by bus-sharing line. From the city planning aspect, it is believed that to design the bus line and decide how the bus line connects the areas into integration is more critical. Therefore, it is better to plan bus line direction before deploying bus stops, which most current optimization models failed to realize.
(2) Recent research on bus-sharing line planning has been mostly restricted to establishing comprehensive optimization models and can hardly be applied to the city-wide range of areas with massive demand input. The existing works usually focused on hundreds of passengers-level cases which is far lower than the case scale of a city-wide analysis.

Chapter 3

Data Preprocessing

3.1 Data Source

The GPS data is extracted from the 'Konzatsu-Tokei(R)'GPS dataset collected by NTT DOCOMO INC. "Konzatsu-Tokei (R)" Data refers to people flows data collected by individual location data sent from a mobile phone with enabled AUTO-GPS function under users' consent, through the "docomo map navi" service provided by NTT DOCOMO, INC. Those data are processed collectively and statistically in order to conceal the private information. Original location data is GPS data (latitude, longitude) sent in about every a minimum period of 5 minutes and does not include the information to specify individual such as gender or age. There is no any processing or filtering done to the data before the usage.

3.2 Preliminaries

Definition 1 (Raw human trajectory): The raw trajectory collected from an individual is a sequence of 3-tuple: (timestamp; latitude; longitude), which can indicate a person's location according to a captured timestamp. Here, I use t to represent timestamp and x for (latitude; longitude). Hence, the raw human trajectory could be presented as $\left(t_{i,j}, x_{i,j}\right)$. Where i is superscript for the set of the participant and j is for the set of GPS point.

In order to keep the dataset more structured, I use discrete-time presentation with fixed time step Δt, that is:

$$\forall i \in [1, mi], \forall j \in [1, mj), \left|t_{i,j+1} - t_{i,j}\right| = \Delta t \tag{3.1}$$

Definition 2 (Speed sequence): Speed information is important for trajectory travel mode detection. Based on Definition 1, speed sequence could be presented by $\left(t_{i,j}, v_{i,j}\right)$, where,

$$\forall i \in [1, mi], \forall j \in [1, mj), v_{i,j} = L\left(x_{i,j+1} - x_{i,j}\right) / \Delta t\,, \tag{3.2}$$

L is the distance function by the difference of latitude-longitude coordinates.

Definition 3 (Rider and driver mobile stream): For riders in the system, OD coordinates, departure, and arrival time information are the main determinants for participants matching. For drivers, their raw human trajectory data will provide much information revealing the uncertainties of travel habits. Here, I use $Rid=\left(ODx_{i,k}, ODt_{i,k}, ur, \Delta t, TRm_{i,k}\right)$ to determine the urban rider mobile stream. Where k is superscript for set of trajectory segment, ODx is OD coordinates. ODt is departure and arrival time information. ur is urban area. TRm is the travel mode. F is human travel mode detection model.

$$\forall i \in I_{\text{Rid}}, \forall k \in [1, mk], TRm_{i,k} = \left\{ mod \mid mod \in \{walk, bike, vehicle, train\} \wedge mod \in F\left(t_i, v_i\right) \right\} \tag{3.3}$$

I use $Dri=\left(x_{i,k}, t_{i,k}, ur, \Delta t, TRm_{i,k}\right)$ to present the urban diver mobile stream, which satisfies:

$$\forall i \in I_{\text{Dri}}, \forall k \in [1, mk], TRm_{i,k} = vehicle \tag{3.4}$$

3.3 Travel Mode Detection

3.3.1 Extraction of Stay Versus Moving Trajectory Segments

A stay segment within a trajectory corresponds to a group of consecutive points representing a user who is stopped at a location. In this study, stay segments were extracted based on thresholds for distance and time. Neighboring points with distances and times lower than the threshold were classified as stay points. In addition to the conventional stay points, some noise points were also featured in the GPS trajectories; those points had a large offset from the true location. Note that it was possible for the distance from the noise points to the neighboring points to be larger than the threshold as if they corresponded to a stay status. In order to detect these outliers, a Gaussian distribution was utilized to represent the GPS points, with the mean μ and standard deviation σ being calculated as follows:

$$\sigma = \sqrt{\frac{1}{N}\sum_{1}^{N}\left(x_i - \mu\right)^2}, where\ \mu = \frac{1}{N}\sum_{1}^{N}\left(x_i\right) \tag{3.5}$$

where x_i is the i th GPS point and N is the total number of points. If the distances from an inner point to its neighbors were larger than $\left(x_i - \mu\right)/\sigma > 2.6$, this point was

labeled an outlier and removed from the trajectory.

3.3.2 Splitting Moving Segments

The moving segments were split through the following procedure. Firstly, moving segments with speeds <100 m/min were extracted according to the pedestrian travel mode. Then, the remaining segments were split according to the change points, at which the users changed transportation mode. The change points were detected using two features: the velocity change rate (VCR) and points on train lines. The VCR was defined as the average speed of the current segment compared with the current observed speed, such that

$$VCR = \left| S.Speed_{average} - P.Speed \right| \Big/ S.Speed_{average} \tag{3.6}$$

where $S.Speed_{average}$ is the average speed of the segment and $P.Speed$ is the speed of one point in the segment. The use of the VCR index was proven to be more effective and more stable than the acceleration index in the case of sparse GPS data.

3.3.3 Traffic Mode Classification

Once the split traffic segments were obtained, the traffic mode was classified using the random forest method, which is an extension of the traditional decision tree and is a supervised learning method that has been shown to perform well for transportation extraction. In the random forest method, multiple trees are constructed to classify a new object, where each tree provides a classification as a vote; then, the forest classifies the object with the most votes.

In this study, the input features used in the random forest method were as follows: the total distance (in meters) and duration (in minutes) of the segment; six-speed features (i.e., minimum speed, maximum speed, average speed, overall average speed, maximum acceleration, and velocity change rate); and the percentage of the points that coincide with a train line or the road network, as extracted by a 50 or 100 m buffer, respectively. Five traffic mode labels were output by the random forest model: train and subway, vehicle, bicycle, walk and stay. Test datasets of the five traffic modes were utilized in the experiment to determine the model performance for the different traffic modes.

3.3.4 Segment Merging

This step reduces the trivial and uncertain segments by applying specific rules. First, consecutive segments with the same traffic mode were extracted. Then, segments for

which the traffic modes were uncertain were merged into neighboring segments. For other complex cases, classifiers with training data were utilized to merge the segments.

Chapter 4

Bicycle-Sharing System: Market-Oriented Sub-Area Division

Bicycle-sharing is an up-to-date travel mode and has been introduced to many cities worldwide [83]. As a popular form of urban transport, public bicycles have the following advantages: (1) As they produce no air or noise pollution, bicycle-sharing systems provide residents and tourists with a convenient, environmentally friendly way to travel and enhance the city's sustainable competitiveness [24]. (2) Meanwhile, cycling also helps to enhance public health and reduce morbidity levels associated with urban disease [29]. (3) Finally, compared with other public transportation modes, public bicycles benefit from small volume, flexible operation, good accessibility, and lower investment cost [84]. Bicycles can be integrated with other modes of public transit, and cover short trips of 1–5 km, thereby helping to improve utilization of the road network [85], ease traffic congestion, improve the efficiency of urban traffic [55], and reducing the transportation emission [86]. Evaluating the potential emissions reduction of adopting the bicycle-sharing system for cities is significant to the local government and bicycle-sharing companies. This could encourage and promote the transportation sector to make sustainable policies about bicycle-sharing and guide companies to formulate the development strategy [87]. However, city-level evaluation models for demand potential usually are not solvable due to the large model scale. We need to divide the city into several sub-area for separately modeling to reduce the model size and make the model solvable for better system design, monitoring & prediction, and rebalancing, as also similar for other socioeconomic division analysis. The current management modes of the existing bicycle-sharing systems are usually administrative division based, but the administrative division is not effective in representing actual mobility boundaries (Figure 4. 1a). Therefore, we need a market-oriented sub-area division.

In this chapter, a market-oriented method of bicycle-sharing system planning is proposed. The proposed method involves three steps: identifying potential demand, dividing sub-service areas, and analyzing potential use patterns and potential emission reduction. The rest of this chapter is organized as follows — Section 4.1 defines the problem that is going to be solved. Section 4.2 presents the methodology of this work, including possible bicycle-sharing behavior detection, sub-area division, bicycle-sharing pattern indicators, hotspot identification, and the emission reduction model; Section 4.3 presents the case study in Tokyo and proposes the contributions and conclusions of this study.

4.1 Market-Oriented Sub-Area Division

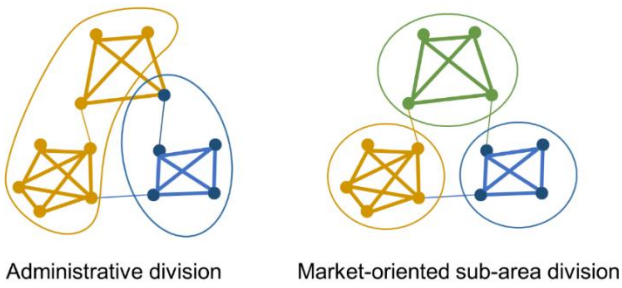
The division of sub-service areas is essential in the market-oriented bicycle-sharing system planning. As is illustrated in Figure 4. 1b, the benefits of sub-service area division can be concluded in three aspects:

- Operation: by dividing the coverage areas into sub-service areas, the potential market share of each subarea can be estimated. Based on the demand and travel patterns, bicycle-sharing operators can estimate the number of bicycles allocated in each area to meet the demand without resource waste.
- Management: for subareas with different demand patterns, corresponding policies can be made, including dynamic pricing, restricting parking areas, etc. The division of subareas can also give suggestions to the infrastructure construction of the bicycle system.
- Rebalance: from the volume of cross-area demand, the penalty of cross-area traveling can be decided without largely influence the whole system, which can primarily reduce the effort of rebalancing and ensure the usage in some specific areas. From the traveling patterns in each subarea, bicycle-sharing operators can decide their rebalancing strategy.

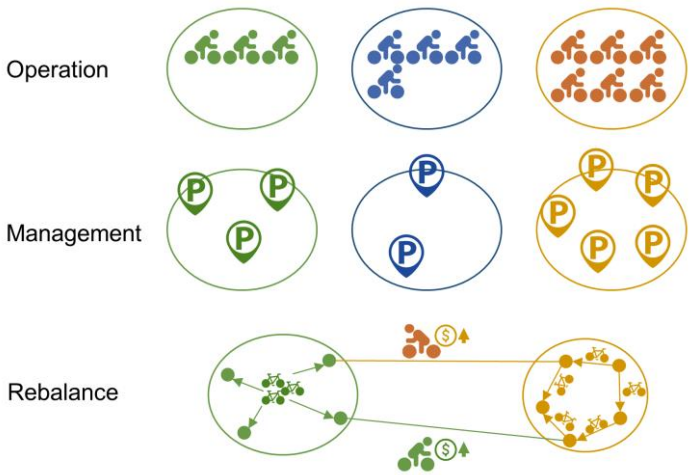
According to the demand of market-oriented sub-area division, making a proper division of coverage area should obey the rules below:

- The subareas should be clusters of regions and continuous in spatial.
- The division of coverage area should follow the bicycle-sharing use patterns by using human mobility data, with a relatively larger volume of intro-area than cross-area travel.
- The size of subareas should be suitable and usable for bicycle-sharing system management without being too large to invalidate the benefit of sub-area division or being too small to increase the difficulty of management.

In general, constructing a network based on human mobility data and applying the network community detection method on the network will be a perfect solution for the market-oriented sub-area division.



(a)



(b)

Figure 4. 2 Problem description of the sub-area division

(a) illustrates the difference between administrative division and market-oriented sub-area division. (b) illustrates the benefits of a market-oriented sub-area division.

4.2 Methodology

4.2.1 Framework

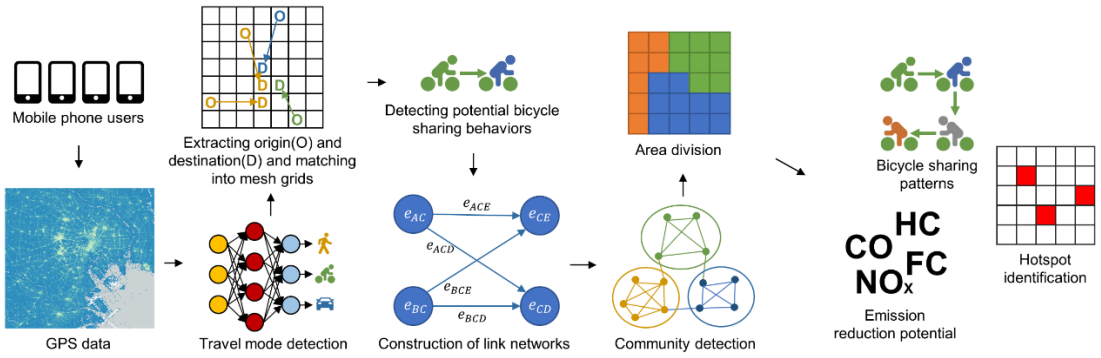


Figure 4. 3 Framework of methods introduced for market-oriented sub-area division

The framework of the methodology is shown in Figure 4. 2. In this chapter, mobile phone GPS data is used to detect the potential bicycle-sharing demand. The sub-service area is divided based on the spatial interaction of the potential bicycle-sharing use demand. In brief, the methodology is as follows: firstly, from the mobile phone data, travel mode is detected by the data mining method proposed in our previous work [88]. The origin(O) and destination(D) information of walking, cycling and short-distance vehicle trips are matched into mesh grids and aggregated by grid cells. Secondly, a link network is generated based on the potential bicycle-sharing behaviors detected in the OD trips. Community detection is then applied to detects the link community, which naturally suggests the sub-divisions of the coverage area. Finally, analysis from three

aspects is conducted based on the sub-areas division, including the sharing behavior pattern, the hotspot, and the potential of emission in sub-areas.

4.2.2 Detecting Potential Bicycle-Sharing Behaviors

Considering the scenario as follows: if a traveler rides a bicycle to his/her destination and parks the bicycle. Soon after that, another traveler starts his/her trip from the exact location using the same bicycle. This scenario can be regarded as one perfect bicycle-sharing procedure.

Based on this scenario, given short distance OD trips (1-3km in this study) inferred from the mobile phone GPS data, I propose a method to identify the potential bicycle-sharing trips. I first define the spatial and temporal resolution for which bicycle-sharing trips can be matched. A mesh grid composed of equal-width grid cells is generated on the spatial aspect, and the OD trips are aggregated by grid cells over the study area. Here, the spatial resolution is set as 500m*500m squares; On the temporal aspect, I assume that trips occur uniformly throughout each hour and compute the number of trips within each time window Δ, for which the timestamp $t_n = t_{n-1} + \Delta$. Δ is the maximum allowable gap between the arrival time of the first trip and the departure time of the second trip to consider as a desired bicycle-sharing behavior. A larger Δ will enable more trips to be matched as bicycle-sharing. However, it implies a lower efficiency of the sharing system and will generate more bicycle parking demand. Here, considering the data volume, Δ is set as 30 minutes.

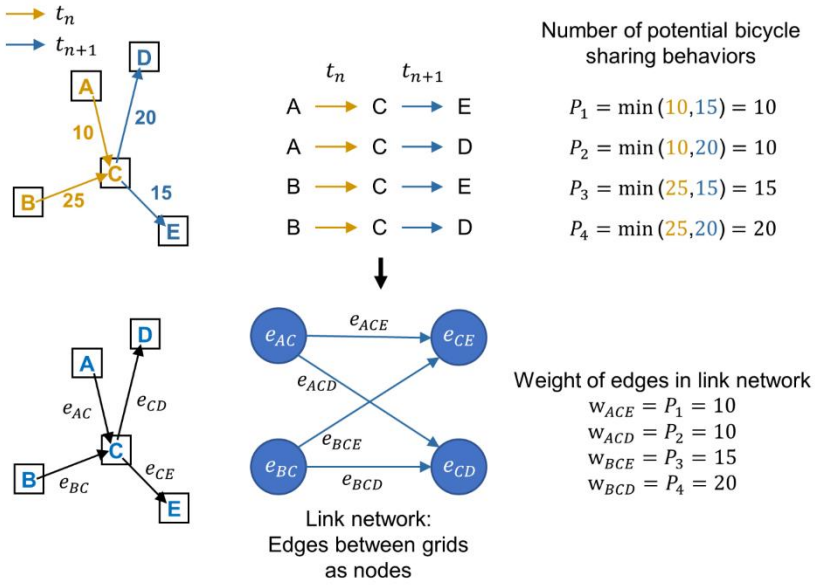


Figure 4. 4 Method of potential bicycle-sharing estimation and link network construction

Consider in timestamp t_n, there are q_{AB} trips from location A to B, and in the next timestamp t_{n+1}, there are q_{BC} trips from location B to C. Users of the bicycle will be changed in location B, and, thus, is the place where the bicycle-sharing behavior is

generated. The maximum of potential bicycle-sharing trips from the edge e_{AB} to edge e_{BC} is $P_{ABC} = min\,(q_{AB}, q_{BC})$. Suppose I consider the edges between the mesh grids (OD) as nodes in a network. In that case, naturally, a link network can be constructed, with the number of potential bicycle-sharing trips between nodes as the weight of edges in this network. This link network can be denoted as $G(V, E, W)$, where $V = \{e_{ij}; i \neq j\}$ is the set of nodes, $E = \{e_{ijk}; i \neq j, j \neq k\}$ is the set of edges, and $W = \{w_{ijk}; i \neq j, j \neq k\}$ is the edge-weights. An example of this method is illustrated in Figure 4. 3.

4.2.3 Market-Oriented Division of Sub Service Areas

In the network analysis, a community is a group of nodes having a higher probability of being connected than to nodes of other groups[89]. Placing each link as a node in the network allows us to reveal hierarchical and overlapping relationships simultaneously. With every link belonging to a single community, the nodes they connect to can belong to several communities and unveil the network's overlapping structure of the network [90].

The principle of defining sub-service areas for the layout of the bicycle-sharing system is to find the sub-area with the strong connection inside each area and the relatively weak connection between different areas, which makes the community detection a perfect method for this purpose. Several methods have been developed to find community structure. For the mega-scale link network of bicycle-sharing, the two measures of importance are efficiency and interpretability. A higher computational efficiency shows that the adopted method is practical for mega-scale networks. A more interpretable outcome proves the method is appropriate for the planning practice and helpful to solve practical issues. Therefore, the fast unfolding algorithm is adopted in this study [91].

The inspiration for the fast unfolding algorithm is the optimization of modularity. Modularity is a benefit function that measures edges inside communities compared to edges between communities [92]. It is defined as follows:

$$Modularty = \frac{1}{2m}\sum_{i,j}\left[w_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \tag{4.1}$$

where w_{ij} represents the edge-weight between nodes v_i, v_j; k_i is the sum of edge-weights attached to the node v_i; c_i denotes the community to which node v_i is assigned; $2m$ is the sum of edge weights in the network; $\delta(u, v)$ is a simple delta function defined as follows:

$$\delta(u, v) = \begin{cases} 1 & u = v \\ 0 & u \neq v \end{cases} \tag{4.2}$$

In order to maximize the modularity efficiently, the fast unfolding algorithm has two steps that are repeated iteratively.

Step 1: modularity optimization. All the nodes in the network are assigned to their own communities. For each node v_i, the change in modularity is calculated by removing v_i from its original community c_i and moving to other neighboring communities. Once modularity is calculated for all communities to which node v_i is connected, v_i is assigned to the community resulted in the highest modularity increase. If no increase is possible, node v_i remains in its original community. This process is applied repeatedly and sequentially to all nodes until the local maximum of modularity is hit.

Step 2: community aggregation. A new network is built where nodes are the communities from the previous step. The new network in each iteration has a significant actual sense to stand for a hierarchical level in the spatial structure. The first step can be re-applied to the new network.

The iteration ends when global modularity can no longer be improved.

By applying community detection on a link network, in the result generated, every link belongs to a link community. When the links from different communities connecting to the same mesh grid, this grid will belong to multiple communities. Using $e_{ij}(k)$ to denote that the edge from grid i to grid j belongs to link community k, and $w_{ij}(k)$ to denote the weight of this link, by adding the weight of links from different communities, the impact of link community k for grid j can be denoted as follows:

$$I_{jk} = \frac{\sum_i w_{ij}(k)}{\sum_i \sum_k w_{ij}(k)} \tag{4.3}$$

To divide sub-areas, grids can be regarded as belonging to the community with the most significant impact.

Because most of the grids are impacted by multiple communities, the link communities also imply the local core-periphery structure for the network. The grids that are most impacted by a single link community as core grids and the grids are simultaneously impacted by several link communities as peripheral grids. The core and peripheral grids can be distinguished by setting a threshold t, and grid j is a core grid if $\max_k\{I_{jk}\} \geq t$, otherwise it is a peripheral grid. Here, I set the threshold $t = 0.8$, which denotes that the impact from the community it belongs to is over 80% and is dominated by the community.

4.2.4 Indicators for Bicycle-Sharing Patterns

In order to describe the different bicycle-sharing patterns in sub-areas, several network

community indicators are proposed here.

- The proportion of cross-area bicycle-sharing trips:

By obtaining the number of intro-trips inside the subareas Q_{inside}, the trips with the origin outside the area and destination inside the area Q_{in}, and the trips with the origin inside the area and destination outside the area Q_{out}, the total number of trips that happened in the area is $Q_{total} = Q_{inside} + Q_{in} + Q_{out}$. The proportion of cross-area bicycle-sharing trips in area k can be calculated as follows:

$$P_k = \frac{Q_{in}+Q_{out}}{Q_{total}} \tag{4.4}$$

The lower of the P_k indicates that the area has less crossing-area bicycle-sharing demand, which will imply the demand for rebalancing bicycles between the communities.

- Graph density:

In network science, graph density is an index to measure the density of edges[93]. For each sub-areas, the graph density provides an indicator of how the grids are densely connected. It is defined as Eq.5:

$$D_k = \frac{m}{n(n-1)} \tag{4.5}$$

where n is the number of grids in sub-area k, m is the number of edges in sub-area k. The higher of the D_k Indicates that the grids in this sub-area are more densely connected by bicycle-sharing, and bicycles can have more possibilities to enter a circulating flow travel pattern. Therefore, graph density represents the demand for bicycle rebalancing inside the community, which is more desirable for the bicycle-sharing system.

- Complementary cumulative distribution function (CCDF):

CCDF is used to find the probability of a variable taking a value greater than x. Comparing the CCDF curves of node strength (the number of potential bicycle-sharing trips normalized by dividing them with the maximum in each sub-areas), the one decrease quickly in the small value of node strength indicates that potential bicycle-sharing trips concentrate on few locations. These locations play an important role in attracting bicycle-sharing behavior.

4.2.5 Hotspot Identification

Identifying the spatial hotspot clusters of potential bicycle-sharing demand in each sub-areas is significantly helpful for system layout design. In order to promote bicycle-sharing service in sub-areas, the hotspot area implies the locations where should be

initially considered. Among the hotspot identification methods, the local Moran's I index is the most popularly used [94]. The local Moran's I index examines the individual locations, enabling hotspots to be identified based on a comparison with the neighboring samples.

Applying the local Moran's I index on the bicycle-sharing scenario, the number of potential bicycle-sharing trips P_{ABC} from the edge e_{AB} to edge e_{BC} implies that there will be P_{ABC} potential hand over of bicycle in location B. The high positive local Moran's I index indicates that the number of potential bicycle-sharing demand in the grid has similarly high or low values as its neighbors. The local Moran's I index can identify two types of spatial clusters: high-high clusters (high values in a high-value neighborhood) and low-low clusters (low values in a low-value neighborhood); And two types of outliers: spatial outliers include high–low (a high-value in a low-value neighborhood) and low–high (a low value in a high-value neighborhood). In bicycle-sharing system layout design, the high-high clusters and high–low clusters are the "regional hotspots" and "individual hotspots" with significantly higher sharing demand but with different patterns.

4.2.6 Emission Reduction Potential Model

Here, COPERT (Computer Programme to calculate the Emissions from Road Transport) model is adapted to calculate fuel consumption (FC) and emissions of carbon monoxide (CO), nitrogen oxides (NOx), and hydrocarbon (HC). COPERT is a widely used emission model developed by the European Environment Agency (EPA). Based on distinguishing vehicle categories, fuel types, road categories, and other parameters, COPERT model determines the emissions of different pollutants and FC by adopting regression analysis for speeds and traveling distance of vehicles [95,96]. The detail of the emission model is shown in Supplementary materials A.

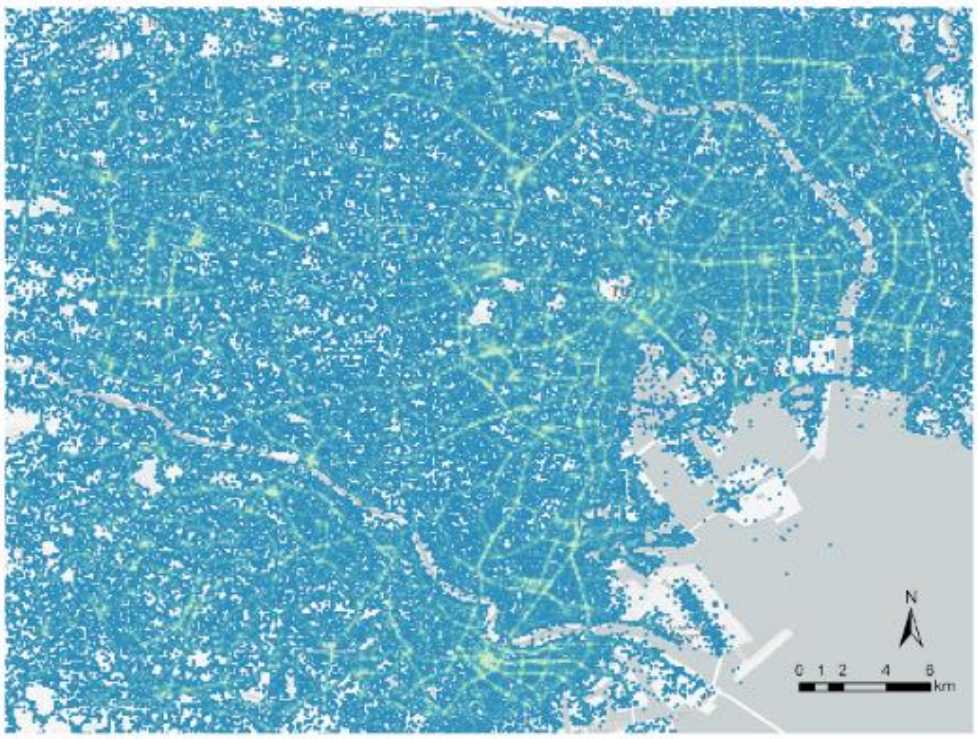
Adding up the emissions generated by all the short car trips in 1-3km connecting to the grids/sub-areas, the potential emission reduction can be calculated. The normalized emissions in grids/sub-areas normalized by dividing them with the sum of all grids/sub-areas can be calculated to be comparable for each type of emissions.

