May 2018

Data-Driven Optimization Models for Feeder Bus Network Design

Yaojun Wang
University of Wisconsin-Milwaukee

Recommended Citation

Wang, Yaojun, "Data-Driven Optimization Models for Feeder Bus Network Design" (2018). *Theses and Dissertations*. 1944.
https://dc.uwm.edu/etd/1944

DATA-DRIVEN OPTIMIZATION MODELS FOR FEEDER BUS NETWORK DESIGN

by

Yaojun Wang

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
in Engineering

at

The University of Wisconsin-Milwaukee
May 2018

ABSTRACT

DATA-DRIVEN OPTIMIZATION MODELS FOR FEEDER BUS NETWORK DESIGN

by

Yaojun Wang

The University of Wisconsin-Milwaukee, 2018
Under the Supervision of Assistant Professor Jie Yu

Urbanization is not a modern phenomenon. However, it is worthwhile to note that the world urban population growth curve has up till recently followed a quadratic-hyperbolic pattern (Korotayey and Khaltourina, 2006). As cities become larger and their population expand, large and growing metropolises have to face the enormous traffic demand. To alleviate the increasing traffic congestion, public transit has been considered as the ideal solution to such troubles and problems restricting urban development. The metro is a type of efficient, dependable and high-capacity public transport adapted in metropolises worldwide. At the same time, the residents from crowded cities migrated to the suburban since 1950s. Such sub-urbanization brings more decentralized travel demands and has challenged to the public transit system. Even the metro lines are extended from inner city to outer city, the commuters living in suburban still have difficulty to get to the rail station due to the limited transportation resources.

It is becoming inevitable to develop the regional transit network such as feeder bus that picks up the passengers from various locations and transfer them to the metro stations or transportation hubs. The feeder bus will greatly improve the efficiency of metro stations whose

service area in the suburban area is usually limited. Therefore, how to develop a well-integrated feeder system is becoming an important task to planners and engineers.

Realizing the above critical issues, the dissertation focus on the feeder bus network design problem (FBNDP) and contributes to three main parts:

1. Develop a data-mining strategy to retrieve OD pair from the large scale of the cellphone data. The OD pairs are able to present the users' daily behaver including the location of residence, workplace with the timestamp of each trip. The spatial distribution of urban rail transit user demand from the OD pair will help to support the establishment and optimization of the feeder bus network. The dissertation details the procedure of data acquisition and utilization. The machine leaning is applied to predict the travel demand in the future.
2. Present a mathematical model to design the appropriate service area and routing plans for a flexible feeder transit. The proposed model features in utilizing the real-world data input and simultaneously selecting bus stops and designing the route from those targeted stops to urban rail stops.
3. Propose an improved feeder bus network design model to provide precise service to the commuters. Considering the commuters are time-sensitive during the peak hours, the time-windows of each demand is taken in to account when generating the routes and the schedule of feeder bus system. The model aims to pick up the demand within the time-windows of the commuters' departure time and drop off them within the reasonable time. The commuters will benefit from the shorter waiting time, shorter walking distance and efficient transfer timetable.

To

my parents,

my wife,

and especially my daughter

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

API	Application Programming Interface
BP	Backpropagation
BRT	Bus Rapid Transit
DRT	Demand Responsive Transport
FBNDP	Feed Bus Network Design Problem
GA	Genetic Algorithm
GIS	Geographic Information System
LSTM	Long Short-Term Memory Network
MIP	Mixed Integer Program
OD	Origin-Destination
RMSE	Root-Mean-Square Error
RNN	Recurrent Neural Network

ACKNOWLEDGMENTS

No man is an island. No thesis is solely the work of one person. I must express my gratitude to many people who have assisted me. First, I must acknowledge my research advisers Prof. Yue Liu and Prof. Jie Yu for all the support and encouragement. The door to Prof. Yue Liu and Prof. Jie Yu office was always open whenever I ran into a trouble spot or had a question about my research or writing. They consistently allowed this paper to be my own work but steered me in the right direction whenever he thought I needed it.

My sincere thanks also go to Prof. Xiao Qin, Prof. Matthew Petering and Prof. Changshan Wu who are on my committee and provide valuable contributions to my dissertation. They are all the outstanding experts in their fields.

I am highly indebted to the UW-Milwaukee Graduate School for their guidance and constant supervision as well as for providing necessary information regarding this research and also for their support in completing this achievement.

I would also like to thank my lab mate Yun Yuan who is a great person to work with and to discuss academic problems.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. My beloved and supportive wife, Bing who is always by my side when I needed her most and helped me a lot in making this study, and my lovable daughter, Emily who served as my inspiration to pursue this undertaking. This accomplishment would not have been possible without them. Thank you.

Chapter 1: Introduction

1.1 Background

In the process of urbanization, more people worldwide migrate to cities, and the World Health Organization announced the world's population living in towns and cities surpassed 50% and predicted that this proportion will continue to increase. Many growing economies are facing two major problems: air pollution and traffic congestion. Public transit is considered as the most efficient way to relieve these issues.

Transit-dependent cities are generally more sustainable than car-dependent cities. They cover less land and tend to have fewer emissions both per capita and per distance traveled. Besides that, fewer emissions, shorter commuting time, less parking spaces required are all the benefit from the public transit. Public transit also plays an important role in helping people commute to work.

Public transit, including city buses, trams, rapid transit (metro/subways/undergrounds, etc.) and ferries, is a shared passenger-transport service which is available for use by general public. Compared with buses and cars, trains can carry far more people with much greater efficiency. One effective solution for alleviating big city malaise is the establishment of metro systems. As of October 2014, metro systems operate in about 157 cities in more than 55 countries (International Association of Public Transport, 2015).

Enormous investments have been put on the facilicity construction of public transit systems, such as urban metro system, Bus Rapid Transit (BRT). Meanwhile, various public transit priority policies have been issued by the government. However, the growing concern for public transit is its inability to provide door-to-door services for passengers, which has been known as "the first/last mile" issue. This is especially acute in the lower-density suburbs where the existing public transportation options are not within walking distance from a traveler's origin/destination. Therefore, transit use in these areas is often less practical. Critics claim this promotes a reliance on cars, which results in more traffic congestion, pollution, and urban sprawl. Feeder bus is considered as the ladder to the predicament. Feeder bus services are designed to pick up passengers in a certain locality and take them to a transfer point where they make an onward journey on a trunk service. This can be another bus, or a rail-based service such as a tram, rapid transit or train. Feeder buses may act as part of a wider local network. A well-working feeder bus system will enhance the utilization of metro systems and greatly encourage people to shift their travel mode and is now widely regarded as an effective tool to improve service efficiency.

Most feeder systems fall into three categories, namely, fixed-route fixed-schedule feeder system, demand responsive feeder system, and hybrid feeder system.

The typical cost efficiency of fixed-route fixed-schedule feeder system is due to the predetermined schedule, the large loading capacity of the vehicles and the consolidation of many passenger trips onto a single vehicle (ridesharing). Many researchers have made significant contributions in pioneering studies on fixed-route fixed-schedule feeder system. (Byrne and Vuchic, 1972; Hurdle, 1973; Byrne, 1976; Chien and Yang, 2000; Wirasinghe,1980; Stanger and Vuchic, 1979; Dunn, 1980; May, 1991; Kuah and Perl, 1989; Matins and Pato, 1998; Gholami and Mohaymany, 2011; Deng et al., 2013; Fu et al., 2003; Sun and Hickman, 2005; Chien et al., 2010). However, the public considers them to be inconvenient because of their lack of flexibility since often the locations of pick-up and/or drop-off points and/or the service's schedule do not match the individual rider's desires. Therefore, an increasingly larger portion of the growing population relies almost exclusively on private automobiles for their transportation needs, causing many urban areas to suffer from increasing congestion and pollution problems.

As an alternative to private automobiles, the fixed-route fixed-schedule feeder system has deficiencies. Thus, there is a need for a transit system that provides flexible service at a cost-efficient price in the area with low density. Demand Responsive Transport (DRT) systems instead provide much of the desired flexibility with a door-to-door type of service, but they are much costlier to deploy. In most cases, the demand responsive feeder system is only discussed and deployed in the low-density area (Ellis and Silva, 1998; Paulley et al., 2004; Fan and Machemehl 2006; Quadrifoglio and Li, 2009; Li and Quadrifoglio, 2010). Hence, transit agencies are facing a growing demand for improved and extended DRT services.

The broad and new category of "flexible" transit services includes all types of hybrid services that combine pure demand responsive and fixed-route features. These services have established stop locations and established schedules, combined with some degree of demand responsive operation. Their characteristics have, in several cases, efficiently responded to some of the needs and want of both the customers and the transit agency as well. However, their use has been quite limited in practice so far, as opposed to regular DRT systems.

Most existing feeder system design research has some technical deficiencies that remain to be overcome. For examples,

- Most pioneer research is assumption-based. The real-world characteristic is ignored, that could be an environmental constraint and law restriction. These deficiencies could lead to inaccuracy and lack of convening of the presented models.
- A majority of the work is focused on the feeder bus network design only. Most of the efforts have been made on the algorithm aspect. However, the trip information including OD-pair, departure time, trip duration is usually inauthentic. Lack of real data will lead to inaccurate

and unreliable results.

- The deficiency of detailed trip information limits the efficiency of feeder bus services. In the low-median demand area, the feeder bus sometimes misses the commuters' departure time. The improper feeder bus schedules reduce the commuters' willingness to take feeder bus service.

1.2 Research Objectives

Realizing above reviewed limitations of existing studies on FBNDP, the research aims to develop an integrated framework for feeder bus network design, which will focus on the following critical research tasks:

1. Develop a cellphone data processing methodology that can extract the spatial distribution of urban rail transit user demand to cope with demand uncertainty issue in existing studies;
2. Develop a neural network model to predict the passenger demand in the future. The trips of potential passengers extracted from the previous steps will be the input to the neural network. The Long Short-Term Memory (LSTM) is adapted and trained to provide the predicted passenger number in the target area based on the historical data.
3. Introduce Online Geographic Information System (GIS) tools (Baidu Map; Google Map) to retrieve the features and travel time information of real traffic network which is able to reflect the network topology and real traffic status;
4. Propose an integrated optimization model that is capable of seamlessly and simultaneously coordinating the passenger boarding guidance and transit routing for fixed-route fixed-schedule feeder bus network with median demand area; and
5. Establish an optimization model to generate the route and schedule of the flexible feeder bus network and provide precise service in which the pickup time and location are

considered. The time-windows of the commuters will be involved in the proposed model.

6. Illustrate the proposed methodology through a real-world case study to best understand and apply the proposed methodology during the design process of feeder bus system.

1.3 Thesis Outline

Base on the proposed research objective, this study is organized into seven chapters. The core of those tasks and their interrelations are illustrated in Figure 1.1.

Chapter 2 presents a comprehensive review of relevant research, including the problem definition, data preparation, data pre-processing, approach models, solution methods of the three types of the feeder bus system, namely, fix-route fixed-schedule feeder bus system, demand-responsive feeder bus system and hybrid feeder bus system. The review focuses on identifying the advantages and limitations of those studies, along with their potential enhancement.

Chapter 3 illustrates the modeling framework of the proposed research, based on critical operational issues that need to be addressed in the design of feeder system. It briefly describes the functions of each principle modeling component and case studies, which provides the foundation for the identification of research tasks for this study.

Chapter 4 describes cellphone data processing strategy that are used to retrieve commuters' trip information. Data pre-processing steps are first applied to fix defect of the dataset such as missing values, impossible data combinations, duplicate records. The quality of data is first and foremost before running an analysis. The big data mining is then introduced to retrieve trip information from the cellphone data. Further, the Online GIS tools will help to obtain the traffic information (e.g., total driving time, total distance, etc.) of potential routes. A neural network is built for predicting future demand, with the Long Short-Term Memory Network (LSTM) being selected as the internal units in the recurrent neural network.

Chapter 5 presents an improved mathematical model to design feeder bus network to access an existing urban rail system in the median travel demand area. The traditional mixed integer model is reconstructed to adapt to the real-world data source (e.g., travel demand information, travel time matrix, etc.). The model is applied in the case study which aims to design a fixed-route fixed-schedule feeder bus system for Jiandingpo Station in Chongqing, China. The model is expected to create a feeder bus system that simultaneously minimizes weighted passengers walking distance and operational cost of feeder bus system.

Chapter 6 further extends the model presented in Chapter 5 by merging the time window of each demand which can provide accurate feeder bus service to the passengers. Unlike the traditional way on solving FBNDP in which only the quantity and the location of demand are met, this model aims to provide precise feeder bus services that transfer passengers to the subway station with consideration of their departure time from the residence. Such a system benefits from using cellphone data to obtain the information of each demand including location and departure time in the peak hours. The service rate is selected as the object which can ensure a feasible solution with a limit budget.

Chapter 7 summarizes the contributions of this dissertation and the directions for future research.

Chapter 2: Literature Review

2.1 Introduction

In the recent decades, a large scale of literature has been published in FNDSP. Many efforts have been devoted to achieving high efficient feeder bus system. The problem consists of different subproblems including data preparation, system policy, feeder bus network design and scheduling. This chapter will focus on performing a review of research works for a specific problem arising in a feeder network design, with the goal of providing readers with a broader and more complete insight on the subject. Descriptive analysis and classification of previous works are presented to highlight the main characteristics and solution methods. The existing research will be organized and discussed in the following content. The purpose is to identify the special characteristics, strengths, and deficiencies of existing studies and thus to define the primary directions for this study.

2.2 Data Preparation

Data preparation includes area's topology, origin-destination (OD) matrices, fleet size, and more information such as bus and train operating costs, route length, speed, demand, and so forth. The area's topology and origin-destination (OD) matrices are the most critical input to the models.

In most previous feeder system optimization models, the environment (e.g., street network, travel time, walking distance, etc.) and the demand (e.g., OD pairs) are over-simplified. The examples with unreality number have low reliability to prove the feasibility and robustness of the presented models. Later with a parcel-level geographic database (i.e., street network and land parcel data), Furth et al. (2007) estimated transit demand based on the land use and development intensity.

Accurate OD pairs are considered as one of the important parts in locating bus stops. High mobile phone penetration makes it possible for us to retrieve authentic OD pair from cellphone data. This methodology is based on the fact that a mobile phone is moving on a particular route always tends to change the base station nearly at the same position. A method has been raised by Caceres and Wideberg (2007) that the population distribution data can be extracted from mobile phone network system. Calabrese et al. (2011) estimated origin-destination flows using mobile phone location data.

Geographic Information Systems (GIS) are computer-based systems designed to support the capture, management, manipulation, analysis, and modeling of spatially referenced data. Yeh and Chow (1996) discussed the integration of GIS and a location-allocation model for public facilities planning, using open space planning in Hong Kong as an example. Lu (2006) developed the web GIS based intelligent transportation application systems with web service technology.

At the same time, online source and cloud service are increasingly accepted. Through online geographical information systems (Online GIS), the spatial data will be accessed by the developers. Furthermore, researchers are able to tap the potential of Online GIS application by running their own codes or upload specific geodata. For example, Ghita (2014) used Open GIS Technology to locate a business in a decision support system. In recent years, cloud technology speeds up the development of Online GIS. The API (Application Programming Interface) of several online maps owned by internet giants such as Google map and Baidu map has been opened to the public, which contributes to the generalization of the Online GIS.

2.3 Passenger Demand Prediction

Over the past few decades, many short-term prediction models have been developed with the huge progressing made on the computation power. Farokhi et al. (2010) used the moving average method with constant weights and two adaptive moving average methods to predict short-term travel times. Min and Wynter (2011) provided predictions of road traffic with spatiotemporal correlations. A Bayesian dynamic linear model approach for real-time short-term freeway travel time prediction was raised by Fei et al. (2011). The empirical mode decomposition and neural networks were applied to forecast the short-term metro passenger flow in Wei and Chen’s report (2012).

With the rapid development of technology, Long Short-Term Memory (LSTM), one of the models in the recursive neural network (RNN) class, has been widely used in numerical prediction (Ma et al.,2015; Tian and Pan, 2015; Zhao et al., 2017; Li et al., 2017). LSTM was proposed by Hochreiter and Schmidhuber (1997) and improved by Ger et al. (1999).

2.4 System Policy

Quadrifoglio and Li (2009) proposed an analytical model to identify the condition of switching operations between a demand responsive and a fixed-route policy. Then the model was extended to the assumption with two buses (Li and Quadrifoglio, 2010).

Regarding passenger demand, it is usually assumed to be fixed or inelastic, for simplicity. Fixed demand may be reasonable for systems at which passengers are insensitive or independent of service quality or price. However, elastic demand can probably be variable, due to sharing or competition of the public transport. Moreover, Jakimaviˇcius and Burinskiene (2009) stated that rising rate of mobility demand will be a significant factor in the efficiency of urban transportation modeling.

For feeder bus, two types of travel demand patterns, namely, many-to-one and many-to-many, are measured. The many-to-one demand pattern has been discussed in most papers (Kuah and Perl, 1989; Kuan et al., 2004; Kuan et al., 2006; Chien and Schonfeld, 1998; Chien and Yang, 2000; Kuah and Perl, 1988; Xiong et al., 2013). This model refers to passengers traveling from multiple origins to a single destination. This is usually more applicable to feeder bus services which carry passengers to a common destination (central business district or a transfer station), and peak hour trips to and from CBD can be considered in this pattern. In most feeder services, many-to-many demand pattern is considered whenever passengers have different origins and destinations.

2.4.1 Fixed-route fixed-schedule Feeder Bus System

The fixed-route fixed-schedule feeder network design has been well studied by a large number of researchers (Newell, 1979; Mandl, 1980; Ceder and Wilson, 1986; LeBlanc, 1988; Chang and Schonfeld, 1991; Bookbinder and Desilets, 1992; Baaj and Mahmassani, 1991, 1995; Chien and Schonfeld, 1997; Shrivastav and Dhingra, 2001; Borndörfer et al., 2005; Bunte et al., 2006; Ceder, 1984, 2001, 2003).

In these problems, the objective is to minimize some function that combines operator and passenger cost. The operating cost is usually represented as a function of the number of vehicles and miles traveled by the vehicles while the passenger cost is a function of their travel times.

The existing solution method is mainly falling into two categories (Deng et al., 2013), namely, analytic approach and network programming. The analytic approach derives the optimal route spacing, operating headway and the optimal stop spacing. A couple of pioneering studies were proposed to optimize the location, headway of feeder buses as well as determining the number and the length of routes (Byrne and Vuchic, 1972; Hurdle, 1973; Byrne, 1976; Chien and Yang, 2000). Wirasinghe (1980) presented an approximate analytical model and corresponding solution algorithm to design a feeder bus system access to a rail station served a peak-period demand. The model was further applied to the Calgary (South Corridor) Light Rail Transit (LRT) system. In addition to optimizing the route spacing, operating headway and the stop spacing, Kuah and Perl (1988) also discussed the influencing factors of bus-stop spacing in three different cases. Chowdhury and Chien (2002) took stochastic feeder vehicle arrivals at transfer stations into account, and then the slack times of coordinated routes were optimized by balancing the savings from transfer delays and additional cost from slack delays and operating costs. Although the analytic models have been widely used to deal with FBNDP, their basic assumptions, the shape of the street geometry and the spatial distribution of demand, are regarded as the significant limitations (Deng et al., 2013).

In recent years, an ever-increasing interest in using the network programming approach to handling FBNDP have been raising. Typically, this approach decomposes the traffic network into a set of nodes and a set of links. Some of the nodes are selected as bus stops while some of them represent urban rail transit stations. The links between stops, and between stop and station are treated as bus route segments. The demand is centered on the nodes and transported from node to node.

Many mathematical programming models, especially, integer and mixed integer programs, are developed to handle FBNDP under either M-to-1 or M-to-M demand pattern (Kuah and Perl, 1989; Martins and Pato, 1998; Gholami, 2011; Chien et al., 2003). Based on those developed methodologies, some more advanced studies are followed. Ciaffi et al. (2012) proposed a two-phase mode to cope with FBNDP. At the first phase, a heuristic algorithm is developed to generate two different and complementary sets of feasible routes and then, with the output from the first phase, a GA-based algorithm is utilized for finding a sub-optimal set of routes with the associated frequencies at the second phase. Considering a multi-level cost structure, including, passengers' cost and operators' cost, Deng et al. (2013) proposed a model to solve the issue of M-to-M feeder bus network. Pan (2014) employed a bi-level model to maximize the number of served passengers by the feeder transit system in the upper level and to minimize the operational cost for transit operators in the lower lever level.

2.4.2 Demand-responsive Feeder Bus System

Demand-Responsive Transit (DRT) services help enhance mobility, especially in low-density areas. Such systems provide personalized service on demand and have flexible routing and scheduling.

Exact approaches to solving DRT systems provide optimal solutions, but the combinatorial nature of the problem limits the applicability of these methods to very small instances; therefore, they provide a good theoretical insight, but practically cannot be used to solve real situations. Psaraftis (1980, 1983) described an exact backward and forward dynamic programming solution approach for the single vehicle Dial-a-Ride problem for static and dynamic environments without time windows. Another dynamic programming approach is described in Desrosiers et al. (1986). Sexton and Sexton and Bodin (1985) and Sexton and Choi (1986) described a Benders' decomposition approach to solve the single-vehicle PDP with time windows and capacity constraints. Desrosiers et al. (1991) present a Dantzig-Wolfe approach for optimally solving the multiple vehicles PDP with time windows and capacity constraints. Savelsbergh and Sol (1998) proposed a branch-and-price based algorithm to solve the dynamic multi-vehicle PDP. Fischetti and Toth (1989) developed an additive bounding procedure suitable for a branch-and-bound algorithm for the single vehicle PDP. An exact algorithm approach is described in Lu and Dessouky (2003). Barra (2007) proposed a comprehensive nouvel model, based on Constraint Satisfaction, with essential and complementary constraints to obtain an optimized public transport network. Lu et al. (2016) presented a flexible feeder transit routing model that can serve irregular-shaped networks.