4.3 Results and Discussion

4.3.1 Study Case

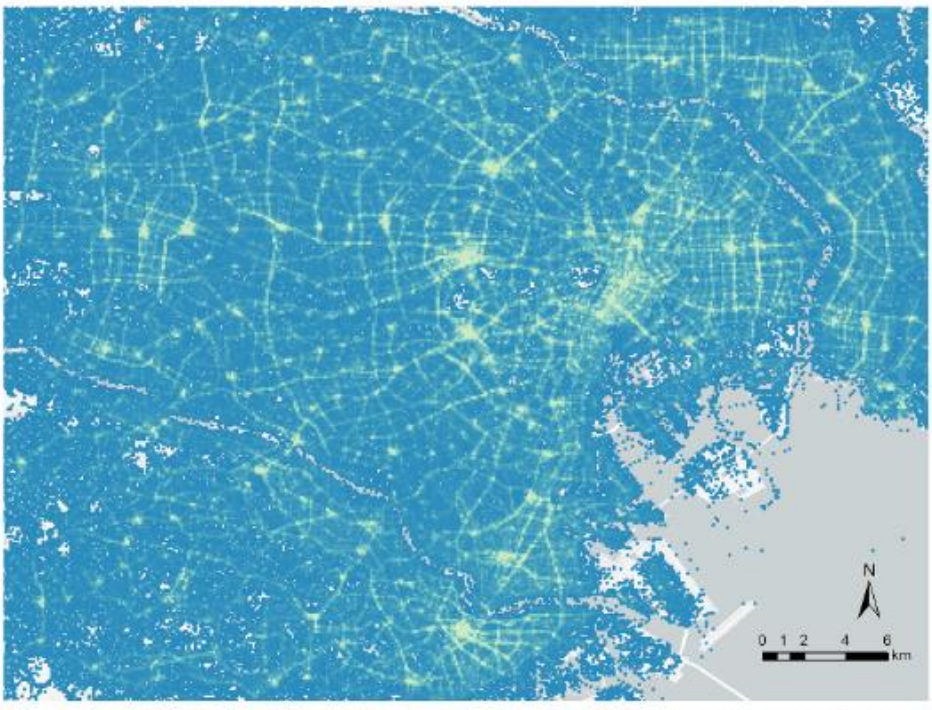
Bicycles usually cover short trips of 1–3 km. Therefore, after identifying the travel modes, this study only extracts the short-distance walk, bicycle, and car travel in 1-3 km as our research subject. Figure 4. 4 shows the spatial distribution of trajectory data in 1 month. The data set comprises 34,181,231 trajectories, including 19,821,894

walking trips, 11,994,279 bicycle trips, and 2,365,058 car trips.



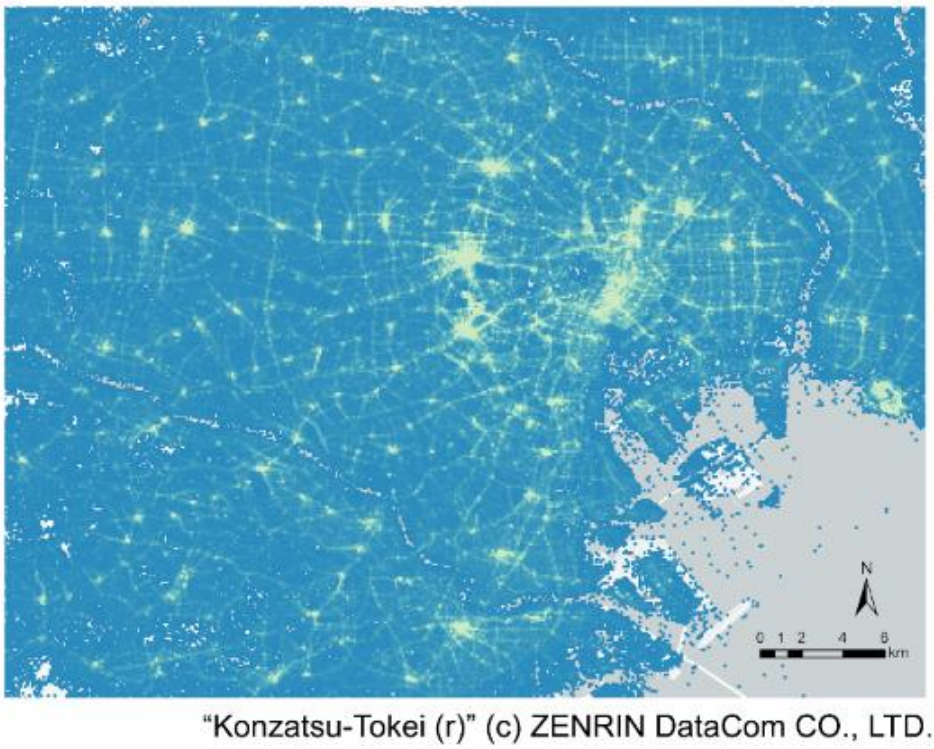
"Konzatsu-Tokei (r)" (c) ZENRIN DataCom CO., LTD.

(a)



"Konzatsu-Tokei (r)" (c) ZENRIN DataCom CO., LTD.

(b)



(c)

Figure 4. 5 Spatial distribution of trajectory dataset in 1 month.

(a) Car trajectories, (b) Bicycle trajectories, (c) Walk trajectories

4.3.2 Result of Market-Oriented Sub-Area Division

By applying the methods above on the data set, 47,666,836 potential bicycle-sharing trips are matched in total. Aggregating all year data, a link network with 587,606 nodes and 9,425,875 edges have constructed. Applying the fast unfolding algorithm on the link network, the algorithm iterates two times. The modularity index increases from 0.836 to 0.842 and hits its maximum. The high value of modularity indicates a fairly good community structure discovered by the algorithm in the link network. The result of link communities of each iteration is shown in Supplementary materials B.

Figure 4. 5 shows the result of community detection on the link graph. Each link community contains 27,948 links on average. By only visualizing the links with the top 5% weight in Figure 4. 5, it can be seen that, in suburban areas, the bicycle-sharing pattern is mainly organized as a radiating shape, connecting adjacent areas to local centers. In urban districts, links are organized with high-density connecting mesh grids distributed around the whole area.

Based on the link communities, the area of Tokyo is divided into 21 sub-areas, with each sub-area contains 216 mesh grids(54 km^2) on average(Figure 4. 6). In the sub-area division, although the community detection method is conducted without inputting any spatial information, the resulting link communities and sub-area divisions can merge the adjacent grids and intertwine with the geographical space. Comparing the sub-area

division with the administration division, most of the area boundaries are different, indicating that the potential bicycle-sharing demand is not organized as administrative units. This result proves that when designing a bicycle-sharing system, simply divide the area based on the administrative division would be inappropriate.

By distinguishing the core and peripheral grids in Figure 4. 6, I can identify the locations in sub-areas that have frequent connections with other areas. These are the locations where the bicycles may be circulated into other sub-areas and causing the demand for rebalancing. In this division of sub-areas, the periphery grids are mostly the mesh grids among the boundary of sub-areas, which can ensure that in most parts of the center of these sub-areas, the balance of the bicycle can be guaranteed.

The sub-area segmentation is influenced mainly by natural geographical barriers, such as rivers, mountains. For some of the sub-areas, subarea 6, 7, 14, 16, and 19, their boundaries are consistent with the administrative division. These areas are primarily in suburban areas. Subarea 1, 2, 3, 4, 5, and 8 subareas are very different from the administrative division in the city center. Small isolated areas with a high density of short-distance travel are also detected, including subarea 17(the Tokyo Disney resort), 20(Odaiba: an artificial island), and 21(Haneda airport). This result indicates that divide the bicycle-sharing market by administration division will be inconsistent with the bicycle-sharing pattern of users.

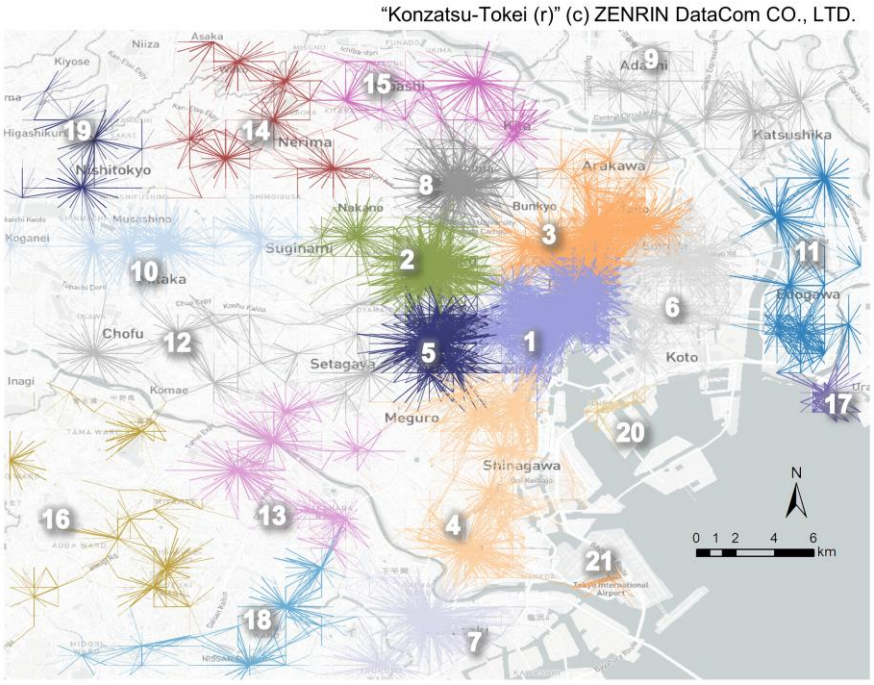


Figure 4. 6 Result of link communities

(only links with top 5% weight are visualized)

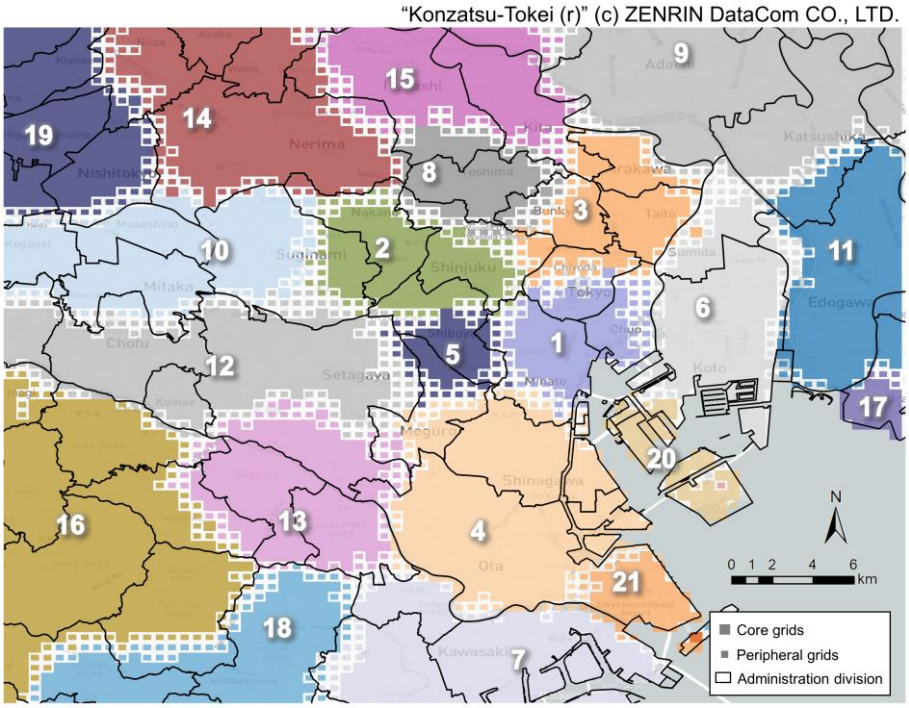


Figure 4. 7 Division of sub-areas based on the link communities.
The different colors represent different communities and sub-areas, which are named in descending order according to the number of the total trips Q_total in the area.

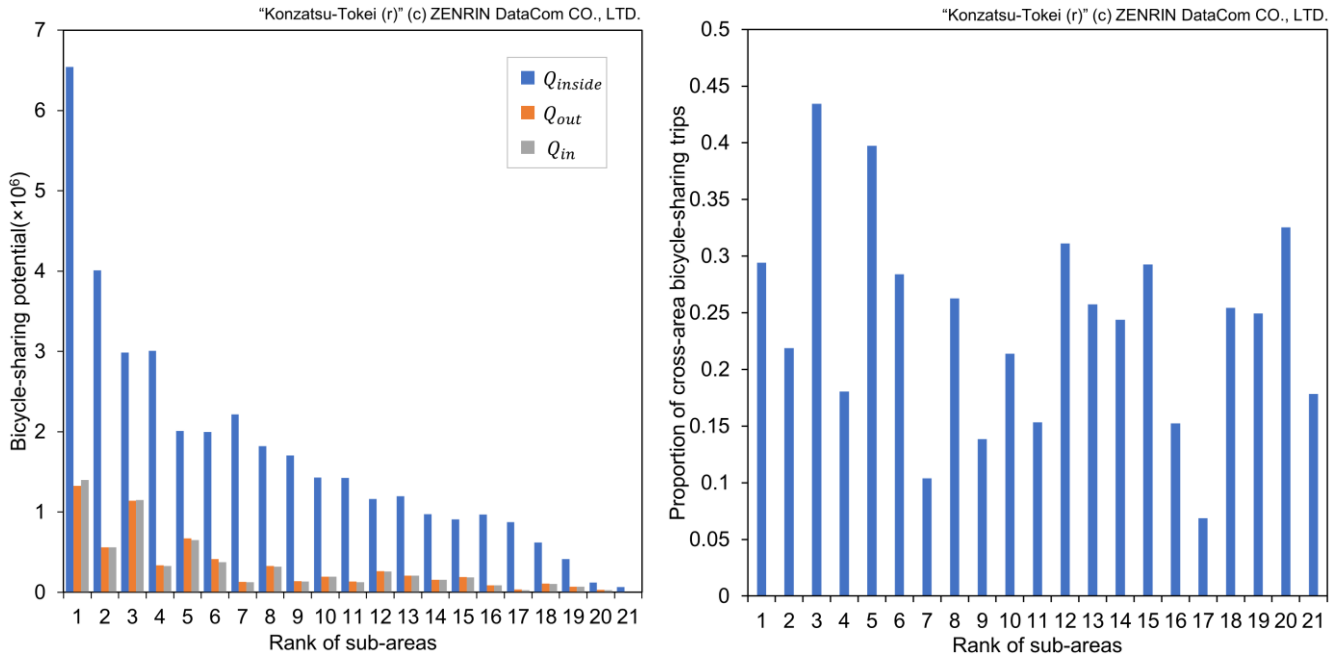
4.3.3 Analysis of Bicycle-Sharing Pattern

Figure 4. 7 shows the indicators of the sub-areas. From Figure 4. 7 (a) and Figure 4. 7 (b), I can see that for some of the sub-areas, although they may have a similar amount of trips inside, they may have different levels of cross-area trip demand. In practice, to promote bicycle-sharing services, these areas should be managed with different policies. For example, in sub-area 7, 9, 11, 16, and 17, they have a lower proportion of cross-area bicycle-sharing demand (under 15%), setting the restriction policy for cross-area bicycle-sharing trips can reduce the effort of bicycle rebalancing and not affecting too much bicycle-sharing demand. But in sub-area 1, 3, 5, 6, 12, 15, and 20, they have a relatively higher proportion of cross-area bicycle-sharing demand (around or over 30%). The same policy may cause inconvenience for users. Based on the proportion of cross-area trips, a dynamic pricing dispatching fee policy for cross-area trips will also be an excellent solution for rebalancing optimization.

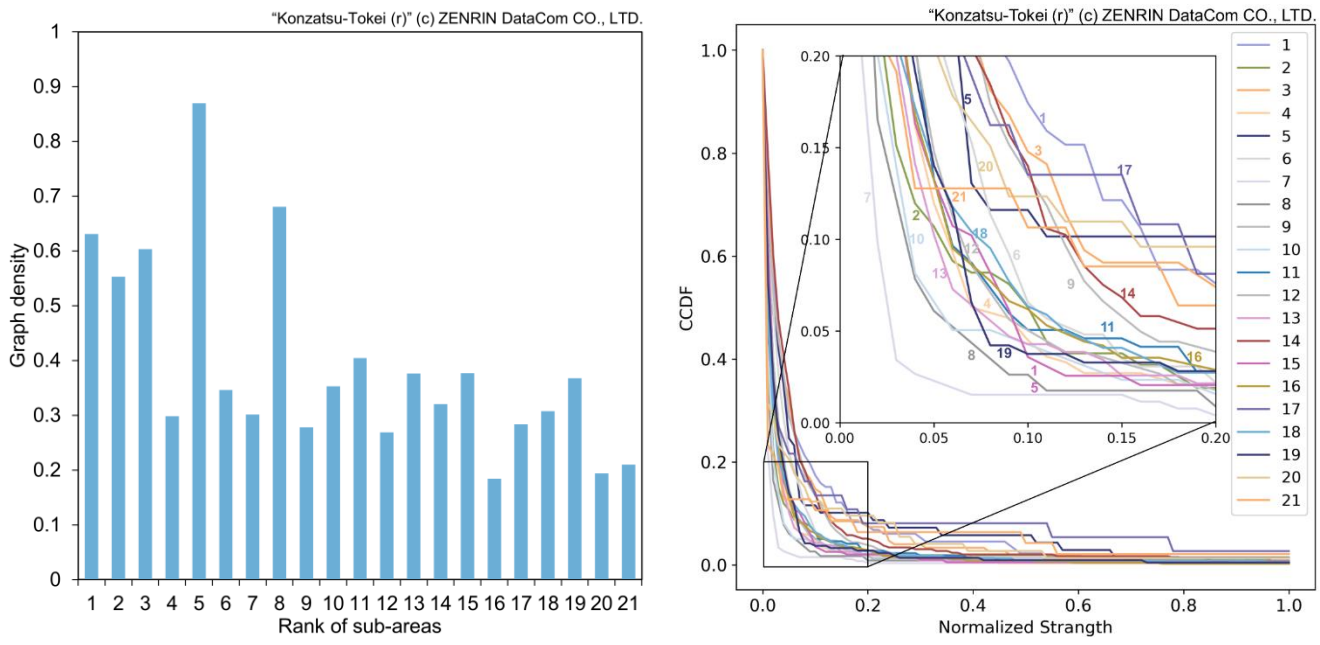
From Figure 4. 7 (c) and Figure 4. 7 (d), the sub-areas 1, 2, 3, 5, 8 are the areas with higher graph density and, spatially, are located in the city center, but the CCDF curves of areas 1, 3, 5 are reducing much slower than area 2 and 8. This result indicates that although these areas are all in the city center, they may share a different traveling pattern of bicycle-sharing, with trips in areas 2 and 8 centralized and trips in areas 1, 3, and 5 more decentralized. In suburban areas, the graph density is among the same level of 0.2-0.3, but the CCDF curves also show the different centralization of bicycle-sharing

trips. In general, trips in suburban areas are more centralized than in urban areas.

From Figure 4. 8, the temporal distribution of bicycle-sharing trips is similar in each sub-areas. The hour distribution shows a characteristic of double peaks in a day, from 7:00 to 9:00 as the morning peak and 17:00 to 19:00 as the evening peak. Besides, more trips are generated during the evening peak, especially for the sub-areas in the city center. In suburban areas, the evening peaks are also with more trips, but almost at the same level as the morning peak. This reason can be in two aspects: 1. In the morning peak, traveling demand is mostly for commuting; But in the evening peak, besides commuting, people will have more diverse activities and generating more travel demand. 2. In urban areas, with denser public transportation networks, higher accessibility, and a better-built environment. People will have more opportunities and generate diverse activities. Bicycles can be used for many purposes, especially in the evening time. In contrast, in suburban areas, the function of bicycle transportation mode is mainly to serve the commuting demand, which makes the morning peak at the same level as the evening peak. In the weekday distribution, the trips on Friday will reach the week's peak and then drop on weekends. In the month distribution, more trips will be generated in spring but fewer in winter. The reason is that bicycle is a travel mode that will be primarily affected by the weather and temperature[29].



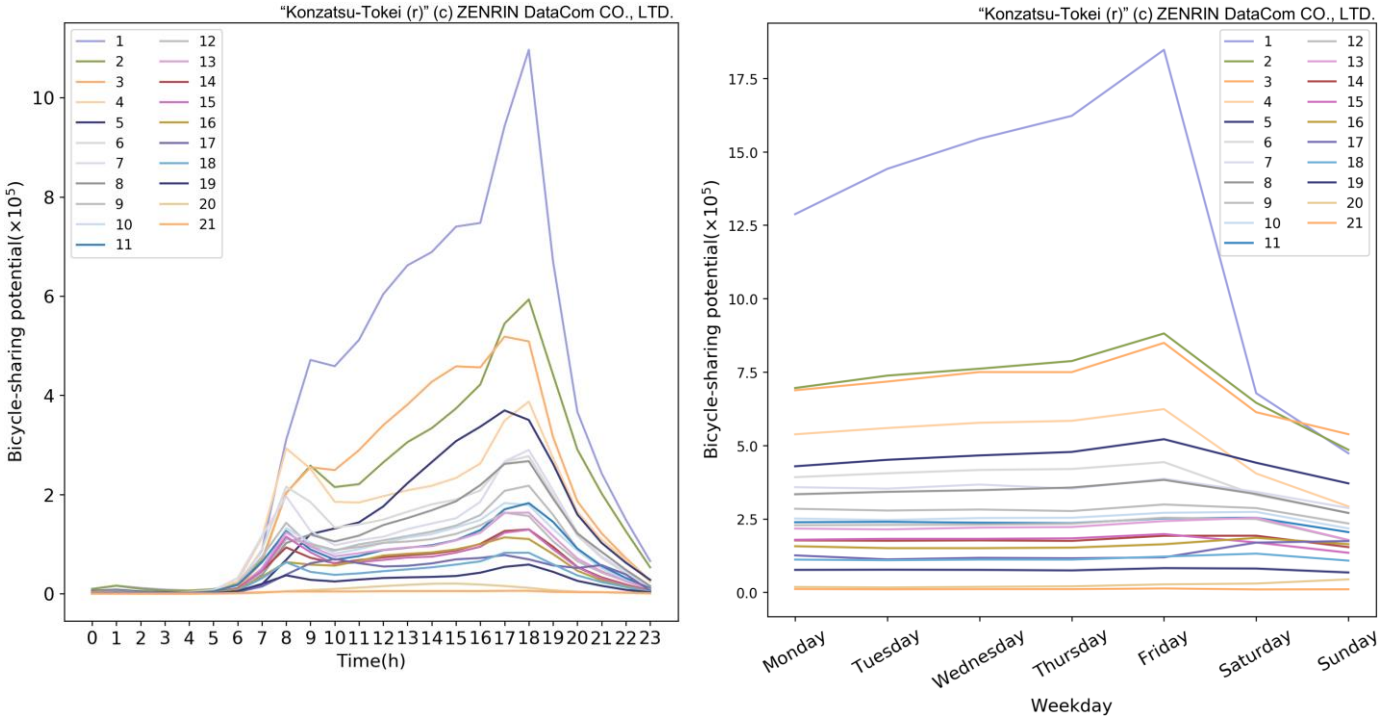
(a) (b)



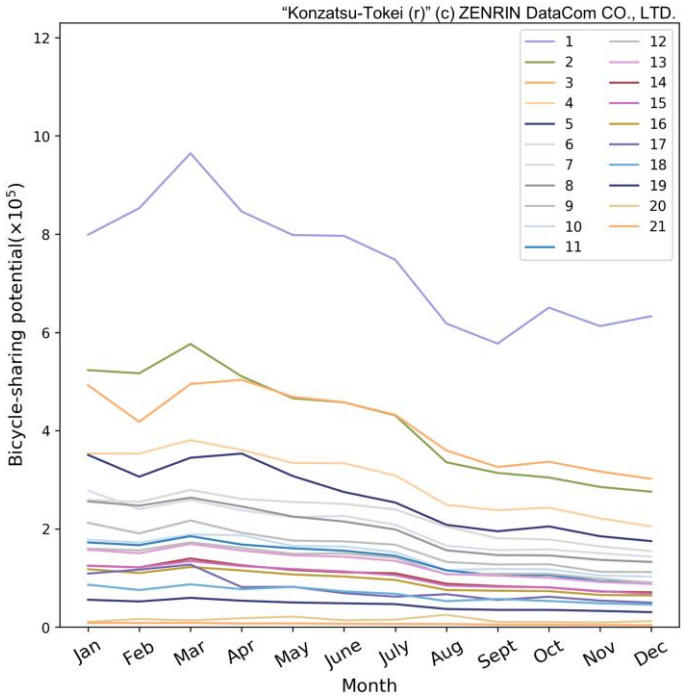
(c) (d)

Figure 4. 8 Indicators for the bicycle-sharing pattern in sub-areas.

(a) The number of intro-trips inside the subareas Q_{inside}, the trips with the origin outside the area and destination inside the area Q_{in}, and the trips with the origin inside the area and destination outside the area Q_{out}. (b) The proportion of cross-area bicycle-sharing trips P_k. (c) Graph density of sub-areas D_k. (d) CCDF of sub-areas.



(a) Hour (b) Weekday



(c) Month

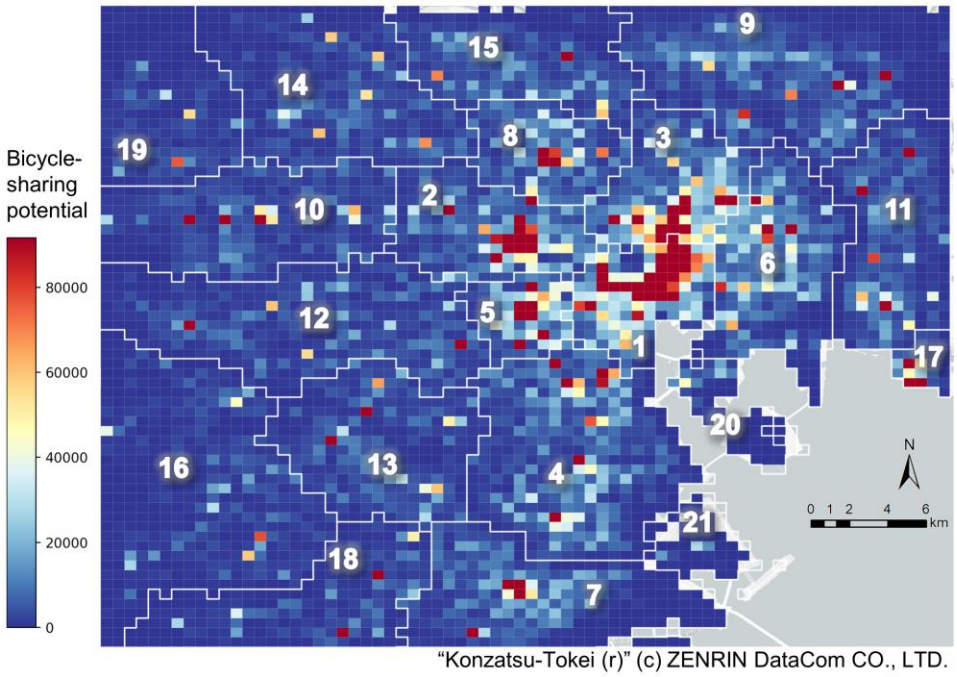
Figure 4. 9 Temporal distribution of bicycle-sharing potential

4.3.4 Result of Hotspot Identification

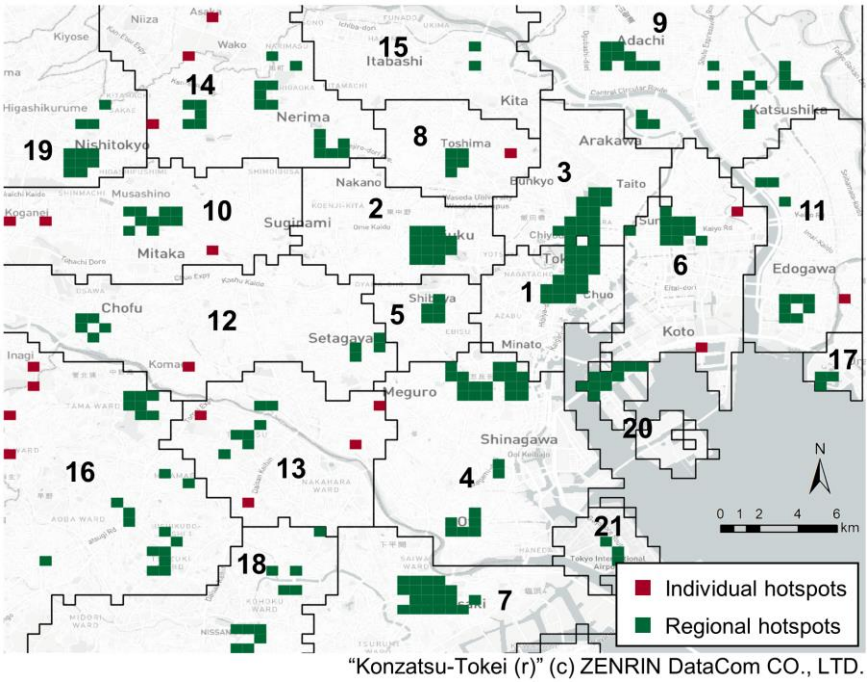
Hotspot identification can identify the critical locations when considering promoting an Internet-based bicycle-sharing service without docking stations. The priority of bicycle management and infrastructure construction should be given in hotspot areas and receive better outcomes.

Figure 4. 9 (a) shows the spatial distribution of the location generating the potential bicycle-sharing behavior, and Figure 4. 9 (b) is the hotspot identified based on it. This study identified two types of hotspots: regional hotspots and individual hotspots. Bicycle-sharing patterns in these two types of hotspots are different, and thus, management in these areas should also be different. Regional hotspots are regions that consist of multiple grids with high bicycle-sharing potential. Sharing behaviors may frequently generate between the grids, and infrastructures like bicycle parking lots should be widely distributed over these regions. The rebalancing of bicycles in these areas will need fewer efforts since the bicycles can circulate by sharing behavior.

Nevertheless, for Individual hotspots, it indicates a high-potential grid in a low-potential neighborhood, which usually happened as an essential spot, such as a point of interest(POI) or a transportation station, attracting bicycle-sharing demands. In these areas, sharing behavior will be more concentrated, and more bicycle parking lots should be placed around the spot. Bicycles in the neighborhood will gradually gather in the individual hotspot and causing an imbalance between the central area and surrounding areas, which may need more rebalancing.



(a) Spatial distribution of bicycle-sharing



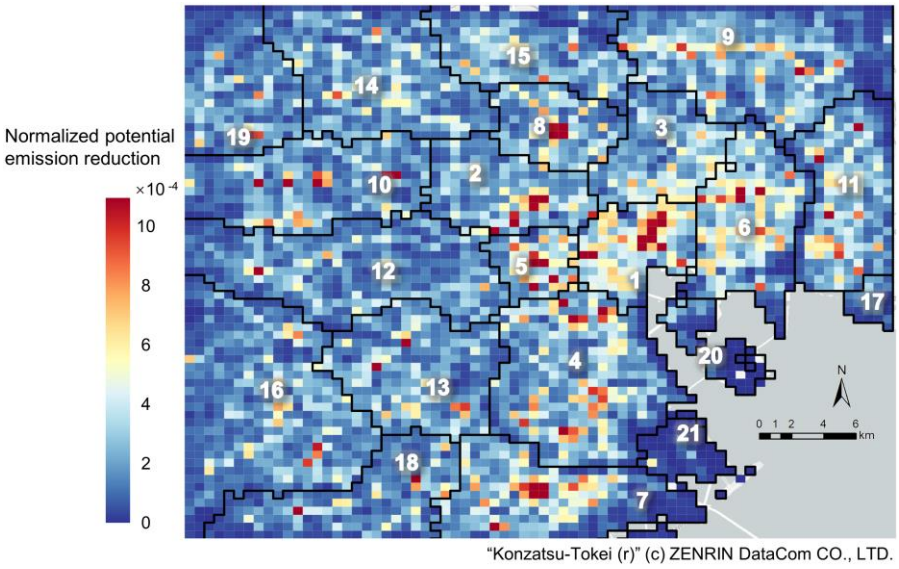
(b) Hotspot identified

Figure 4. 10 Result of hotspot identification

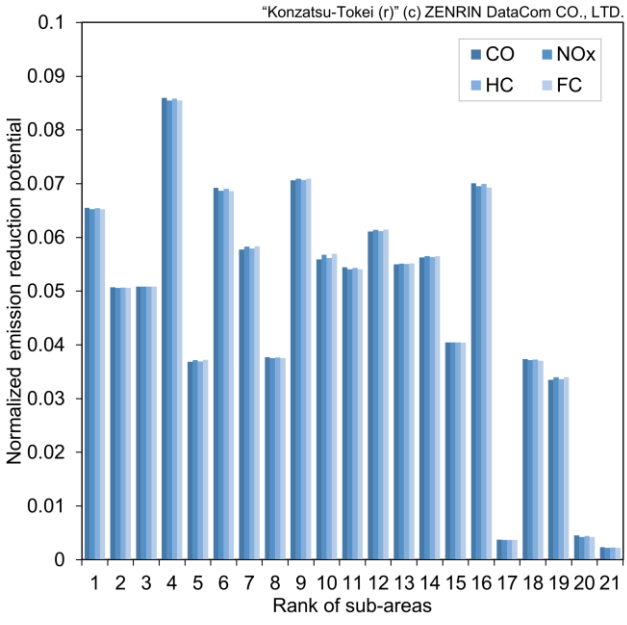
The spatial patterns of bicycle-sharing are different between urban and suburban areas. Similarly, the shape of hotspots in these areas is also different. In the city center, hotspots are all regional hotspots forming consecutive regions with high potential demand. Most of these areas are the central urban districts or popular tourist attractions. Hotspots in the suburban area are more scattered, with relatively more minor regional hotspots and more individual hotspots. Moreover, many of these spots are the metro

stations or core areas of suburban satellite cities. From this result, bicycle management should be more centralized and focus on some unique spots in suburban areas and be more decentralized and widely distributed among the regional hotspots.

4.3.5 Analysis of Emission Reduction Potential



(a)



(b)

Figure 4. 11 Estimation of potential emission reduction in sub-areas

(a) Spatial distribution of the normalized potential emission reduction in mesh grid level. (b) Normalized potential emission reduction in sub-areas

After normalization, the four types of potential emission reduction are in the value with the same level (see Supplementary materials C). Thus, the average value of 4 types of

emission can be an index for emission reduction. Figure 4. 10 (a) shows the spatial distribution at the mesh grid level. This result also implies how many short-distance trips connecting the area depend on the car model. Significant emissions of short-distance car trips are handled by the central part of the urban area. Most emissions occur within the commercial districts, such as Ginza, Shinjuku, Shibuya, and Ikebukuro. In suburban areas, there are also some clustering regions with high emission reduction potential sparsely distributed.

Figure 4. 10 (b) shows the four types of normalized potential emissions in sub-areas. In suburban areas, although with a lower number of potential bicycle-sharing trips when compared with urban areas, they still have a high potential for emission reduction. For most of the existing bicycle-sharing programs, the marketing areas are restricted to urban centers and neglecting suburban areas. However, as shown in the result, with proper management and promotion of bicycle-sharing services, sub-urban areas can also have high performance in emission reduction. The result also shows that small isolated areas with a high density of short-distance travel, subarea 17, 20, and 21 have less emission reduction potential, indicating that the short trips in these areas depend less on cars, which is reasonable considering the land use of these subareas.

4.4Conclusions

City-level evaluation models for demand potential usually are not solvable due to the large model scale. We need to divide the city into several sub-area for separately modeling to reduce the model size and make the model solvable. However, few studies aimed at sub-service area division driven by the bicycle traveling pattern, which is the most fundamental part of system design, monitoring & prediction, and rebalancing.

Therefore, a market-oriented method was proposed in this chapter. The proposed method involves identifying potential bicycle-sharing demand, the division of sub-service area, and the analysis of potential use patterns and potential emission reduction. The potential bicycle-sharing demand is detected using mobile phone GPS data. Based on the link network constructed from potential bicycle-sharing demand, the community detection method is applied to discover the densely connected modes, define the division of sub-service areas and distinguish the core and peripheral areas. Then, several network community indicators are proposed for subareas to understand the potential patterns of bicycle-sharing use. Hotspots of subareas are identified for the priority of bicycle management and infrastructure construction. The emission reduction model is used to measure the potential emission reduction. A case study is conducted by using the mobile phone data set in Tokyo and divides the area of Tokyo into 21 sub-areas. Suggestions for bicycle management, infrastructure development, and bicycle-sharing system planning are given regarding subareas with different properties. Potential emission reduction for applying a bicycle-sharing system is calculated in the area of Tokyo.

Chapter 5

Bicycle-Sharing System: Layout Optimization

The layout design of the bicycle-sharing system will have a great influence on the system's implementation performance and the effect of emission reduction. In recent years, many real-world cases have shown pervasive problems of public bicycle schemes, such as abandoning bicycles, uncoordinated rationing, and low public acceptance. Those problems mainly result from the inconvenience caused by the inappropriate layout of bicycle-sharing systems. Therefore, when evaluating the potential emissions reduction of adopting a new bicycle-sharing system, the pragmatic and performance-oriented layout optimization would be primary and necessary.

Nevertheless, pragmatic and performance-oriented layout optimization is a highly complex issue [97]. The demand prediction will be the first difficulty. Generally, bicycle docking stations are co-located with points of interest (PoI), such as subway stations, hospitals, universities, and shopping malls [98]. The people density of each PoI is counted to estimate potential bicycle-rental demand and to decide the locations and capacity of the required bicycle-sharing stations [99]. This type of method has several limitations. Firstly, the method cannot detect personal travel modes, which results in the inaccurate prediction of bicycle rental demand. Secondly, origin-destination (OD) information is absent, and therefore cannot guide the rationing of public bicycles [100]. The second difficulty is the uncertainty of construction conditions at specific locations. At the layout design stage, it is difficult to collect detailed information on conditions for the potential construction of bicycle docking stations at multiple locations throughout the city [101]. Theoretically optimal locations may have unsuitable conditions for the construction of bicycle-sharing infrastructure [102]. In the absence of such information, it is pretty challenging to identify the real optimal locations for bicycle-sharing stations [103]. Consequently, there is considerable uncertainty when attempting to optimize the layout of bicycle-sharing systems.

The remainder of this chapter is organized as follows: Section 5.1 presents the methodology of this work, including geometry-based probability model, improved PSO method, and an integer linear programming (ILP) model for rebalancing optimization; Section 5.2 presents the results of the real-world case study. Finally, the conclusions are presented.

5.1 Methodology

5.1.1 Framework

In this chapter, a method (see Figure 5. 1) is proposed, which employs mobile phone GPS data to optimize the layout of the bicycle-sharing system and analyze the potential reduction in emissions of CO2. The proposed method shows a relatively universal computation framework and implementation methodology. The method comprises four steps: (1) Detection of human travel mode: developing a data mining method to detect trajectories and travel modes from GPS data (comprising users' ID, longitude, latitude, and time information) [88]; (2) Layout Optimization: here, a geometry-based probability model is proposed to deal with the uncertainty of construction conditions at candidate locations for bicycle docking stations. According to the origin and destination (OD) information of the walking, cycling, and short-distance vehicle trips, which are calculated in step 1, an improved particle swarm optimization (PSO) method is introduced to solve the model; (3) Rebalancing optimization: with considering demand uncertainty, a multi-scenario integer linear programming (ILP) model is proposed for optimal scheduling of rebalancing operations (i.e., dynamic relocation of bicycles between docking stations, in response to demand), to determine detailed design-scale information for potential reduction analysis; (4) Potential emissions reduction analysis: current days, emission are great complexes and affected by anthropogenic disturbance [104]. As a matter of reality, total emission potentiality could be attenuated by bicycle production-related emission and affected by truck type, truck fuel selection, and truck loading capacity. Therefore, based on the layout and rebalancing results, the study evaluates the potential for bicycle-sharing systems to reduce emissions.

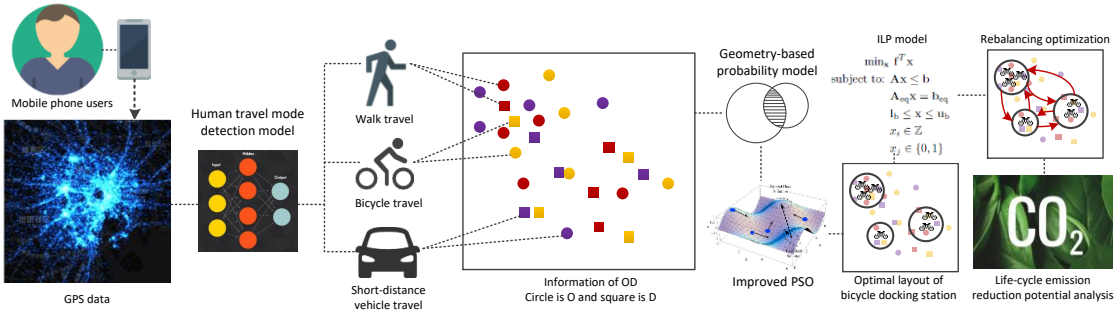


Figure 5. 1 Framework of model utilizing mobile phone GPS data

5.1.2 Geometry-Based Probability Model

Generally, when initially planning the layout of a bicycle-sharing network, it is difficult to obtain detailed information on the construction conditions of multiple locations throughout the city. Therefore, if the final solution is presented as a series of optimal location points, the results might be unworkable for project administrators due to

inadequate real-world construction conditions at some locations. Here, a new definition, termed probability area, is introduced, a specific range for constructing each bicycle-sharing station. At each location in this area, the probability of constructing the bicycle docking station is the same. Therefore, when constructing the system, it is only necessary to find the most suitable place in this broader area to build the station, thereby providing a very flexible approach to guiding the construction of the system. Working out a series of optimal probability areas will also be more practical in initial layout design[105]. However, this output form makes the model more complex, as both the central location and area range need to be decided. Moreover, this also introduces more significant uncertainty concerning the potential for the adoption of public bicycles.

In order to address these problems, a geometry-based probability model of the adoption potential is proposed. In Figure 5. 2, the solid circle represents the probability area, N_S is the area's central location, and R is the radius. N_t is the origin or destination location of the walk, bicycle, and short-distance vehicle trajectories. r is acceptable distance, which represents the maximum distance that users will tolerate to find the bicycle-sharing station. The imaginary circle shows the acceptable area. d is the distance between N_S and N_t. This method provides probability areas but not the exact locations of the stations; therefore, although r is known, whether a particular N_t location would be covered by a station is still cannot be ascertained. However, based on the definition of the probability area, the covering probability can be calculated by geometric analysis.

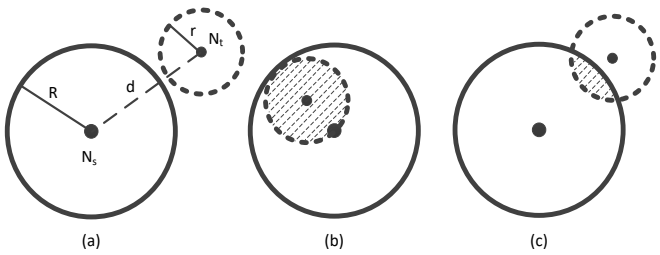


Figure 5. 2 Geometry-based probability model

There will be three kinds of relationships between the probability area and acceptable area: In case 1, there is no intersection between two areas (Figure 5. 2 (a)). In this case, no matter where the station is, the covering probability will always be zero. In case 2, the acceptable area is totally covered by the probability area (Figure 5. 2 (b)). In this case, if the station is set in the range of the acceptable area, N_t would be covered. Therefore, the covering probability equal to the acceptable area divides the probability area (if $R<r$, covering probability = 1). In case 3, part of the acceptable area is covered by the probability area (Figure 5. 2 (c)). Here, only when the station is set within the range of the shaded area, will the N_t location be covered. Thus, covering probability equal to the shaded area S_S divides probability area. The expression of covering probability P_C is shown as follows:

$$P_{\mathrm{C}}=\begin{cases}0 & \text{if } (r+R\le d)\\ r^2/R^2 & \text{if } (R-r\ge d)\\ S_{\mathrm{S}}/\pi R^2 & \text{else}\end{cases} \tag{5.1}$$

Where,

$$S_{\mathrm{S}}=a_1\cdot r^2+a_2\cdot R^2-r\cdot d\cdot\sin(a_1) \tag{5.2}$$

a_1 and a_2 are the intermediate variables:

$$a_1=\mathrm{acos}\left[\left(r^2+d^2-R^2\right)/\left(2\cdot r\cdot d\right)\right] \tag{5.3}$$

$$a_2=\mathrm{acos}\left[\left(R^2+d^2-r^2\right)/\left(2\cdot R\cdot d\right)\right] \tag{5.4}$$

To avoid repeat counting, each N_{t} belongs to a maximum of one station. The superscript i represents trajectory, and the superscript j represents the probability area of a station. The objective is to find the optimal center locations $N_{\mathrm{S}j}$ and these radiuses R_j to maximize the total covering probability of each trajectory. Since the bicycle-sharing system would be useful for a trip only when its origin and destination are both covered by stations, the covering probability of the trajectory is equal to the covering probability of its origin multiplied by that of its destination, as shown as follows:

$$\max F=\sum_i \alpha_i \max_j\left(P_{\mathrm{C}i,j}\left(O_i,N_{\mathrm{s}j},R_j\right)\right)\max_j\left(P_{\mathrm{C}i,j}\left(D_i,N_{\mathrm{s}j},R_j\right)\right), \tag{5.5}$$

where, α_i is a coefficient for the travel mode of the origin or destination location. The potential rates of substitution by public bicycle differ between the other travel modes. If the estimated substitution potential rate of travel mode i is large, this travel mode would have a greater impact on the final optimal layout, therefore α_i will be set to a high value. O_i and D_i represent the OD coordinates of trajectory *i*.

According to Eq. 5.1 and Eq. 5.5, to obtain the optimal value of F, each probability area radius R_j should be carefully designed. When R_j is set overlarge, the probability area j could cover a larger N_{t} as this reduces the probability of the first condition ($r+R\le d$); however, the value of $P_{\mathrm{C}i,j}$ the two other conditions would be decreased. Therefore, the value of F may also be reduced. On the other hand, if R_j is set over-small, the value of $P_{\mathrm{C}i,j}$ in the second and third conditions would be increased, but a larger number of $N_{\mathrm{t}i}$ would meet the first condition and also cause a decrease of

F. According to the above analysis, for locations with a high density of N_t, such as subway stations, hospitals, and shopping malls, the corresponding N_S should be closer to these PoIs and R should be small. However, for areas with a lower density of N_t, such as communities and business districts, the corresponding N_S will trend close to the center of this area and R should be larger. This qualitative discussion is very similar to the process of PoI-based subjective analysis and closer to a convenience-oriented design. Furthermore, the detailed layout design parameters are formulated for the theoretical optimum-oriented design. Therefore, the proposed model incorporates the advantages of both types of design.

5.1.3 Improved Particle Swarm Optimization Method

PSO was first proposed by Kennedy and Eberhart [106]. Owing to its high convergence accuracy and searching ability, PSO has been widely used in many engineering research fields [107]. Each member or particle in the swarm represents a feasible solution, and their velocities and positions are updated during iterations according to Eq. 5.6 and Eq. 5.7:

$$v_k(\tau+1) = v_k(\tau) + c_1 r_1 (p_k(\tau) - x_k(\tau)) + c_2 r_2 (p_g(\tau) - x_k(\tau)) \tag{5.6}$$

$$x_k(\tau+1) = x_k(\tau) + v_k(\tau+1) \tag{5.7}$$

where v_k and x_k are the velocity and position of particle k, respectively, c_1 and c_2 are the respective acceleration parameters, r_1 and r_2 represent two randomly generated numbers within the range [0, 1], p_g denotes the best position in the swarm during the search period, and p_k is the best position for a particle k at the τ^{th} iteration.

Research into improving PSO has burgeoned in the past two decades. Consequently, a great many improved PSOs have been proposed, aimed at models with differing complexity, scale, and computational efficiency demand. For a new raised optimization problem, a comparative analysis is necessary to determine which improved algorithm is most suitable. Here, a virtual case study with an area of 10×10 km and 10000 OD coordinates is generated. The number of stations (N) is set to 10,30, 50, 70, and 90, respectively, and chose the two classical algorithms (GPSO [106] and LFIPSO [108]), a well-known algorithm (COM-MCPSO [109]) and a recent algorithm (PP-PSO [110]) for the comparison.

The initial version of PSO, named global particle swarm optimization (GPSO), gained popularity mainly due to its speed of convergence and ease of use. However, in GPSO, each individual is influenced by the best performer among his neighbors. Thus it receives inadequate information.

Subsequently, the fully informed particle swarm optimization (FIPSO) model adopts a mechanism to make ensure that the individuals are "fully informed," among which the Ring topology version (LFIPSO) is conceptually more concise and promises more effective performance than the traditional particle swarm algorithm. In this new version, the particle uses information from all of its neighbors rather than just the best one. Despite the improvement in designing topologies, there are difficulties in controlling the balance between exploration (global investigation of the search place) and exploitation (the refined search around a local optimum).

In order to balance the exploration and exploitation in PSO, a multi-swarm cooperative scheme, named multi-swarm cooperative PSO, was introduced, comprising one master swarm and several slave swarms. The slave swarms can supply many new promising particles (the position giving the best fitness value) to the master swarm as the evolution proceeds. The master swarm updates the particle states based on the best position discovered by all the particles in the slave swarms and their own. The interactions between the master swarm and slave swarms influence the balance between exploration and exploitation and maintain a suitable diversity in the population even when approaching the global solution, thus reducing the risk of converging to a local sub-optimal solution. Despite the improvement of adding new strategies, a mass of particles with low velocity may concentrate in some locations, especially in the mid-late iterations, and there is the poor capability of exploring new search regions and finding better solutions for such particles.

Ultimately, considering the "slothful particles" with low velocities, which contribute little to the optimization and impact the computation speed, the prey-predator PSO (PP-PSO) was proposed, which employs the three strategies of catch, escape, and breeding. In PP-PSO, slothful particles can be deleted or transformed, and while the former helps to speed up convergence and computation speed, the latter improves the optimization results.

Since it cannot be proved which algorithm will be most appropriate to this issue simply by theoretical derivation, detailed stability and convergence tests are necessary. First, N is set to 50 and repeated the computation four times by each improved algorithm. The resulting stability analysis is shown in Figure 5. 3 (F means the value of the objective function). The results show that COM-MCPSO and PP-PSO performed approximately 40% better than GPSO and LFIPSO in this scenario. Meanwhile, although COM-MCPSO has more excellent stability than PP-PSO (the difference among the four times' results solved by COM-MCPSO is lower than that of PP-PSO), PP-PSO could see convergence to a better solution. To further indicate which algorithm is best suited to this issue, a convergence comparison analysis with N= 10, 30, 70, and 90 is shown in Figure 5. 4. F is the value of the objective function (shown in Eq. 5.5). The results show that PP-PSO is superior for this case, providing approximately 10% gain over the other methods, and hence PP-PSO was chosen as the most suitable algorithm. In order to avoid errors associated with the instability of the algorithm, when experimenting, the

computation was repeated four times, and the best result was selected as the output of the algorithm.