Figure 2.1 Classification of The Literature Based on Problem Perspective

Reference	Year	Decision Variable	Approach model	solution Method	Objective
Kuah and Perl	1989	Bus route location; Bus frequency	Network	Heuristic	Optimal route
Martins and Pato	1998	Bus route location; Bus frequency	Network	Heuristic	Optimal route
Xiong et al.	2013	Route location; Bus headway		Heuristic	Optimal route
Chien & Schonofeld	1998	Bus headway; BS; Bus route location; RS	analytical model	Metaheuristic	Optimal route
Chien & Yang	2000	Bus route location; Bus headway	analytical model	Metaheuristic	Optimal route; headway
Shrivastav and Dhingra	2002	Travel time	Network	Heuristic	Optimal route; schedules
Chien et al.	2004	Bus route location; Bus headway	analytical model	Metaheuristic	Total welfare
Chowdhury and Chien	2001	Bus headway; Bus travel time	analytical model	Heuristic	Optimal route
Kuan et al.	2004	Bus route location; Bus frequency	Network	Metaheuristic	Optimal route
Chien	2001	Bus headway; Fleet size; Bus route location	analytical model	Heuristic	Total welfare
Kuan and Perl	1989	Bus route location; Bus frequency	Network	Metaheuristic	Optimal route
Shrivastava and O'mahony	2006	Bus route location; Bus frequency	Network	Metaheuristic	Optimal route; schedules
Shrivastava and O'mahony	2007	Bus route location; Bus frequency	Network	Hybrid	Optimal route; schedules
Shrivastava and O'mahony	2009	Bus route location; Bus frequency	Network	Hybrid	Optimal route; schedules

Gholami and Mohaymany	2011	Bus route location; Bus frequency; Mode	Network	Metaheuristic	Optimizing multimode feeder
Sivakumaran et al.	2012	Bus headway; Bus route length	analytical model	mathematical	Optimal schedules
Ciaffi et al.	2012	Bus route location; Bus frequency	Network	Hybrid	Optimal route; schedules

Clustering approaches use the intuitive idea of merging in a single point requests that are physically close to each other. The problem instances are reduced in size, and exact approach can then be applied efficiently. Ioachim et al. (1995) developed a clustering algorithm to solve the multi-vehicle PDP with time windows. The work of Daganzo (1984) described a checkpoint DRT system that combines the characteristics of both fixed route and door to door service.

2.4.3 Hybrid Feeder Bus System

Both the fixed-route fixed-schedule feeder bus system and demand-responsive feeder bus system are well studied and applied worldwide. However, they have a limitation on serving the disabilities and reducing the system cost, respectively.

In the recent years, the concept of the hybrid system is raised and developed. The hybrid system is an innovative concept that merges the flexibility of DRT systems with the low-cost operability of fixed-route bus systems.

A decision support system which automatically constructs efficient paratransit vehicle routes and schedules was designed to improve the accessibility and efficiency of a dial-a-ride system with fixed-route buses (Liaw et al., 1996). Hickman and Blume (2000) developed an insertion heuristic and test it on a dataset from Houston, Texas. A tabu search heuristic was employed and tested on a dataset from Antelope Valley in California by Aldaihani and Dessouky (2003). The case study showed that the vehicle distance will be reduced by 16.6%, compared to the on-demand system, with shifting some of the demand to a hybrid service route (18.6% of the requests). Aldaihani et al. (2004) developed an analytical model that aids decision-makers in designing a hybrid network that integrates a flexible demand responsive service with a fixed route service.

Large amount of work has been presented in feeder bus network design. In most presented studies, routes and schedules are discussed independently.

Most of the researchers have developed routes from a given initial skeleton using a heuristic approach (Bullnheimer et al., 1999; Quadrifoglio, et al., 2007;) Different methods have been developed on solving feeder bus route design problem. Kuah and Perl (1989) further extend the heuristic method to many-to-many (M-to-M) FBNDP. A Tabu search algorithm was adapted by Martins and Pato (1998) to solve the FBNDP, and the computational results from a set of problems simulating real-life situations were given. Later, Gendreau, et al. (1999) is able to route and dispatch the vehicles in real-time through parallel tabu search. Kuan et al. (2006) proposes genetic algorithms and ant colony optimization for solving the feeder bus network design problem. Several tests were randomly generated to verify the method.

There are also large amount studies working on the feeder bus scheduling problem. Chien and Schonfeld (1998) developed a model for jointly optimizing bus schedules in an urban corridor. Li et al. (2009) discussed analytical models for optimizing a bus and rail transit system with feeder bus services under different market regimes. The time window was considered in the Diana et al. (2006)'s report, and a model was designed for demand responsive transportation services. Hickman and Blume (2001) present a heuristic for scheduling integrated transit trips that accommodates both passenger and vehicle scheduling objectives.

With the development of the intelligent algorithm, a combined routing and scheduling problem was attempted in the recent decades. A total cost minimization model was developed by Chien and Yang (2000) for finding the optimal bus route location and its operating headway in a heterogeneous service area in which numerical examples were analyzed. Shrivastava and O'Mahony (2006) developed feeder routes and frequencies leading to schedule coordination of feeder buses with main transit are using genetic algorithms. Ciaffi et al. (2012) developed a new metaheuristic procedure that simultaneously generates routes and frequencies of the feeder bus network in an actual size large urban area.

High mobile phone penetration makes it possible for us to retrieve authentic OD pair from cellphone data. This methodology is based on the fact that a mobile phone is moving on a particular route always tends to change the base station nearly at the same position. Compare to the traditional OD survey; the cellphone data mining method has several advantages: a). The OD pairs from the cellphone data mining are more complete and accurate; b). It is budget-friendly and efficient. The high-performance computer is required instead of the labor on the street. c). The OD pairs generated are easy to keep up-to-date. Regarding the utilizing of the cellphone data, several studies have been reported on different fields. A method has been raised by Caceres and Wideberg (2007) that the population distribution data can be extracted from mobile phone network system.

Enormous amounts of work focus on the optimization model and neglect the importance of data part. The algorithm for solving the model has evolved with the development of the mathematics and operations research. However, the inaccurate data leads to the improper result. The bias of data comes from the outdated data collection method and ignore the importance of the data. The feeder bus is time-sensitive and easily affected by the traffic. The methods presented usually lack robustness when dealing with the unstable traffic conditions. Furthermore, the commuters are time-sensitive, and the schedule of the feeder bus is crucial in the peak hours. The time-window of different commuters is not considered in the models.

2.5 Summary

Although most of the studies above have successfully handled a variety of feeder bus system, the following critical issues deserve further investigation during the process of design feeder bus network:

1. Most studies have neglected the integrated operation of pedestrian guidance (from home addresses to candidate bus stops) and transit routing (from selected bus stops to urban rail transit stop). Since the stops are the first point to connect bus services with passengers, the ignorance of passengers' interests easily results in decreasing the attractiveness of transit riders (Fu et al., 2003; Sun and Hickman, 2005; Chien et al., 2010).

2. Most of the existing studies have set the shape of the street geometry and spatial distribution of demand as the basic assumptions for model development. The first assumption is unable to reflect the characteristic of real traffic network, such as a one-way street or left-turn only intersection. Considering realistic network has been widely regarded as a vital role in bus network design problem as an accurate traffic network contributes to the reliability of the result (Chien, 2003; DiJoseph and Chien, 2013). Moreover, the second one directly leads to an ambiguity in determining the location of bus stops. Both of them could make the case study fail in practicability and authenticity.

3. The real traffic status, especially the travel time which changes dramatically under different network conditions, likely, peak period, non-peak period and weather condition, has not been included when proposing the methodology for FBNDP.

4. The commuters which are the feeder system major passengers have their own time pattern. Moreover, the commuters are time sensitive. However, either the fix-route fix-schedule feeder system, the demand-responsive feeder system, or the hybrid system in current studies does not take such characteristic into account. The failure in providing accurate services will cause losing commuter passengers and increasing waiting time.

Chapter 3: A Systematic Research Framework

3.1 Introduction

This chapter illustrates the modeling framework of the proposed research and the interrelations between its principal components. The brief description of proposed models and expected output are also included.

3.2 Key Research Issues and Primary Research Tasks

To address the critical issues listed in Chapter 2, this dissertation has divided the research efforts into the following primary tasks:

Task 1: Raise a method to obtain the travel demand information (e.g., the location of the demand, OD pairs, etc.) for the certain area from cellphone signaling dataset. The real-world dataset from the data processing method will ensure the accuracy and the reliability of the proposed feeder bus network design.

Task 1.1: Develop a strategy to retrieve OD pair from the large scale of the cellphone data. The OD pairs are able to present users' daily behaver including the location of residence and workplace with the timestamp of each trip. **Note that** cellphone data has never been used to obtain trip information and travel demand distribution on such a large scale. Such strategy will greatly improve the accuracy and feasibility of the proposed model and will diminish the effect of assumptions.

Task 1.2: Develop and train the neural network to predict the passenger demand in the future based on the cellphone data. The demand will be imported to the feeder bus network design model to provide a reliable plan for the target area. **Note that** this research mainly contributes to constructing the neural network to make a prediction based on the cellphone dataset within a reasonable time.

Task 1.3: Present a method to obtain the traffic status through Online GIS tools. The traffic status such as total walking distance, total driving time and distance will be the inputs as the proposed models. Note that, traffic information is an indispensable factor in the FBNDP. The online GIS tools are employed to obtain real travel time in which street topology (e.g., one-way street) and traffic operation issues (e.g., left turn delays) are considered in the generated travel time matrix rather than the trip distance in most previous researches.

Task 2: Develop a data-driven mathematical model to design feeder bus network to access an existing urban metro system for the area with low and medium travel demand. Two mix integer models will be proposed individually.

Task 2.1: A mix integer optimization model is proposed to feature an integrated operational framework, which can simultaneously select bus stops, dispatch and route buses from those targeted stops to urban rail stops. Note that, the proposed model is capable of maximally utilizing real-world data to limit the bias from some unrealistic assumptions that have been made in those previous research.

Task 2.2: By extending the task 2.1 with time dimension, an enhanced mix integer optimization model is further proposed, in which the departure time of the commuters is considered when the feeder bus network is designed. Note that, this novel model features the capability of providing punctual service with regard to the commuters' departure time window has never been taken serious considertion in the previous research on feeder bus services.

3.3 Modeling Framework

In view of the above research tasks, Figure 3.1 depicts the framework of the proposed method for this dissertation.

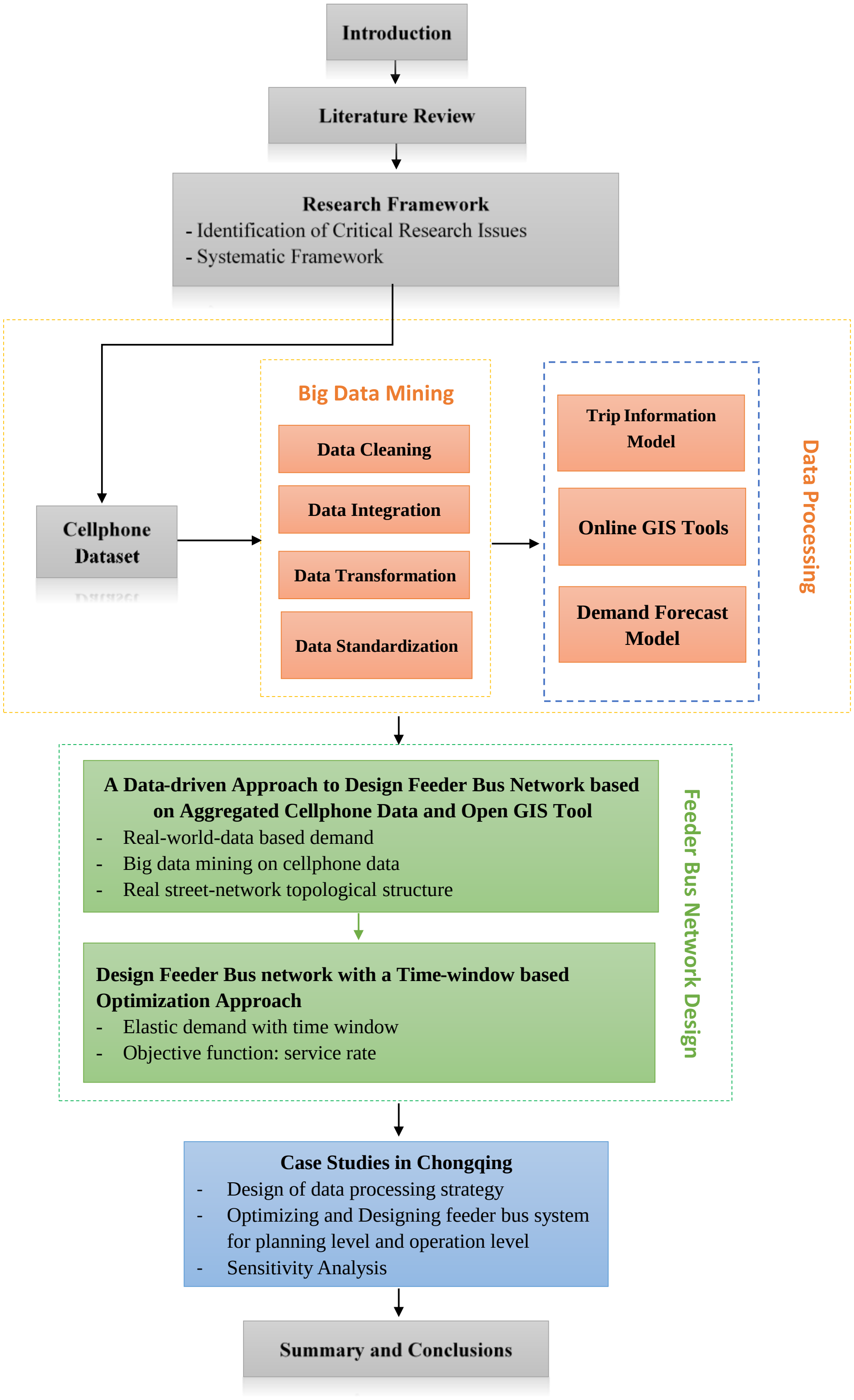


Figure 3.1 Dissertation Organization

Chapter 4: Data Processing for Cellphone Dataset

4.1 Dataset Descriptions

Mobile phone data used in this research was collected by the base transceiver stations (BTS), where BTS is treated as the fixed traffic detector to monitor the movement of mobile users. It should be noted that the communication record (associating with commutating event) typically not directly contains geo-location information but only provides a unique ID of the communicating tower (a combination of LACID and CellID). Therefore, a matching job between communication records and BTS geo-location table is required to obtain positioning information for each record. The data format of communication records used in this study is shown in table 1.

Table 4.1 Format of communication record

Fields	Description
MSID	The identity of a mobile user
MSTIME	Timestamp of the acquired communicating event
LACID	The ID of location area code of connected BTS
CellID	The ID of cell identifier of the connected BTS
EventID	The ID of the acquired communication event
LON	Longitude of the geographic coordinates of the connected BTS
LAT	Latitude in the geographic coordinates of the connected BTS

The daily data size is about 2GB or 2.2×10^7 records per 1 million population. For some megalopolis with dozens of million population in China, more than 3 billion records are generated during each month. The regular ways we used to analyze the data, such as SQL or MATLAB, will be not able to handle the cellphone data efficiently. In this research, Python is selected to do the data processing work.

4.2 Data mining procedure

The popularity of cell phones has soared in the most recent decade and generate massive amounts of signaling data. The regular signaling data includes MSID, CellID, LacID, Timestamp, etc. The MSID is the identification code of the device; it is unique and unable to convert back to a cellphone number. The timestamp indicates when the activities happen. The CellID and LacID show the tower that the device connects to. Any action on the cellphone such as turning on/off, texting, making a phone call or connecting to the internet will trigger the signaling record and be stored in the carriers' database. **Figure 4.1** states that cellphone will communicate to the closest base station along the route and the moving path of the cellphone user can be retrieved from the complete signaling data. The complete signaling dataset reflects the population distribution and the human movement patterns.

Figure 4.1 The travel path of the cellphone user

The cellphone signaling dataset is required to be decoded to the exact value for each parameter prior to the data mining procedure. After the train station is selected, the service area is able to be circled. The high density of cellphone base station is required to provide high network speed in recent years. The extra benefit of it is that base stations in the study area are now able to be considered as the location of the cellphone users with high accuracy.

The position of the device on the early morning is considered as the original position. By tracking user's moving path, it can be inferred whether the user takes the subway. Once the user moving path is along the subway line, he/she would be considered as one of the potential users of the feeder bus network. **Figure 4.2** indicates the flow chart of data mining process to retrieve OD pairs of the study area.

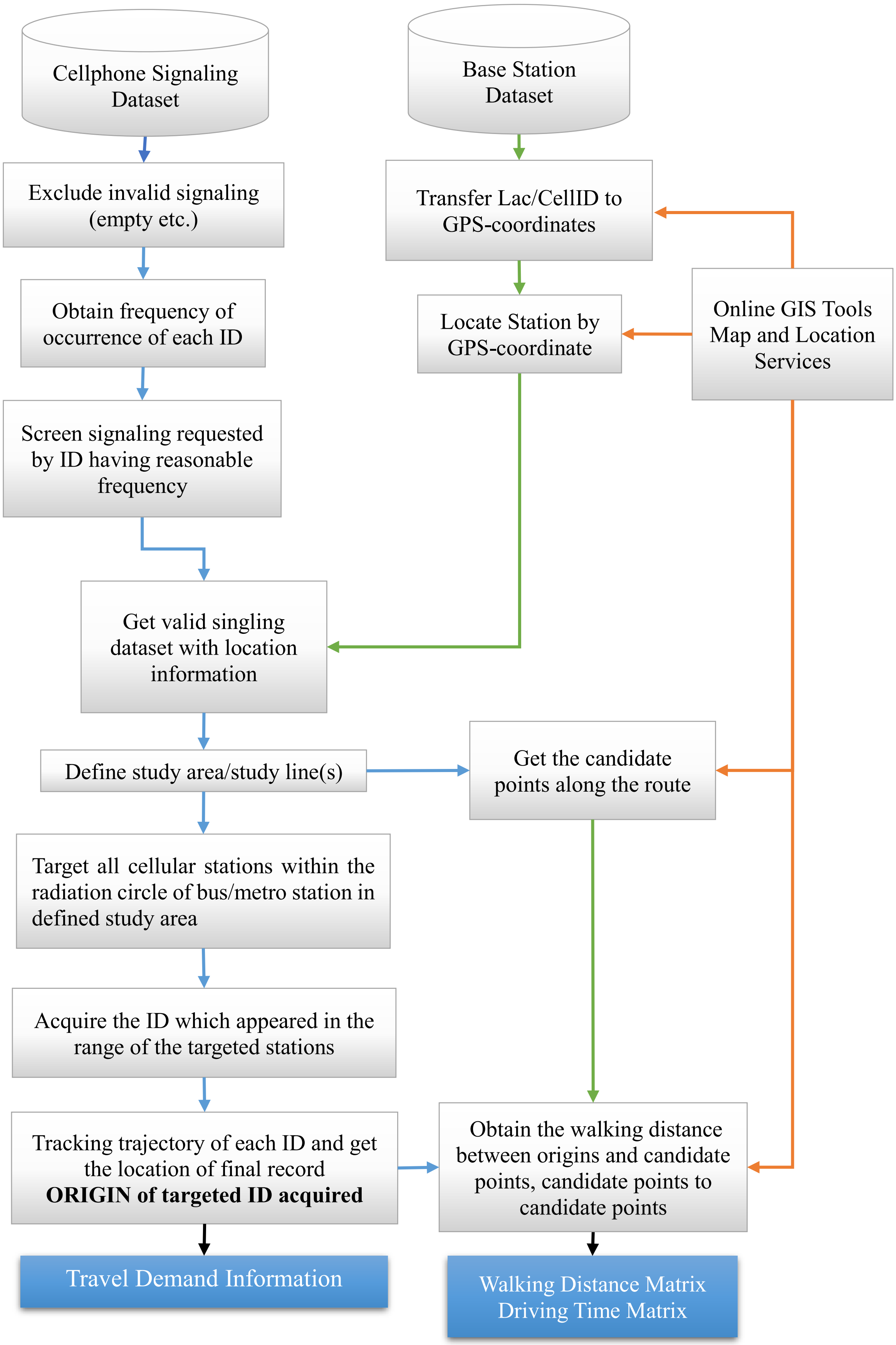


Figure 4.2 Flow chart of data mining process

4.3 Online GIS Tools

With rapid development of network technology, the online giants are more open and flexible. A lot of map-engine providers, such as Google Map, Open Street, and Baidu Map, etc., offer full access to their GIS database by calling defined Application Program Interface (API).

The Google Maps Distance Matrix API is a service that provides travel distance and time for a matrix of origins and destinations, based on the recommended route between starting and ending points. Google Maps Distance Matrix API can be accessed through an HTTP interface, with requests constructed as a URL string, using origins and destinations, along with API key. For example, in Google Map APIs, one can easily request the distance matrix data between Washington, DC and New York City by calling https://maps.googleapis.com/maps/api/distancematrix/json?units=imperial&origins=Washington, DC&destinations=New+York+City,NY&key=YOUR_API_KEY" in JASON, JavaScript or Python format.