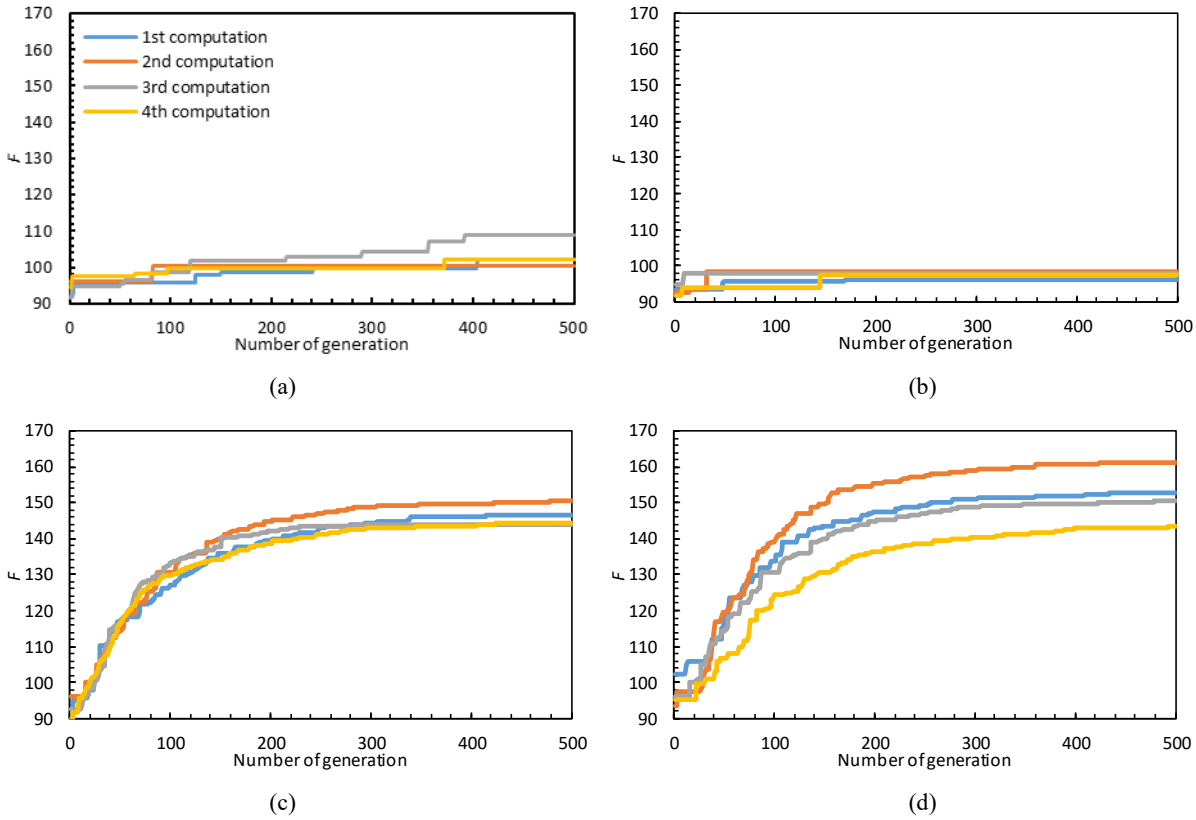
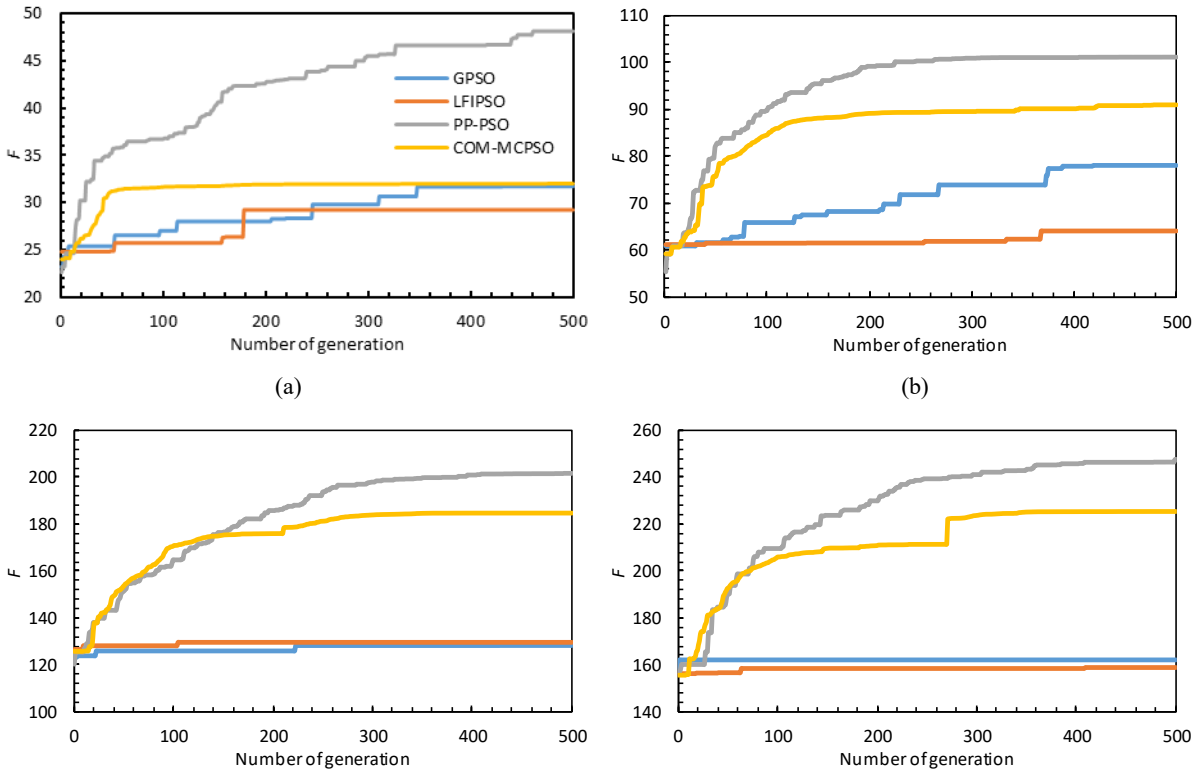


Figure 5. 3 Stability analysis with N=50
by (a) GPSO, (b) LFIPSO, (c) COM-MCPSO, and (d) PP-PSO



(c) (d)

Figure 5. 4 Comparison of convergence performance with N= (a) 10, (b) 30, (c) 70, and (d) 90

5.1.4 Rebalancing Optimization

Although the optimal layout for bicycle-sharing stations could be obtained through the geometry-based probability model and improved PSO, the detailed design-scale information still cannot be obtained (such as how many docking facilities need to be provided at each station, how many bicycles need be put into the sharing system, and even how much energy will be consumed when rebalancing the numbers of bicycles across the system). These are the significant parameters that determine the potential reduction in emissions that may be achieved by bicycle-sharing. Therefore, a rebalancing optimization model is necessary to determine those decision parameters.

Since there must be uncertainties in the usage demand of each bicycle-sharing station and the probability distribution of those uncertainties could be indicated by statistical analysis of daily trajectory data, the stochastic optimization method can be used to solve this issue. Based on the Monte Carlo method, a multi-scenario ILP model is developed for the rebalancing optimization issue, based on the optimal layout derived previously.

5.1.4.1 Model Requirements

The model is formulated as a multi-scenario ILP and is solved using the Gurobi optimization mathematical programming solver.

Inputs:

- Studied horizon.
- Cost information: Unit purchasing and maintenance (P&M) costs of the bicycle, unit P&M costs of docking facility, unit rebalancing cost.
- Demand information: Probability distribution of the usage demand at each station, containing the number probability of people who will select/return a bicycle at this station during a given time window (this may be obtained by statistical analysis of daily trajectory data for the whole year)
- Station information: Layout of the station network (solved by geometry-based probability model and improved PSO), actual road distance between each station (between paired centers of probability areas).
- Technical parameters: Turnover coefficients.

Determine:

- The number of docking facilities at each station.
- The number of bicycles needs to be purchased.

- Detailed rebalancing schedule (how many bicycles need to be transferred between stations during each time window).
- Total construction and operating costs of the bicycle-sharing system.

Objective:

Minimize the total cost to design and operate the bicycle-sharing system under various operational and technical constraints. Total cost includes P&M costs of bicycles, P&M costs of docking facilities, and cost of rebalancing operations.

Model assumption:

In order to build and solve the model effectively, it is assumed that the cost of rebalancing is simplified to be linear with the product of the number of transferring bicycles and the road distance between stations; that is, the detailed vehicle routing problem (VRP) is not taken into consideration.

5.1.4.2 Mathematical Model

Based on the Monte Carlo simulation, a multi-scenario ILP model is developed to transfer the uncertain optimization to a specific two-stage optimization. Specifically, since the demand parameters of each station are uncertain, the issue could be divided into many specific scenarios. In each scenario, by sampling method, one set of specific demand parameters is generated based on their probability distributions. Therefore, with generating enough scenarios and set the average costs of all scenarios as the objective, according to Monte Carlo theory, the result is equivalent to that of the uncertain model. The proposed mathematical model is shown as follows.

Objective function:

$$\begin{aligned}\min\ f = {} & Cb \cdot \varpi \cdot \sum_{i} BNb_i + Cd \cdot \sum_{i} Nd_i \\ & + Co \cdot \sum_{t}\sum_{i}\sum_{i'} L_{i,i'} Un_{i,i',t} + Co \cdot \sum_{s}\sum_{i}\sum_{i'} L_{i,i'}\, fUn_{s,i,i'} \Big/ sm\end{aligned}$$

$$\forall i, i' \in I\ ,\ \forall s \in S\ ,\ \forall t \in T \tag{5.8}$$

Constraints:

$$Nb_{s,i,t+1} = Nb_{s,i,t} + \left(Tin_{s,i,t} - Tout_{s,i,t}\right) + \sum_{i'}\left(Un_{i',i,t} - Un_{i,i',t}\right)$$

$$\forall i, i' \in I\ ,\ \forall s \in S\ ,\ \forall t \in T \tag{5.9}$$

$$BNb_i = Nb_{s,i,0} \qquad \forall i \in I, \forall s \in S \tag{5.10}$$

$$BNb_i - Nb_{s,i,tm} = \sum_{i'} \left(fUn_{s,i',i} - fUn_{s,i,i'} \right) \qquad \forall i, i' \in I, \forall s \in S \tag{5.11}$$

$$Nb_{s,i,t} \geq 0 \qquad \forall i \in I, \forall s \in S, \forall t \in T \tag{5.12}$$

$$Nb_{s,i,t} \leq \lambda Nd_i \qquad \forall i \in I, \forall s \in S, \forall t \in T \tag{5.13}$$

$$Nd_i \leq Ndm_i \qquad \forall i \in I \tag{5.14}$$

where, $i \in I$ is the subscript and set of bicycle-sharing stations. $s \in S$ is the subscript and set of scenarios, the maximum number of scenarios is sm. $t \in T$ is the subscript and set of time windows, the maximum number of time windows is tm. Cb is unit P&M costs per bicycle (0.11 $ per day). Cd is unit P&M costs per docking facility (0.091 $ per day). Co is unit rebalancing cost (0.029 $/km). ϖ is the turnover coefficient for purchasing bicycles (0.15). λ is the turnover coefficient for setting docking facilities (0.9). $L_{i,i'}$ is the road distance between station i and i'. $Tin_{s,i,t}$ and $Tout_{s,i,t}$ are the return and picking number of bicycles at station i during time window t in scenario s. Ndm_i is the maximum number of docking facilities that can be set at station i. $Nb_{s,i,t}$ is the number of bicycles held at station i during time window t in scenario s. Nd_i is the number of docking facilities set at station i. $Un_{i,i',t}$ is the number of bicycles transferred from station i to i' during time window t, $t \neq tm$. $fUn_{s,i',i}$ is the number of bicycles transferred from station i to i' during the last time window in scenario s. BNb_i is the initial number of bicycles set at station i.

The objective function of Eq. 5.8 contains four terms. The first is the P&M costs of bicycles, which equals the product unit cost, turnover coefficient, and the total initial number of bicycles. The second is the P&M costs of docking facilities. The equation is similar to that for bicycles. The third and fourth terms are rebalancing cost, which is designed based on the model assumption. Considering the demand uncertainty, the rebalancing during the final time window in different scenarios will not be the same.

Constraint (5.9) is the quantity conservation equation. Here, the discrete-time is adopted to present the model. The number of bicycles held at a station during the following time window is equal to that during the last time window plus the net flow number by the client's usage and the net flow number by rebalancing.

Constraint (5.10) refers to the number of bicycles held at a station during the initial time window, which must be the initial setting number.

Constraint (5.11) refers to the rebalancing during the final time window, which is dependent on the difference between the initial setting number and the holding number during the second to last time window.

Constraints (5.12) and (5.13) determine that the holding number at each station must not be less than zero or more significant than the product of the number of docking facilities and its turnover coefficient.

Constraint (5.14) means the set number of a docking facility cannot be larger than the maximum capacity of the station.

5.1.4.3 Emission Reduction Potentiality Analysis

The potentiality of total emission reduction generated by previous calculations could be affected by other production and system operation processes. Considering those influential potential variables, emission reduction potentially analysis is helpful to reveal the emission potentiality difference, which could also provide valuable insight into the approaching bicycle-sharing system regulations and policies. Here, it is assumed that emission reduction performance will be attenuated by the bicycle production process and affected by truck types, truck fuel selection, and loading capacity. Assuming there are n types of bicycle rebalancing need (Reb_n), Emission reduction potentiality ($Emission$) can be described as:

$$Emission = E_{total} - E_{BikePro}/Y - \frac{Reb_n}{Cap_{i,j}^{max} * Rt_{i,j}^{m}/Bwt} * T_{i,j} * EI_i \quad (5.15)$$

where E_{total} is a full year of travel emission reduction potentiality, $E_{BikePro}$ is the emission emitted by bicycle production, and Y is the bicycle durable years. Here, bicycle production-related emission is set up of 0.456 tones CO_2 per bicycle [111]. $Cap_{i,j}^{max}$ is the maximum loading capacity of truck type j that adopting fuel type i. T refers to energy consumption (l/km) and EI is emission intensity. Here, it is set that truck types include both light and medium trucks. Light trucks (LGT), used in the rebalancing process, have an average maximum loading capacity of around 0.35t and 0.5t. Based on the different loading capacities of the LGT, energy consumption distribute from 2.74 l/km (10%) to 0.324 l/km (100%)[112]. Moreover, medium truck (MD) indicates those trucks of average maximum loading capacity at 2t. Therefore, each loading capacity could generate eight forms of emission reduction potentiality by considering truck types and truck fuel selection.

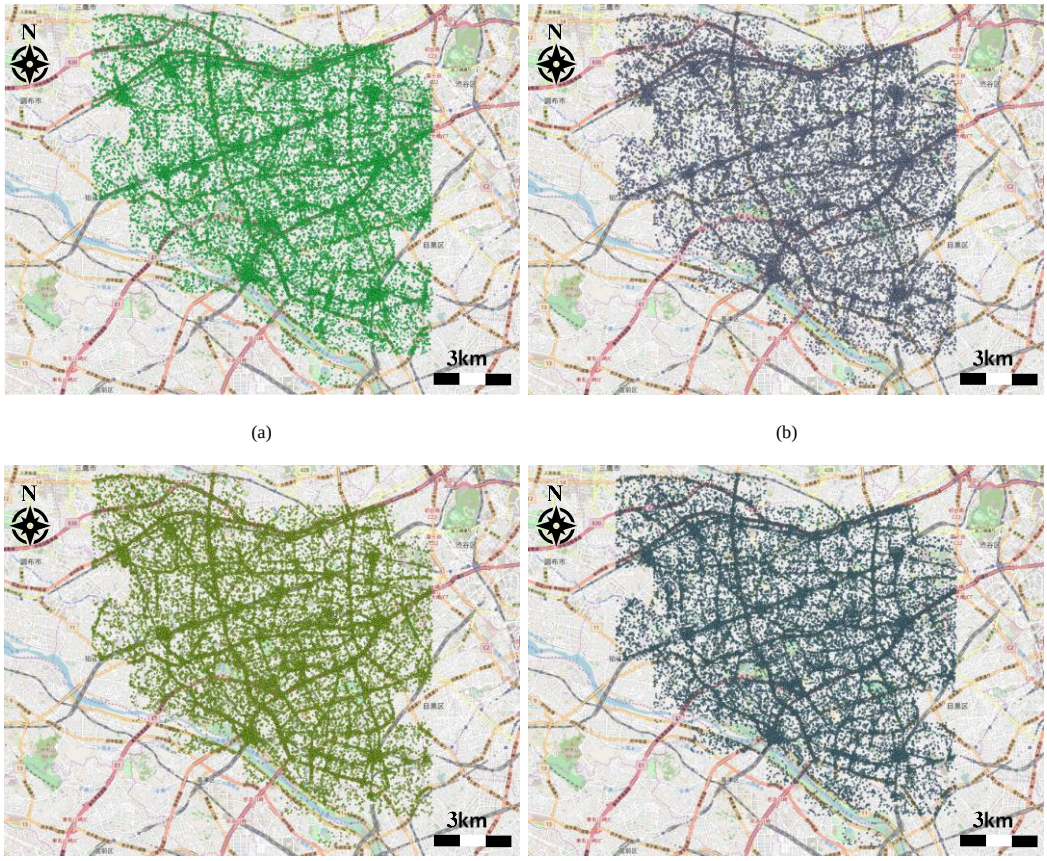
5.2 Results and Discussion

5.2.1 Study Case

In order to apply our model to real-world human activity, the GPS trajectories in Setagaya Ward, Tokyo, throughout 2012 are utilized. As introduced in Chapter 2, the GPS data is extracted from the 'Konzatsu-Tokei(R)'GPS dataset collected by NTT DOCOMO INC. The GPS trajectories are classified into different transportation modes (stay, walk, bicycle, car, and train) using the method introduced above and utilize the GPS trajectories representing walk, bicycle, and car modes in the subsequent experiment. Here, since the stay and train modes have nothing to do with our subject, it will not be considered in following. The data set comprises 3659703 trajectories, of which the walk, bicycle, and car modes account for 2904820, 481660, and 273223 trips, respectively. The OD information of these trajectories is shown in Figure 5. 5. In order to indicate the macro results for this area, the input parameters are modified when calculating energy consumption and optimal rebalancing schedule, according to statistics on the actual population of this area.

5.2.2 Layout Result

The PP-PSO is used for the layout optimization and set N=30, 50, 70, and 90, respectively. The optimal layout of the bicycle-sharing system is shown in Figure 5. 6. The blue circles represent the probability area of setting a bicycle-sharing station, with circle size-adjusted proportionally to the probability area. The results show that most stations are at locations with a high density of OD information and along main roads. Meanwhile, the probability area would be small if the station is set at a position with a high OD density. This is because: from Eq. 5.5, the relationship between the size of the probability area and its contribution to the objective function is nonlinear. The parameters of this nonlinear function are dependent on the area's OD density. If the OD density is high, the corresponding optimal size of the probability area will tend to be small, which could increase the contribution to the model's objective. Meanwhile, there are interactions among the stations. When N is increased, some alternative stations will be divided into two or three adjacent stations. The probability area size of those divided stations will be small to avoid inter-competition.



(a) (b)

(c) (d)

Figure 5. 5 OD information for different travel modes

(a) O for walk, (b) D for walk, (c) O for bicycle, (d) D for bicycle

Figure 5. 7 shows the replacing travel distance (km) of each transport mode by sharing bicycle in the presence of an optimally laid out and operated bicycle-sharing system. From the results, the adoption of a bicycle-sharing system can decrease vehicle travel, especially during peak weekday commuting periods and on weekend afternoons. Most stations are at locations with a high density of OD information and along main urban roads.

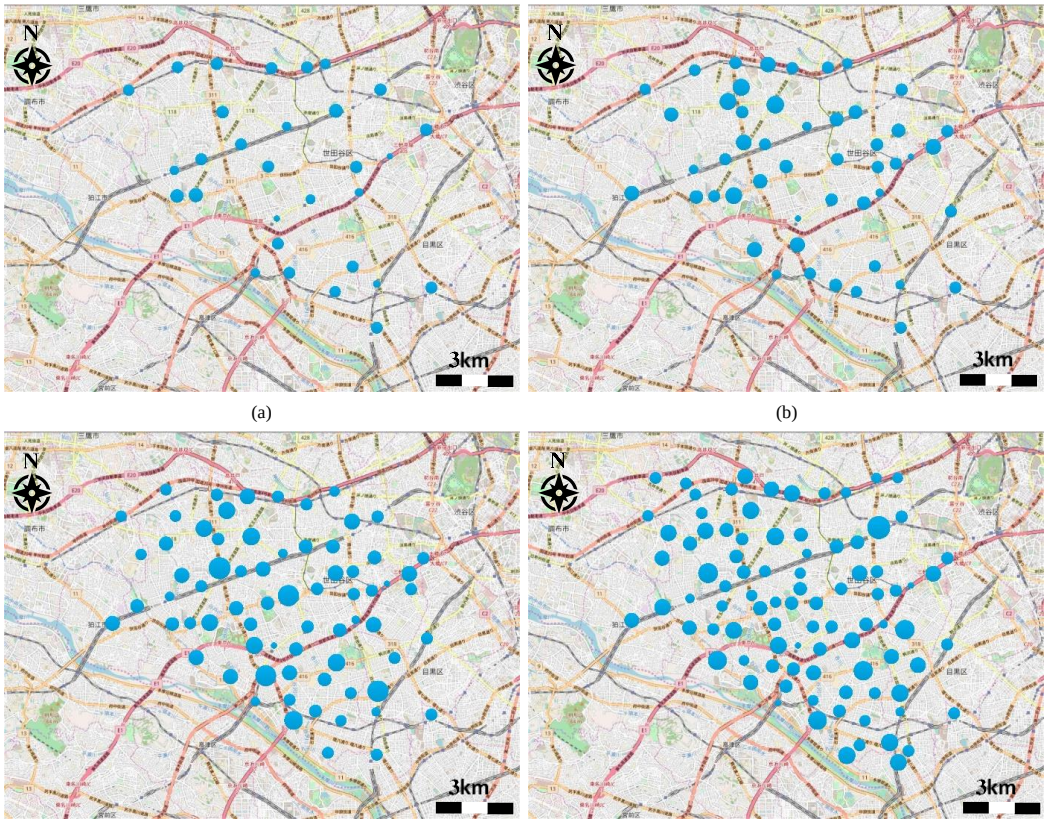


Figure 5. 6 Optimal layout of bicycle-sharing system
with *N*= (a) 30, (b) 50, (c) 70, and (d) 90 docking stations.

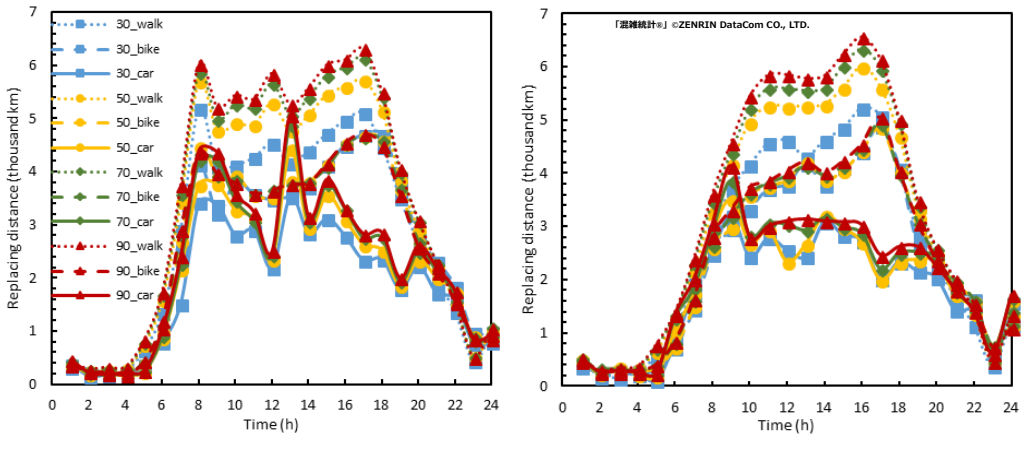


Figure 5. 7 Replacing distance by sharing bicycle
(a) weekday, (b) weekend with 30 to 90 bicycle-sharing stations.

5.2.3 Emission Reduction Potential Analysis

Here, three scenarios are presented for analyzing the potential reduction in emissions of a bicycle-sharing scheme.

Scenario 1: analyzing the emission reduction associated with modal shift (from vehicle

to bicycle-sharing) and assuming that all the people whose journey could be served by bicycle-sharing will adopt this travel mode. Since few previous studies on layout design for bicycle-sharing stations considered rebalancing, this scenario enables comparative analysis of the optimality of the layout results with this kind of study.

Scenario 2: analyzing both the emission reduction caused associated with modal shift (from vehicles to bicycle-sharing) and the emissions associated with the production, maintenance, and recycling of bicycles and facilities as well as rebalancing. Meanwhile, it is assumed that all of the people whose journeys could be served by bicycle-sharing will adopt this travel mode. The purpose of setting this scenario is to determine the impacts of emissions of bicycles and facilities and the emissions of rebalancing and how these affect the final reduction in emissions.

Scenario 3: As in Scenario 2, but it is assumed that just 10% of people whose trips could be served by bicycle-sharing will adopt this mode. This scenario is intended to test the sensitivity of the model to the rate of bicycle-sharing adoption. In both scenario 2 and scenario 3, the truck loading capacity adopted in the bicycle rebalancing system is 80%.

For Scenario 1 (see Figure 5. 8), based on the TripEnergy model [113], car travel speed, and ambient temperature of this year, the CO_2 emission of each car journey trajectory is calculated. The detailed technological data sources, description, and sample results were reported previously [88]. The results show that with an increase in docking stations, from 30 to 90, the adoption of a bicycle-sharing system can reduce CO_2 emissions by approximately 3.1–3.8 thousand tonnes annually. Here, the methods proposed by Chen and Sun [84] and García-Palomares, et al. [114] are taken as baselines for comparative analysis. The results show that the proposed model performs better than other methods and that with the increase in the number of stations, the performance gap becomes more considerable. When N=90, the results obtained by the proposed method could reduce emissions by a further 6.4% and 4.4% compared with the results of the other two methods. The reason is that the previous method [114] took the density of generated trips within a given area as the metric for deciding where bicycle-sharing stations should be located, which is a type of greedy method. However, subsequent adoption occurs only when the network of stations can cover both the origin and destination of a trip. Some trips' origins may be at locations with a high density of generated trips (such as a subway station). At the same time, the destinations may be low-density locations (such as residential areas). If the density of generated trips is taken as the sole metric, the final designed bicycle-sharing system may not cover such trips. Meanwhile, with the increased scale of the system, this disadvantage will become more prominent. In addition, although the other previous method [84] avoided this disadvantage by taking the transformation of trips into consideration, the method was based on mesh modeling, which cannot represent the trajectory information accurately if the grid size is large, and become computationally time-consuming if the grid size is small.

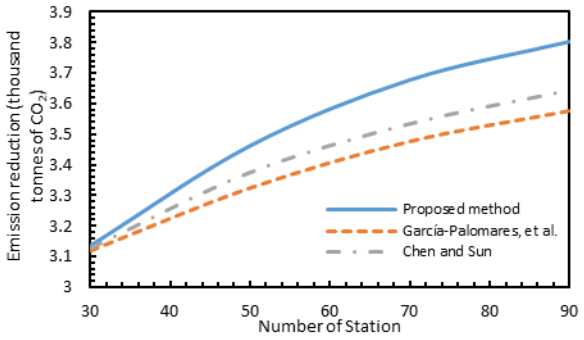


Figure 5. 8 The relationship between station number and emission reduction in Case 1

For Scenario 2 (see Figure 5. 9), consideration of energy consumption resulting from bicycle rebalancing and bicycle production is included. To begin with, it is assumed that all of the people whose travel could be served by bicycle-sharing will adopt this travel mode. Moreover, by adopting the different types of truck and fuels, the emission variation could be found in Figure 5. 9. Compared with scenario 1, bicycle production emission and truck-generated emission resulting from bicycle rebalancing need to attenuate reduction emission performance around 11.30% to 21.26%. The offset emission increases along with the number of stations. Medium-sized truck shows lower offset volume; whist light truck is not environmentally preferable. Under this circumstance, policymakers are recommended to consider the emission offset brought by bicycle rebalancing. In Figure 5. 9, the abbreviation in the legend is defined as: the first term represents the number of the scenario. The second term represents the type of transportation truck, and the third term represents the type of consuming oil.

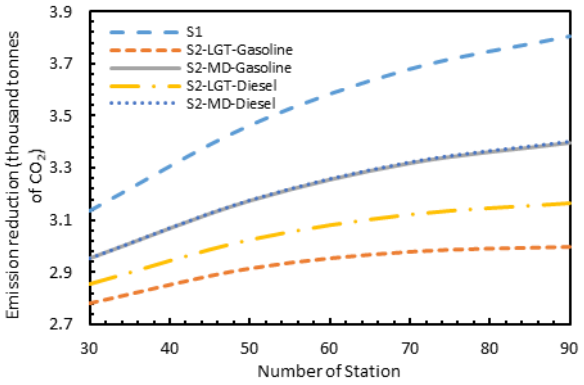


Figure 5. 9 Emission reduction potentiality of scenario 2 at the 80% truck locating capacity

Scenario 3 (see Figure 5. 10) assumes that just 10% of the people whose travel could be served by bicycle-sharing will adopt this travel mode. Different from scenario 2, the offset emissions do not show a distinct gap among four truck-related forms, and the emission variation trend is close to that of scenario 1. Therefore, when the reality is close to scenario 3, the environmental restrictions of the truck and its fuel selection could have much more freedom than that of scenario 2.