After import the OD pair from data mining to the google map engine, the Google Maps Distance Matrix API returns information based on the recommended route between starting and ending points, as calculated by the Google Maps API, and consists of rows containing duration and distance values for each pair.

In this study, the Python Script is used to develop the interactive tool connecting with Online GIS engine which aims to collect map-based travel distance and time data based on the input of OD pairs.

4.4 Passenger Demand Prediction

The artificial neural network is widely used in the various fields to conduct prediction, estimation or analysis just like a human. There are two types of neural networks being developed in the recent years with the progress in computer science.

CNN (Convolutional Neural Network) is a class of deep, feed-forward artificial neural networks. The convolutional layer is used connect the input layer and the output layer. Compare to the regular neural network, the amount of the parameters is greatly reduced. The CNN has successfully been applied to analyze visual imagery, such as image and video recognition, recommender systems and natural language processing. (Van den Oord, et al., 2013; Collobert and Weston, 2008)

RNN (Recurrent Neural Network) is designed to sequential data because each neuron or unit can use its internal memory to maintain information about the previous input. One of the most successful application is natural language processing in which the context of each word is required to considered and will help to determine the word coming next. In addition, the Long Short-Term Memory (LSTM), a special kind of RNN, capable of learning long-term dependencies. The model raised by Hochreiter and Schmidhuber (1997) can make a prediction based on both short-term and long-term data.

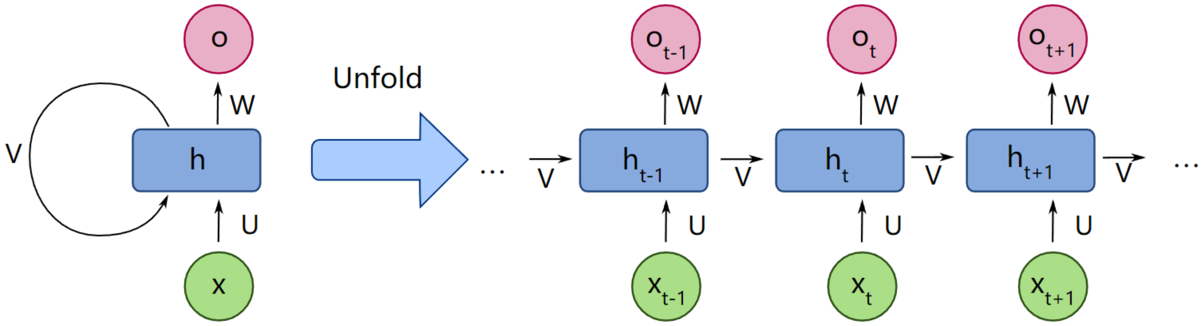


Figure 4.3 Structure of the recurrent neural network

In this research, considering the main characteristic of passenger demand in the area which is continuous and affected by both time of the day and day of the week, the recurrent neural network (RNN) which allows the network to exhibit dynamic temporal behavior for a time sequence is adopted. The structure recurrent neural network is described in **Figure 4.3**. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs.

In the recurrent neural network, each node is connected a weighted neural. The LSTM units are selected for the nodes. **Figure 4.4** illustrates the internal structure of the LSTM unit. The mathematic model of LSTM unit can be described as:

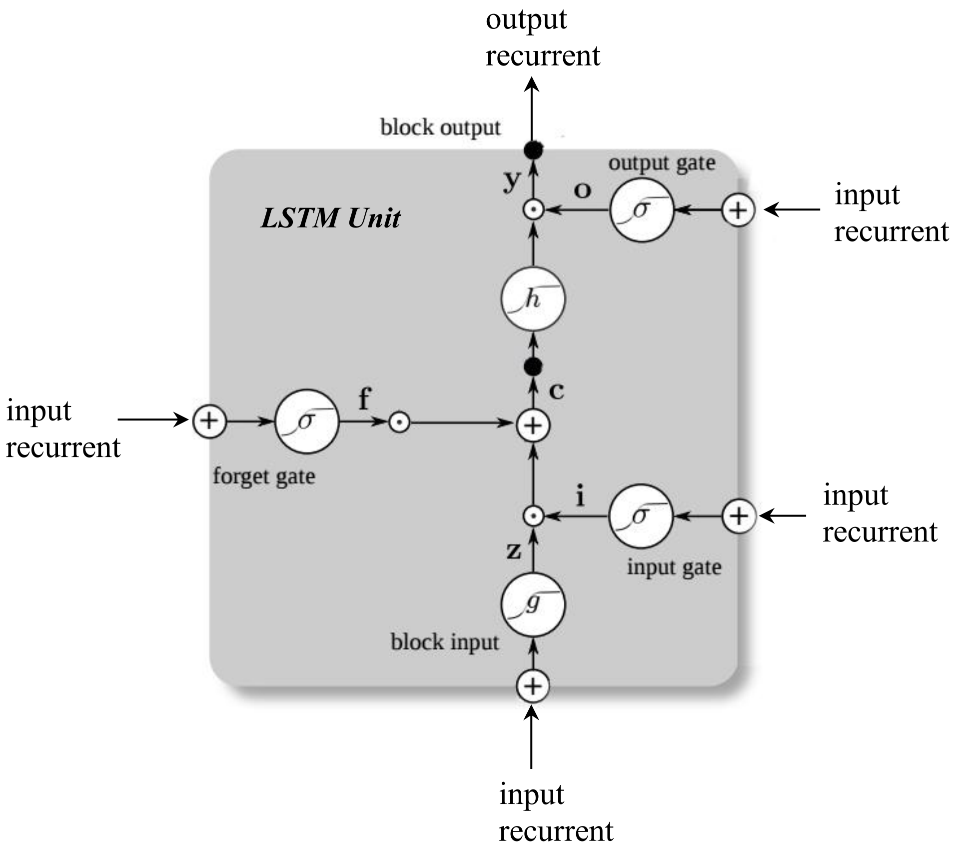


Figure 4.4 Structure of LSTM unit

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \circ \sigma_h(c_t)$$

where "◦" denotes the Hadamard product.

Variables of the LSTM with a forget gate:

$$x_t \in R^d: input\ vector\ to\ the\ LSTM\ unit$$

$$f_t \in R^h: forget\ gate's\ activation\ vector$$

$$i_t \in R^h: input\ gate's\ activation\ vector$$

$$o_t \in R^h: output\ gate's\ activation\ vector$$

$$h_t \in R^h: output\ vector\ of\ the\ LSTM\ unit$$

$$c_t \in R^h: cell\ state\ vector$$

$$W \in R^{h\times d}, U \in R^{h\times d}\ and\ b \in R^{h\times d}: weight\ matrices\ and\ bias\ vector\ parameters$$

In this research, the passenger demand of each period will be described as a vector

$$x_t = (d_{1,t}, d_{2,t}, \dots \dots, d_{s,t})$$

$d_{n,t}$ indicates the number of demand of nth demand point at time t.

The expected output from the RRN can be described as:

$$x_{t+q} = (d_{1,t+q}, d_{2,t+q}, \dots \dots, d_{s,t+q})$$

4.5 Case study

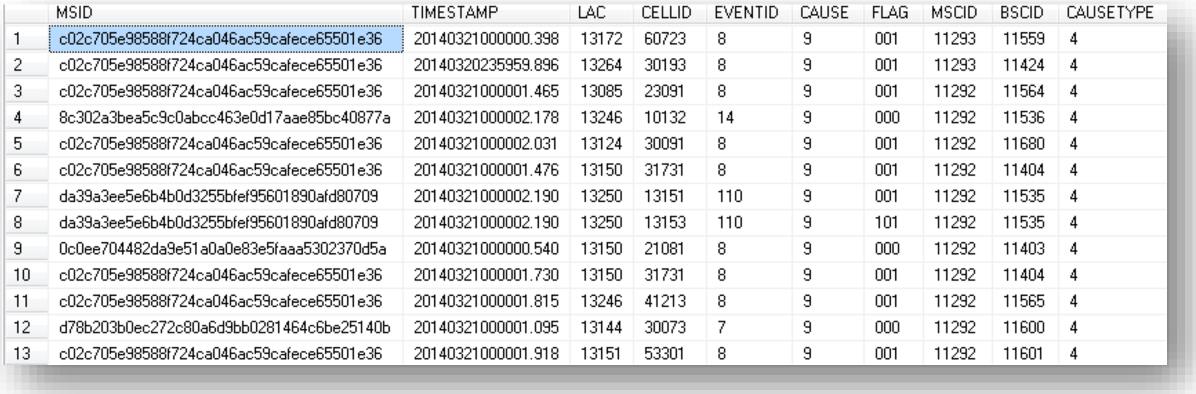
Regarding the special geography environment of Chongqing, a specific area is chosen in the case study. The terminal of the metro line is located at a residential quarter, the other lies at Yuzhong peninsula, one of the largest business districts. In the case study, the trip of the commuters who have an original location in the residential area and take the metro line to the

Yuzhong peninsula will be retrieved from the cellphone data.

It has to be made clear that the method for estimating travel demand by analyzing the signaling data do not infringe the privacy of phone users. The unique equipment identification numbers cannot be converted back to the phone number. Furthermore, the methodology is based on macroscopic scale, and it will not be used to trace the specific cellphone user. Hence, the privacy is protected by the anonymity and inconvertibility of CellIDs.

4.5.1 Cellphone Data Processing

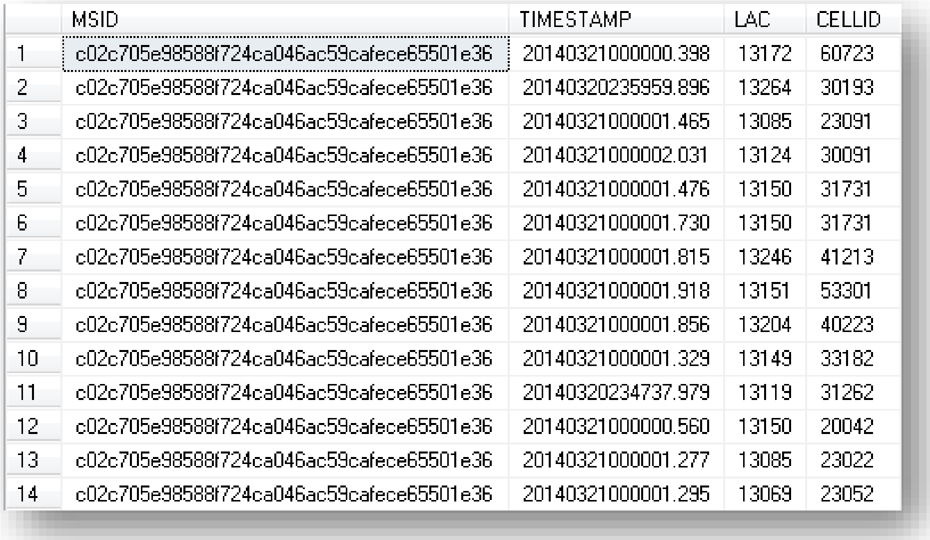
Starting with identifying and eliminating the records of non-human behaviors in the data, cellphone dataset has to be filtered and analyzed. Every mobile telecom carrier has a larger size of the industrial users which will have a negative effect on the accuracy of the study. To exclude industry users such as GPS Tracker and information recipient, an appropriate statistical method has to be applied to screen the original cellular dataset. Each ID will be checked if its frequency comes within the range of the reasonable frequency according to the statistical approach.



	MSID	TIMESTAMP	LAC	CELLID	EVENTID	CAUSE	FLAG	MSCID	BSCID	CAUSETYPE
1	c02c705e98588f724ca046ac59cafece65501e36	20140321000000.398	13172	60723	8	9	001	11293	11559	4
2	c02c705e98588f724ca046ac59cafece65501e36	20140320235959.896	13264	30193	8	9	001	11293	11424	4
3	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.465	13085	23091	8	9	001	11292	11564	4
4	8c302a3bea5c9c0abcc463e0d17aae85bc40877a	20140321000002.178	13246	10132	14	9	000	11292	11536	4
5	c02c705e98588f724ca046ac59cafece65501e36	20140321000002.031	13124	30091	8	9	001	11292	11680	4
6	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.476	13150	31731	8	9	001	11292	11404	4
7	da39a3ee5e6b4b0d3255bfef95601890afd80709	20140321000002.190	13250	13151	110	9	001	11292	11535	4
8	da39a3ee5e6b4b0d3255bfef95601890afd80709	20140321000002.190	13250	13153	110	9	101	11292	11535	4
9	0c0ee704482da9e51a0a0e83e5faaa5302370d5a	20140321000000.540	13150	21081	8	9	000	11292	11403	4
10	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.730	13150	31731	8	9	001	11292	11404	4
11	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.815	13246	41213	8	9	001	11292	11565	4
12	d78b203b0ec272c80a6d9bb0281464c6be25140b	20140321000001.095	13144	30073	7	9	000	11292	11600	4
13	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.918	13151	53301	8	9	001	11292	11601	4

Figure 4.4 Sample of Cellphone Dataset

The industrial devices can account for 10 percent of total cellular devices in the Chinese carriers' network. After analyzing the frequency of occurrence of each ID counted, 5th and 95th percentile are set as upper bound and lower bound. Therefore, the IDs whose activities are less than 5 or more than 154 are considered as industry devices and will be excluded in our case.



	MSID	TIMESTAMP	LAC	CELLID
1	c02c705e98588f724ca046ac59cafece65501e36	20140321000000.398	13172	60723
2	c02c705e98588f724ca046ac59cafece65501e36	20140320235959.896	13264	30193
3	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.465	13085	23091
4	c02c705e98588f724ca046ac59cafece65501e36	20140321000002.031	13124	30091
5	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.476	13150	31731
6	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.730	13150	31731
7	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.815	13246	41213
8	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.918	13151	53301
9	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.856	13204	40223
10	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.329	13149	33182
11	c02c705e98588f724ca046ac59cafece65501e36	20140320234737.979	13119	31262
12	c02c705e98588f724ca046ac59cafece65501e36	20140321000000.560	13150	20042
13	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.277	13085	23022
14	c02c705e98588f724ca046ac59cafece65501e36	20140321000001.295	13069	23052

Figure 4.5 Daily Activities of One User

The base stations located in Yuzhong peninsula have to be separated out to find the users appearing in the specific area. The list of targeted stations can be obtained according to the longitude and latitude converted from LacID and CellID of each station. If the user has activities thought the targeted base station, the location of the user can be considered as the location of the base station.

The beginning of office hour in the most company is ranged from 8 am to 10 am in Chongqing. The users who request services during the period in the specific and connect to the base stations in the list from previous become surfaced. These users are considered as having the potential demand to Yuzhong peninsula in the peak hours.

As regular commuters, the residences of them are always fixed. Trace back the activities of the targeted users having potential demand traveling to the Yuzhong peninsula in the morning peak though unique MSID, the base station of the last service request from is considered as the residence as the commuter. In another word, the position of the residence is the origin of each

OD pair. **Table 4.2** presents the number of demand within the selected residential area in the peak hour on Mar 20, 2014, Monday.

Table 4.2 The Original Location of Commuters on 3/3/2014 7am-8am

lac	CELLID	lon	lat	USERS
13122	30122	106.3121	29.61905	53
13122	32251	106.2971	29.61793	51
13122	33021	106.3158	29.6118	50
13122	31121	106.3066	29.61786	47
13122	30142	106.2941	29.61117	39
13122	31143	106.2995	29.61279	35
13122	30073	106.3119	29.61538	22
13122	30123	106.3173	29.62278	21
13122	32101	106.3097	29.6296	19
13122	33082	106.3149	29.61408	18
13122	30143	106.2965	29.61536	17
13122	31072	106.3114	29.61227	16
13122	33022	106.315	29.60886	14
13122	30133	106.295	29.61433	13
13122	31122	106.3098	29.61788	13
13122	33023	106.3059	29.60852	12
13122	32362	106.3116	29.61229	11
13122	31141	106.302	29.61425	11
13122	31252	106.2995	29.61578	11
13122	33072	106.3046	29.61932	11
13122	30121	106.3168	29.6248	11
13173	16021	106.2985	29.60671	10
13173	16251	106.2912	29.61049	10
13122	30132	106.3013	29.61041	9
13122	30373	106.3092	29.62988	3

4.5.2 Open GIS Dataset

Considering the service provided by Google is restricted in China, the Baidu Map APIs (similar to Google Map APIs) is adopted here to offer both traffic status and network information, specifically, the travel time of the shortest path in the morning peak hour. Specifically, in **Table 4.3,** 1075 pairs (25 demand points x (42 candidate bus stops+1 station)) of walking distance between demand points and candidate bus stops are extracted while another 1849 ((1 station + 42 candidates) x ((1 station + 42 candidates)) pairs of shortest travel time during morning peak hour (7:00 am – 8:00 am) among candidate bus stops and between candidate stops, and urban rail transit station are also obtained in **Table 4.4**.

Table 4.3 Walking Distance Matrix (Demand Points to Bus Stop Candidates)

Distance		Candidate Stops																																									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
Demand Points	1	4300	1100	3000	2800	2500	3500	1200	840	530	360	800	650	420	480	730	990	1200	1918	1852	1768	1624	1473	1226	1110	1134	1241	1438	1573	1676	1803	2143	1954	1604	1600	1457	1545	1665	1390	1606	1700	1397	1465	1000
	2	3332	3227	3016	2827	3607	3409	1152	822	570	344	801	647	416	474	727	986	1212	1916	1850	1766	1622	1471	1224	1108	1132	1239	1436	1571	1674	1801	2141	1952	1602	1597	1437	1525	1645	1388	1604	1706	1395	1463	1000
	3	2907	2802	2591	2675	3182	3033	1400	1107	714	824	376	222	8	370	302	561	787	1812	1746	1662	1518	1367	1120	1004	1028	1135	1330	1465	1568	1697	2035	1848	1498	1723	1640	1810	1930	1284	1500	1602	1291	1359	930
	4	3284	3179	2968	2306	3037	2888	1348	1018	900	1270	1193	1039	808	462	1092	847	623	1395	1329	1245	1101	950	703	587	611	718	910	1045	1148	1280	1615	1431	1081	1306	1407	1558	1678	867	1083	1185	874	942	510
	5	2972	2867	2656	2740	3247	3098	1697	1331	1020	889	441	287	313	676	268	258	481	2139	2073	1989	1845	1694	1411	1295	1355	1462	1652	1787	1890	2024	2357	2175	1825	2050	2151	2266	2386	1611	1827	1929	1618	1686	930
	6	3152	3047	2836	2870	3427	3278	1841	1511	1200	1069	621	467	493	856	197	187	301	1959	1893	1809	1665	1514	1267	1151	1175	1282	1472	1607	1710	1844	2177	1995	1645	1870	1971	2122	2242	1431	1647	1749	1438	1506	750
	7	1733	1628	1417	1522	2033	1884	1436	1583	1894	2061	1768	1922	2153	2516	2205	2464	2692	2463	2493	2577	2687	2842	3089	3150	3174	3281	3476	3611	3714	3843	4181	3233	2889	2129	2028	2116	1798	3430	3646	3538	3183	3251	3100
	8	1932	1616	1182	729	580	457	290	1079	716	1059	910	392	62	1169	1002	1308	1213	1158	1100	1592	1827	1193	1567	1510	1645	677	1862	1204	1554	2215	1063	1748	778	1285	1449	1767	1524	885	1665	1456	1500	1500	1100
	9	1967	1870	1657	1223	530	621	342	267	578	745	1193	1347	1367	1004	1632	1855	1880	882	816	732	671	827	1074	1190	1481	1588	1797	1932	2035	2150	2502	1223	873	447	346	434	554	1737	1624	1522	1168	1236	1400
	10	1852	1747	1536	1061	434	404	698	623	934	1101	1549	1703	1723	1360	1988	2211	2024	865	799	715	654	810	1057	1173	1626	1733	1943	2078	2181	2295	2648	1206	856	430	329	283	403	1718	1607	1506	1151	1219	1500
	11	1270	1165	954	520	343	194	879	1026	1337	1504	1952	2106	2126	1763	2437	2570	2340	973	1039	1031	970	1126	1373	1489	1942	2049	2259	2394	2497	2611	2964	1522	1172	746	645	494	374	2034	1924	1822	1467	1535	1900
	12	525	420	209	813	1338	1189	1517	1664	1975	2142	2325	2479	2710	3081	2764	3023	3249	1738	1804	1888	1996	2151	2400	2514	2968	3075	3285	3420	3523	3637	3990	2544	2198	1805	1704	1553	1433	3060	2949	2847	2492	2560	2900
	13	671	566	355	350	988	839	1167	1314	1625	1792	1975	2129	2360	2658	2414	2673	2899	1271	1337	1421	1529	1684	1931	2047	2501	2608	2818	2953	3056	3170	3523	2077	1731	1338	1237	1086	966	2593	2482	2380	2025	2093	2400
	14	1205	1100	889	255	1000	851	1179	1326	1637	1804	2509	2663	2894	2567	2948	3038	2808	1180	1246	1330	1438	1593	1840	1956	2410	2517	2727	2862	2965	3079	3432	1986	1640	1247	1146	995	875	2502	2391	2289	1934	2002	2300
	15	3496	3391	3180	2856	3587	3438	2185	1855	1544	1413	965	811	837	1200	531	272	262	1933	1867	1783	1639	1483	1236	1120	1144	1251	1440	1575	1678	1813	2243	1969	1619	1856	1957	2108	2228	1400	1616	1718	1407	1475	740
	16	4085	3980	3769	3104	3838	3689	2455	2126	2008	2175	2303	2149	1918	1569	2084	1825	1586	2142	2076	2073	1889	1733	1486	1370	911	804	595	460	357	242	449	1402	1752	2100	2208	2359	2479	840	1056	1158	1430	1362	1300
	17	2836	2731	2520	1858	2589	2440	1945	1615	1497	1664	1807	1653	1422	1059	1731	1530	1283	929	863	779	635	479	309	425	645	538	635	770	873	987	1339	965	615	858	960	1111	1231	442	658	648	293	361	820
	18	3169	3064	2853	2191	2922	2773	1842	1512	1394	1561	1704	1550	1319	956	1625	1420	1180	1262	1196	1112	968	812	642	757	298	191	287	422	525	640	992	924	948	1192	1292	1443	1563	95	311	413	626	617	720
	19	2754	2600	2438	1776	2507	2358	1568	1493	1415	1582	1725	1571	1340	977	1649	1448	1201	847	781	697	553	397	227	343	714	607	704	839	942	1056	1408	883	533	776	877	1029	1149	511	668	566	211	279	740
	20	2714	2609	2398	1736	2467	2318	1528	1453	1375	1542	1685	1531	1300	937	1609	1400	1161	807	741	657	513	357	187	303	762	721	818	953	1056	1170	1523	804	493	736	837	989	1109	573	439	337	16	51	700
	21	3062	2957	2746	2084	2815	2666	1876	1777	1659	1826	1969	1815	1584	1221	1894	1684	1445	1155	1089	1005	861	705	535	651	563	456	553	688	791	905	1257	797	841	1084	1185	1336	1456	308	92	10	364	296	980
	22	3167	3062	2851	2189	2920	2771	2002	1672	1554	1721	1864	1710	1479	1116	1785	1579	1340	1260	1194	1110	966	810	640	756	458	351	448	583	686	800	1152	802	946	1189	1290	1441	1561	203	13	115	469	401	880
	23	2851	2746	2535	1836	2602	2453	1668	1593	1904	2184	2327	2173	1942	1579	2253	2047	1808	907	841	798	654	587	834	950	1409	1516	1738	1603	1500	1365	1592	206	451	877	978	1129	1249	768	754	726	760	692	1300
	24	2416	2311	2100	1438	664	634	932	752	1063	1230	2194	2040	1809	1446	2120	1917	1687	528	462	378	317	473	720	836	1289	1396	1606	1741	1844	1958	2311	869	519	93	137	288	408	1381	1270	1168	814	882	1200
	25	2272	2167	1956	1302	903	873	967	892	1203	1370	1818	1972	1808	1445	2119	1916	1686	527	461	377	316	472	719	835	1288	1395	1605	1740	1843	1957	2310	868	518	261	362	513	633	1381	1269	1167	812	880	1200

Table 4.4 Travel Time Matrix (Candidates to Candidates)

sec	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	MS
1	0	27	69	190	296	269	329	344	398	431	424	451	491	615	505	555	600	348	363	379	404	433	476	498	589	611	651	675	693	726	794	496	436	381	361	330	305	609	587	564	495	509	571
2	27	0	42	163	269	242	302	316	371	404	396	424	464	587	478	527	573	321	335	352	376	405	449	471	562	584	624	647	665	699	766	469	409	354	334	303	277	582	560	536	467	482	544
3	69	42	0	121	227	200	260	275	329	362	355	382	422	545	436	485	531	279	294	310	335	364	407	429	520	542	582	605	624	657	725	427	367	312	292	261	235	540	518	495	425	440	502
4	190	163	121	0	144	116	176	209	264	296	424	451	466	426	559	510	465	158	173	189	214	243	286	308	399	421	461	485	503	536	604	306	246	191	171	140	115	419	397	374	305	319	381
5	296	269	227	144	0	27	33	91	145	178	256	284	289	222	336	385	391	198	184	167	160	189	233	255	345	367	407	431	449	483	550	253	193	115	95	80	105	365	344	320	251	265	327
6	269	242	200	116	27	0	60	118	173	205	284	311	316	249	364	413	418	195	180	164	156	185	229	251	342	364	404	427	445	479	546	249	189	111	91	75	100	245	340	316	247	262	324
7	329	302	260	176	33	60	0	58	113	145	224	251	256	189	304	353	358	227	213	196	189	218	262	284	276	298	338	362	380	414	481	282	222	144	124	113	138	395	373	349	280	295	258
8	344	316	275	209	91	118	58	0	55	87	165	193	198	131	245	295	300	215	200	184	176	284	240	218	218	240	280	304	322	355	423	269	209	131	111	127	153	267	306	325	267	282	200
9	398	371	329	264	145	173	113	55	0	33	111	138	127	138	175	224	282	351	336	320	295	266	223	201	201	223	263	286	305	338	405	352	292	185	165	182	207	250	289	308	250	265	182
10	431	404	362	296	178	205	145	87	33	0	78	105	136	147	160	209	255	384	369	353	328	299	255	234	234	255	295	319	337	371	438	385	325	218	198	215	240	283	322	341	283	297	215
11	424	396	355	424	256	284	224	165	111	78	0	27	67	135	82	131	176	403	388	372	347	318	275	253	253	275	315	338	356	390	457	404	344	296	276	293	318	302	341	360	302	316	240
12	451	424	382	451	284	311	251	193	138	105	27	0	40	107	55	104	149	375	361	345	320	291	247	225	225	247	287	311	329	363	430	376	316	324	304	320	345	275	314	333	275	289	213
13	491	464	422	466	289	316	256	198	127	136	67	40	0	67	47	96	162	335	321	305	280	251	207	185	185	207	247	271	289	323	390	336	276	325	345	376	402	235	274	293	235	249	173
14	615	587	545	426	222	249	189	131	138	147	135	107	67	0	115	164	191	268	254	237	213	184	140	118	118	140	180	204	222	255	323	269	209	258	278	309	335	167	206	225	167	182	105
15	505	478	436	559	336	364	304	245	175	160	82	55	47	115	0	49	95	383	368	352	327	298	255	233	233	255	295	318	336	370	437	384	324	373	393	424	449	282	321	340	282	296	183
16	555	527	485	510	385	413	353	295	224	209	131	104	96	164	49	0	45	352	337	321	296	267	224	202	202	224	264	287	305	339	406	353	293	342	362	393	418	251	290	309	251	265	134
17	600	573	531	465	391	418	358	300	282	255	176	149	162	191	95	45	0	306	292	275	251	222	178	156	156	178	218	242	260	294	361	307	247	296	316	347	373	205	245	264	205	220	88
18	348	321	279	158	198	195	227	215	351	384	403	375	335	268	383	352	306	0	15	31	55	85	128	150	241	263	303	326	345	378	445	148	88	84	104	135	165	261	239	215	146	161	223
19	363	335	294	173	184	180	213	200	336	369	388	361	321	254	368	337	292	15	0	16	41	70	114	135	226	248	288	312	330	364	431	134	74	69	89	120	145	246	225	201	132	146	208
20	379	352	310	189	167	164	196	184	320	353	372	345	305	237	352	321	275	31	16	0	25	54	97	119	210	232	272	295	314	347	415	117	57	53	73	104	129	230	208	185	115	130	192
21	404	376	335	214	160	156	189	176	295	328	347	320	280	213	327	296	251	55	41	25	0	29	73	95	185	207	247	271	289	323	390	93	33	45	65	96	122	205	184	160	91	105	167
22	433	405	364	243	189	185	218	284	266	299	318	291	251	184	298	267	222	85	70	54	29	0	44	65	156	178	218	242	260	294	361	85	25	75	95	125	151	176	155	131	62	76	138
23	476	449	407	286	233	229	262	240	223	255	275	247	207	140	255	224	178	128	114	97	73	44	0	22	113	135	175	198	216	250	317	129	69	118	138	169	195	162	120	96	27	42	95
24	498	471	429	308	255	251	284	218	201	234	253	225	185	118	233	202	156	150	135	119	95	65	22	0	91	113	153	176	195	228	295	151	91	140	160	191	216	140	142	118	49	64	73
25	589	562	520	399	345	342	276	218	201	234	253	225	185	118	233	202	156	241	226	210	185	156	113	91	0	22	62	85	104	137	205	200	182	231	251	282	307	49	88	107	140	155	76
26	611	584	542	421	367	364	298	240	223	255	275	247	207	140	255	224	178	263	248	232	207	178	135	113	22	0	40	64	82	115	183	178	204	253	273	304	329	27	66	85	142	127	98
27	651	624	582	461	407	404	338	280	263	295	315	287	247	180	295	264	218	303	288	272	247	218	175	153	62	40	0	24	42	75	143	193	244	293	313	344	369	42	81	100	156	142	138
28	675	647	605	485	431	427	362	304	286	319	338	311	271	204	318	287	242	326	312	295	271	242	198	176	85	64	24	0	18	52	119	297	267	316	336	367	393	65	105	124	180	165	162
29	693	665	624	503	449	445	380	322	305	337	356	329	289	222	336	305	260	345	330	314	289	260	216	195	104	82	42	18	0	34	101	279	285	335	355	385	411	84	123	142	198	184	180
30	726	699	657	536	483	479	414	355	338	371	390	363	323	255	370	339	294	378	364	347	323	294	250	228	137	115	75	52	34	0	67	245	319	368	388	419	445	117	156	175	232	217	214
31	794	766	725	604	550	546	481	423	405	438	457	430	390	323	437	406	361	445	431	415	390	361	317	295	205	183	143	119	101	67	0	300	360	435	455	486	512	185	224	243	299	285	281
32	496	469	427	306	253	249	282	269	352	385	404	376	336	269	384	353	307	148	134	117	93	85	129	151	200	178	193	297	279	245	300	0	60	138	158	189	215	151	144	140	147	345	224
33	436	409	367	246	193	189	222	209	292	325	344	316	276	209	324	293	247	88	74	57	33	25	69	91	182	204	244	267	285	319	360	60	0	78	98	129	155	231	180	156	87	102	164
34	381	354	312	191	115	111	144	131	185	218	296	324	325	258	373	342	296	84	69	53	45	75	118	140	231	253	293	316	335	368	435	138	78	0	20	51	76	251	229	205	136	151	213
35	361	334	292	171	95	91	124	111	165	198	276	304	345	278	393	362	316	104	89	73	65	95	138	160	251	273	313	336	355	388	455	158	98	20	0	31	56	271	249	225	156	171	233
36	330	303	261	140	80	75	113	127	182	215	293	320	376	309	424	393	347	135	120	104	96	125	169	191	282	304	344	367	385	419	486	189	129	51	31	0	25	331	280	256	187	202	264
37	305	277	235	115	105	100	138	153	207	240	318	345	402	335	449	418	373	165	145	129	122	151	195	216	307	329	369	393	411	445	512	215	155	76	56	25	0	356	405	381	312	326	289
38	609	582	540	419	365	245	395	267	250	283	302	275	235	167	282	251	205	261	246	230	205	176	162	140	49	27	42	65	84	117	185	151	231	251	271	331	356	0	39	58	115	100	125
39	587	560	518	397	344	340	373	306	289	322	341	314	274	206	321	290	245	239	225	208	184	155	120	142	88	66	81	105	123	156	224	144	180	229	249	280	405	39	0	24	93	78	165
40	564	536	495	374	320	316	349	325	308	341	360	333	293	225	340	309	264	215	201	185	160	131	96	118	107	85	100	124	142	175	243	140	156	205	225	256	381	58	24	0	69	55	184
41	495	467	425	305	251	247	280	267	250	283	302	275	235	167	282	251	205	146	132	115	91	62	27	49	140	142	156	180	198	232	299	147	87	136	156	187	312	115	93	69	0	15	122
42	509	482	440	319	265	262	295	282	265	297	316	289	249	182	296	265	220	161	146	130	105	76	42	64	155	127	142	165	184	217	285	345	102	151	171	202	326	100	78	55	15	0	136
MS	571	544	502	381	327	324	258	200	182	215	240	213	173	105	183	134	88	223	208	192	167	138	95	73	76	98	138	162	180	214	281	224	164	213	233	264	289	125	165	184	122	136	0

4.5.3 Travel Demand Prediction

By applying the data processing methodology explained in section 4.4, the RNN provides the perdition of the number of potential passengers on each demand point for the coming week. In **figure 5.4**, the blue line indicates the sum of the number of passengers obtained from cellphone dataset in the current week. The total number of the potential passengers in the coming week from the RNN is presented as the yellow line.

The RNN network has one input layer, a hidden layer with 4 LSTM units, and a Dense layer connects all LSTM units then makes a prediction. The default sigmoid activation function is used for the LSTM units. The network is trained for 100 epochs, and a batch size of 1 is used.

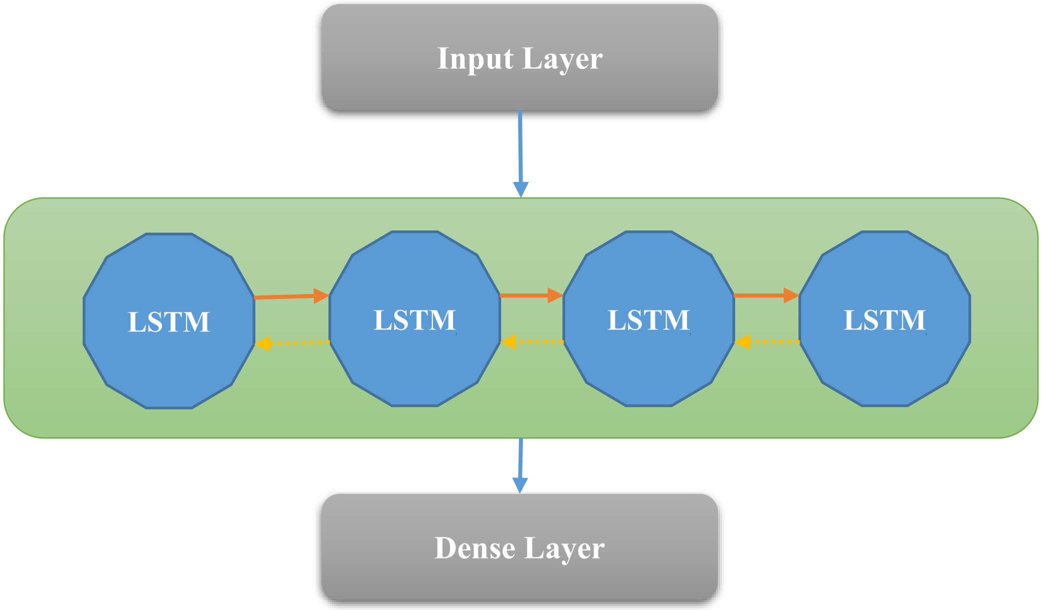


Figure 4.6 The Structure of Proposed RNN

The output from RNN (**Table 4.5**) shows that total 513 passengers in 25 demand points who will catch Line 1 at Jiandingpo Station in morning peak hour on the coming Monday. Moreover, after detection of demand distribution, we pick up 42 candidate bus stops surrounding those 25 demand points based on local traffic network. The prediction of total demands for the coming week in the residential area is presented in **figure 4.7**.

Table 4.5 No. Passengers corresponding to demand points on the coming Monday morning, 7 am – 8 am

Demand Point	No. Passengers	Demand Point	No. Passengers
D1	23	D16	6
D2	36	D17	12
D3	50	D18	25
D4	7	D19	12
D5	22	D20	8
D6	70	D21	6
D7	11	D22	7
D8	25	D23	6
D9	32	D24	5
D10	21	D25	5
D11	30	D14	24
D12	17	D15	25
D13	28		

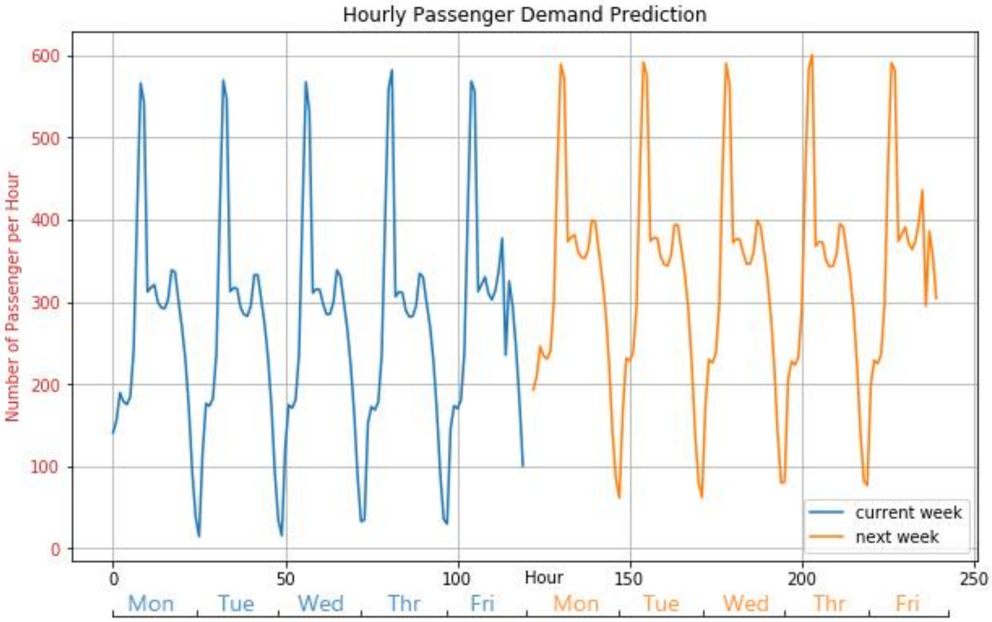


Figure 4.7 Hourly Passenger Demand Prediction

Design Feeder Bus Network based on Aggregated Cellphone Data

5.1 Introduction

This chapter presents a mathematical model to design feeder bus network to access an existing urban rail system. A method was introduced to obtain the real distribution of passenger demand mined from cellular data and travel impedance matrix calculated from an Online GIS tool. The LSTM is applied to estimation the passenger demand in the future. The proposed model features an integrated operational framework, which can simultaneously select bus stops, and design the route from those targeted stops to urban rail stops. This chapter further presents an improved GA–based heuristic approach to yield acceptable solutions to the model in a reasonable amount of time. The model is applied to a real-world case which aims to design a feeder bus system for Jiandingpo Station in Chongqing, China. More than 3.51×10^8 cellular records were filtered and aggregated to obtain the associated demand patterns, and more than 2500 pairs of walking distances, travel time and vehicle distance between demand points and candidate bus stops, among candidate bus stops, were calculated with Online GIS tools to reflect real traffic status and network topology within study areas. Sensitivity analyses were also performed to investigate the impact of the number of designed bus routes on the model performance. The clarity of model inputs and its seamless integration with the commonly used Online GIS offer its best potential to be used as an effective tool for transit authorities to design and refine feeder bus network.

5.2 Research Motivation

Realizing above reviewed limitations of existing studies on FBNDP, this chapter will focus on the following critical research tasks:

1. Propose an integrated optimization model that is capable of seamlessly and

simultaneously coordinating the passenger boarding guidance and transit routing process when the development of feeder bus network;

2. Develop a heuristic solution algorithm to yield the acceptable solution to the proposed model efficiently;
3. Illustrate the proposed methodology through a real-world case study to best understand and apply the proposed methodology during the design process of feeder bus system.

5.3 Research Framework

A feeder bus system is proposed that provide services to transport passengers to urban rail transit station conveniently. Such a system benefits from using cellphone data to extract real distribution of passenger demand which is expected to efficiently solve the issue of demand uncertainty in traditional feeder bus design problem. Moreover, an open GIS tool is further developed to offer traffic status and topological features within study area when designing feeder bus network. With aggregated demand patterns, traffic, and network information, mixed-integer programming is formulated to select the most appropriate locations of bus stops and guide passengers from demand points to their associated stops. Then the model dispatches and routes buses from selected stops to urban rail transit station. Those key components can be better illustrated by a research framework graph, as shown in **Figure 5.1**.

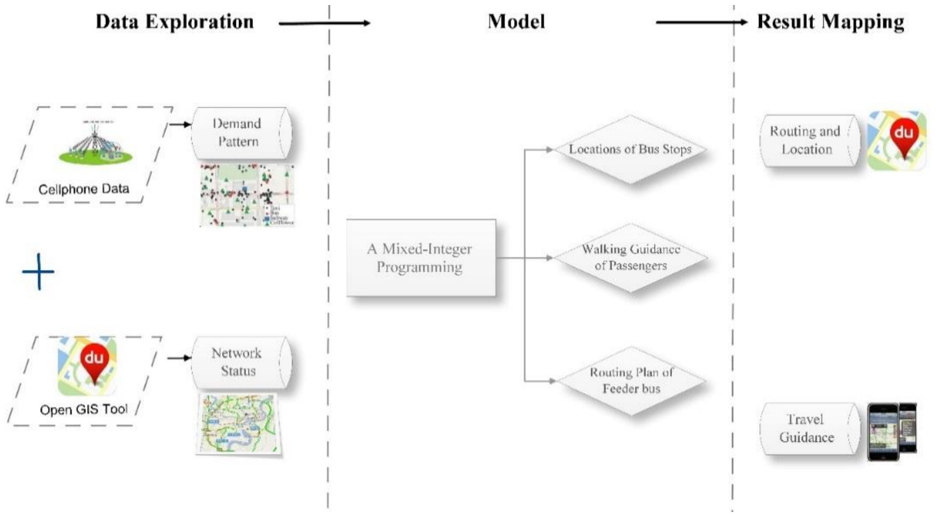


Figure 4.8 Research Framework

In **Figure 5.2**, five candidates (red dots) are finally selected as feeder bus stops due to high accessibility to surrounding demands points (black dots). Three example bus routes are highlighted and shown in solid red line (node 1-node 2- urban rail transit station, node 3-node 4- urban rail transit station and node 5-node 6- urban rail transit station). The aim of the proposed mixed integer program is to find a sub-graph that simultaneously minimize weighted passengers walking distance and operational cost of feeder bus system. In this illustrated network, passengers are assigned to stops concerning minimizing their walking distances that is also the principle of determining the locations of bus stops in our study. Once the demand is assigned at the stops, the feeder route for each bus will be constructed to transfer passengers waiting at the stops to their connected urban rail transit station by finding out the shortest path. Different from phase-based or stage-based approaches in existing studies, we formulate it as a combined stop location and vehicle routing problem.