Scenarios 2 and 3 are calculated at the truck loading capacity of 80%. Based on our observation, higher loading capacity offset fewer reduction potentiality in all the tuck and fuel selection forms, while the substantial reduction potentiality varies greatly. In the supporting material, the reason for how the truck loading capacity rate impacts emission reduction potentiality is given out.

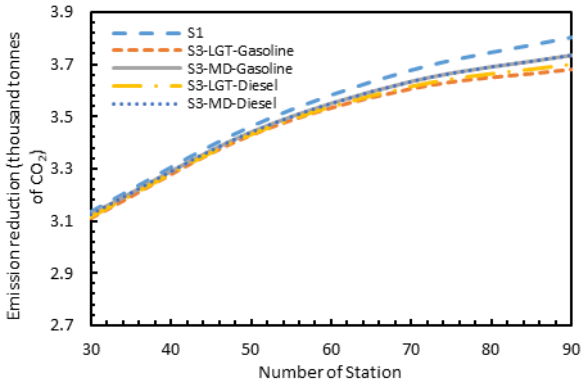


Figure 5. 10 Emission reduction potentiality of scenario 3 at the 80% truck locating capacity

5.3 Conclusions

This chapter analyzes the potential reduction in CO_2 emissions for adopting a bicycle-sharing system. A method for detecting human travel modes and a geometry-based probability model is proposed to support particle swarm optimization. Then, based on the resulting optimal layout of bicycle docking stations and considering demand uncertainty, a multi-scenario mixed-integer linear programming model is proposed for optimal rebalancing to obtain the detailed design-scale information required for potential reduction analysis. The proposed model is tested using a data set of approximately 3 million GPS trajectories obtained in Setagaya Ward, Tokyo. The results obtained by the proposed method could reduce emissions by a further 6.4% and 4.4% compared with previous methods. Three scenarios are selected to simulate emission reduction volume after an offset by bicycle production and bicycle rebalancing. In the bicycle rebalancing process, a different number of station truck loading capacity rates, truck types, and truck fuel selection will affect the total emission reduction potentiality. Therefore, the emission reduction volume by considering those influential factors can be evaluated. From purely environmental insight, results show that the medium size of truck and fuel selection on diesel is preferable. For example, in scenario 2, the offset phenomena are significantly alleviated based on selecting medium trucks, especially when diesel medium trucks are adopted. To maintain the emission reduction brought by the bicycle-sharing system, two sets of policies are taken into consideration. Firstly, if scenario 2 is expected as much closer to reality, the usage of the bicycle-sharing system should be closely monitored to generate the corresponding rebalancing transportation form. Here, diesel medium trucks will help alleviate the offset emission, which is the priority for environmental consideration. If scenario 3 is much closer to reality, truck selection could consider the economic cost. Both sets of

policies aimed at realizing the bicycle-sharing system with a lower environmental or economic cost under the realistic constitutions. As a representation of intelligent and sustainable city development, the bicycle-sharing system is one of the hottest topics in transportation, public health, urban planning. Additionally, the bicycle-sharing system has become increasingly popular in many countries. Our findings are not limited to Japan but can be theoretical guidelines for bicycle-sharing companies and policymakers to promote an intelligent transportation grid.

Chapter 6

Ride-Sharing System: City-level Potential Analysis

Spreading green and low-consumption transportation methods is becoming an urgent priority. Ride-sharing, which refers to the sharing of car journeys so that more than one person travels in a car, and prevents the need for others to have to drive to a location themselves, is a critical solution to this issue. Before being introduced into one place, it needs a potential analysis. However, current studies did this analysis based on home and work locations or social ties between people, which is not precise and straight enough.

This chapter proposed a new MaaS framework for ride-sharing and emission reduction potential analyses. This work's contributions are shown as follows: (1) The potential analysis of ride-sharing is based on the actual travel demand, which is more precise and reliable than the base of urban population distribution. (2) I proposed a deep learning-based method for real-time matching feasibility estimation. (3) Based on the proposed high-effective matching algorithm, the big data-based potential and emission reduction analyses are carried out. (4) Tokyo is taken as the case study with mining millions of trajectories data of mobile phone users in one year.

The remainder of this chapter is organized as follows. In Section 6.1, the methodology is proposed. Then, the case study is introduced in Section 6.2. Section 6.3 presents the experimental methodology and the result. Finally, conclusions are given in Section 6.4.

6.1 Methodology

6.1.1 Framework

The proposed framework shown in Figure 6. 1 provides a comprehensive methodology for the comprehensive potential analyses of urban ride-sharing and emission reduction based on smartphone GPS data. The method is decomposed into three parts: (1) Human travel mode detection: detecting the travel mode of trajectories from GPS data (contain longitude, latitude, and time information); In this chapter, I utilized the travel mode detection method, which was proposed in our previous work [88]. The process contains four steps: extraction of staying and moving segments, splitting travel segments, traffic mode classification, and segment merging. After these processing steps, the trajectories can be clustered into four groups by travel modes: walking, bike, public transit and

vehicle. (2) Matching feasibility estimation: considering the, estimate the matching feasibility of each two trips. (3) Potential of ride-sharing and emission reduction potential analyses: based on the estimate results, indicating the relation among rate, emission reduction, and policy support.

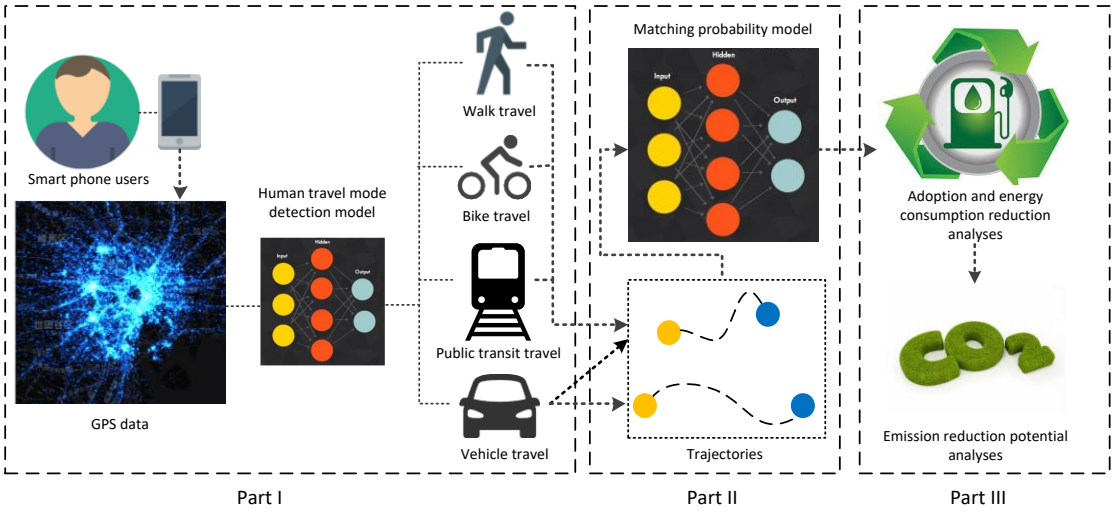


Figure 6. 1 Method framework for ride-sharing system city-level potential analysis

6.1.2 Matching Feasibility Estimation Model

6.1.2.1 Elaboration on Model

The model I use for matching feasibility estimation is the deep learning method, the structure of model training is shown in Figure 6. 2.

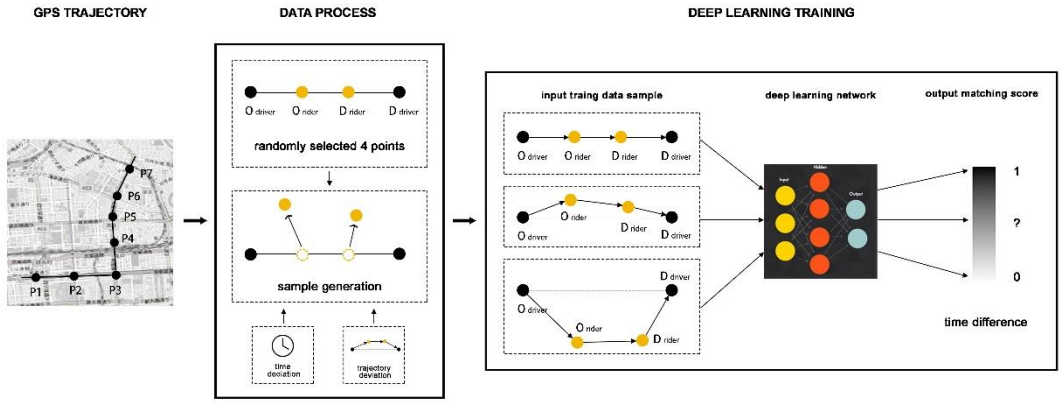


Figure 6. 2 Model training structure

I used the GPS trajectory and did the training data preparation process as the base of a deep learning model. It includes the process of marking the matching feasibility of a pair of ODs. A comprehensive training dataset with different matching feasibility value is needed to make the deep learning model well trained. However, in the original trajectory set, there is no direct mark of feasibility. Thus, a reliable strategy to generate

training data is necessary. Finally, the generated data will be fed to the deep learning model to finish the training.

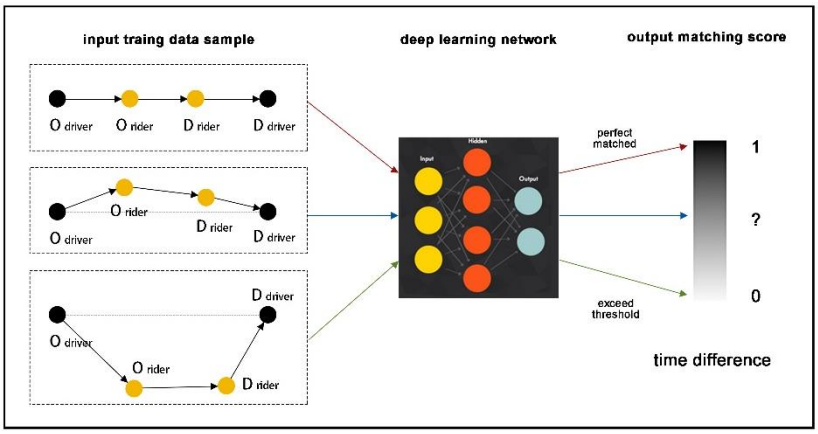


Figure 6. 3 Generation and training process of deep learning

The training data process is shown in Figure 6. 3. We extract a piece of complete and continuous trajectory record from the dataset, randomly choose two records (at least with two other records in the interval of these two records) in this trajectory as the OD of the driver. Then we randomly choose two other points at the interval of these points as the perfect matched rider (Rider*). Taking the OD information of the original trajectory (driver) $\left(ODx_{i,k}, ODt_{i,k}\right)$ and perfect matched OD information (rider) $\left(ODx, ODt\right)^{*}_{i,k}$ as the input, the corresponding feasibility output is $P_{(i,k),(i',k')}=1$. Then, We generate some noises z into $\left(ODx, ODt\right)^{*}_{i,k}$ both from the aspect of location and time as the "not perfectly matched samples." In these samples, the feasibility output is a value between 0 and 1 based on the noise. We generate this kind of training data because comprehensive training data assures the accuracy of the deep learning model. When the noises are more significant than a threshold, these samples can be the "bad matched sample" ($P_{(i,k),(i',k')}=0$). After that, We will put these training samples into the deep learning model. Finally, this model could be used to operate feasibility estimation for real-world samples when finishing the training.

Definition 1 (Perfect matching): To a matching issue for the ride-sharing system, if the rider's OD exactly belongs to a driver's trajectory, meanwhile the departure and arrival time are the same with the passing time of the driver to those two points, I would say that the rider and driver are perfectly matched, and the matching feasibility is 1 (introduced in Section 5).

$$P_{(i,k),(i',k')}=\left\{1 \mid \left(ODx_{i',k'}, ODt_{i',k'}\right) \in \left(x_{i,k}, t_{i,k}\right)\right\} \tag{6.1}$$

Here the set of OD coordinates, departure and arrival time information for perfect matching is denoted as $\left(ODx, ODt\right)^{*}_{i,k}$

Definition 2 (Noise matching): To label some "bad matching samples", here I added an

unacceptable noise into the $(ODx, ODt)^*_{i,k}$ and set the matching feasibility of this sample to 0.

$$P_{(i,k),(i',k')} = \left\{ 0 \mid \left(ODx_{i',k'}, ODt_{i',k'} \right) \in \left\{ \left(ODx, ODt \right)^*_{i,k} + z \right\} \right\} \tag{6.2}$$

Where z is an unacceptable noise.

6.1.2.2 Deep Learning Architecture

Deep Learning is a branch of machine learning, the main idea of simulating the mechanics of the human brain. Therefore, deep learning applies layer-based construction to propagate information, just like the electricity passing among brain neurons. Deep learning algorithms process data and imitate the thinking process of understanding the hidden key factors of output. Therefore, it is widely applied in complex data computation, image processing, natural language processing, and other fields.

In deep learning, motivated by the human brain, information is passed at a layer level constructed by neuron units. In detail, the general architecture of deep learning contains three parts: input layer, hidden layers, and output layer. The input data are put into the input layer. Each input value is treated as a neuron unit, such as $x_1, x_2, x_3, \dots, x_n$. Then, these input neurons are connected with hidden layers, which contain hidden neuron units $h_1, h_2, h_3, \dots, h_n$. The number of hidden layers is not limited to one. There are usually well-designed and determined sequences connecting properly hidden layers for better training or learning of the model in many cases and these hidden layers. One hidden layer will receive the input from the hidden layer before it. These hidden layers can extract the features of their inputs and turn these learned features into the coefficients of layer computation matrix to best fit the desired output in the training dataset. Finally, the neuron units in the last hidden layers are also connected with the output layer, representing the results I want. Simply speaking, the hidden layers learn the latent relations from input data to output data.

In general, the output of adjacent layers is defined as:

$$A_n = W_n X_n + B_n \tag{6.3}$$

where W_n is the layer computation matrix and B_n is the bias of n-th layer. A_n is the output of neurons at n-th layer, and the input to $(n+1)$-th layer.

Once we have designed the construction of deep learning, we need to train the network and optimize parameters W and B. Combining the Forward and Backpropagation method is widely used in training neural networks with the gradient-based optimization technique. However, the error cannot be correctly propagated with the increase of layer

number when the number of hidden layers is larger than two, making the backpropagation method has bad performance for deep construction. Recently, a greedy layer-wise approach has been proposed to tackle this problem. The main idea is to train the neural network in a bottom-up way. Once the first n layers are trained, the $(n+1)$-th layer is trained as the latent features are now computed from the layers below.

Optimization algorithms play an essential role in deep learning. They help us to find a proper set of parameters for our model. In order to achieve this purpose, the optimization algorithm will try to minimize the loss function. The problem of minimizing loss function is generally expressed as:

$$\text{Min } (\mathrm{L}(\theta) = \frac{1}{n}(\sum_{i=1}^{n}(y - h_\theta))) \tag{6.4}$$

$$L(\theta) = \underset{i}{\text{mean}}\left(L_i(\theta)\right) \tag{6.5}$$

Where y is the true output in the training set; h_θ is the hypothesize of output computed by the deep learning network under the parameter set of θ including W_n and B_n in each layer; n is the size of the training set.

In this study, I adopt the input layer x_1 as $(O_x^d, O_y^d, \Delta t_O, O_x^r, O_y^r, D_x^d, D_y^d, D_x^r, D_y^r)$ according to the principle of training data generation and the output layer y as $P_{(i,k),(i',k')}$. Where, O_x and O_y refer to the longitude and latitude of origin; D_x and D_y refer to the longitude and latitude of destination of destination; d and r refer to driver and rider respectively; Δt_O refers to the time difference between departure time.

In this model, given a set of $Rid_{i',k'}$ and $Dri_{i,k}$, the matching feasibility estimation model will construct a model $P_\theta\left(\tilde{P}_{(i,k),(i',k')} \mid \left(Dri_{i,k}, Rid_{i',k'}\right)\right)$, in which $\tilde{P}_{(i,k),(i',k')}$ is the predicted matching feasibility. It will be built as a regression model, and its parameters θ can be obtained by minimizing the prediction error $L\left(\tilde{P}_{(i,k),(i',k')}, P_{(i,k),(i',k')}\right)$ with generated training samples:

$$\theta = \text{argmin } L\left(\tilde{P}_{(i,k),(i',k')}, P_{(i,k),(i',k')}\right) = \text{argmin } \frac{1}{n}(\tilde{P}_{(i,k),(i',k')} - P_{(i,k),(i',k')})^2 \tag{6.6}$$

$$\theta = \underset{\theta}{\arg\min}\, L\left(\tilde{P}_{(i,k),(i',k')}, P_{(i,k),(i',k')}\right) = \underset{\theta}{\arg\min} \left\| \tilde{P}_{(i,k),(i',k')} - P_{(i,k),(i',k')} \right\|^2 \tag{6.7}$$

In the Result and Discussion Section, I will illustrate and elaborate on the accuracy of the trained model.

6.2 Case Study

The training data is one of the essential factors that influence the performance of the deep learning model. In this case, we want the deep learning model to compute the matching feasibility of two trajectories, which is a highly complex case. So, many factors should be considered. The first one is that the training cases should be comprehensive and cover various situations, which is the basic rule of the training set. As said in the previous section, we chose the generate bias to the original trajectories to make the imperfect matching cases. Here we will give a more comprehensive explanation. Those who serve as drivers in the ride-sharing decide whether to pick someone up or not mainly based on a detour or further, we can say the time shift they have to make on their original schedule if they accept the job. So, the more biased the passenger's trajectory is from the driver's, the less possible the matching stands. By the same principle, if a passenger's trajectory coincides with drivers, the more feasibly the matching stands. To make the perfect matching cases, we choose the first and last GPS trajectory point as the OD of a driver and randomly choose another two points in the trajectory as the OD of the passenger, then mark them with the score of 1. Then, to make corrupted matching cases, we consider adding bias from two aspects: time and location. Location bias will bring the detour time, and time bias will bring schedule change to driver. After that, we will estimate the extent of bias and mark the cases with a value between 0 and 1 based on the extent.

6.3 Results and Discussion

6.3.1 Model Accuracy Verification

After completing the training of our deep learning model, we need to verify our model. A part of the training set is selected to test the accuracy. We randomly choose 29,138 groups of data as the test set. In each group, there is a random quantity of individual trajectories. A metric is set for the verification.

Verification Metric: In each group of test data, we randomly chose one piece of individual trajectory and assumed it as the driver, and match it to the rest of the individual travels. The deep learning model will give the matching score to each pair of matching. Then, we will rank the matching pairs and treat the one with the highest matching score as the best matching. Simultaneously, we also compute the detour time that the driver has to take to pick up the passenger in each matching pair by Google API to evaluate how well the driver is matched to each passenger. The shorter the detour time is, the better the matching is. Thus, the matching pair with the least detour time is the best matching in a real case. After that, we will check if the passenger in the best matching selected by the deep learning model is the same as the one selected by Google API. If they are the same, the computation can be treated as an accurate one. Then, after

testing all 29,138 groups in the test set, we will calculate the accurate computation rate among all groups.

We chose this metric mainly because when matching passengers to a driver, only the most suitable passenger will be assigned to the driver, and the rest shall be left for the next time. Thus, we only need to find the most suitable passenger for the driver and neglect the rest. The accurate calculation rate among all groups is 87.93%, which means 25,621 groups are accurate by the verification metric out of 29,138.

Here, we provide one sample in our test set for a more comprehensive presentation to understand the metric better. We separately took one driver OD and time stamp and eight-passenger ODs and time stamps and put them into our model and Google API separately. Our model would give the matching possibility(score) of these eight passengers related to the diver. Then, we will rank these scores from high to low. On the other hand, the Google API would compute the detour time the driver had to pick up the corresponding passenger. Then, we would rank these detour times from low to high.

The visualization of OD of these eight passengers and one driver are shown in Figure 6. 4:

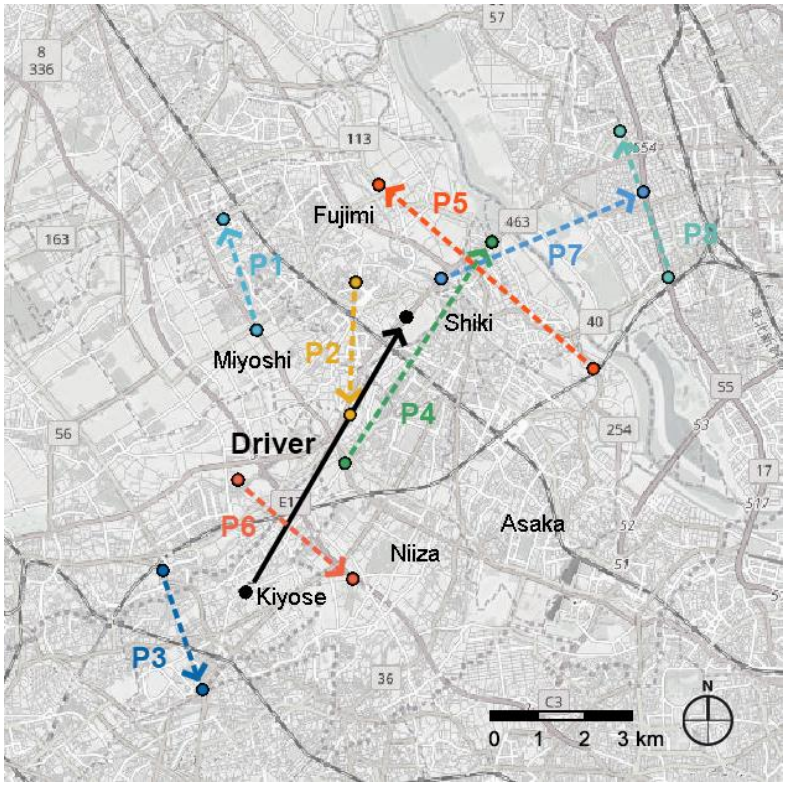


Figure 6. 4 The visualization of ODs of passengers and driver

The details of their ODs are shown in Table 6. 1:

Table 6. 1 ODs of passengers and driver

	time O	Longitude O	Latitude O	time D	Longitude D	Latitude D
Driver	6:47:00	139.526603	35.784649	7:07:00	139.564562	35.837553
P1	7:00:00	139.529303	35.83515	7:10:00	139.521191	35.856586
P2	6:42:00	139.55291	35.844473	6:57:00	139.551194	35.818874
P3	6:44:00	139.50699	35.789039	6:59:00	139.516379	35.766091
P4	6:47:00	139.584785	35.852181	7:02:00	139.550269	35.809756
P5	6:44:00	139.558051	35.862995	6:59:00	139.608957	35.827895
P6	6:49:00	139.551752	35.787352	6:59:00	139.52508	35.806289
P7	6:54:00	139.620814	35.861773	7:10:00	139.573013	35.845066
P8	6:43:00	139.615117	35.873407	6:53:00	139.626719	35.845575

The results from Google API and our deep learning model are shown in Table 6. 2:

Table 6. 2 Computation result comparison between Google API and Deep Learning Model

Ranking	Google API (detour time)	Deep Learning Model (score)
1	P6 (10 mins)	P6 (0.10935)
2	P1 (29 mins)	P1 (0.07304)
3	P3 (32 mins)	P3 (0.01724)
4	P4 (34 mins)	P4 (0.00563)
5	P7 (38 mins)	P7 (0.00557)
6	P5 (58 mins)	P2 (0.00526)
7	P2 (70 mins)	P8 (0.00229)
8	P8 (79 mins)	P5 (0.00116)

The ranking of the first five ranked passengers by the deep learning model is precisely the same as Google API. During the matching, we only choose the first one as the matched passenger to the driver.

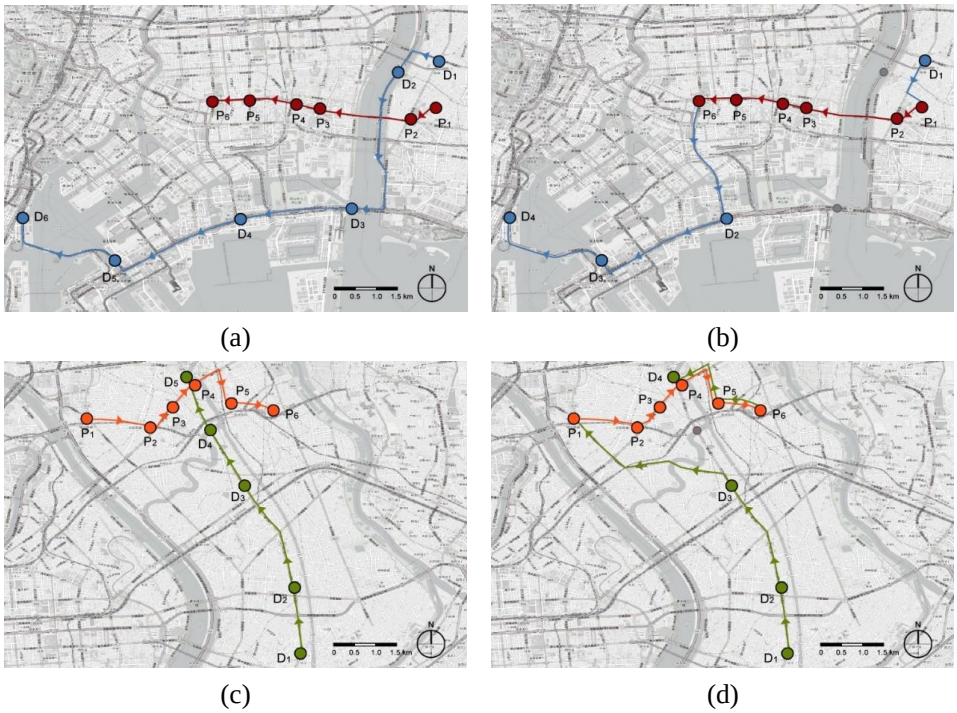


Figure 6. 5 Sample cases from matching result

a. Initial trajectories of passenger and driver in case 1 b. Matched trajectories in case 1 c. Initial trajectories of passenger and driver in case 2 d. Matched trajectories in case 2

6.3.2 Result of Matching and Spatial Analysis

In the dataset, we detected the travel mode of collected GPS trajectories, then took the first record in one complete piece of trajectory as the origin and the last destination. There are a total of 1,046,190 pieces of trajectories. This study would use the proposed deep learning model to find the best match for all these trajectories and calculate the emission and travel distance that could be potentially reduced by shared transportation. From the matching result, nearly 81.29% of trajectories, which is said to be 850,400 pieces, can be matched to another proper trajectory. Averagely, each matching pair can reduce the travel distance by 26.97% compared to their actual travel distance. It tells us that there are relatively high similarities in the travel pattern of people in Tokyo. The potential of ride-sharing in the Tokyo area is considerable. In this case, a total of over 135 million kilometers of travel distance can be reduced. Here are two samples from the results of matching to illustrate how ride-sharing saves travel distance.

Figure 6. 5 shows the probable trajectories if they are matched by ride-sharing, and their original trajectories, are shown in b and d. There are some similarities between their trajectories. In both cases, one can detour a little distance or change the initial routine to simultaneously pick the other up and complete both travels. According to the computation, if this ride-sharing matching is adopted, the travel distance can be reduced by 26.95% and 26.94%, respectively.