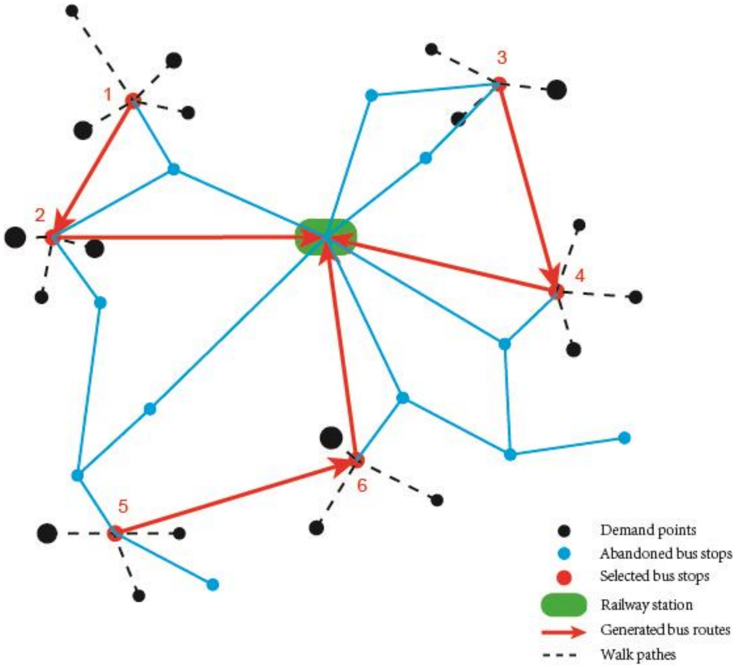


Figure 4.9 Graphical representation of the integrated FBDNP problem

5.4 Model Formulation

5.4.1 Notation

To facilitate the model presentation, all definitions and notations used hereafter are summarized in **table 5.1**.

Table 4.6 Parameters and variables in the mathematical model

	Indices
i	Demand point index
j,m,p	Vehicular node (bus stop candidates and urban rail transit station) index
k	Bus route index
	Sets
I	Set of demand points
M	Set of bus stop candidates
MS	Set of urban rail transit stations, without loss of generality we are assuming a single station in this model
K	Set of bus routes
Parameters	

$Demand_i$	Number of passengers at demand point *i*; $i \in I$
P	Maximum number of designed bus stops;
Q_k	The capacity of bus route *k*; $k \in K$
T_{max}	Maximum travel time;
L_{min}	Minimum route length;
d_{ij}	Map-based walking distance from demand point *i* to bus stop candidate *j*; $i \in I, j \in M$
t_{jm}	Map-based vehicular distance from vehicular node *j* to vehicular node *m*; $j, m \in M \cup MS$
C_m	Operational cost per km (unit: dollar);
C_h	Operational cost for drivers per operating hour (unit: dollar);
C_p	The value of the passenger's walking time per hour (unit: dollar);

Decision Variables

c_{jm}^k	Number of passengers at stop j assigned to route k traveling from j to m (unit: person);
U_{ik}	An auxiliary (real) variable for sub-tour elimination constraint in the route of bus *k*;
$z_{jm}^k = \begin{cases} 1 \\ 0 \end{cases}$	If stop *j* precedes point *m* on the route *k*; Otherwise
$x_{ij} = \begin{cases} 1 \\ 0 \end{cases}$	If demand point *i* is assigned to stop candidate *j*; Otherwise;
$y_j = \begin{cases} 1 \\ 0 \end{cases}$	If candidate node *j* is selected as a stop; Otherwise;

5.4.2 Formulation

The proposed problem can be formulated as the following mixed integer program (MIP):

Minimize:

$$(C_m \sum_{j\in M}\sum_{m\in M\cup MS}\sum_{k\in K} Z_{jm}^k d_{jm} + C_h \sum_{j\in M}\sum_{m\in M\cup MS}\sum_{k\in K} Z_{jm}^k t_{jm}) + C_p \sum_{i\in I}\sum_{j\in M} demand_i\, d_{ij} x_{ij}$$

(5-1)

Subject to:

$\sum_{j\in M} y_j \le P;$ (5-2)

$x_{ij} \le y_j \qquad \forall i \in I, \forall j \in M;$ (5-3)

$\sum_{j\in M} x_{ij} = 1 \qquad \forall i \in I;$ (5-4)

$2 * z_{jm}^k \le y_j + y_m \qquad \forall k \in K, \forall j \in M, \forall m \in M \cup MS;$ (5-5)

$\sum_{j\in M}\sum_{k\in K} z_{jm}^k \le 1; \quad \forall m \in M;$ (5-6)

$\sum_{p\in M}\sum_{k\in K} z_{pj}^k \le 1; \quad \forall j \in M;$ (5-7)

$\sum_{m\in M\cup MS} z_{jm}^k - \sum_{p\in M} z_{pj}^k \ge 0; \qquad \forall j \in M, \forall k \in K;$ (5-8)

$U_{ik} - U_{jk} + \left(|H| * z_{jm}^k\right) \ge |H| - 1, \forall j, m \in M \cup MS, \forall k \in K;$ (5-9)

$\sum_{j\in M}\sum_{m\in M\cup MS} c_{jm}^k \le Q_k; \qquad \forall k \in K;$ (5-10)

$\sum_{m\in M\cup MS}\sum_{k\in K} c_{jm}^k = \sum_{i\in I} Demand_i * x_{ij}; \qquad \forall j \in M;$ (5-11)

$c_{jm}^k - z_{jm}^k \ge 0; \qquad \forall j \in M, \forall m \in M \cup MS, \forall k \in K;$ (5-12)

$c_{jm}^k \le Q_k * z_{jm}^k; \qquad \forall j \in M, \forall m \in M \cup MS, \forall k \in K;$ (5-13)

$\sum_{j\in M}\sum_{m\in M\cup MS} z_{jm}^k * t_{jm} \le T_{max}; \ \forall k \in K;$ (5-14)

$\sum_{j\in M}\sum_{m\in M\cup MS} z_{jm}^k * d_{jm} \ge L_{min}; \ \forall k \in K;$ (5-15)

$\sum_{j\in M}\sum_{m\in MS} z_{jm}^k = 1; \qquad \forall k \in K;$ (5-16)

$\sum_{j\in M}\sum_{m\in MS} z_{mj}^k \le 0; \qquad \forall k \in K;$ (5-17)

In this formulation, the objective function is given by Eq. (1), which includes two terms: the first term is dealing with routing and the second one is related to location selection of bus stops and assigning passengers to those targeted stops. The first term minimizes the total operational cost for designed feeder bus system and the second term minimizes the total equal value of the walking distances from demand points to selected stops.

Constraint (5-2) indicates that the number of selected bus stops should be no more than the allowed maximum number. Constraint (5-3) and (5-4) guarantee the demand points can only be assigned to those selected stop candidates and each demand point must be matched and served by only one bus stop. Constraint (5-5) specifies the bus route links may exist between two candidate nodes only if both candidates are selected as stops. Constraint (5-6) guarantees each selected bus stop can only serve one bus route which aims to avoid undesirable competition among bus routes and further increase the whole system efficiency. Constraint (5-7) and (5-8) set each bus stop (except urban rail station) being served to have the same incoming and outgoing arcs. Constraint (5-9) is used for sub-tour elimination in the vehicle routing problem and is a constraint with polynomial cardinality (Miller, 1995). Constraint (5-10) guarantees the number of passengers in each route boarding from selected bus stops and transported to the urban rail transit station must be less than the vehicle capacity during each route. Constraint (5-11) and (5-12) ensure that all the passengers are picked up. Constraint (5-13) guarantees that passengers are assigned to the route only if this route serves the selected stop candidates. Constraint (5-14) and (5-15) are used to limit the minimum length and maximum travel time for each route. Constraint (5-16) and (5-17) ensure that each route is eventually ended at the urban rail transit station.

5.5 A GA-based Heuristic Algorithm

The proposed optimization model is an extension of the vehicle routing problem that has been proved to be non-deterministic polynomial-time hard (NP-hard). In case of small-scale networks, some powerful solvers such as IBM ILOG CPLEX may be able to find an optimal solution for proposed problems. However, those tools are intractable for those large-scale network problems.

Genetic algorithm (GA) based heuristics algorithm have been widely used to find the exact or approximate solution for optimization problems and made great achievements. (Chakroborty et al.,1995, 1997, 2001, 2002, 2003; Bielli et al., 2002) John Holland introduced Genetic algorithm in the early 1970s (Hines, et al., 1997). A cumulative genetic algorithm was applied to design transportation network in Vignaux's research (1991). Ceylan et al. (2004) used genetic algorithm approach to solve traffic signal control and traffic assignment problem and show that the GA approach is efficient and much simpler than the previous heuristic algorithms. Lin et al. (2014) proposed a GA-based optimization model for designing green transportation schemes and provide a guidance of implementing green transportation for the logistics service providers. Hua et al. (2014) proposed a customized genetic algorithm (GA) with a specially designed mutation mechanism is designed to solve the model efficiently. Ardjmand et al. (2016) applied a genetic algorithm to a new bi-objective stochastic model for transportation, location, and allocation of hazardous materials and finds optimum and high-quality solutions for both small and large problems.

Thus, a GA-based heuristic approach is further developed to efficiently yield acceptable solutions to the model in a reasonable amount of running time.

5.5.1 Coding of GA chromosomes

An efficient coding of GA chromosomes which is able to capture the characteristics of the solution structure plays a key role in the process of GA searching. In our study, the main body of proposed optimization model is composed of selecting stop locations, matching demand points with candidate bus stops, and designing routing plans that are corresponding to $y_j, x_{ij}\ and\ z_{jm}^k$ respectively. Therefore, if we use vectors $U = (u_1, u_2, \dots, u_M, u_{M+1}, \dots, u_{2M}, u_{2M+1}, \dots, u_{2M+I})$ to represent solutions to this model, each vector consists of $(2M + I)$ binary strings, it could be further decomposed into three parts, as explained follows:

The first part of GA chromosomes $(u_1, u_2, \dots, u_M)$ (the vector of binary variables) is used to represent the decision of location selection of bus stops. If $u_j = 1$, then the corresponding candidate nodes *j* is targeted as a feeder bus stop;

The second part of GA chromosomes $(u_{M+1}, u_{M+2}, \dots, u_{2M})$ (the vector of integer variables) is used to assign the selected stops to different routes, thus each *u* ranges from 1 to *k* where $k \in K$ represents the number of designed bus routes. For example, $u_{M+2} = 3$ indicates that bus stop *(M+2)* is assigned to route 3. A Dijkstra algorithm is then implemented to search the shortest bus route dispatching from urban rail transit station so as to order the sequence of targeted stops for each bus route;

The third part of GA chromosomes $(u_{2M+1}, u_{2M+2}, \dots, u_{2M+I})$ (the vector of integer variables) is used to match demand points to the closest stops for each bus route. Each *u* ranges from 1 to *k* where $k \in K$ represents the number of designed bus routes. The function of $(\sum_{i \in I} \sum_{j \in M} demand_i\, d_{ij} x_{ij})$ in model objective is utilized to obtain the distance matrix so as to determine the value of $x_{ij} = \{i | min \sum_i demand_i d_{ij}\}$ which help each demand point find its most appropriate boarding stop.

5.5.2 Fitness Evaluation

Note that the candidate solutions may violate constraints (5-10), (5-14) and (5-15). To deal with this problem, we include those constraints as penalty terms into the function of fitness evaluation. Thus, the modified fitness function in our study is given by:

$$F = f + M_1 \cdot \sum_{k \in K} \left(\max \left(\sum_{j \in M} \sum_{m \in M \cup MS} c_{jm}^{k} - Q_k, 0 \right) \right)^2 +$$

$$M_2 \cdot \sum_{k \in K} \left(\max \left(T_{max} - \sum_{j \in M} \sum_{m \in M \cup MS} z_{jm}^{k} * t_{jm}, 0 \right) \right)^2 +$$

$$M_3 \cdot \sum_{k \in K} \left(\max \left(\sum_{j \in M} \sum_{m \in M \cup MS} z_{jm}^{k} * d_{jm} - L_{min}, 0 \right) \right)^2$$

Where f is the objective function (Equation 5-1) of the proposed model; F is the function used in fitness evaluation. And M_1, M_2, M_3 are large positive penalty coefficients.

5.5.3 A Heuristic Algorithm of Generating Initial Population

As it has been widely recognized, the quality of the solution found, or the computational resources required by applying the GA-based algorithm highly depends on the selection of initial population to the proposed problem. Thus, to better solve the presented model and improve computational efficiency, a heuristic algorithm for generating feasible initial population is further developed to embed into GA process. The procedures are explained as follows:

Step 1. Input parameters defined within the proposed model, namely, *M* (a set of candidate bus stops), *MS* (urban rail transit station) and *K* (a set of feeder bus routes), etc.;

Step 2. For each route $k \in K$, starting with the node of rail station: 1) initial a feasible set of candidate bus stops $M' \in M$ in which distance between selected node and rail station is less than L_{min}; 2) randomly select a candidate bus stop from M'; 3) repeat searching next node from the rest nodes in M' until violating the constraints (5-2), (5-14) and (5-15);

Step 3. Set $y_j = y_m = 1$ in case of $z_{jm}^{k} = 1$ for each route $k \in K$;

Step 4. Calculate the distance between each demand point $i \in I$ and all confirmed bus stops, and then determine the values of x_{ij}which is for assigning the most appropriate bus stop to each demand point with considering the capacity constraints and minimum walking distance;

Step 5. Use the obtained decision of $y_{j,}x_{ij}, z_{jm}^{k}$ from Step 1 to 4 to generate initial population U;

5.6 Genetic Operators

5.6.1 Selection

Selection operators give preference to better solutions (chromosomes), allowing them to pass on their 'genes' to the next generation of the algorithm. In our study, uses both of random competition and elitist selection strategies to ensure that individuals with the highest fitness in the previous population are retained in the next population.

5.6.2 Crossover and Mutation

Crossover operator simulates recombination for exchange part of genes in two individuals to produce new individuals in evolutionary processes. In our study, we use a one-point method which randomly selects an integer P between 1 and $(2M + I)$, and exchange the front and the rear portions of two parent $U_1\ and\ U_2$ to generate new offspring chromosome $U_1{}'and\ U_2{}'$. Mutation operator in this study is also implemented with one-point mea thod where we define mutation fraction to 0.15. If gene $u_j \in U$ has been selected as a mutation point, and then u_j is set to 1 or 0 in case of $j \in [1, M]$ while u_j randomly takes value range from 1 to $|K|$.

5.6.3 Stopping Criteria

The GA stops to evolve until the following criteria are met:

(1).$\left|\frac{\hat{\mathcal{F}}^{n}_{min}-\hat{\mathcal{F}}^{n-1}_{min}}{\hat{\mathcal{F}}^{n}_{min}}\right| < \epsilon$, i.e., the difference between the minimum evaluation values between two adjacent generations is less than a threshold ϵ; or

(2). A pre-set maximal number of generations are reached.

5.7. Case Study

To illustrate the applicability of the proposed Data-driven framework and models in designing feeder bus network access to the urban rail transit, this study has selected Jiandingpo Station at Metro Line 1 in Chongqing (the biggest municipality under direct administration by the Chinese central government) for a case study, aiming to design a feeder system for transporting those detected demand to Jiandingpo Station at morning peak hour from 7:00 am to 8:00 am. As the first stop of Line 1, Jiandingpo station is located at the west part of Chongqing. Its geographical location is given in **Figure 5.3**

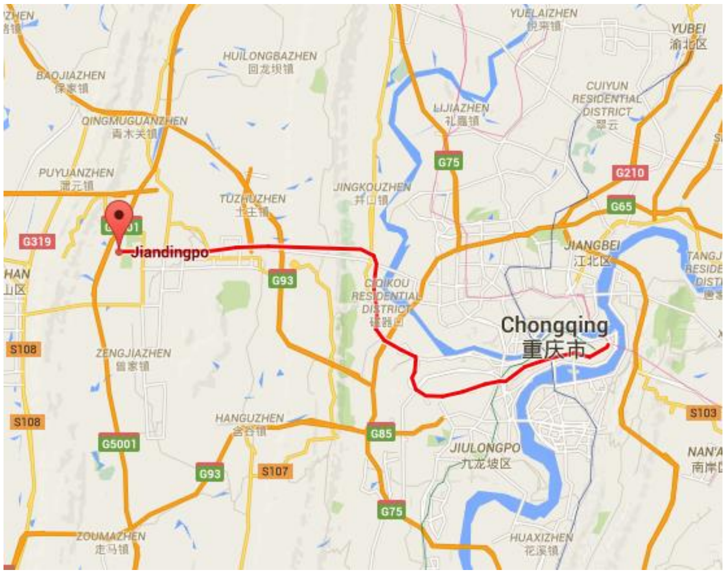


Figure 4.10 Geolocation of Jiandingpo Station (Source: Google Map)

5.7.1 Cellphone Dataset

The cellphone dataset used in this case study to extract the spatial distribution of demand in relation to Jiandingpo station is collected from both China Mobile and China Unicom (the two biggest communication operators in China) during March 3rd 2015 and March 7th, 2015. **Table 5.2** describes some basic information about dataset used in this study.

Table 4.7 Cellphone Dataset Description

Day	Size	No. Records	No. Devices
March 3rd 2015	15.7GB	1.78×10^8	7.61×10^6
March 4th 2015	15.2GB	1.73×10^8	7.58×10^6
March 5th 2015	15.2GB	1.73×10^8	7.58×10^6
March 6th 2015	15.2GB	1.73×10^8	7.58×10^6
March 7 h 2015	15.5GB	1.76×10^8	7.60×10^6

Figure 5.4 maps both inferred demand points and chose candidate stops where larger size of the blue dot represents a larger demand size and red dots represent candidate bus stops. Thus, the

main objective of this case study is to design a convenient feeder system for transporting those detected demand to Jiandingpo Station at morning peak hour from 7:00 am to 8:00 am.

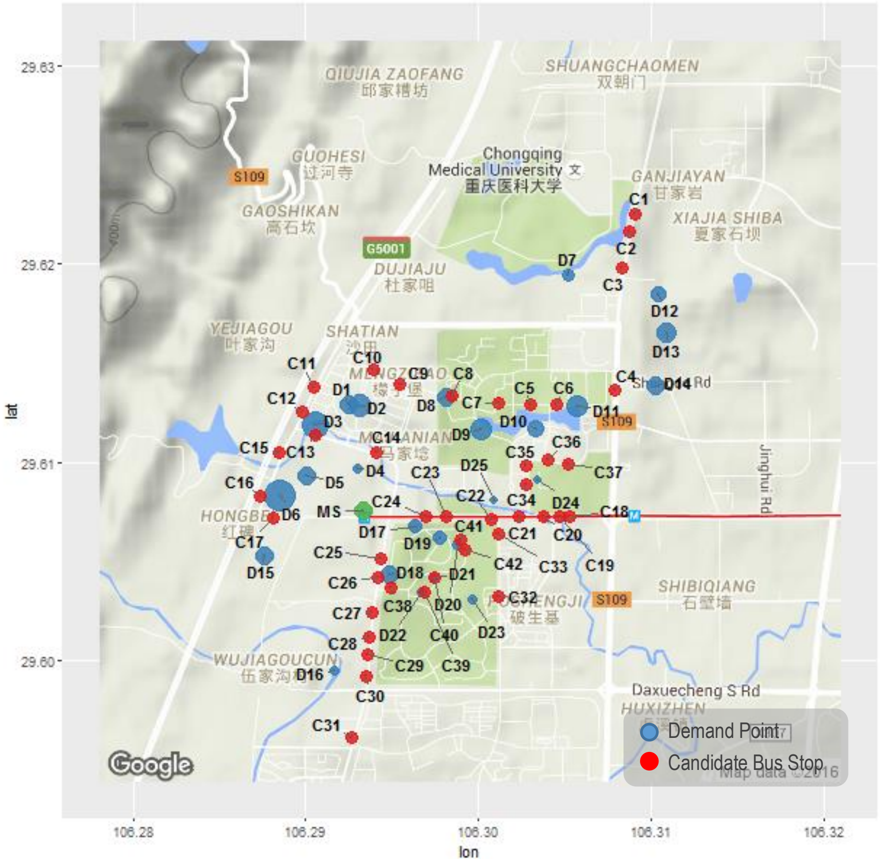


Figure 4.11 Spatial distribution of demand and candidate bus stops

5.7.2 Results Analysis

Key parameters and assumptions used in the case study are given as follows:

- No. of Bus routes: 3;
- Vehicle capacity: 180 persons;
- No. of maximum allowed Stops: 16;
- Maximum allowed travel time for each route: 20 min;
- Minimum route length: 2 km;

- Operational cost for feeder buses: \$3 per km;
- Operational cost for drivers: \$5 per hour.
- The value of passenger's walking time: \$1 per hour

Using the demand collected from cellphone data exploration, distance and time matrix generated from Open GIS tool, the proposed model was firstly solved in *CPLEX 12.6*. **Table 5.3** summarizes the assignment results of passengers from demand points to selected candidate bus stops which is also the decision of location selection of bus stops.