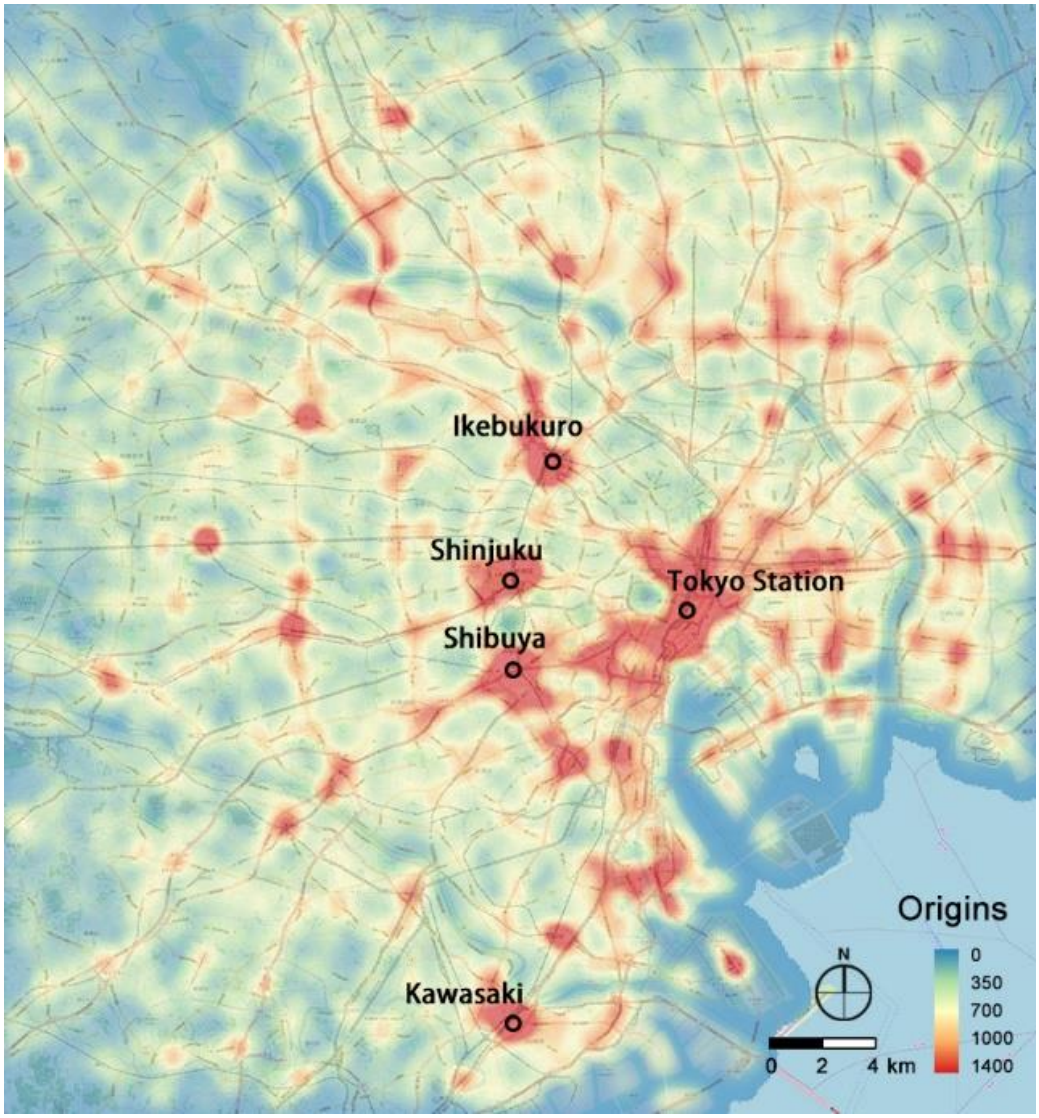


Figure 6. 6 Heat map of origins of trajectories

In order to illustrate the case study, we plot a heat map of all the origins of matchable trajectories in the study area, as shown in Figure 6. 6. We found that most of the cases were distributed in the center of Tokyo and Tokyo station, where many government offices and financial facilities are located around it and the subcenters, typically the area of Ikebukuro Shinjuku, Shibuya, and Kawazaki. They are the central business districts distributed around the city center. Commonly in Japan, more mobility flows appear in these places because of activities, such as commuting and daily shopping, thus causing more travel demands around these areas. Here is the detailed explanation. According to the investigation conducted by the Ministry of Land, Infrastructure, Transport, and Tourism, in the Tokyo area, a considerable part of people travel every day with the railway. The stations with extremely high traffic volumes include the Ikebukuro, Shinjuku, Shibuya, Tokyo, and Kawasaki station. So, these areas become a large distribution center of origins of travels. Therefore, it is evident that more ride-sharing potentials around these areas and ride-sharing service providers like Uber can consider paying more attention to these places if they want to develop ride-sharing

services in the Tokyo area.

6.3.3 Emission Analysis

After obtaining all the computation results, we can compute the emission proportion that can be reduced if users adopt ride-sharing. In the computation results, there are about 22.10% of trajectories use public transit. In the following analysis, we consider two scenarios. The first one assumes all the travelers in the original dataset traveling with private cars will all adopt ride-sharing. Half of the users who initially used public transit adopt ride-sharing. The second one assumes all the travelers in the dataset adopt ride-sharing. The computation method of emission from on-road vehicles is mainly based on the test on the CO_2 emission of on-road Japanese vehicles operated by the National Institute for Land and Infrastructure Management. After investigating their test results, I found that the CO2 emission of vehicles depends on the average spend of the entire trip. According to the result, the relationship between the average speed and CO_2 emission can be fitted as a polynomial shown as Eq. (6.8).

$$Q = \frac{1611.23}{v} + 96.62 - 1.10v + 0.01v^2 \tag{6.8}$$

where Q (g/km) is the quantity of CO_2 emission per kilometer, and v (km/h) is the average velocity

For people who adopted public transit, the emission by statistics result conducted by the Ministry of Land, Infrastructure, Transport, and Tourism show that the emission quantity of CO_2 is 19 g/km*person. Next, we will discuss the emission results into two scenarios introduced above.

Scenario 1: In this scenario, we assume that half of the original public transit riders will use the ride-sharing system, which refers to about 11.05% trajectories. Initially, these trips are completed on their original travel mode. The total emission would be 20425.2134 tons of CO_2. That would be an average of 12.0 kg CO_2 per trip. If that 11.05% of trajectories that initially used public transit turn to adopt ride-sharing and all the trajectories that originally traveled with private vehicles turn to adopt ride-sharing, the total emission would be averagely reduced by 6478.6 tons of CO_2, which would be 6.9 kg of CO_2 per trip.

Scenario 2: In this scenario, we assume that all original public transit riders will use the ride-sharing system, which refers to about 22.10% trajectories. If that 22.10% of the trajectories that initially used public transit turn to adopt ride-sharing and all the trajectories that originally traveled with private vehicles turn to adopt ride-sharing, the total emission would be averagely reduced 2179.2 tons of CO_2, which would be 2.1 kg of CO_2 per trip.

Thus, we can see that ride-sharing can truly bring emission reduction of CO_2 to city air. In our case, when half of the public transit riders turn to adopt ride-sharing and all private car drivers adopt ride-sharing, around 68.2813% of CO_2 emission can be reduced; when all the travelers adopt ride-sharing, around 89.3311% of total CO_2 emission can be reduced. Combining the result that 81.2854% of trajectories can be feasibly matched with another trajectory, there is much potential for ride-sharing in the Tokyo area.

Here, we plot the spatial distribution of CO_2 that can be reduced in Scenarios 2 by ride-sharing in Figure 6. 7.

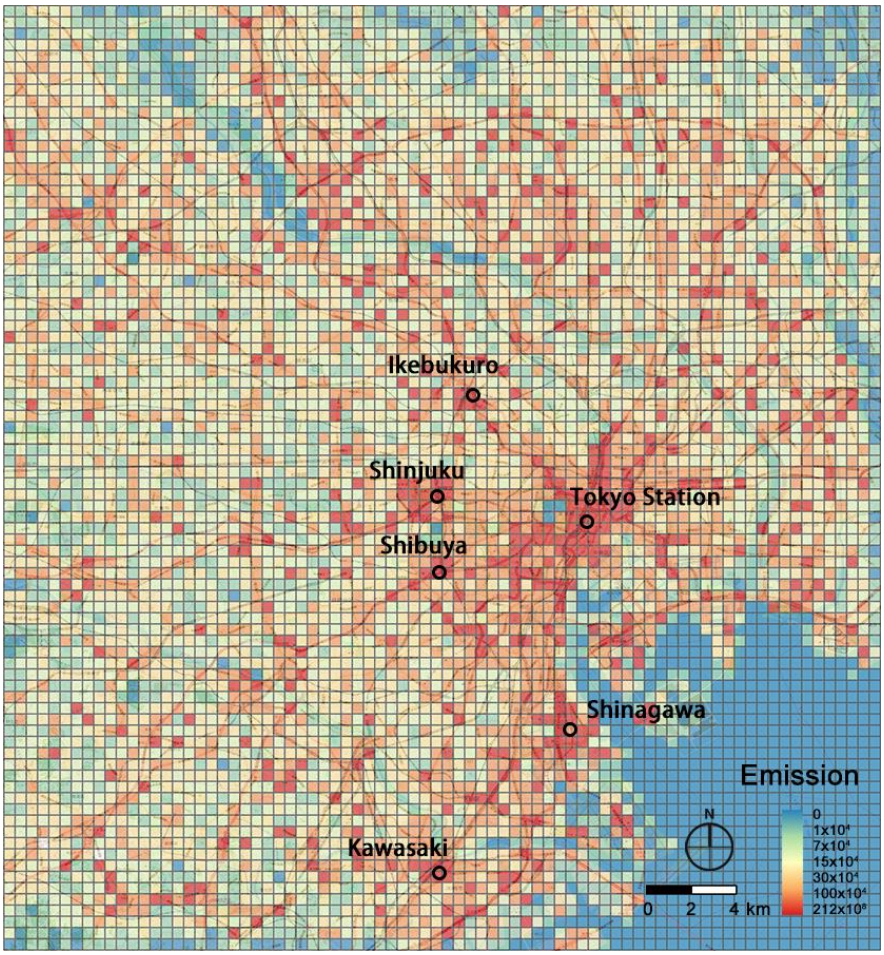


Figure 6. 7 Heat map of the distribution of the quantity of emission that can be reduced

The figure mainly shows the origins of exhaust emission and the total quantity of CO_2 that can be reduced by ride-sharing in the study area. According to the map, the main areas like Ikebukuro, Shinjuku, Shibuya, and Tokyo station can be improved much more in exhaust emission. This is because of the high concentration of travel origins in these areas explained in the previous subsection. A relatively high quantity of travels in these areas can be matched with other trips and a high potential for ride-sharing. Thus, ride-sharing can bring a lot of emission reduction caused by vehicle traffic in these areas. This also indicates that the population of ride-sharing can primarily alleviate the

heavy traffic congestion and exhaust emission in populated areas.

6.4 Conclusions

Ride-sharing based on MaaS achieves the purpose of saving energy and cutting exhaust emissions from transportation by allowing people to provide mobility services and share trips with others. The introduction of ride-sharing into places with high car exhaust emission areas can be an urgent issue. In the meantime, the service provider should also consider the regional potential of adopting ride-sharing from making a profit. Combining two factors, an analysis method for mining the potential of a place to adopt ride-sharing and exhaust emission estimation is necessary. The spread of ride-sharing in an area depends mainly on people's travel patterns over an extended period.

In this chapter, we proposed an analysis method to mine the potential of ride-sharing in an area based on quantitative historical GPS trajectories. The deep learning model in the method aims to find out how well the trajectories can be matched, thus sharing. Then we chose the case study of the Tokyo area because Tokyo is a city faced with heavy traffic congestion and exhaust emission under the background of rising usage of fossil fuel in Japan. By the computation of the deep learning model and exhaust emission, we found that travels in the Tokyo area mainly concentrated in the center and subcenters of Tokyo, like Shibuya, Shinjuku, and Ikebukuro, which makes these places the sources of exhaust emission. If ride-sharing is adopted, the exhaust emission reduction effect by ride-sharing can perform well in these areas.

Chapter 7

Bus-Sharing System: City-level Dynamic Lines Design

Bus-sharing (Customized Bus) is a new type of Internet-supported public transportation mode that can be one of the major strategies to reduce the usage of private cars and mitigate greenhouse gas emissions from road traffic. For a customized bus system, a dynamic bus line planning system based on the demand can primarily improve the performance and the public acceptance of customized bus service.

This chapter introduced a method to generate planning suggestions for bus-sharing lines and stops based on massive demand data. Using the car trajectory extracted from the mobile phone dataset as the input of the method, a case study is conducted in Tokyo and generates 29 bus-sharing lines.

The rest of this chapter is organized as follows —Section 7.1 defines the problem that is going to be solved. Section 7.2 presents the detailed methodology of this work, including the bus-sharing line extraction and hotspot analysis for bus-sharing stop deployment. Section 7.3 presents the result of the case study in Tokyo using car travel demand extracted from mobile phone data. Section 7.4 proposes the conclusions of this study.

7.1 Problem Description

A dynamic bus-sharing line planning system is a crucial element for a bus-sharing service, Figure 7. 1 shows an overview of the system. First of all, passengers will submit their demands via the online platform. After each specific period, the demand will be upload into the cloud server and aggregated into origin(O) and destination(D) as the input for the dynamic bus-sharing line planning system. Then, the system will search for the best route choice for each OD pairs. Based on the routes, the system will bus-sharing lines in the city to meet the demand. The stops of the bus-sharing lines will then be deployed so that the bus-sharing system can publish bus-sharing riding information to passengers and telling the location and time to get on the bus.

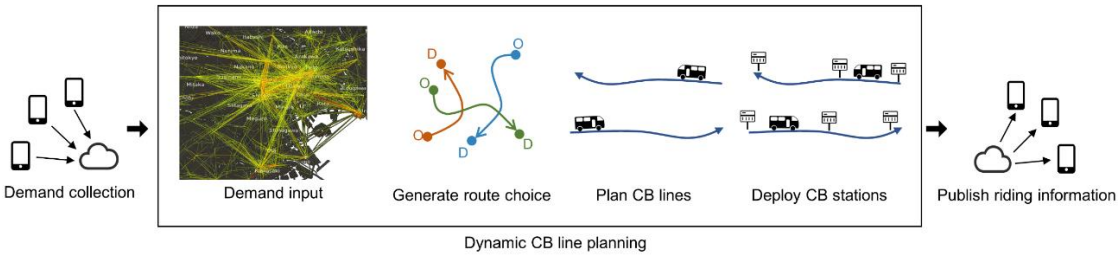


Figure 7. 1 Overview of a dynamic customized bus line planning system

In this system, the most challenging part is the bus-sharing line planning. The aim is to generate bus-sharing lines and their directions base on the massive volume of travel routes. The system has to decide how many bus-sharing lines to deploy and how is the direction of each bus-sharing line. Naturally, the more bus-sharing lines deployed as possible, the more demands will be served, but at the same time, the more of the bus-sharing lines will be uneconomical with less demand for each bus-sharing line. So the solution to this problem involves the balance between two objectives: 1. The lesser number of bus-sharing lines with higher demand assembled for each bus-sharing line yielding higher profits. 2. More demand can be effectively served.

Planning a bus line is a highly complex task in real-world situations. Many factors will affect the chosen bus line, including the difficulty of dispatch and management, the accessibility of the stop, the preference and acceptance of passengers, etc. The resulting bus lines are usually a compromise of interest from three parties: government, bus company, and passenger. Thus, to generate a result with too detailed route choice of the bus line and the exact location of bus stops from an optimization model may not be suitable in real-world situations and challenging to implement.

The methodology aims to generate suggestions of the line direction and a range of regions to set bus stops with possibilities for further detailed design. The bus line planner can further design and adjust the detailed route choice and bus stop location according to the real-world situation. The result generated by the methodology will answer the following questions: 1. the bus-sharing line will serve which part of the transportation demand and connect which part of the city? Moreover, based on that, how the direction of the bus-sharing line should be? 2. which range of areas along the bus-sharing line is suitable to arrange the bus stops?

7.2 Methodology

7.2.1 Framework

The framework of this chapter design is shown in Figure 7. 2. Mobile phone data is used here to offer suggestions for bus-sharing lines design. The input of our methodology is the demand routing trajectories of bus-sharing services. In this study,

the trajectory of car mode is identified from the mobile phone GPS data by using the data mining method introduced in our previous work [8,88]. A basic assumption is that urban car travel demand shares the same distribution as bus-sharing demand. Here, the car trajectories sampled from the dataset are used as the dynamic input for the bus-sharing line design methodology.

The dynamic, customized bus line design methodology is as follows: After matching the input trajectories into mesh grids, a link network is constructed to represent the route sharing of trajectories. The network community detection method is applied to segment the link network into communities that reflect travel demand clusters with similar traveling patterns. By detecting the core-peripheral structure of communities, the core part of communities is extracted to suggest the direction of bus-sharing lines. Boarding and alighting hotspots of potential demand are also identified to suggest choosing the location of bus-sharing stops. Finally, the bus-sharing lines' potential travel demand is identified to analyze the travel pattern and emission reduction potential.

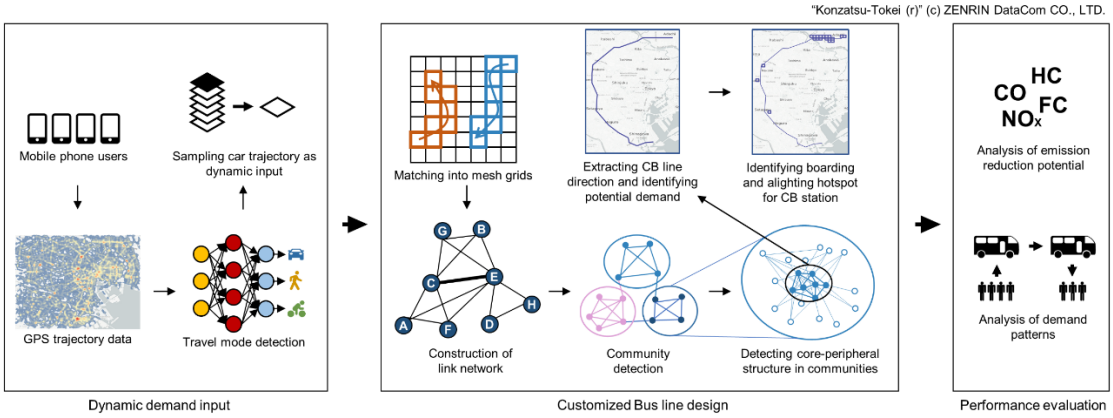


Figure 7. 2 Framework of bus-sharing lines design

7.2.2 Construction of Link Network

The basic concept of bus-sharing line design is to discover the clusters of trajectories sharing their routes. Moreover, the critical routes in each cluster indicate the direction of bus-sharing lines. To reach this goal, instead of matching and clustering the trajectories, this study aims to generate a link network to describe the sharing routes from the trajectory data set and segment the network nodes into network communities that indicate the travel demand clusters.

First of all, the spatial resolution is decided for matching the sharing route of trajectory. Here, the spatial resolution is set as 500m*500m mesh grids. Considering a set of trajectory $T = \{Traj_1, Traj_2, \ldots, Traj_{num_{tra}}\}$, each trajectory is denoted as $Traj_i = \{p_j(loc_j)\}(1 < i \leq num_{tra}, 1 \leq j \leq len_i)$. Where num_{tra} is the number of

trajectories, len_i is the length of $Traj_i$, loc_j is the index of the mesh grid representing the location of point p_j. Each trajectory can also be transformed into the set of links $Traj_i = \{l(p_1, p_2), l(p_2, p_3), \dots l(p_j, p_{j+1}) \dots l(p_{len_i-1}, p_{len_i})\}$. If we consider the links between mesh grids as vertices in a network, each trajectory indicates a set of connections between nodes. Moreover, a weighted undirected link network can be constructed. The detail of link network construction is shown in Algorithm 1.

The advantage of constructing a link network from trajectory is that the most important routes which the trajectories shared the most will be the links with the strongest connection in the link network. For example, Figure 7. 3 shows a link network constructed from three trajectories. $Traj_1$ and $Traj_2$ share the links C and E, and $Traj_1$, $Traj_2$ and $Traj_3$ share the link E. It is evident that link E is the most critical link and link C is the second important, and, respectively, vertices C and E are in the core position of the link network. Thus, analyzing the structure of the link network allows us to discover routes sharing patterns in trajectories.

Algorithm 2 Generating link network from trajectories

Input: A set of trajectories $T = \{Traj_1, Traj_2, \dots, Traj_{num_{tra}}\}$

Output: A link network $G(V, E)$

Algorithm:

1. Set the network vertex set $V = \emptyset$, and the edge set $E = \emptyset$
2. For each $Traj \in T$ do
3. For each $l(p_j, p_{j+1}) \in Traj$ do
4. If $l(p_j, p_{j+1})$ do not exist in V
5. Append $l(p_j, p_{j+1})$ into V
6. End if
7. For each $l(p_k, p_{k+1}) \in Traj, k > j$ do
8. If $l(p_j, p_{j+1}) \neq l(p_k, p_{k+1})$
9. If $edge\left(l(p_j, p_{j+1}), l(p_k, p_{k+1})\right)$ do not exist in E
10. Append $edge\left(l(p_j, p_{j+1}), l(p_k, p_{k+1})\right)$ into E, and set the weight of this edge $w = 1$
11. Else
12. Set the weight w of $edge\left(l(p_j, p_{j+1}), l(p_k, p_{k+1})\right)$ to be $w = w + 1$
13. End if
14. End if
15. Construct the link network $G(V, E)$

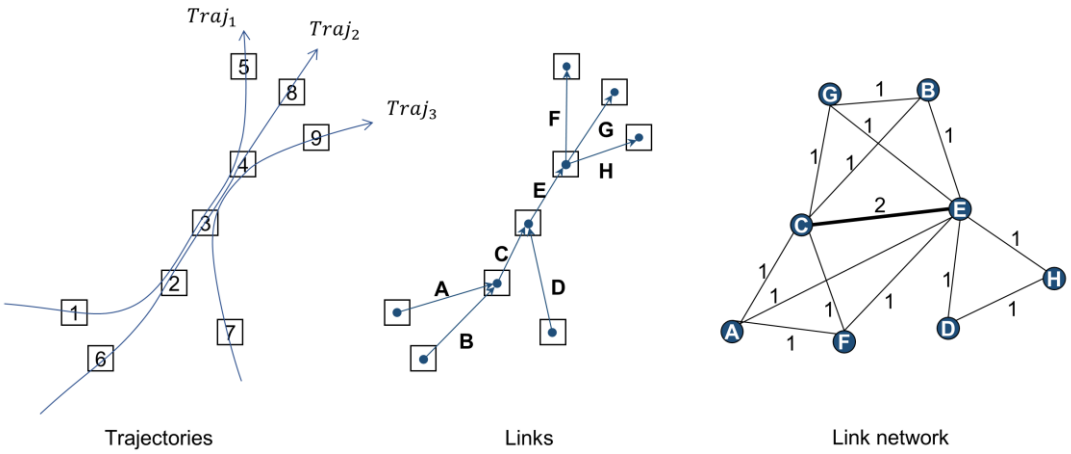


Figure 7. 3 Construction of link network

7.2.3 Community Detection

Real-world networks are usually global community structures with local core-periphery structures (such as many social networks and the World Wide Web)[115,116]. Trajectories construct the link network. In the link network constructed by trajectory data, using community detection to segment the network into densely intro-connected communities can reach the goal of discovering travel demand clusters and yield the suggestion for bus-sharing line planning.

In graph theory, a community is a collection of highly interconnected nodes. The nodes belonging to different communities are sparsely connected[89]. Studies of community structure have been very successful, and methods have been developed to find community structure. Here, the fast unfolding algorithm based on modularity optimization is adopted to decompose the link network into sub-communities. The modularity index of a partition is a scalar value between -1 and 1 that measures the density of links inside communities compared to links between communities [92]. It is defined as follows:

$$M = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \tag{7.6}$$

where A_{ij} represents the weight of the edge between i and j, k_i is the sum of the weights of the edges attached to vertex i, c_i is the community to which vertex i is assigned, the function $\delta(c_i, c_j)$ is defined as follows:

$$\delta(c_i, c_j) = \begin{cases} 1 & c_i = c_j \\ 0 & c_i \neq c_j \end{cases} \tag{7.2}$$

In order to maximize the modularity efficiently, the fast unfolding algorithm repeatedly

iterate modularity optimization and community aggregation to obtain the maximum of global modularity. For detailed information on the algorithm, please refer to [117]. We adopted the fast unfolding algorithm provided in igraph python package[118] for this study.

7.2.4 Detecting Core-Peripheral Structure in Communities

After detecting link communities from the link network, determining which nodes are part of a densely connected core and part of a sparsely connected periphery in link communities allows us to extract the critical sharing routes in the travel demand clusters. Here, the *Rombach's* algorithm is used to find continuous core-periphery structure in each community. The algorithm is a quantitative method to investigate the core-periphery structure, which computes a constant value called 'coreness' for each node in the network. The objective of the algorithm proposed 'coreness' measurement maximizes the core quality of the network. The core quality is formulated as follows:

$$R = \sum_{i,j} A_{ij} C_i C_j \tag{7.7}$$

where A_{ij} is the weight of the edge between nodes i and j, and it equals to 0 if nodes i and j are not connected and C_i denotes the local coreness of the i^{th} node, which is given by follows:

$$C_i = \begin{cases} \frac{i(1-\alpha)}{2\beta} & , i\epsilon\{1, \dots, \beta N\} \\ \frac{(i-\beta)(1-\alpha)}{2(N-\beta)} + \frac{1+\alpha}{2}, & i\epsilon\{\beta N + 1, \dots, N\} \end{cases} \tag{7.8}$$

where N is the total number of nodes, parameter α sets the size of the score jump between the highest-scoring periphery node and the lowest scoring core node and the parameter β sets the size of the core. The objective of the algorithm is to find a shuffle of node that yield C_i for each node which maximizes the core quality R of the network. To mitigate the computational cost, a label switching algorithm is implemented in the algorithm. Here, we set the parameters to be $\alpha = 1$ to classify each node to either the core or the periphery, $\beta = 0.8$, and extract the nodes with the top 20% coreness as core nodes in a community. For detailed information on the algorithm, please refer to [116].

7.2.5 Extracting bus-sharing Line Direction and Identifying Potential Demand

The densely connected core nodes of link communities are the set of links connecting mesh grids. Because car trajectories construct the link network, the core links are mainly distributed along trunk roads in the city. The geographic Information System(GIS) based approach is applied here to extract bus-sharing lines directly from the links. The extracting method contains three steps: 1. Select the communities with an a-line shape and capable of extracting bus-sharing lines. 2. Merge the core links

within a tolerance distance and simplified into a single line. In this study, the tolerance distance is set to be 500m. 3. Match the simplified line into the road network as the bus line direction.