Table 4.8 Assignment of passengers from demand points to selected bus stops

Demand Point	Candidate Bus Stop	Served Demand	Bus Route	Walking Distance (meter)
D7	C3	11	R1	1417
D12	C3	17		209
D13	C4	28		255
D14	C4	24		350
D11	C6	30		434
D23	C32	6		206
D24	C34	5		93
D25	C34	5		261
D10	C36	21		283
D17	C41	12		293
D19	C41	12		211
D20	C41	8		16
D9	C8	32	R2	267
D1	C10	23		360
D2	C10	36		344
D3	C13	50		8
D8	C8	25		62
D4	C14	7		462
D5	C16	22	R3	187
D6	C16	70		258
D15	C16	25		272
D16	C30	6		242
D18	C38	25		95
D22	C39	7		13
D21	C40	6		10

Table 5.4 details the routing plans for each bus route. Due to the fact that the inputs are generated from Open-GIS tool, the result of vehicle travel time/distance and walking distance is

capable of representing the real traffic status and network topology within the study area. A map-based graphical illustration of bus routing plans, as well as passenger guidance, is shown in **figure 5.5**. The solid red line represents Route 1, the blue solid line represents Route 2, and black dash line represents Route 3. In addition to bus route plan, those solid green lines indicate the walking paths of passengers.

Table 4.9 Routing Plans and Number of Passengers Served

Route Index	Routes	Length(km)	Travel Time(min)	Served Demand	Weighted Average Walking Distance(m)
1	C3-C4-C6-C36-C34-C32-C41-MS	5.07	15.5	179	351
2	C10-C8-C14-C13-C16-MS	3.90	11.5	290	219
3	C30-C39-C40-C38-MS	2.05	6.0	44	90

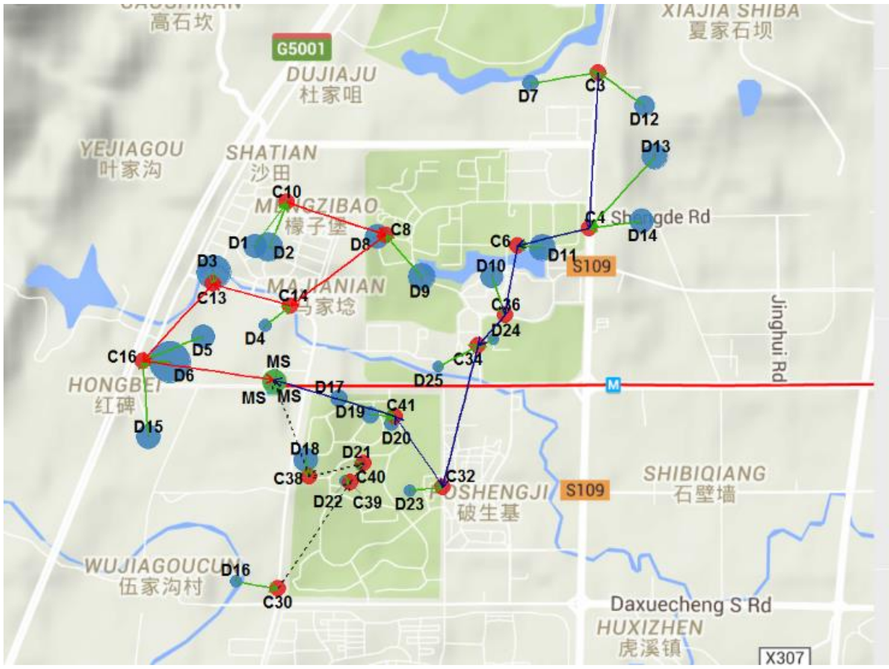


Figure 4.12 Case Study Result (Map source: Google)

CPLEX 12.6 has successfully found a global optimization for this case. However, it takes about 390-second computing time even for this small-scale test network, which may raise an issue of computation efficiency and hinder its application in large-scale or more complicated cases. Therefore, we also implement the GA-based heuristic algorithm developed in this chapter to solve the scenario of 3 routes, 4 routes and 5 routes respectively. **Figure 5.6** shows the convergence process of GA algorithm for three scenarios. As more routes are designed, more iterations are required for convergence. A result comparison of CPLEX and GA algorithm is unfolded regarding computation efficiency and solution difference, as recorded in **Table 5.5**. One can observe that the computation time of CPLEX solved for global optimization is up to more than one hour in case of designing 5 routes while the developed heuristic algorithm takes about 82 seconds to find an acceptable near-optimal solution. The error of average bus route length and travel time is controlled under 20% only except for scenario of 4 routes problem. The difference of weighted average walking distance is varied from 15% to 30% which increases along with the number of routes.

Considering the feasible solution set grows exponentially with the increase of the number of routes. In the case study, the time consumption to get the optimal solution of 5 route plan is becoming painful in the CPLEX. However, the GA shows its capability of getting the near-optimal solution within acceptable time, meanwhile, the error level is still under control. The threshold can be set as 4 routes in this case. Once the number of routes is less or equal than 4, CPLEX may be productive to obtain the optimal solution, otherwise, GA is the better way to get the near-optimal solution.

Table 4.10 Comparison of CPLEX Solution and Heuristic Solution

	CPLEX Results				**Heuristic Results**				**Difference**		
Scenario	Solved Time (s)	Average Route Length (km)	Average Travel Time (min)	Average Weighted Walking Distance (m)	Solved Time (s)	Average Route Length (km)	Average Travel Time (min)	Average Weighted Walking Distance	Average Route Length	Average Travel Time	Weighted Walking Distance
3 Routes	390	3.67	11.00	253.74	48	4.17	12.44	292.94	14%	13%	15%
4 Routes	573	2.61	7.85	253.74	66	3.15	9.63	307.13	21%	23%	21%
5 Routes	3813	2.29	6.72	253.74	82	2.64	7.89	329.59	15%	17%	30%

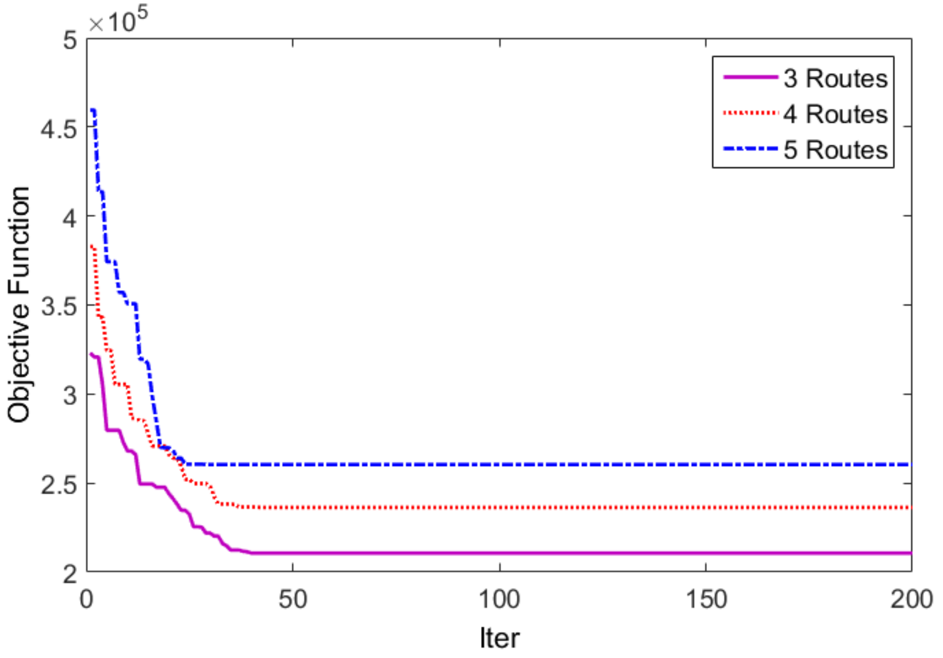


Figure 4.13 Convergence Process of GA Algorithm of Three Scenarios

5.8 Conclusion

This chapter presents a data-driven approach for designing feeder bus network connecting to the urban rail transit station. Different from existing studies, the proposed methodology features in: 1) developing a mixed integer programming to offer an interactive process of pedestrian guidance (from home addresses to candidate bus stops) and transit routing (from selected bus stops to urban rail transit stop, such integration will significantly improve the performance of the feeder bus system; 2) introducing a big data processing technology for extracting aggregated-level spatial distribution of demand with using cellphone dataset to solve the issue of demand uncertainty; 3) retrieving map-based travel distance and time information to include the network characteristics and traffic status by using Open GIS tool; 4) developing an improved GA-based heuristic algorithm in which a heuristic algorithm of generating the initial population is further proposed and embedded. The feasibility and applicability of the proposed model are illustrated with a real-world example, Jiandingpo Station of Chongqing Metro Line1, solved to optimality. Results show that the proposed model can yield valid and detailed passenger walking guidance and transit routing plans for feeder bus system. To validate the performance of the developed heuristic algorithm, a comparison of CPLEX solutions and heuristic algorithm solutions in case of designing 3 routes, 4 routes and 5 routes respectively, the results suggest that the proposed algorithm is able to yield effective solutions to the proposed problem in acceptable computation time.

Note that the problem studied in this chapter is static in the way that the inferred OD table and the number of bus routes are all stable. Determination of bus stops locations and assignment of demands and routing also use a static representation of the network. Therefore, this model is very useful at the initial stage of strategic feeder bus network planning. Especially, because of a good connection with map engine, this model also has a potential to embed into transit APPs' development which is able to guide passengers starting trips from home. Extending the model to an explicitly dynamic setting with time-varying demand generation rates and travel times is a worthwhile direction for further work and future research.

Chapter 5: Design Feeder Bus Network with a Time-window based Optimization Approach

6.1. Introduction

Due to the uncertain demand, insufficient budget and improper routes and schedule, the feeder bus network usually runs insufficiently and hard to be the top of the commuters' travel choice. In this chapter, a data-mining approach for feeder bus network design and scheduling is proposed to provide precise service to the commuters. Consider the commuters are time-sensitive during the peak hours, a well-designed and appropriate feeder services will be a boost for the willing of commuters to take public transit. The data-mining process is applied to cellphone data to retrieve the location information and time-window of each potential demand. The online GIS tool is employed to obtain route with real-time traffic status. An optimization model is constructed to establish the feeder bus network.

6.2 Objectives and contributions

Realizing above reviewed limitations of existing studies on FBNDP, this chapter will primarily focus on the following research tasks to:

1. Develop a cellphone data mining methodology to retrieve trip information, including original, destination and path. Also, departure time and arrival time are able to be mined out in the data procedure.
2. Introduce a method to obtain traffic information from an online GIS tool. The total driving time, walking time, route information and traffic condition outputted from the GIS will help the optimization model to generate the route and the schedule of the feeder bus network.

3. Establish an optimization model to generate the route and schedule of the feeder bus network and provide precise service in which the pickup time and location are considered.

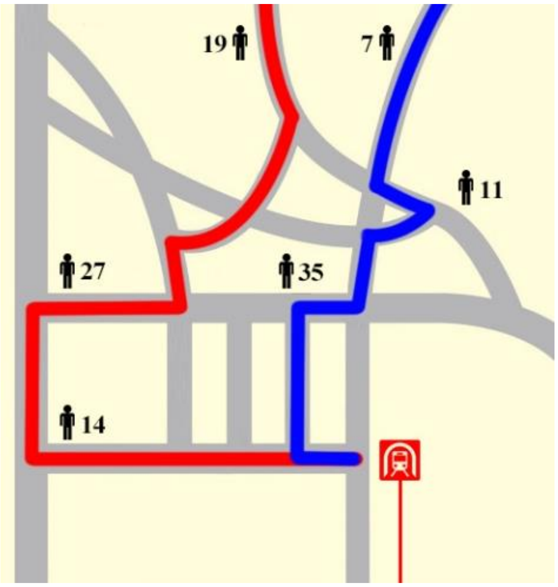


Figure 5.2 Traditional solution for FBNDP

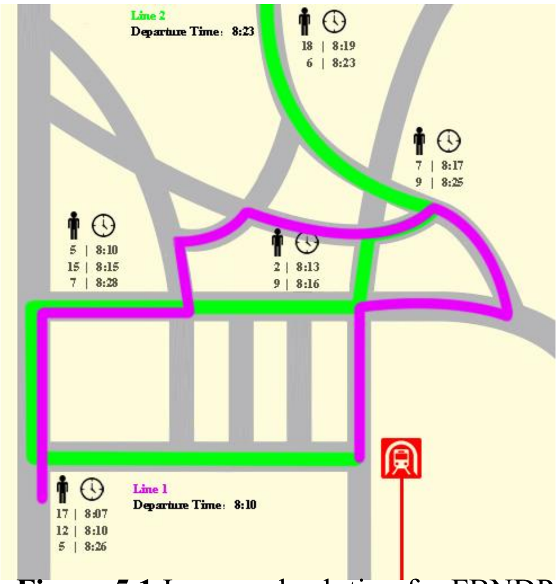


Figure 5.1 Improved solution for FBNDP

The time-windows of the commuters will be considered in the proposed model. As shown below (**Figure 6.1** and **Figure 6.2**), the routes generated by the system will be different when the time-window of the passengers is considered.

6.3 Research Framework

In this research, the feeder buses are dispatched to provide feeder bus services considering the time-window of the commuters. Unlike the traditional way on solving FBNDP in which only the quantity and the location of demand is met, this research aims to provide precise feeder bus services that transfer the passenger to the subway station with consideration of their departure time from the residence. Such a system benefits from using cellphone data to obtain the information of each demand including location and departure time in the peak hours. Moreover, the GIS tools are further developed to acquire route length, total driving time of each section of potential bus routes and total walking time of each possible walking path. The integer programming model is applied in the optimization model to select the proper bus routes and schedule. Such model is extended from the transitional feeder bus network design model. The time-windows constraints are added to yield the time-window of the commuters. The objective function is aimed to maximize the total services rate yield the budget limit. Those key components can be better illustrated by a research framework graph, as shown in **Figure 6.3**.

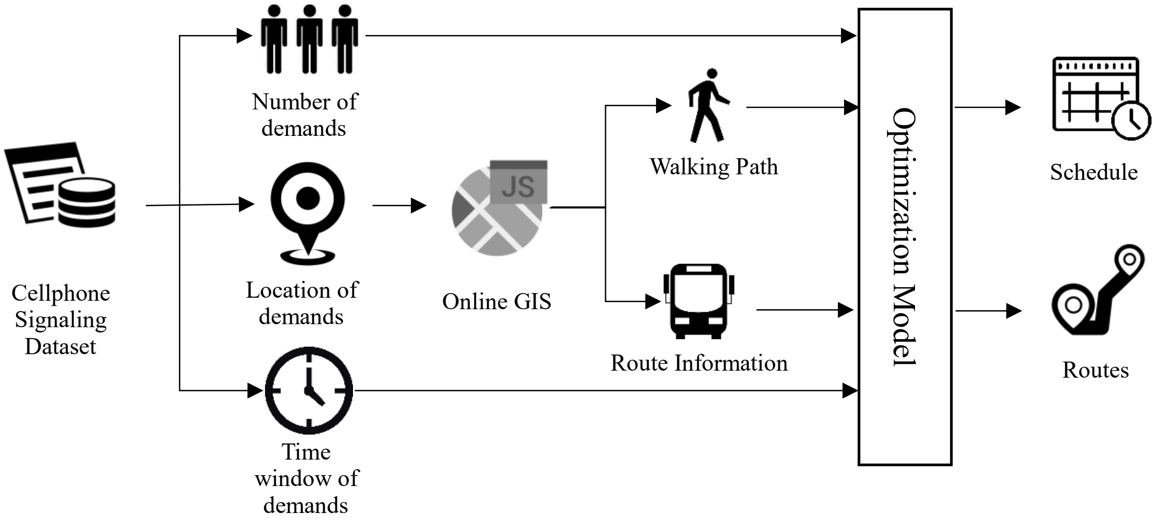


Figure 6.3 Research Framework of time-windowed FBNDP

6.4 Model Formulation

6.4.1 Notation

Decision Variables

x_{ijt}^{k}: Demand point i is assigned to bus route k stop j at time t

u_{jmt}^{k}: Route k start from i to j at time t

y_{jt}: The candidate is chosen at time t

Indices

i	Demand point index
j,m,p	Vehicular node (bus stop candidates and urban rail transit station) index
k	Bus route index
t,r	Time

Sets

I	Set of demand points
M	Set of bus stop candidates
MS	Set of urban rail transit stations, without loss of generality we are assuming a single station in this model
K	Set of bus routes
T	Set of morning peak time period

Parameters

TD	Number of total demand in the system
r_{it}	Number of passengers on demand point i at time t
Q_k	The capacity of bus route *k*; $k \in K$
T_{max}	The maximum travel time of the bus;
D_{walk}	Acceptable walking distance
d_{ij}	Map-based walking distance from demand point *i* to bus stop candidate *j*; $i \in I, j \in M$
t_{jm}	Map-based vehicular distance from vehicular node *j* to vehicular node m; $j, m \in M \cup MS$
$t_{acp-stop}$	The acceptable dwell time at the bus station
$t_{waiting}$	The acceptable waiting time of the passenger after assigned to the bus stop

6.4.2 Formulation

The proposed problem can be formulated as the following mixed integer program (MIP):

Objective function:

$$\text{Maximize } Z = \sum_{i \in I} \sum_{j \in M} \sum_{t \in T} \sum_{k \in K} \frac{x_{ijt}^{k} \cdot r_{it}}{TD} \qquad (6\text{-}1)$$

The objective function is given by Eq. (6-1) which maximizes the services rate of the feeder bus system in the service area. The numerator is the total number of passengers served, and denominator indicates the total potential passengers. It aims to find the optimal feeder bus network under the budget while serving the maximum commuters.

Subject to:

$$\sum_{k\in K}\sum_{j\in M} x_{ijt}^{k} \le 1 \qquad \forall i\in I, \forall t\in T; \qquad (6\text{-}2)$$

$$\sum_{k\in K}\sum_{j\in M} x_{ijt}^{k} \le r_{it} \qquad \forall i\in I, \forall t\in T; \qquad (6\text{-}3)$$

$$x_{ijt}^{k} d_{ij} \le D_{walk} \qquad \forall i\in I; \forall j\in M, \forall t\in T, \forall k\in K \qquad (6\text{-}4)$$

$$\sum_{r\in[t-t_{waiting},t+t_{waiting}]}\sum_{m\in M} u_{jmr}^{k} \le x_{ijt}^{k} \qquad \forall i\in I; \forall j\in M, \forall t\in[5,T-5], \forall k\in K \qquad (6\text{-}5)$$

$$\sum_{t\in T}\sum_{m\in M} u_{jmt}^{k} \le 1 \qquad \forall j\in M;\ \forall k\in K \qquad (6\text{-}6)$$

$$\sum_{j\in M}\sum_{t\in T} u_{jmt}^{k} \le 1 \qquad \forall m\in M;\ \forall k\in K \qquad (6\text{-}7)$$

$$\sum_{j\in M}\sum_{m\in M} u_{jmt}^{k} \le 1 \qquad \forall t\in T;\ \forall k\in K \qquad (6\text{-}8)$$

$$u_{jmt}^{k} \le \sum_{p\in M}\sum_{t'\in[t+t_{jm},t+t_{jm}+t_{acp-stop}]} u_{mpt'}^{k} \qquad \forall k\in K\ ; \forall t\in T \qquad (6\text{-}9)$$

$$\sum_{i\in I}\sum_{j\in M} x_{ijt}^{k} \le \sum_{j\in M}\sum_{t'\in[t,t+T_{max}]} u_{jmt'}^{k} \qquad \forall k\in K\ ; \forall t\in T \qquad (6\text{-}10)$$

$$\sum_{i\in I}\sum_{j\in M}\sum_{t\in T}\left(x_{ijt}^{k}\cdot r_{it}\right) \le Q_k \qquad \forall k\in K; \qquad (6\text{-}11)$$

$$\sum_{j\in M}\sum_{m\in MS}\sum_{t\in T} u_{jmt}^{k} = 1; \qquad \forall k\in K; \qquad (6\text{-}12)$$

$$\sum_{j\in MS}\sum_{m\in M}\sum_{t\in T} u_{jmt}^{k} = 0; \qquad \forall k\in K; \qquad (6\text{-}13)$$

Constraint (6-2) ensures that each demand point only can be assigned once for each period. Constraint (6-3) states that the demand point is activated only there is a passenger at period t.

Constraint (6-4) is used to ensure the assigned bus stop is within the passenger's acceptable walking distance. Constraint (6-5) ensures the bus visits the station within *t_waiting* seconds when the passenger is assigned. Constraint (6-6) and (6-7) guarantee the bus can departure from and arrive at each station at most once. Constraint (6-8) ensures the bus only can appear at one place in each period t. Constraint (6-9) limits the bus will depart the bus stop within the acceptable dwell time after it arrives. Constraint (6-10) restrict the total driving time of each feeder bus. Constraint (6-11) guarantees the number of passengers in each route boarding from selected bus stops must be less than the vehicle capacity on each route. Constraint (6-12) and Constraint (6-13) ensures that the metro station is the destination of each route.

6.5 Heuristic Algorithm

The Genetic algorithm (GA) based heuristics algorithm is applied to solve the proposed optimization model as well. However, several differences in coding of GA chromosomes and fitness evaluation among two models have to be pointed out.

In the study, we generate vectors of variables $u_1 = \{(x^1_{1,1}, x^1_{2,1}, \dots x^1_{I,1})_1, (x^1_{1,2}, x^1_{2,2}, \dots x^1_{I,2})_1, \dots (x^1_{1,M}, x^1_{2,M}, \dots x^1_{I,M})_1, \dots (x^1_{1,M}, x^1_{2,M}, \dots x^1_{I,M})_T, \dots (x^K_{1,M}, x^K_{2,M}, \dots x^K_{I,M})_T\}$ represent the decisions of demands allocation; $u_2 = \{(u^1_{1,1}, u^1_{2,1}, \dots u^1_{M,1})_1, (u^1_{1,2}, u^1_{2,2}, \dots u^1_{M,2})_1, \dots (u^1_{1,M}, u^1_{2,M}, \dots u^1_{M,M})_1, \dots (u^1_{1,M}, u^1_{2,M}, \dots u^1_{M,M})_T, \dots (u^K_{1,M}, u^K_{2,M}, \dots u^K_{M,M})_T\}$ represent the decisions of link activation; $u_3 = \{y_{1,1}, y_{1,2}, \dots, y_{1,T}, \dots y_{M,T}\}$ represent the selection of candidate stops. Thus, each chromosome encoding a possible solution to the proposed model will take the following form: $u = \{u_1, u_2, u_3\}$.

Note that the candidate solutions may violate constraints (6-3), (6-4) - (6-11). To deal with this problem, we include those constraints as penalty terms into the function of fitness evaluation. Thus, the modified fitness function in our study is given by:

$$
\begin{aligned}
F = f + M_1 \cdot \sum_{i \in I} \sum_{t \in T} \max\{\left(r_{it} - \sum_{k \in K} \sum_{j \in M} x_{ijt}^{k}\right), 0\} \; + M_2 \\
\cdot \sum_{i \in I} \sum_{j \in M} \sum_{t \in T} \sum_{k \in K} \max\{(D_{walk} - x_{ijt}^{k} d_{ij}), 0\} \; + M_3 \\
\cdot \sum_{k \in K} \max\{(Q_k - \sum_{i \in I} \sum_{j \in M} \sum_{t \in T} (x_{ijt}^{k} \cdot r_{it})), 0\}
\end{aligned}
$$

Where f is the objective function (Equation 6-1) of the proposed model; F is the function used in fitness evaluation. And M_1, M_2, M_3 are large positive penalty coefficients.

6.6 Case Study

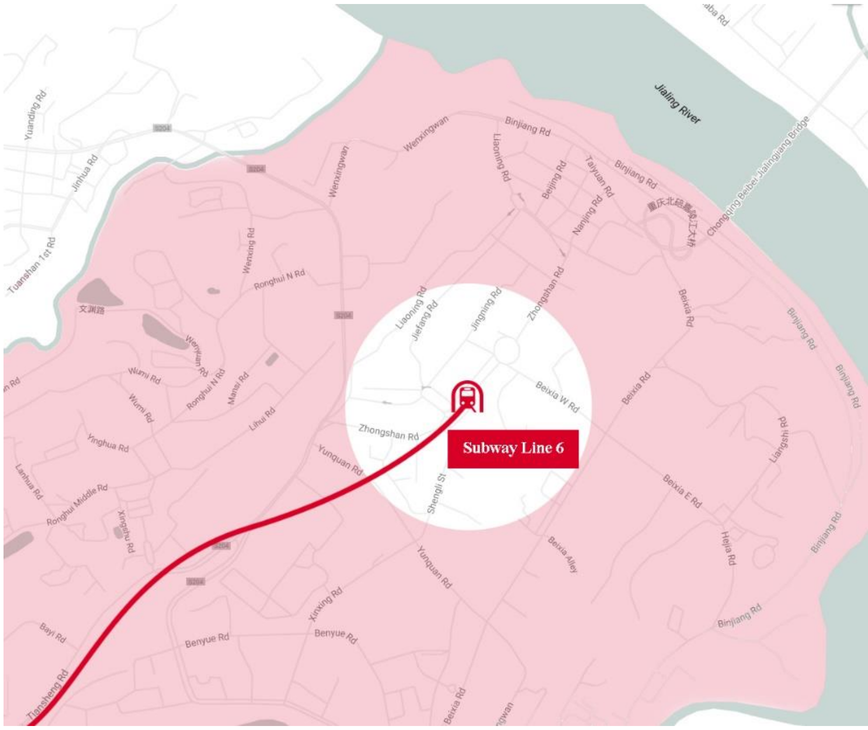


Figure 5.3 Service area of feeder bus network

To verify the feasibility of the proposed method, the area around the destination of Metro line 6 in Chongqing (the biggest municipality under direct administration by the Chinese central government) is selected to test the procedures. The Metro line 6 has a total length of 47.2 mi, and it connects the suburban area to Jiangbei districts in central Chongqing. The daily ridership is up to 452,300, and the lines 6 is also considered as one of the commuters' primary transportation options.

As the terminal of subway line 6 in the suburban area, the Beibei station servers several nearby residential areas. The study area is defined in **figure 6.4**. The shadowed area indicates the service area of proposed feeder bus network in which the distance from other subway stations is considered. In addition, consider the scope of influence of adjacent subway station, the range of service area is restricted to 3 km. Further, the scope within the walkable distance to the metro station is excluded.

The cellphone dataset used in the case study to locate the potential passengers of the feeder bus around the Beibei subway station is collected from one of the three biggest carriers in China. After the cellphone data is processed in the data mining method, the location and the time windows of 57 passengers in the study are extracted. The period selected in the study area is from 7:30 am to 8:00 am in the morning peak. The potential passengers depart from 23 locations. Moreover, 28 bus stop candidates are pre-selected to serve around the demand points with the consideration of the restriction of the street network. Note the bus stop candidates can be randomly distributed yield the required resolution of the network. The computational time will be growing exponentially with the increase of the amount of bus stop candidates and demand locations. Fig 7 presents the spatial distribution of demand using Beibei Station and candidate bus stops

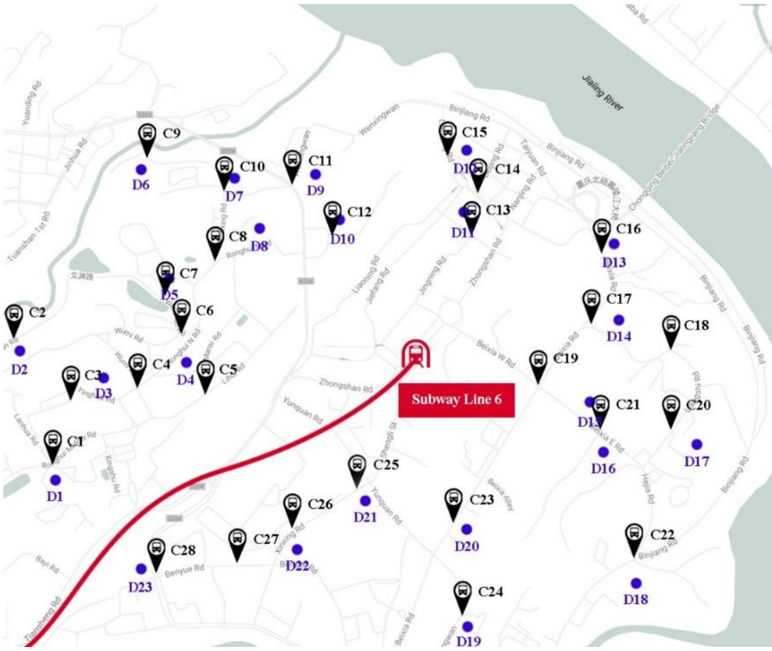


Figure 5.4 Spatial distribution of demand using Beibei Station and candidate bus stops

Table 5.1 Passenger list around the subway station

Passenger	Location	Time	Passenger	Location	Time	Passenger	Location	Time
1	D22	7:01 AM	**20**	D13	7:11 AM	**39**	D9	7:19 AM
2	D16	7:02 AM	**21**	D14	7:12 AM	**40**	D1	7:19 AM
3	D4	7:02 AM	**22**	D7	7:12 AM	**41**	D3	7:19 AM
4	D23	7:03 AM	**23**	D8	7:12 AM	**42**	D13	7:20 AM
5	D20	7:04 AM	**24**	D9	7:12 AM	**43**	D5	7:20 AM
6	D22	7:04 AM	**25**	D16	7:12 AM	**44**	D21	7:21 AM
7	D17	7:05 AM	**26**	D17	7:13 AM	**45**	D2	7:21 AM
8	D18	7:05 AM	**27**	D4	7:13 AM	**46**	D13	7:22 AM
9	D11	7:05 AM	**28**	D8	7:13 AM	**47**	D19	7:22 AM
10	D21	7:06 AM	**29**	D4	7:14 AM	**48**	D13	7:23 AM
11	D15	7:07 AM	**30**	D13	7:15 AM	**49**	D6	7:23 AM
12	D5	7:08 AM	**31**	D12	7:15 AM	**50**	D21	7:24 AM
13	D20	7:08 AM	**32**	D11	7:15 AM	**51**	D7	7:25 AM
14	D1	7:08 AM	**33**	D22	7:15 AM	**52**	D16	7:26 AM
15	D5	7:09 AM	**34**	D7	7:16 AM	**53**	D9	7:27 AM
16	D2	7:10 AM	**35**	D9	7:16 AM	**54**	D20	7:28 AM
17	D4	7:11 AM	**36**	D12	7:17 AM	**55**	D1	7:28 AM
18	D6	7:11 AM	**37**	D11	7:18 AM	**56**	D10	7:29 AM
19	D7	7:11 AM	**38**	D6	7:18 AM	**57**	D21	7:30 AM

The google map engine is adapted to obtain the travel information including total driving time, total driving distance, total walking distance, and total walking time during the morning peak. There are three OD matrix involved in the model, which are demand points to bus stop candidates (23 demand points x 28 bus stop candidates); bus stop candidates to others (28 bus stop candidates x 28 bus stop candidates); bus candidates to subway station (28 bus stop candidates x 1 subway station). After generating the API calls from the OD matrixes, the google map engine returns the travel information as required.

Key parameters and assumptions used in the case study are given as follows:

- No. of Bus routes: 3 routes;
- Route capacity: 35 persons;
- Maximum allowed travel time for each route: 20 minutes;
- Acceptable walking distance: 0.2 km;
- Acceptable waiting time at the bus stop: 5 minutes;
- The acceptable dwell time at the bus station: 2 minutes

Using the demand collected from cellphone data exploration, distance and time matrix generated from online GIS tool, the proposed model was solved in CPLEX 12.6.3 to optimality within 35 minutes. Table 5 summarizes the assignment results of passengers from demand points to selected candidate bus stops which is also the decision of location selection of bus stops. The services rate of the optimal solution is 0.63.

Table 3 Assignment of passengers from demand points to selected bus stops

Bus Route	Demand Point	Candidate Bus Stop	Departure Time	Pick Up Time
R1	D23	C28	7:03 AM	7:03 AM
	D22	C26	7:01 AM	7:05 AM
	D22	C26	7:04 AM	7:05 AM
	D21	C25	7:06 AM	7:06 AM
	D20	C23	7:04 AM	7:09 AM
	D20	C23	7:08 AM	7:09 AM
	D15	C21	7:07 AM	7:12 AM
	D16	C21	7:12 AM	7:12 AM
	D17	C20	7:13 AM	7:13 AM
	D14	C17	7:12 AM	7:15 AM
	D13	C16	7:11 AM	7:16 AM
	D13	C16	7:15 AM	7:16 AM
TERMINAL				7:20 AM
R2	D2	C2	7:10 AM	7:10 AM
	D5	C7	7:08 AM	7:13 AM
	D5	C7	7:09 AM	7:13 AM
	D4	C6	7:11 AM	7:14 AM
	D4	C6	7:13 AM	7:14 AM
	D4	C6	7:14 AM	7:14 AM
	D7	C10	7:12 AM	7:17 AM
	D7	C10	7:16 AM	7:17 AM
	D9	C11	7:19 AM	7:19 AM
	D21	C25	7:21 AM	7:24 AM
	D21	C25	7:24 AM	7:24 AM
TERMINAL				7:26 AM
R3	D6	C9	7:11 AM	7:11 AM
	D7	C10	7:11 AM	7:13 AM
	D8	C8	7:12 AM	7:14 AM
	D8	C8	7:13AM	7:14 AM
	D9	C11	7:12AM	7:16 AM
	D9	C11	7:16 AM	7:16 AM
	D12	C15	7:15 AM	7:19 AM
	D12	C15	7:17 AM	7:19 AM
	D11	C14	7:15 AM	7:20 AM
	D11	C14	7:18 AM	7:20 AM
	D13	C16	7:20 AM	7:23 AM
	D13	C16	7:22 AM	7:23 AM
	D13	C16	7:23 AM	7:23 AM
TERMINAL				7:27 AM

Table 5.2 Routing Plans and Number of Passengers Served

Route Index	Routes	Length(km)	Departure Time	Travel Time(min)	Served Demand
1	C28-C26-C25-C23-C21-C20-C17-C16-T	4.2	7:03 AM	17	12
2	C2-C7-C6-C10-C11-C15-C14-C16-C17-T	3.3	7:10 AM	16	11
3	C9-C10-C8-C11-C15-C14-C16-C17-T	3.8	7:11 AM	16	13

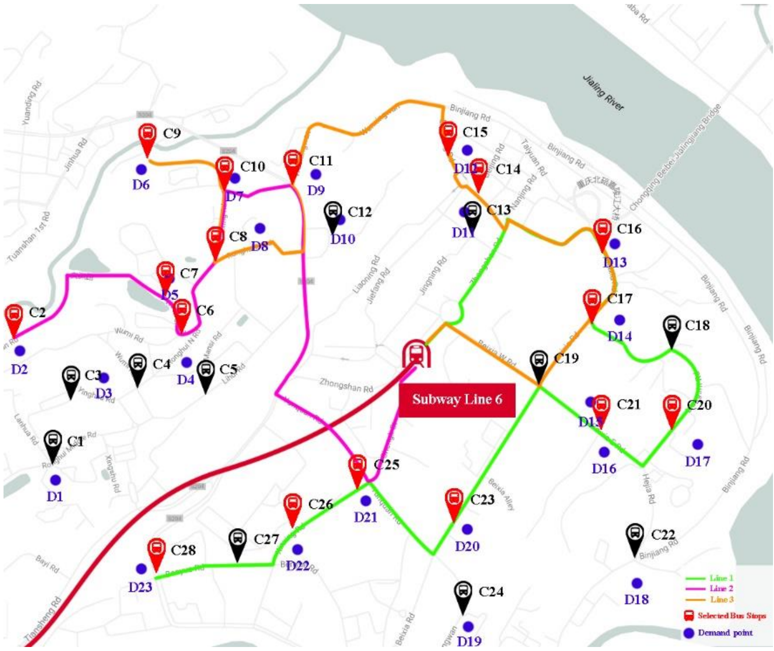


Figure 5.5 Case Study Result

Table 6.2 details the routing plans for each bus route. The inputs are generated from the online GIS tool, so the result of vehicle travel time/distance and walking distance can represent the real traffic status and network topology within the study area. A map-based graphical illustration of bus routing plans is shown in **Figure 6.6**. The green line represents Route 1, purple line represents Route 2, and orange line represents Route3. In addition to bus route plan, the red bus stop indicates that passengers are assigned to the bus stop.

6.7 Conclusion

In this chapter, a data-mining approach to design feeder bus network to access an existing urban rail system with efficient and precise service toward the commuters in which the objective is the maximize the services rate in the operation area. Compare to the existing solution for feeder bus network design problem the proposed methodology contribute to 1) developed a cellphone data mining methodology to retrieve trip information, including original, destination and path. In addition, departure time of the passenger is retrieved and considered when dispatching the feeder bus. 2) Established an optimization model to generate the route and schedule of the feeder bus network and provide precise service in which the pickup time for the passenger with different time window and location are considered. 3) Introduce a method to obtain traffic information from an online GIS tool. The real-world case study verifies the feasibility and applicability of the proposed model which can be solved optimally. Results show that the proposed model can generate feeder bus routes and schedule with the consideration of real-world traffic status and commuters' time-window for feeder bus system.

Chapter 6: Summary and Conclusions

This dissertation develops a data-driven method to the FBNDP. The data processing strategies and optimization models are presented. The cellphone dataset is used as the data source to obtain the OD pair and demand information. The procedure of data acquisition and utilization are detailed, and the big data mining algorithm and up-to-date machine learning model are developed to provide high-quality trip information of the commuters. A neural network RNN is adopted to estimate the future passenger demand. The online GIS tools help to obtain the trip information with real traffic status for the potential routes.

The first designed model features an integrated operational framework, which is able to simultaneously select bus stops, and dispatch and route buses from those targeted stops to urban rail stop. This research further presents an improved GA–based heuristic approach to yield acceptable solutions to the model in a reasonable amount of time. The model is applied to a real-world case which aims to design a feeder bus system for Jiandingpo Station in Chongqing, China. More than 3.51×10^8 cellular records were filtered and aggregated to obtain the associated demand patterns, and more than 2500 pairs of walking distances, travel time and vehicle distance between demand points and candidate bus stops, among candidate bus stops, were calculated with Open GIS tools to reflect real traffic status and network topology within study areas. Sensitivity analyses were also performed to investigate the impact of the number of designed bus routes on the model performance. The clarity of model inputs and its seamless integration with the commonly used Open GIS offer its best potential to be used as an effective tool for transit authorities to design and refine feeder bus network.

Considering the departure time of feeder bus is sensitive to the commuters in the median-low demand area. The model purposed in Chapter 5 is modified and extended in Chapter

6. The optimization model is established to generate the route and schedule of the feeder bus network and provide precise service in which the pickup time for the passenger with different time window and location are considered. The area around the terminal of subway line 6 in the suburban is selected in the case study, the Beibei station servers several nearby residential areas.

Further research on feeder bus network design could focus on improving the data integrity to obtain the high-quality OD pairs; introduce the approximation algorithms to get the solution for thousands of demands retrieved from complete cellphone data from all carriers; generate recurrent dispatching strategy to server more passenger in the peak hours.

References

Aldaihani, M., & Dessouky, M. M. (2003). Hybrid scheduling methods for paratransit operations. Computers & Industrial Engineering, 45(1), 75-96.

Aldaihani, M. M., Quadrifoglio, L., Dessouky, M. M., & Hall, R. (2004). Network design for a grid hybrid transit service. Transportation Research Part A: Policy and Practice, 38(7), 511-530.

Ardjmand, E., Young II, W. A., Weckman, G. R., Bajgiran, O. S., Aminipour, B., & Park, N. (2016). Applying genetic algorithm to a new bi-objective stochastic model for transportation, location, and allocation of hazardous materials. Expert systems with applications, 51, 49-58.

Baaj, M. H., & Mahmassani, H. S. (1991). An AI-based approach for transit route system planning and design. Journal of advanced transportation, 25(2), 187-209.

Baaj, M. H., & Mahmassani, H. S. (1995). Hybrid route generation heuristic algorithm for the design of transit networks. Transportation Research Part C: Emerging Technologies, 3(1), 31-50.

Barra, A., Carvalho, L., Teypaz, N., Cung, V.D., Balassiano, R., (2007). Solving the transit network design problem with constraint programming. In: Proceedings of the 11th World Conference in Transport Research, University of California, Berkeley, USA, June 24–28.

Bielli, M., Caramia, M., Carotenuto, P., (2002). Genetic algorithms in bus network optimization. Transportation Research Part C 10 (1), 19–34.

Bookbinder, J. H., & Desilets, A. (1992). Transfer optimization in a transit network. Transportation science, 26(2), 106-118.

Borndörfer, R., Grötschel, M., & Pfetsch, M. E. (2005). A path-based model for line planning in public transport. Konrad-Zuse-Zentrum für Informationstechnik Berlin.

Bullnheimer, B., Hartl, R. F., & Strauss, C. (1999). Applying the ant system to the vehicle routing problem. In Meta-heuristics (pp. 285-296). Springer, Boston, MA.

Bunte, S., Kliewer, N., Suhl, L. (2006). An overview on vehicle scheduling models in public transport. In: Proceedings of the 10th International Conference on Computer-Aided Scheduling of Public Transport, Leeds, UK. Springer-Verlag, 2006.

Byrne B.F., Vuchic V. (1972). Public transportation line positions and headways for minimum cost. Traffic flow and transportation, 347-360.

Byrne B.F. (1976). Cost minimizing positions, lengths and headways for parallel public transit lines having different speeds. Transport Research, 10(3), 209-214.

Caceres, N., Wideberg, J. P., & Benitez, F. G. (2007). Deriving origin destination data from a mobile phone network. Intelligent Transport Systems, IET, 1(1), 15-26.

Calabrese, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. IEEE Pervasive Computing, 10(4), 0036-44.

Ceder, A. (1984). Bus frequency determination using passenger count data. Transportation Research Part A 18, 453–469.

Ceder, A. (2003). Designing Public Transport Network and Routes. Pergamon Imprint/Elsevier Science Ltd.. pp. 59–91.