After extracting bus-sharing lines, we set a rule to identify the potential demand for each line: 1. Generate a buffer area with a tolerance distance for each bus-sharing line. Here, we set the tolerance distance to be the same value in bus-sharing line extracting 2. For each trajectory, the proportion of the trajectory shared with the bus-sharing line can be calculated as follows:

$$p = \frac{l_{shared}}{l_{total}} \tag{7.9}$$

where l_{shared} is the length of the sub trajectory inside the buffer area, and l_{total} is the total length of the trajectory. The trajectory with the p over 80% can be regarded as the potential travel demand of the bus-sharing line, indicating that over 80% part of the trajectory is sharing the same route with the bus-sharing line.

To evaluate the operational benefits of each bus-sharing line, I propose a measure for each bus-sharing line to access the potential demand trajectories per kilometer as follows:

$$O_i = \frac{n_i}{d_i} \tag{7.10}$$

where n_i is the number of potential travel demand (trajectories here) of bus-sharing line i, d_i is the length of the bus-sharing line. A higher value of O_i indicates that with the exact operation cost, the bus-sharing line can attract more potential travel demand with better operational benefits.

7.2.6 Identifying Boarding and Alighting Hotspot for bus-sharing Stop

For determining the locations most suitable for setting the bus stop for each bus-sharing line, it is significantly helpful to identify the spatial hotspot clusters of the potential boarding and alighting location. After extracting the potential travel demand of bus-sharing lines, the first point and the last point of the sub trajectory inside the buffer area can be regarded as the potential boarding and alighting location of this trajectory. Here, I apply the local Moran's I index to examines the boarding and alighting locations [94], enabling hotspots to be identified based on a comparison with the neighboring samples.

The high positive local Moran's I index indicates that the number of potential bicycle-sharing behaviors in the grid has similarly high or low values as its neighbors. The local

Moran's I index can identify two types of spatial clusters: high-high clusters (high values in a high-value neighborhood) and low-low clusters (low values in a low-value neighborhood); And two types of outliers: Spatial outliers include high–low (a high-value in a low-value neighborhood) and low–high (a low value in a high-value neighborhood). In the scenario of bus-sharing line boarding and alighting hotspot identification, the high-high clusters and high–low clusters are the “regional hotspots” and “individual hotspots” with significantly higher demand for boarding and alighting along the bus-sharing lines but with different demand patterns.

7.2.7 Emission Reduction Potential Model

In order to estimate the potential of emission reduction for each bus-sharing line, an assumption is presented: assuming that all the potential travel demand can be served by bus-sharing travel mode, the potential reduction of emission for bus-sharing lines will be the emission of all the potential car travel in the area.

Here, COPERT (Computer Programme to calculate the Emissions from Road Transport) model is adapted to calculate fuel consumption (FC) and emissions of carbon monoxide (CO), nitrogen oxides (NOx), and hydrocarbon (HC). COPERT is a widely used emission model developed by the European Environment Agency (EPA). Based on distinguishing vehicle categories, fuel types, road categories, and other parameters, COPERT model determines the emissions of different pollutants and FC by adopting regression analysis for speeds and traveling distance of vehicles [95,96]. For the detail of the emission model, please refer to [95]. Adding up the FC and emissions generated by each bus line's potential car trips, the potential FC and emission reduction can be calculated.

7.3 Results and Discussion

7.3.1 Data and Study Area

Therefore, after identifying the travel modes, we sample the car trajectory data as the simulation of dynamic demand to input our algorithm. Figure 7. 4 shows the spatial distribution of car trajectory data. The data set comprises 1,409,451 car trips. We implement the methodology on the platform with Intel i7-8650U CPU and 16GB RAM. For a one-day sample data taken from the dataset (32,647 car trajectories in total), our algorithm takes 18.55s to construct the link network, 9.36s to community detection and 43.57s core-peripheral structure of the network. The total computation time to generate the result is approximately 1 minute.

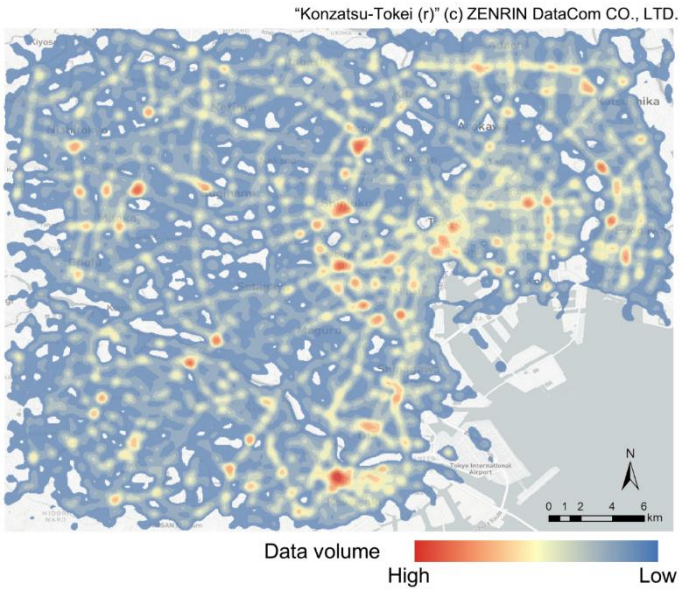


Figure 7. 4 Spatial distribution of trajectory dataset

7.3.2 Result of Community Detection and Core-Peripheral Structure in Communities

From the car trajectory dataset, a link network with 189,829 nodes and 2,336,738 edges is generated. By applying the fast unfolding algorithm on the link network, the algorithm produces community segmentation with the modularity of 0.73, indicating that there is a community structure in the link network. As a result of community segmentation, there are 10,313 communities in total, 851 communities have over ten links, and only 152 communities have over 100 links. The complementary cumulative distribution function (CCDF) of the number of links in link communities is shown in Figure 7. 5. The CCDF curve decreases quickly in the small number of links, indicating that a relatively small number of communities contain a significantly large number of links and play an essential role in the link network. Examine the core-peripheral structure of the large communities. Most of the core links of link communities are in a line shape and suitable to extract bus lines. Figure 7. 6 shows two examples of the core-peripheral structure of link communities and the bus line extracted.

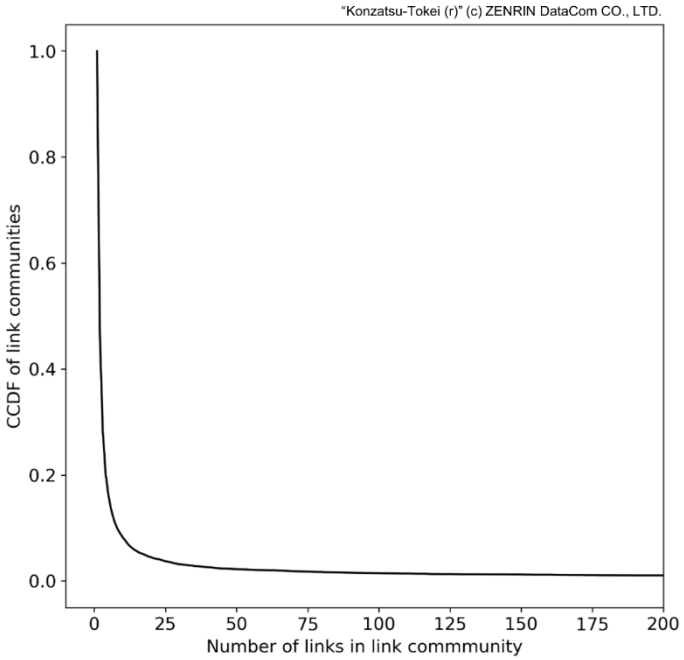


Figure 7. 5 Complementary cumulative distribution function (CCDF) of the number of links for link communities

(a)

(b)

Figure 7. 6 Example of Link communities and bus-sharing line extracted

7.3.3 Result of bus-sharing Line Direction and Bus Stop Hotspots

From the core-peripheral structure of 152 link communities with over 100 links, 29 bus-sharing lines are extracted in Tokyo. The average length of the bus-sharing lines is 18.7 km, with the longest of 38.1 km and the shortest of 11.6 km. The number of bus-sharing lines is less than the number of link communities, and the reason is that the bus-sharing lines extracted from some of the small communities are much shorter. Most of them are in the same direction as the larger communities. The travel demand from these small link communities can be satisfied by the bus-sharing lines extracted from large communities if the location of bus stops are appropriately arranged. The core-peripheral structure of link communities that extract the 29 bus-sharing lines are shown in Supplementary materials A.

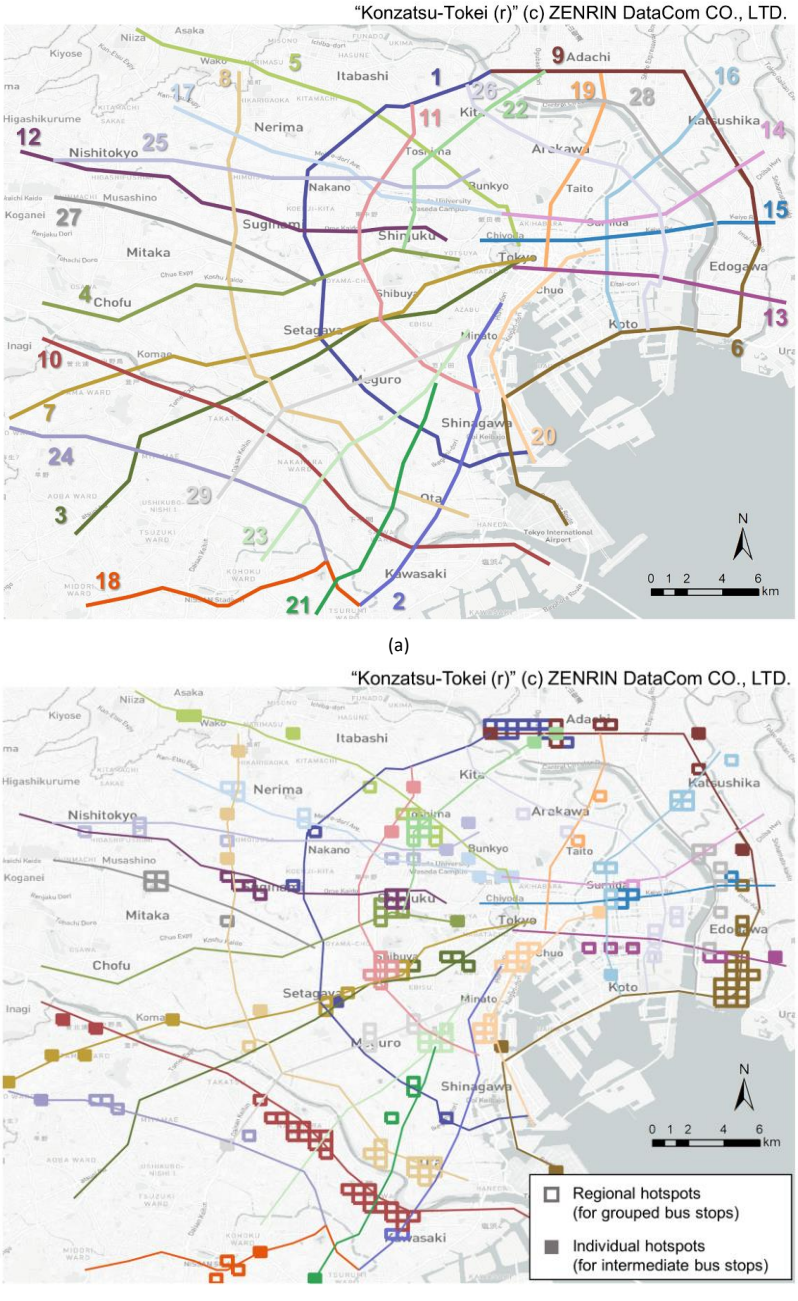
After extracting bus-sharing lines, we identify the potential demand for each line. The bus-sharing lines are named in descending order from 1 to 29 according to the potential demand. Figure 7. 7 (a) shows the spatial distribution of bus-sharing lines direction. According to the shape and spatial location of the bus-sharing lines, they can be classified into three types:

- Radiation type lines: This type of bus-sharing line include line 2, 3, 4, 5, 7, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, 25 and 29. These bus-sharing lines extend from the city center to suburban area in all directions. They offer an efficient transportation method for passengers from some specific suburban residential areas to rapidly enter the city center or sub-center. These bus-sharing lines are mostly in the straight-line shape as the shortest path from the origin to the destination, ensuring the efficiency of the bus-sharing lines.
- Ring-type lines: This type of bus-sharing line include line 1, 6, 8, 9, 11, 22, 26 and 28. This bus-sharing line is on the edge of the central urban district connecting subcenters of the city.
- Suburban lines: This type of bus-sharing line include line 10,18,24 and 27. These bus-sharing lines are connecting multiple suburban centers.

These three types of bus-sharing lines are reasonable and in line with bus-sharing lines as a supplement in the urban transportation system. Notice that the presenting bus-sharing lines network does not connect multiple Central Business Districts(CBDs) in one line. The reason is that urban subway lines can well serve the travel demand from one CBD to another.

Figure 7. 7 (b) shows the regional hotspots and individual hotspots identified from boarding and alighting potential demand from each bus-sharing line. Regional hotspots indicate the cluster of high-demand locations along the bus-sharing line, which can be the location for grouped bus stops. Grouped stops in a range of areas can assemble

passengers and reduce their walk distance to the bus stops; Individual hotspots indicate a high demand location in a low demand neighborhood. Intermediate bus stops can be arranged in these hotspots. Arranging these two types of stops according to the hotspots can ensure the satisfaction of demand without devastating the efficiency and directness of the bus-sharing line.



(a)

(b)

Figure 7. 7 Bus-sharing lines planning suggestions from link communities

(a) Direction of 29 bus-sharing lines extracted (b) Hotspot identified for bus stop

7.3.4 Evaluation of Travel Demand for bus-sharing Lines

To sum up the number of potential demand, there are 290,465 trajectories in total, which

can be potentially replaced by the bus-sharing travel mode (20.6% of the total number of all car trajectories and 13.1% of the total mileage).

Figure 7. 8 shows the operation benefit of O_i of each bus-sharing line. The O_i of bus-sharing lines are around 400 to 600 trajectories per kilometer, indicating that with proper line detail design and operation management, the bus-sharing lines will have a similar level of operational performance. Among all the bus-sharing lines, line 2 is the one with the highest operation benefit, connecting Kawasaki city directly to the city center.

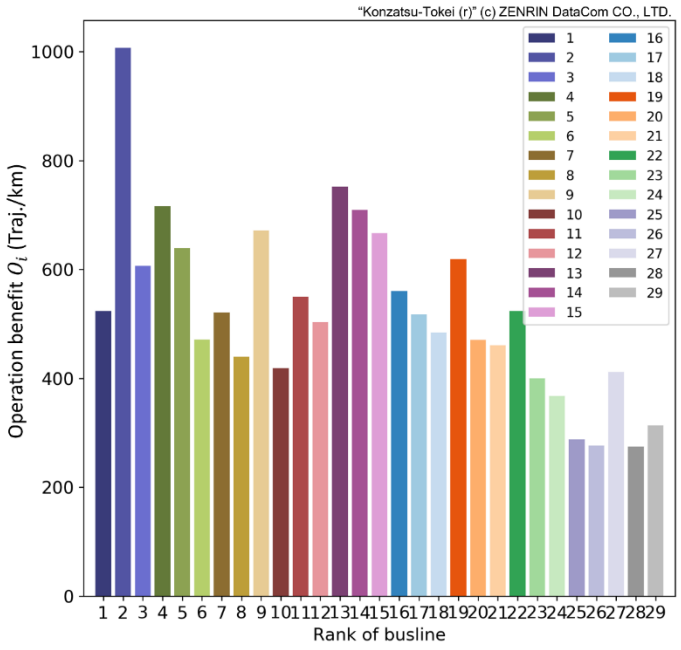


Figure 7. 8 Operation benefit of each bus-sharing line

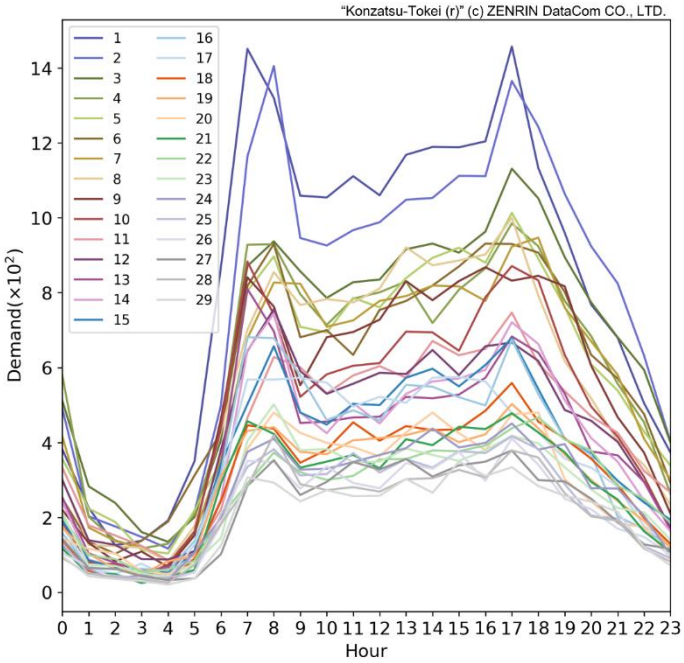
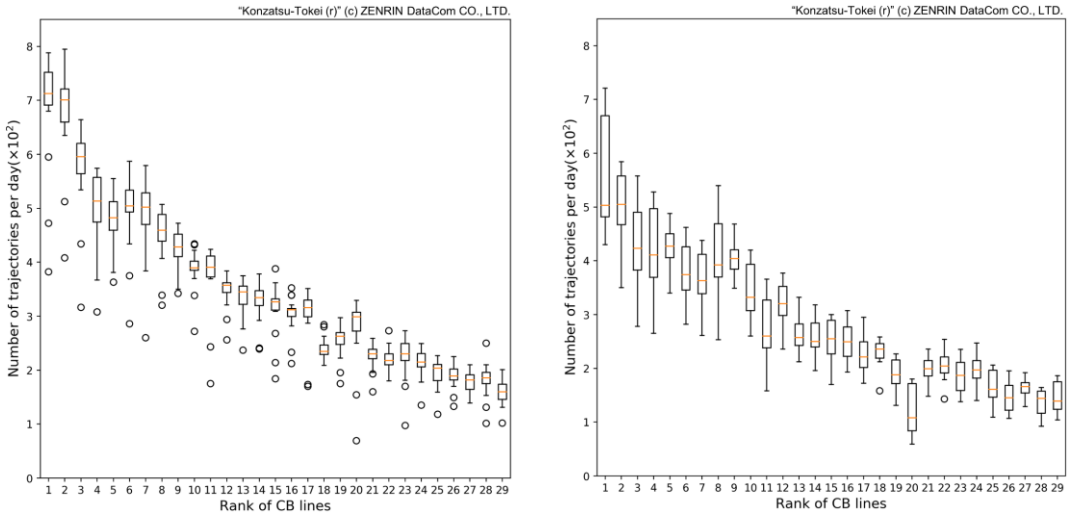


Figure 7. 9 Hourly demand change of bus-sharing lines

Figure 7. 9 shows the hourly demand change of the bus-sharing lines, which can suggest the bus schedule arrangement and pricing strategy. In the hourly demand, the morning peak is around 6:00 to 8:00, and the evening peak is around 16:00 to 18:00. However, the hourly demand is distributed differently among bus-sharing lines. One of the demand patterns is the morning peak and evening peak can be identified, including bus-sharing lines 1, 2, 3, 4, 5, 6, 10, 13, 14, 15, 16, 20, 21, and 28. These bus-sharing lines are in a tide traffic demand pattern with a high proportion of commuting trips. On the contrary, the morning and evening peak hours of bus-sharing lines 7, 8, 9, 11, 12, 17, 18, 19, 22, 23, 24, 25, 26, 27, and 29 are not precise, with high demand during all day. This pattern indicates that the travel demand of these lines may vary with less repeatability.



(a) Weekdays

(b) Weekends & Holidays

Figure 7. 10 Number of potential demand trajectories per day for each bus-sharing lines

Figure 7. 10 shows the box plot of the number of potential demand for each bus line with a comparation of weekdays and weekends & holidays. The potential demand fluctuates within a small range, indicating that the bus-sharing lines will have sustainable demand. Comparing the weekdays with weekends & holidays, in general, the potential demand for bus-sharing lines are more significant and more stable on weekdays. Travel demand on weekdays is more stable with high repeatability.

7.3.5 Result of Emission Reduction Potential

Based on the speed and travel length of potential trips, the FC and emission are calculated. The potential FC and emission reduction can be calculated by summing the emission produced by the car trajectories covered by each bus-sharing line. Figure 7. 11 shows the summing of the proportion of the potential emission reduction of each bus-sharing line to the total car emission in the whole city. The result shows that by implementing a total of 29 bus-sharing lines, there will be 13.6%, 13.4%, 13.0%, and

12.8% potential reduction of NOX, FC, HC, and CO.

Figure 7. 12 shows the spatial distribution of the emission reduction potential. The spatial heat map shows the relative values ranging from 0 to 1, compared to the maximum volume, averaged by four types of emissions. As is shown in Figure 7. 12, a large part of potential emissions handled by bus-sharing lines is distributed on the periphery area of the city center, especially the southwest part of the city. Most emissions occur within the Central Business Districts(CBD), such as Setagaya, Shibuya, Shinagawa, and Toshima. This distribution pattern is that the three types of bus-sharing lines developed by our methodology can serve the travel demand from suburban areas to CBD areas in the city center. Such travel highly depends on the urban expressway. Therefore, bus-sharing lines can potentially reduce emission pressure on urban expressways by replacing private car travel.

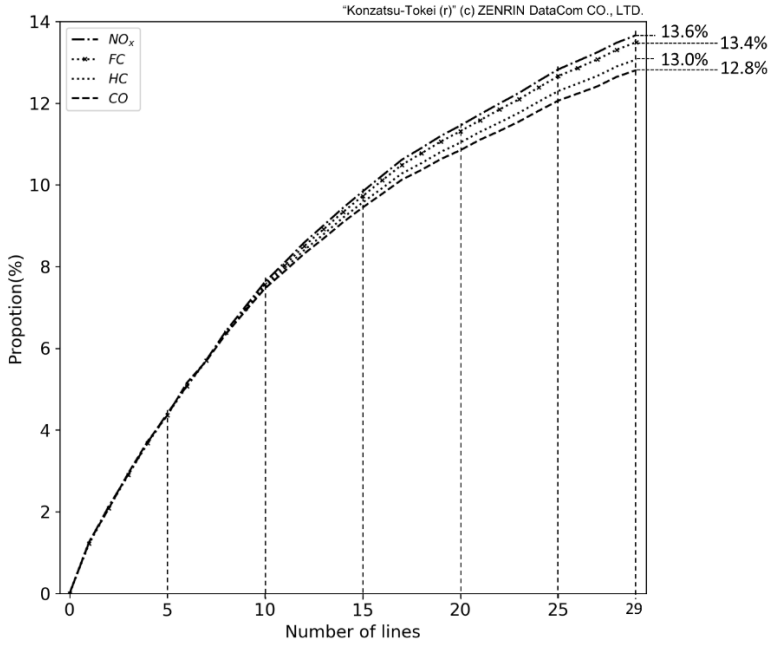


Figure 7. 11 Estimation of potential emission reduction of each bus-sharing lines

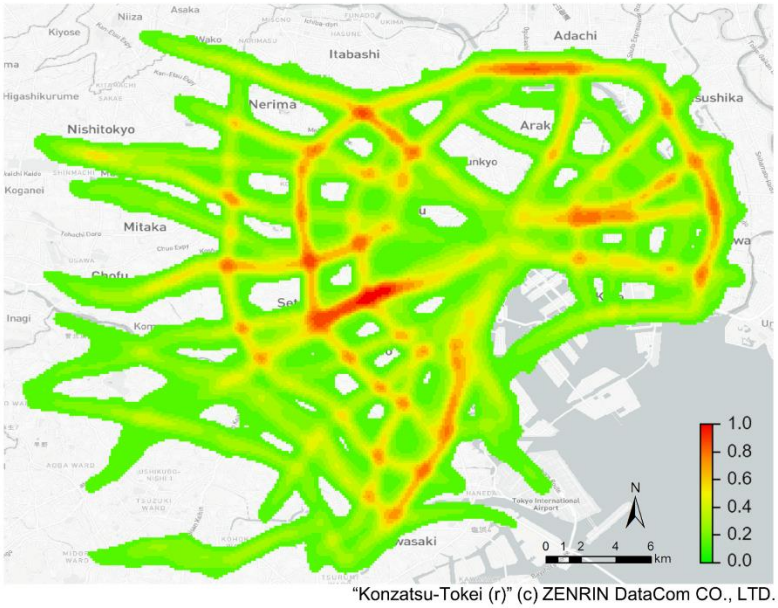


Figure 7. 12 Spatial distributions of emission reduction potential

7.4 Conclusions

As a new type of Internet-based transportation mode in the public transportation system, the bus-sharing system has a high potential to replace the private car and reduce road traffic GHG emissions in the urban area. However, designing a demand-oriented and efficient system is critical and highly related to the bus-sharing system's public acceptance and operation performance. Under this background, this chapter proposes a methodology to generate planning suggestions for bus-sharing lines and stops based on massive demand data with high computing speed capable of integrating into a dynamic bus-sharing planning system. The proposed method involving the following steps:

The car trajectories are extracted from mobile phone data as the input of potential bus-sharing travel demand. From the demand input, a link network is constructed to represent the sharing route of the demand. Community detection is applied on the link network to segment the link network into network communities with similar travel routes, and the core-peripheral structure of each community is examined. The bus-sharing lines are generated by extracting the core part of the link network communities and match into the road network. Potential demand for bus-sharing lines is identified, and boarding and alighting hotspots are extracted as the suggestion for bus-sharing bus stops. A case study is conducted using mobile phone data in Tokyo. 29 bus-sharing lines are extracted. According to the shape and spatial location of the bus-sharing lines, three types of bus-sharing lines serving different travel patterns are classified, including radiation type lines, ring-type lines, and suburban lines. By analyzing the emission reduction potential of the bus-sharing lines extracted, the bus-sharing lines generated by the proposed method can reduce emission pressure on urban expressways and mitigate approximately 13% of road traffic emissions.

Chapter 8

Conclusions and Future Directions

This study reviews some state-in-art articles on MaaS regardless of policy, transportation, energy, computer science, economics, and even society filed journals. In that case, we could find that utilizing big data to analyze their hot issues is a noticeable trend. More and more researchers realized that big data could provide solid supports for their theories and conclusion. However, due to the technology gaps between data mining and other research backgrounds, how to scientifically and effectively utilize the data becomes an obstacle to the researchers.

Due to the lack of works that summarized the development and frontiers of this new field, researchers are hard to get comprehensive and high-dimensional information about TST. This research aims to focus on the following questions: a) how to define and reinvent data-driven mobility models by studying urban dynamics, urban mobility, transportation behavior, and sharing potential. b) within a city-level urban mobility framework, how can we characterize the nature of data-enabled shared transportation services among different modes, and what are the similarities and differences. c) the existing positive and successful STSs that can be identified in the studied domain and how they can be best applied for practical success.

To answer these questions, we took three shared transportation modes: bicycle-sharing, ride-sharing, bus-sharing, as the study cases and tried to discuss the issue of STS under the framework of MaaS, and developed a series of methods. The contributions of this study include:

For bicycle-sharing system: market-oriented sub-area division

- Using a dataset generated by real-world travel flow, the data-based methodology ensures the credibility of the link network constructed and the market-oriented subarea division.
- The market-oriented sub-divisions of the coverage area are suggested by the community detection method, which generates the sub-communities by maximizing the connections inside communities and minimizing connections between communities. The sub-regions generated by this method can primarily reduce the effort of bicycle rebalancing.
- The hotspot identification and the network indicators in the subarea can identify places with frequent bicycle-sharing handovers and their traveling patterns, giving suggestions for constructing bicycle-friendly facilities and infrastructures.
- The emission reduction potential analysis will hopefully help promote bicycle-sharing services to replace car trips and reduce the total emission in the city.

For bicycle-sharing system: layout optimization

- The demand prediction is strictly based on mobility information, which ensures the credibility of the input parameters for the optimization model;
- Uncertainty concerning site-specific construction conditions is taken into consideration, which is rarely discussed in previous studies and hence facilitates more practical solutions for real-world bicycle-sharing projects;
- An integrated method is proposed which combines the layout optimization and the rebalance optimization for bicycle-sharing systems;
- The method enables multi-sided sensitivity analysis of the potential for bicycle-sharing to replace walking and vehicular trips. This can inform further research on potential emissions reduction analysis.

For ride-sharing system: city-level potential analysis

- The potential analysis of ride-sharing is based on the actual travel demand, which is more precise and reliable compared with the base of urban population distribution;
- This study proposes a deep learning-based method for real-time matching feasibility estimation;
- Based on the proposed high-effective matching algorithm, the big data-based potential and emission reduction analyses are carried out;
- Tokyo is taken as the case study with mining millions of trajectories data of mobile phone users in one year.

For bus-sharing system: city-level dynamic lines design

- The method proposed can automatically generate bus line direction and stops location suggestions based on massive demand data input, which is more demand-oriented and efficient than the existing traditional bus-sharing line planning method.
- The algorithm can process massive demand data and generate results in time, which has a high potential of integrating into a dynamic bus-sharing line planning system.
- By analyzing the case study in Tokyo, the bus-sharing lines generated by the proposed method have the potential to reduce emission pressure on urban expressways and to mitigate approximately 13% of road traffic emissions.

In our analyses, only travel distance and travel mode were taken as the indicators to screen the potential shared transportation users. The individual adoption attitudes on each shared transportation mode were not considered. To evaluate the impacts of this introduced assumptions on the results, in future studies, field investigations (questionnaire investigations) on the adoption attitudes with demographic information fusion would be carried out to enable a more comprehensive conclusion. The future quantitative studies will be significant for building infrastructure, policymaking, potential market measurement, and further detailed analysis of environmental benefits.

For bicycle-sharing system: market-oriented sub-area division

An optimization model can be considered based on the sub-area to give strategies and quantitative analysis for the accurate bicycle rebalancing demand. Our method only captures the possibility of bicycle-sharing but does not simulate the bicycle-sharing trips in a city. A simulation system for bicycle-sharing services considering human mobility and trip chain can be built to enable a more comprehensive conclusion in future studies. The quantitative studies are significant for building infrastructure, policymaking, potential market measurement, and further detailed analysis of emission reductions.

For bicycle-sharing system: layout optimization

Data acquisition in this study was affected by several factors, including loss of signal or battery power and the difficulty of discriminating between public and private vehicle travel trajectories based on mobile phone GPS data. In addition, many further studies are still required to enable a more comprehensive conclusion. For example, the vehicle routing problem should be considered in the rebalancing model, and a detailed survey is required on people's attitudes to adopting bicycle-sharing as a travel mode. Such quantitative studies will play essential roles in infrastructure development, policymaking, and further detailed quantitative analysis of emission reductions. In the emission reduction potentiality analysis, a complete life-cycle assessment of bicycle-sharing is not included in the current analysis due to data imperfection. The offset emission volume could be increased by introducing more influential factors such as biking recycling and infrastructure investment. In future studies, a complete life-cycle of the bicycle-sharing system is expected to conduct to provide more accurate environmental information to effective operation and promotion of the bicycle-sharing system.

For ride-sharing system: city-level potential analysis

In our analysis, the individual attitude on ride-sharing was not considered. This limitation may introduce an error in the results. Nevertheless, the analysis method I proposed is a configuration tool. A complete result can be obtained via our framework with a further work of attitude survey on ride-sharing.

For bus-sharing system: city-level dynamic lines design

The bus and crew dispatching model can be developed based on the bus lines generated from our method. In the real-world application, an integrated bus-sharing system with a dynamic bus line planning system, dispatching system, and real-time controlling system can be developed with our proposed method as a critical element.

Bibliography

1 Li, P., Zhao, P. & Brand, C. Future energy use and CO2 emissions of urban passenger transport in China: A travel behavior and urban form based approach. *Applied Energy* **211**, 820-842, doi:https://doi.org/10.1016/j.apenergy.2017.11.022 (2018).

2 Cerutti, P. S., Martins, R. D., Macke, J. & Sarate, J. A. R. "Green, but not as green as that": An analysis of a Brazilian bike-sharing system. *Journal of Cleaner Production* **217**, 185-193, doi:https://doi.org/10.1016/j.jclepro.2019.01.240 (2019).

3 Qiu, L.-Y. & He, L.-Y. Bike sharing and the economy, the environment, and health-related externalities. *Sustainability* **10**, 1145 (2018).

4 Shaheen, S. & Chan, N. Mobility and the sharing economy: Potential to facilitate the first-and last-mile public transit connections. *Built Environment* **42**, 573-588 (2016).

5 Liu, L., Sun, L., Chen, Y. & Ma, X. Optimizing fleet size and scheduling of feeder transit services considering the influence of bike-sharing systems. *Journal of Cleaner Production* **236**, 117550, doi:https://doi.org/10.1016/j.jclepro.2019.07.025 (2019).

6 Yu, D.-s. & Shang, L.-c. Opportunities and Challenges Faced by Share Economy: Taking Sharing Bicycle as an Example. *DEStech Transactions on Environment, Energy and Earth Sciences* (2017).

7 Ma, Y., Lan, J., Thornton, T., Mangalagiu, D. & Zhu, D. Challenges of collaborative governance in the sharing economy: The case of free-floating bike sharing in Shanghai. *Journal of Cleaner Production* **197**, 356-365, doi:https://doi.org/10.1016/j.jclepro.2018.06.213 (2018).

8 Zhang, H. *et al.* Mobile phone GPS data in urban bicycle-sharing: Layout optimization and emissions reduction analysis. *Applied Energy* **242**, 138-147, doi:https://doi.org/10.1016/j.apenergy.2019.03.119 (2019).

9 Zhang, L., Zhang, J., Duan, Z.-y. & Bryde, D. Sustainable bike-sharing systems: characteristics and commonalities across cases in urban China. *Journal of Cleaner Production* **97**, 124-133 (2015).

10 García-Palomares, J. C., Gutiérrez, J. & Latorre, M. Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography* **35**, 235-246, doi:https://doi.org/10.1016/j.apgeog.2012.07.002 (2012).

11 Hietanen, S. Mobility as a Service. *the new transport model*, 2-4 (2014).

12 Sochor, J., Strömberg, H. & Karlsson, M. A. Implementing Mobility as a Service: Challenges in Integrating User, Commercial, and Societal Perspectives. *Transportation Research Record Journal of the Transportation Research Board* **No. 2536, Vol. 4**, 1-9 (2015).

13 Sui, Y. *et al.* GPS data in urban online ride-hailing: A comparative analysis on fuel consumption and emissions. *Journal of Cleaner Production* **227**, 495-505 (2019).

14 Sochor, J., Karlsson, M. A. & Strömberg, H. Trying Out Mobility as a Service: Experiences from a Field Trial and Implications for Understanding Demand. *Transportation Research Record Journal of the Transportation Research Board* **No. 2542**, 57-64 (2016).

15 Henderson, A., Patty, G. & Herzberg, A. Benefits of Sharing Economies of Ride-Sharing. *IEEE Transactions on Industry Applications* **51**, 1920-1927 (2015).

16 Teubner, T. & Flath, C. M. The Economics of Multi-Hop Ride Sharing Creating New Mobility Networks Through IS. *Business & Information Systems Engineering* **57**, 311-324 (2015).

17 Meng, F. *et al.* Energy efficiency of urban transportation system in Xiamen, China. An integrated approach. *Applied Energy* **186**, 234-248 (2017).

18 Sun, H., Wang, H. & Wan, Z. Model and analysis of labor supply for ride-sharing platforms in the presence of sample self-selection and endogeneity. *Transportation Research Part B: Methodological* **125**, 76-93 (2019).

19 Dong, Y., Wang, S., Li, L. & Zhang, Z. An empirical study on travel patterns of internet based ride-sharing. *Transportation research part C: emerging technologies* **86**, 1-22 (2018).

20 Zhong, Y., Lin, Z., Zhou, Y.-W., Cheng, T. & Lin, X. Matching supply and demand on ride-sharing platforms with permanent agents and competition. *International Journal of Production Economics* **218**, 363-374 (2019).

21 Jin, S. T., Kong, H., Wu, R. & Sui, D. Z. Ridesourcing, the sharing economy, and the future of cities. *Cities* (2018).

22 Li, P., Zhao, P., Brand, C. & Yan, J. Future energy use and CO2 emissions of urban passenger transport in China: A travel behavior and urban form based approach. (2018).

23 Tachet, R. *et al.* Scaling law of urban ride sharing. *Scientific reports* **7**, 42868 (2017).

24 Long, J., Tan, W., Szeto, W. & Li, Y. Ride-sharing with travel time uncertainty. *Transportation Research Part B: Methodological* **118**, 143-171 (2018).

25 Liu, T. & Ceder, A. Analysis of a new public-transport-service concept: Customized bus in China. *Transport Policy* **39**, 63-76, doi:https://doi.org/10.1016/j.tranpol.2015.02.004 (2015).

26 Piramuthu, O. B. & Zhou, W. in *2016 49th Hawaii International Conference on System Sciences (HICSS).* 2078-2083 (IEEE).

27 Lin, L., He, Z. & Peeta, S. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies* **97**, 258-276 (2018).

28 Sundfør, H. B. & Fyhri, A. A push for public health: the effect of e-bikes on physical activity levels. *BMC public health* **17**, 809 (2017).

29 El-Assi, W., Mahmoud, M. S. & Habib, K. N. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation* **44**, 589-613 (2017).

30 Giot, R. & Cherrier, R. in *2014 IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems (CIVTS).* 22-29 (IEEE).

31 Li, Y., Zheng, Y., Zhang, H. & Chen, L. in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems.* 33 (ACM).

32 Chen, Q. & Sun, T. A model for the layout of bike stations in public bike-sharing systems. *Journal of Advanced Transportation* **49**, 884-900 (2015).

33 Rybarczyk, G. & Wu, C. Bicycle facility planning using GIS and multi-criteria decision analysis. *Applied Geography* **30**, 282-293 (2010).

34 Landis, B. W. Bicycle system performance measures. *ITE journal* **66**, 18-26 (1996).

35 Kabak, M., Erbaş, M., Çetinkaya, C. & Özceylan, E. A GIS-based MCDM approach for the evaluation of bike-share stations. *Journal of cleaner production* **201**, 49-60 (2018).

36 Lopez Gonzalez, L. Optimal Location for Bike Sharing Stations in Downtown Kalamazoo. (2016).

37 Xia, X. & Wu, J. in *2nd International Conference on Humanities Science and Society Development (ICHSSD 2017).* (Atlantis Press).

38 DOCOMO BIKE SHARE, I. *bike share service*, <http://docomo-cycle.jp/?lang=ja> (2019).

39 Frade, I. & Ribeiro, A. Bike-sharing stations: A maximal covering location approach. *Transportation Research Part A: Policy and Practice* **82**, 216-227, doi:https://doi.org/10.1016/j.tra.2015.09.014 (2015).

40 Long, Y. *et al.* Unequal age-based household emission and its monthly variation embodied in energy consumption–A cases study of Tokyo, Japan. *Applied energy* **247**, 350-362 (2019).

41 Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779-782 (2008).

42 Xia, T. *et al.* Measuring spatio-temporal accessibility to emergency medical services through big GPS data. *Health & Place* **56**, 53-62, doi:https://doi.org/10.1016/j.healthplace.2019.01.012 (2019).

43 Shaheen, N. D. C. & A., S. Ridesharing in North America: Past, Present, and Future. *Transport Reviews* **32**, 93-112 (2012).

44 Wolfson, O., Zheng, Y. & Ma, S. in *IEEE International Conference on Data Engineering.* 410-421.

45 Carrese, S., Giacchetti, T., Patella, S. M. & Petrelli, M. in *IEEE International Conference on MODELS and Technologies for Intelligent Transportation Systems.* 721-726.

46 Xu, H., Ordóñez, F. & Dessouky, M. A traffic assignment model for a ridesharing transportation market. *Journal of Advanced Transportation* **49**, 793-816 (2015).

47 Na, T. *et al.* An Efficient Ride-Sharing Framework for Maximizing Shared Route. *IEEE Transactions on Knowledge & Data Engineering* **PP**, 1-1 (2018).

48 Belz, N. P. & Lee, B. H. Y. Composition of Vehicle Occupancy for Journey-to-Work Trips: Evidence of Ridesharing from the 2009 National Household Travel Survey Vermont Add-on Sample. *Transportation Research Record Journal of the Transportation Research Board* **2322**, 1-9 (2012).

49 Agatz, N., Erera, A., Savelsbergh, M. & Wang, X. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research* **223**, 295-303 (2012).

50 Ghoseiri, K., Haghani, A. & Hamedi, M. Real-Time Rideshare Matching Problem. *Optimization* (2010).

51 Giannantonio, R., Claudio, B., Gargiulo, E., Guercio, E. M. & Zenezini, G. in *International Conference on Applied Human Factors and Ergonomics.* 777-784.

52 Stiglic, M., Agatz, N., Savelsbergh, M. & Gradisar, M. Enhancing urban mobility: Integrating ride-sharing and public transit. *Computers & Operations Research* **90**, 12-21 (2018).

53 Mourad, A., Puchinger, J. & Chu, C. A survey of models and algorithms for optimizing shared mobility. *Transportation Research Part B: Methodological* (2019).

54 Amey, A.

55 Bei, X. & Zhang, S. in *Thirty-Second AAAI Conference on Artificial Intelligence.*

56 Singh, A., Alabbasi, A. & Aggarwal, V. A Distributed Model-Free Algorithm for Multi-hop Ride-sharing using Deep Reinforcement Learning. *arXiv preprint arXiv:1910.14002* (2019).

57 Chen, R. & Cassandras, C. G. Optimization of ride sharing systems using event-driven receding horizon control. *arXiv preprint arXiv:1901.01919* (2019).

58 Naoum-Sawaya, J. *et al.* Stochastic optimization approach for the car placement problem in ridesharing systems. *Transportation Research Part B Methodological* **80**, 173-184 (2015).

59 Li, B., Krushinsky, D., Woensel, T. V. & Reijers, H. A. The Share-a-Ride problem with stochastic travel times and stochastic delivery locations. *Transportation Research Part C Emerging Technologies* **67**, 95-108 (2016).

60 Huang, H. J., Yang, H. & Bell, M. G. H. The models and economics of carpools. *Annals of Regional Science* **34**, 55-68 (2000).

61 Ma, R. *et al.* Greenhouse gas emission savings with dynamic ride-sharing. *Revista De La Facultad De Ingenieria* **31**, 152-162 (2016).

62 Zhu, G., Li, H. & Zhou, L. Enhancing the development of sharing economy to mitigate the carbon emission: a case study of online ride-hailing development in China. *Natural Hazards* **91**, 611-633 (2018).

63 Santos, D. O. & Xavier, E. C. Taxi and Ride Sharing: A Dynamic Dial-a-Ride Problem with Money as an Incentive. *Expert Systems with Applications* **42**, 6728-6737 (2015).

64 Yu, B. *et al.* Environmental benefits from ridesharing: A case of Beijing. *Applied Energy* **191**, 141-152 (2017).

65 Yin, B., Liu, L., Coulombel, N. & Viguié, V. Appraising the environmental benefits of ride-sharing: The Paris region case study. *Journal of cleaner production* **177**, 888-898 (2018).

66 Liu, X., Yan, X., Liu, F., Wang, R. & Leng, Y. A trip-specific model for fuel saving estimation and subsidy policy making of carpooling based on empirical data. *Applied Energy* **240**, 295-311 (2019).

67 Tu, W. *et al.* Acceptability, energy consumption, and costs of electric vehicle for ride-hailing drivers in Beijing. *Applied Energy* **250**, 147-160 (2019).

68 Jalali, R., Koohi-Fayegh, S., El-Khatib, K., Hoornweg, D. & Li, H. Investigating the Potential of Ridesharing to Reduce Vehicle Emissions. *Urban Planning* **2**, 26 (2017).

69 Gurumurthy, K. M. & Kockelman, K. M. Analyzing the dynamic ride-sharing potential for shared autonomous vehicle fleets using cellphone data from Orlando, Florida. *Computers, Environment and Urban Systems* **71**, 177-185 (2018).

70 Pettigrew, S., Cronin, S. L. & Norman, R. Brief report: the unrealized potential of autonomous Vehicles for an aging population. *Journal of aging & social policy* **31**, 486-496 (2019).

71 Lyu, Y. *et al.* CB-Planner: A bus line planning framework for customized bus systems. *Transportation Research Part C: Emerging Technologies* **101**, 233-253 (2019).

72 Chuanyu, Z., Wei, G., Jie, X. & Shixiong, J. in *2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE).* 751-755 (IEEE).

73 Cao, Y. & Wang, J. The key contributing factors of customized shuttle bus in rush hour: a case study in Harbin city. *Procedia engineering* **137**, 478-486 (2016).

74 Kirby, R. & Bhatt, K. An analysis of subscription bus experience. *Traffic Quarterly* **29** (1975).

75 McCall Jr, C. H. Com-bus: a Southern California subscription bus service. (United States. Urban Mass Transportation Administration, 1977).

76 Bautz, J. A. Subscription service in the United States. *Transportation* **4**, 387-402 (1975).

77 McKnight, C. E. & Paaswell, R. E. The Potential of Private Subscription Bus to Reduce Public Transit Subsidies. (1985).

78 Chang, S. K. & Schonfeld, P. M. Optimization models for comparing conventional and subscription bus feeder services. *Transportation Science* **25**, 281-298 (1991).

79 Potts, J. F., Marshall, M. A., Crockett, E. C. & Washington, J. A guide for planning and operating flexible public transportation services. (2010).

80 Lyu, Y., Chow, C.-Y., Lee, V. C., Li, Y. & Zeng, J. in *2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS).* 441-446 (IEEE).

81 Li, Z., Song, R., He, S. & Bi, M. Methodology of mixed load customized bus lines and adjustment based on time windows. *PloS one* **13**, e0189763 (2018).

82 Ma, J. *et al.* A model for the stop planning and timetables of customized buses. *PloS one* **12**, e0168762 (2017).

83 Yahya, B. Overall Bike Effectiveness as a Sustainability Metric for Bike Sharing Systems. *Sustainability* **9**, 2070 (2017).

84 Chen, Q. & Sun, T. A model for the layout of bike stations in public bike-sharing systems. *Journal of Advanced Transportation* **49**, 884-900 (2016).

85 Shekarchian, M. *et al.* Impact of infrastructural policies to reduce travel time expenditure of car users with significant reductions in energy consumption. *Renewable and Sustainable Energy Reviews* **77**, 327-335 (2017).

86 Mrkajic, V., Vukelic, D. & Mihajlov, A. Reduction of CO2 emission and non-environmental co-benefits of bicycle infrastructure provision: the case of the University of Novi Sad, Serbia. *Renewable and Sustainable Energy Reviews* **49**, 232-242 (2015).

87 Sun, L. *et al.* A completive research on the feasibility and adaptation of shared transportation in mega-cities–A case study in Beijing. *Applied Energy* **230**, 1014-1033 (2018).

88 Zhang, H. *et al*. Battery electric vehicles in Japan: Human mobile behavior based adoption potential analysis and policy target response. *Applied Energy* **220**, 527-535 (2018).

89 Fortunato, S. & Castellano, C. Community structure in graphs. *Computational Complexity: Theory, Techniques, and Applications*, 490-512 (2012).

90 Ahn, Y.-Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **466**, 761 (2010).

91 Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of community hierarchies in large networks. *J Stat Mech* **abs/0803.0476** (2008).

92 Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proceedings of the national academy of sciences* **99**, 7821-7826 (2002).

93 Wasserman, S. & Faust, K. *Social network analysis: Methods and applications*. Vol. 8 (Cambridge university press, 1994).

94 Anselin, L. Local indicators of spatial association—LISA. *Geographical analysis* **27**, 93-115 (1995).

95 Sui, Y. *et al.* GPS data in urban online ride-hailing: A comparative analysis on fuel consumption and emissions. *Journal of Cleaner Production* **227**, 495-505, doi:https://doi.org/10.1016/j.jclepro.2019.04.159 (2019).

96 Lang, J., Cheng, S., Zhou, Y., Zhang, Y. & Wang, G. Air pollutant emissions from on-road vehicles in China, 1999–2011. *Science of The Total Environment* **496**, 1-10, doi:https://doi.org/10.1016/j.scitotenv.2014.07.021 (2014).

97 Erdoğan, G., Battarra, M. & Calvo, R. W. An exact algorithm for the static rebalancing problem arising in bicycle sharing systems. *European Journal of Operational Research* **245**, 667-679 (2015).

98 Kloimüllner, C. & Raidl, G. R. Full-load route planning for balancing bike sharing systems by logic-based benders decomposition. *Networks* **69** (2017).

99 González, F., Melo-Riquelme, C. & Grange, L. D. A combined destination and route choice model for a bicycle sharing system. *Transportation* **43**, 407-423 (2016).

100 Boyacı, B., Zografos, K. G. & Geroliminis, N. An optimization framework for the development of efficient one-way car-sharing systems. *European Journal of Operational Research* **240**, 718-733 (2015).

101 Pedroso, F. E., Angriman, F., Bellows, A. L. & Taylor, K. Bicycle Use and Cyclist Safety Following Boston's Bicycle Infrastructure Expansion, 2009-2012. *American Journal of Public Health* **106**, e1 (2016).

102 Aziz, H. M. A. *et al.* Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: a random parameter model using New York City commuter data. *Transportation*, 1-23 (2017).

103 Rissel, C., Greaves, S., Wen, L. M., Crane, M. & Standen, C. Use of and short-term impacts of new cycling infrastructure in inner-Sydney, Australia: a quasi-experimental design. *International Journal of Behavioral Nutrition and Physical Activity* **12**, 129 (2015).

104 Long, Y. & Yoshida, Y. Quantifying city-scale emission responsibility based on input-output analysis–Insight from Tokyo, Japan. *Applied Energy* **218**, 349-360 (2018).

105 Lee, C.-H., Shih, C.-Y. & Chen, Y.-S. Stochastic geometry based models for modeling cellular networks in urban areas. *Wireless networks* **19**, 1063-1072 (2013).

106 Kennedy, J. & Eberhart, R. in *IEEE International Conference on Neural Networks, 1995. Proceedings.* 1942-1948 vol.1944.

107 Jordehi, A. R. Particle swarm optimisation (PSO) for allocation of FACTS devices in electric transmission systems: a review. *Renewable and Sustainable Energy Reviews* **52**, 1260-1267 (2015).

108 Mendes, R., Kennedy, J. & Neves, J. The fully informed particle swarm: simpler, maybe better. *IEEE Transactions on Evolutionary Computation* **8**, 204-210 (2004).

109 Niu, B., Zhu, Y. L., He, X. X. & Wu, H. MCPSO: A multi-swarm cooperative particle swarm optimizer. *Applied Mathematics & Computation* **185**, 1050-1062 (2007).

110 Zhang, H., Yuan, M., Liang, Y. & Liao, Q. A Novel Particle Swarm Optimization Based on Prey-Predator Relationship. *Applied Soft Computing Journal* (2018).

111 Kato, H., Yamamoto, M. & Shibahara, N. Life Cycle Assessment of CO2 Emissions from Intra-urban Transport Modes :
Evaluation of Environmental Friendliness of Bicycles. *Japan Society of Material Cycles and Waste Management* **22**, 220-227 (2011).

112 MOE. Green Value Chain Platform. *Climate Change Policy Division, Global Environmental Bureau, Ministry of the Environment, Japan* (2016).

113 Needell, Z. A., McNerney, J., Chang, M. T. & Trancik, J. E. Potential for widespread electrification of personal vehicle travel in the United States. *Nature Energy* **1**, 16112 (2016).

114 García-Palomares, J. C., Gutiérrez, J. & Latorre, M. Optimizing the location of stations in bike-sharing programs: a GIS approach. *Applied Geography* **35**, 235-246 (2012).

115 Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: Structure and dynamics. *Physics reports* **424**, 175-308 (2006).

116 Rombach, P., Porter, M. A., Fowler, J. H. & Mucha, P. J. Core-periphery structure in networks (revisited). *SIAM Review* **59**, 619-646 (2017).

117 Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**, P10008 (2008).

118 Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**, 1-9 (2006).

Related Publications

- Yu, Q., **Zhang, H.**, Li, W., Sui, Y., Song, X., Yang, D., ... & Jiang, W. (2020). Mobile phone data in urban bicycle-sharing: Market-oriented sub-area division and spatial analysis on emission reduction potentials. *Journal of Cleaner Production*, *254*, 119974. **(corresponding author)**
- **Zhang, H.**, Song, X., Long, Y., Xia, T., Fang, K., Zheng, J., ... & Liang, Y. (2019). Mobile phone GPS data in urban bicycle-sharing: Layout optimization and emissions reduction analysis. *Applied Energy*, *242*, 138-147.
- **Zhang, H.**, Chen, J., Li, W., Song, X., & Shibasaki, R. (2020). Mobile phone GPS data in urban ride-sharing: An assessment method for emission reduction potential. *Applied Energy*, *269*, 115038.
- Yu, Q., **Zhang, H.**, Li, W., Song, X., Yang, D., & Shibasaki, R. (2020). Mobile phone GPS data in urban customized bus: dynamic line design and emission reduction potentials analysis. *Journal of Cleaner Production*, Accepted. **(corresponding author)**

Acknowledgements

At first, I would like to give my sincere gratitude to my supervisor Prof. Ryosuke Shibasaki, for his kind support and conduction of my Ph.D. study and related research. His patience and immense knowledge enlighten me on exploring new contributions. His guidance helped me in all my Ph.D. period and writing of this thesis.

My sincere thanks also go to Associate Prof. Xuan Song, who allowed me to join his team and supported me in conducting my research. Also, I want to thank lab members, especially Zipei Fan, Qing Yu, Tianqi Xia, Renhe Jiang and Zhiling Guo, for helping my research and the happy time we have together in the last three years.

Last but not least, I would like to thank my parents for supporting me spiritually throughout writing this thesis and my life in general.