Ceder, A., Golany, B., Tal, O. (2001). Creating bus timetables with maximal synchronization. Transportation Research Part A 35, 913–928.

Ceder, A., & Wilson, N. H. (1986). Bus network design. Transportation Research Part B: Methodological, 20(4), 331-344.

Ceylan, H., & Bell, M. G. (2004). Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing. Transportation Research Part B: Methodological, 38(4), 329-342.

Chakroborty, P. (2003). Genetic algorithms for optimal urban transit network design. Journal of Computer Aided Civil and Infrastructure Engineering 18, 184–200.

Chakroborty, P., Deb, K., Porwal, H. (1997). A genetic algorithm based procedure for optimal transit systems scheduling. In: Proceedings of the Fifth International Conference on Computers in Urban Planning and Urban Management, Mumbai, India, pp. 330–341.

Chakroborty, P., Deb, K., Sharma, R.K. (2001). Optimal fleet size distribution and scheduling of urban transit systems using genetic algorithms. Transportation Planning and Technology 24 (3), 209–226.

Chakroborty, P., Deb, K., Subrahmanyam, P.S. (1995). Optimal scheduling of urban transit systems using genetic algorithms. Journal of Transportation Engineering 121 (6), 544–553.

Chakroborty, P., Dwivedi, T. (2002). Optimal route network design for transit systems using genetic algorithms. Engineering Optimization 34 (1), 83–100.

Chang, S. K., & Schonfeld, P. M. (1991). Multiple period optimization of bus transit systems. Transportation Research Part B: Methodological, 25(6), 453-478.

Chien, S., Byun, J., & Bladikas, A. (2010). Optimal stop spacing and headway of congested transit system considering realistic wait times. Transportation planning and technology, 33(6), 495-513.

Chien, S., & Schonfeld, P. (1997). Optimization of grid transit system in heterogeneous urban environment. Journal of Transportation Engineering, 123(1), 28-35.

Chien, S., Tsai, F. M., & Hou, E. (2003). Optimization of multiple-route feeder bus service: application of geographic information systems. Transportation Research Record: Journal of the Transportation Research Board, (1857), 56-64.

Chien, S., & Schonfeld, P. (1998). Joint optimization of a rail transit line and its feeder bus system. Journal of Advanced Transportation, 32(3), 253-284.

Chien, S., & Yang, Z. (2000). Optimal feeder bus routes on irregular street networks. Journal of Advanced Transportation, 34(2), 213-248.

Chowdhury, S. M., & I-Jy Chien, S. (2002). Intermodal transit system coordination. Transportation Planning and Technology, 25(4), 257-287.

Ciaffi, F., Cipriani, E., & Petrelli, M. (2012). Feeder bus network design problem: a new metaheuristic procedure and real size applications. Procedia-Social and Behavioral Sciences, 54, 798-807.

Guihaire, V., & Hao, J. K. (2008). Transit network design and scheduling: A global review. Transportation Research Part A: Policy and Practice, 42(10), 1251-1273.

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning (pp. 160-167). ACM.

Daganzo, C. F. (1984). Checkpoint dial-a-ride systems. Transportation Research Part B: Methodological, 18(4-5), 315-327.

Deb, K., Chakroborty, P. (1998). Time scheduling of transit systems with transfer considerations using genetic algorithms. Evolutionary Computation 6 (1), 1–24.

Deng, L. B., Gao, W., Zhou, W. L., & Lai, T. Z. (2013). Optimal design of feeder-bus network related to urban rail line based on transfer system. Procedia-Social and Behavioral Sciences, 96, 2383-2394

Desrosiers, J., Dumas, Y., & Soumis, F. (1986). A dynamic programming solution of the large-scale single-vehicle dial-a-ride problem with time windows. American Journal of

Mathematical and Management Sciences, 6(3-4), 301-325.

Desrosiers, J., Dumas, Y., Solomon, M. M., & Soumis, F. (1995). Time constrained routing and scheduling. Handbooks in operations research and management science, 8, 35-139.

Dessouky, M.M., Hall, R., Nowroozi, A., Mourikas, K. (1999). Bus dispatching at timed transfer transit stations using bus tracking technology. Transportation Research Part C 7 (4), 187–209.

Dhingra, S.L., Shrivastava, P. (1999). Modelling for coordinated bus train network. In: Proceedings of the Sixth International Conference on Computers for Urban Planning and Urban Management (CUPUM99), Venice, Italy.

Diana, M., Dessouky, M. M., & Xia, N. (2006). A model for the fleet sizing of demand responsive transportation services with time windows. Transportation Research Part B: Methodological, 40(8), 651-666.

Dunn J.A. (1980). Coordination of urban transit services: the German model. Transportation, 9, 33-43.

DiJoseph, P., & Chien, S. I. J. (2013). Optimizing sustainable feeder bus operation considering realistic networks and heterogeneous demand. Journal of Advanced Transportation, 47(5), 483-497.

Ellis, C. J., & Silva, E. C. (1998). British bus deregulation: Competition and demand coordination. Journal of Urban Economics, 43(3), 336-361.

Eranki, A. (2004). A model to create bus timetables to attain maximum synchronization considering waiting times at transfer stops. Master's Thesis, University of South Florida.

Everett, S. (2006). Deregulation and reform of rail in Australia: some emerging constraints. Transport Policy 13, 74–84.

Fan, W., Machemehl, R. (2004). Optimal transit route network design problem: algorithms, implementations, and numerical results. Tech. Rep. SWUTC/04/167244-1, Center for Transportation Research, University of Texas.

Fan, W., & Machemehl, R. B. (2006). Optimal transit route network design problem with variable transit demand: genetic algorithm approach. Journal of transportation engineering, 132(1), 40-51.

Fan, W., Machemehl, R. (2006). Using a simulated annealing algorithm to solve the transit route network design problem. Journal of Transportation Engineering 132 (2), 122–132.

Fei, X., Lu, C. C., & Liu, K. (2011). A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. Transportation Research Part C: Emerging Technologies, 19(6), 1306-1318.

Fischetti, M., & Toth, P. (1989). An additive bounding procedure for combinatorial optimization problems. Operations Research, 37(2), 319-328.

Friedrich, M., Haupt, T., Nökel, K. (1999). Planning and analyzing transit networks - an integrated approach regarding requirements of passengers and operators. Journal of Public Transportation 2 (4), 19–39.

Fu, L., Liu, Q., & Calamai, P. (2003). Real-time optimization model for dynamic scheduling of transit operations. Transportation Research Record: Journal of the Transportation Research Board, (1857), 48-55.

Furth, P., Mekuria, M., & SanClemente, J. (2007). Stop spacing analysis using geographic information system tools with parcel and street network data. Transportation Research Record: Journal of the Transportation Research Board, (2034), 73-81.

Furth, P.G., Wilson, N.H.M. (1982). Setting frequencies on bus routes: theory and practice.

Transportation Research Record 818, 1–7.

Fusco, G., Gori, S., Petrelli, M. (2002). An heuristic transit network design algorithm for medium size towns. In: Proceedings of the 13th Mini-EURO Conference, Bari, Italy.

Gao, Z., Sun, H., Shan, L. (2003). A continuous equilibrium network design model and algorithm for transit systems. Transportation Research Part B 38, 235–250.

Gendreau, M., Guertin, F., Potvin, J. Y., & Taillard, E. (1999). Parallel tabu search for real-time vehicle routing and dispatching. Transportation science, 33(4), 381-390.

Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.

Ghita, C. (2014). A Decision Support System for Business Location Based on Open GIS Technology and Data. Managing Global Transitions, 12(2), 101.

Gholami, A., & Mohaymany, A. S. (2011). Economic conditions for minibus usage in a multimodal feeder network. Transportation Planning and Technology, 34(8), 839-856.

Goodwin, P., 1999. Transformation of transport policy in Great Britain. Transportation Research Part A 33, 655–669.

Guan, J.F., Yang, H., Wirasinghe, S.C. (2003). Simultaneous optimization of transit line configuration and passenger line assignment. Transportation Research Part B 40 (10), 885–902.

Guihaire, V., & Hao, J. K. (2008). Transit network design and scheduling: A global review. Transportation Research Part A: Policy and Practice, 42(10), 1251-1273.

Han, A.F., Wilson, N. (1982). The allocation of buses in heavily utilized networks with overlapping routes. Transportation Research Part B 16, 221–232.

Hasselström, D. (1979). A method for optimization of urban bus route networks. Tech. Rep., Volvo Bus Corporation, Göteborg.

Hickman, M., Blume, K., (2000). A method for scheduling integrated transit service. In: 8th International Conference on Computer-Aided Scheduling of Public Transport (CASPT).

Hickman, M., & Blume, K. (2001). Modeling cost and passenger level of service for integrated transit service. In Computer-aided scheduling of public transport (pp. 233-251). Springer, Berlin, Heidelberg.

Hines, J., Tsoukalas, L. H., & Uhrig, R. E. (1997). MATLAB supplement to fuzzy and neural approaches in engineering. John Wiley & Sons, Inc..

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Hu, H., Gao, Y., Yu, J., Liu, Z., & Li, X. (2016). Planning Bus Bridging Evacuation during Rail Transit Operation Disruption. Journal of Urban Planning and Development, 142(4), 04016015.

Hurdle V.F. (1973). Minimum cost locations for parallel public transit lines. Transport Science, 7, 340 350.

Ioachim, I., Desrosiers, J., Dumas, Y., Solomon, M. M., & Villeneuve, D. (1995). A request clustering algorithm for door-to-door handicapped transportation. Transportation science, 29(1), 63-78.

Jakimavičius, M., & Burinskiene, M. (2009). Assessment of Vilnius city development scenarios based on transport system modelling and multicriteria analysis. Journal of Civil Engineering and Management, 15(4), 361-368.

Jansen, L.N., Pedersen, M.B., Nielsen, O.A. (2002). Minimizing passenger transfer times in public transport timetables. In: Proceedings of the Seventh Conference of the Hong Kong

Society for Transportation Studies: Transportation in the Information Age, Hong Kong, pp. 229–239.

Kim, D., Barnhart, C. (1999). Transportation service network design: models and algorithms. In: Lectures Notes in Economics and Mathematical Systems, vol. 471, Proceeding of the Seventh International Workshop on Computer-Aided Scheduling of Public Transport. Springer-Verlag, pp. 259–283.

Klemt, W.D., Stemme, W. (1988). Schedule synchronization for public transit networks. In: Computer-Aided Transit Scheduling, Proceedings of the Fourth International Workshop on Computer-Aided Scheduling of Public Transport. Springer-Verlag, Hamburg, Germany, pp. 327–335.

Knoppers, P., Muller, T. (1995). Optimized transfer opportunities in public transport. Transportation Science 29 (1), 101–105.

Korotaev, A. V., & Khaltourina, D. (2006). Introduction to social macrodynamics: Secular cycles and millennial trends in Africa. Editorial URSS.

Koutsopoulos, H.N., Odoni, A., Wilson, N.H.M. (1985). Determination of Headways as Function of Time Varying Characteristics on a Transit Network. North- Holland, Amsterdam. pp. 391–414.

Kuan, S. N., Ong, H. L., & Ng, K. M. (2004). Applying metaheuristics to feeder bus network design problem. Asia-Pacific Journal of Operational Research, 21(04), 543-560.

Kuah G.K., Perl J. (1988). Optimization of feeder bus routes and bus-stop spacing. Transport Engineering, 114(3), 341-54.

Kuah G.K., Perl J. (1989). The feeder-bus network-design problem. Journal of the Operational Research Society, 40, 751-767.

Kuan, S. N., Ong, H. L., & Ng, K. M. (2004). Applying metaheuristics to feeder bus network design problem. Asia-Pacific Journal of Operational Research, 21(04), 543-560.

Kuan, S. N., Ong, H. L., & Ng, K. M. (2006). Solving the feeder bus network design problem by genetic algorithms and ant colony optimization. Advances in Engineering Software, 37(6), 351-359.

LeBlanc, L. J. (1988). Transit system network design. Transportation Research Part B: Methodological, 22(5), 383-390.

LeBlanc, L., Morlok, E., Pierskalla, W. (1975). An efficient approach to solving the road network equilibrium traffic assignment problem. Transportation Research 9 (5), 309–318.

Lee, Y.J., Vuchic, V.R. (2005). Transit network design with variable demand. Journal of Transportation Engineering 131 (1), 1–10.

Li, J., Mei, X., Prokhorov, D., & Tao, D. (2017). Deep neural network for structural prediction and lane detection in traffic scene. IEEE transactions on neural networks and learning systems, 28(3), 690-703.

Li, X., & Quadrifoglio, L. (2010). Feeder transit services: choosing between fixed and demand responsive policy. Transportation Research Part C: Emerging Technologies, 18(5), 770-780.

Li, Z. C., Lam, W. H., & Wong, S. C. (2009). Optimization of a bus and rail transit system with feeder bus services under different market regimes. In Transportation and traffic theory 2009: Golden Jubilee (pp. 495-516). Springer, Boston, MA.

Liaw, C. F., White, C. C., & Bander, J. (1996). A decision support system for the bimodal dial-a-ride problem. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 26(5), 552-565.

Lin, C., Choy, K. L., Ho, G. T., & Ng, T. W. (2014). A Genetic Algorithm-based optimization model for supporting green transportation operations. Expert systems with applications, 41(7), 3284-3296.

Lu, Q., & Dessouky, M. M. (2006). A new insertion-based construction heuristic for solving the pickup and delivery problem with time windows. European Journal of Operational Research, 175(2), 672-687.

Lu, X. (2006). Develop web gis based intelligent transportation application systems with web service technology. In ITS Telecommunications Proceedings, 2006 6th International Conference on (pp. 159-162). IEEE.

Lu, X., Yu, J., Yang, X., Pan, S., & Zou, N. (2016). Flexible feeder transit route design to enhance service accessibility in urban area. Journal of Advanced Transportation, 50(4), 507-521.

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies, 54, 187-197.

Magnanti, T.L., Wong, R.T. (1984). Network design and transportation planning: models and algorithms. Transportation Science 18 (1), 1–55.

Mandl, C. E. (1980). Evaluation and optimization of urban public transportation networks. European Journal of Operational Research, 5(6), 396-404.

Martins C.L., Pato M.V. (1998). Search strategies for the feeder bus network design problem. Europe Journal Operational Research, 106, 425-440.

May A.D. (1991). Integrated transport strategies: a new approach to urban transport policy formulation in the UK. Transport Reviews, 2(3), 233-247.

Miller, D. L. (1995). A matching based exact algorithm for capacitated vehicle routing problems.

ORSA Journal on Computing, 7(1), 1-9.

Min, W., & Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. Transportation Research Part C: Emerging Technologies, 19(4), 606-616.

Murray, A.T. (2003). A coverage model for improving public transit system accessibility and expanding access. Annals of Operations Research 123, 143–156.

Murray, A.T., Davis, R., Stimson, R.J., Ferreira, L., 1998. Public transportation access. Transportation Research Part D 3, 319–328.

Newell, G. F. (1979). Some issues relating to the optimal design of bus routes. Transportation Science, 13(1), 20-35.

Ngamchai, S., Lovell, D., 2003. Optimal time transfer in bus transit route network design using a genetic algorithm. Journal of Transportation Engineering 129 (5), 510–521.

Pan, S., Yu, J., Yang, X., Liu, Y., & Zou, N. (2014). Designing a Flexible Feeder Transit System Serving Irregularly Shaped and Gated Communities: Determining Service Area and Feeder Route Planning. Journal of Urban Planning and Development, 141(3), 04014028.

Park, S.J., 2005. Bus network scheduling with genetic algorithms and simulation. Master's Thesis, University of Maryland.

Pattnaik, S.B., Mohan, S., Tom, V.M., 1998. Urban bus transit route network design using genetic algorithm. Journal of Transportation Engineering 124 (4), 368–375.

Paulley, N., Balcombe, R., Mackett, R. L., Preston, J., Wardman, M., Shires, J., ... & White, P. (2004). The demand for public transport.

Peng, Fan, 2004. A decision support system for the design of urban inter-modal public transit network. In: Proceedings of Codatu XI, Bucarest.

Potter, S., & Enoch, M. (1997). Regulating transport's environmental impacts in a deregulating world. Transportation Research Part D: Transport and Environment, 2(4), 271-282.

Psaraftis, H. N. (1980). A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem. Transportation Science, 14(2), 130-154.

Psaraftis, H. N. (1983). An exact algorithm for the single vehicle many-to-many dial-a-ride problem with time windows. Transportation science, 17(3), 351-357.

Quadrifoglio, L., Dessouky, M. M., & Palmer, K. (2007). An insertion heuristic for scheduling mobility allowance shuttle transit (MAST) services. Journal of Scheduling, 10(1), 25-40.

Quadrifoglio, L., & Li, X. (2009). A methodology to derive the critical demand density for designing and operating feeder transit services. Transportation Research Part B: Methodological, 43(10), 922-935.

Quak, C.B., 2003. Bus line planning. Master's Thesis, Delft University of Technology, The Netherlands.

Rapp, M. H., & Gehner, C. D. (1967). Transfer optimization in an interactive graphic system for transit planning (No. Intrm Rpt.).

Rea, J.C., 1972. Designing urban transit systems: an approach to the route-technology selection problem. Highway Research Record 417, 48–59.

Reeves, C.R., 1993. Modern Heuristic Techniques for Combinatorial Problems. Blackwell Scientific Press, Oxford.

Salzborn, F.J.M., 1972. Optimum bus scheduling. Transportation Science 6 (2), 137–148.

Salzborn, F.J.M., 1980. Scheduling bus systems with interchanges. Transportation Science 14 (3), 211–220.

Savelsbergh, M., & Sol, M. (1998). Drive: Dynamic routing of independent vehicles. Operations Research, 46(4), 474-490.

Sexton, T. R., & Bodin, L. D. (1985). Optimizing single vehicle many-to-many operations with desired delivery times: I. Scheduling. Transportation Science, 19(4), 378-410.

Sexton, T. R., & Bodin, L. D. (1985). Optimizing single vehicle many-to-many operations with desired delivery times: II. Routing. Transportation Science, 19(4), 411-435.

Sexton, T. R., & Choi, Y. M. (1986). Pickup and delivery of partial loads with “soft” time windows. American Journal of Mathematical and Management Sciences, 6(3-4), 369-398.

Scheele, S., 1980. A supply model for public transit services. Transportation Research Part B 14, 133–146.

Shih, M., Mahmassani, H.S., 1994. A design methodology for bus transit networks with coordinated operations. Tech. Rep. SWUTC/94/60016-1, Center for Transportation Research, University of Texas, Austin.

Shih, M., Mahmassani, H.S., Baaj, M., 1998. A planning and design model for transit route networks with coordinated operations. Transportation Research Record 1623, 16–23.

Shrivastava, P., & O’Mahony, M. (2006). A model for development of optimized feeder routes and coordinated schedules—A genetic algorithms approach. Transport policy, 13(5), 413-425.

Shrivastava, P., & O’Mahony, M. (2009). Use of a hybrid algorithm for modeling coordinated feeder bus route network at suburban railway station. Journal of Transportation Engineering, 135(1), 1-8.

Shrivastava, P., & O'Mahony, M. (2007). Design of feeder route network using combined genetic algorithm and specialized repair heuristic. Journal of public transportation, 10(2), 7.

Shrivastav, P., & Dhingra, S. L. (2001). Development of feeder routes for suburban railway stations using heuristic approach. Journal of transportation engineering, 127(4), 334-341.

Silman, L.A., Barzily, Z., Passy, U., 1974. Planning the route system for urban buses. Computers & OR 1 (2), 201–211.

Sivakumaran, K., Li, Y., Cassidy, M. J., & Madanat, S. (2012). Cost-saving properties of schedule coordination in a simple trunk-and-feeder transit system. Transportation Research Part A: Policy and Practice, 46(1), 131-139.

Stanger, R. M., & Vuchic, V. R. (1979). The design of bus-rail transit facilities. Transit Journal, 5(4).

Sun, A., & Hickman, M. (2005). The real–time stop–skipping problem. Journal of Intelligent Transportation Systems, 9(2), 91-109.

Tian, Y., & Pan, L. (2015, December). Predicting short-term traffic flow by long short-term memory recurrent neural network. In Smart City/SocialCom/SustainCom (SmartCity), 2015 IEEE International Conference on (pp. 153-158). IEEE.

Ting, C.J., Schonfeld, P., 2005. Schedule coordination in a multiple hub transit network. Journal of Urban Planning and Development 131 (2), 112–124.

Tom, V.M., Mohan, S., 2003. Transit route network design using frequency coded genetic algorithm. Journal of Transportation Engineering 129 (2), 186–195.

Union Internationale des Transports Publics (UITP) (International Association of Public Transport) (2015) , Statistics Brief World Metro Figures, October 2015. (http://www.uitp.org/sites/default/files/cck-focus-papers-files/UITP-Statistic%20Brief-Metro-A4-WEB_0.pdf)

Van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. In Advances in neural information processing systems (pp. 2643-2651).

Van Nes, R., Hamerslag, R., Immers, B.H., 1988. Design of public transport networks.

Transportation Research Record 1202, 74–83.

Vignaux, G. A., & Michalewicz, Z. (1991). A genetic algorithm for the linear transportation problem. IEEE transactions on systems, man, and cybernetics, 21(2), 445-452.

Wan, Q.K., Lo, H.K., 2003. A mixed integer formulation for multiple-route transit network design. Journal of Mathematical Modelling and Algorithms 2 (4),299–308.

Wei, Y., & Chen, M. C. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C: Emerging Technologies, 21(1), 148-162.

White, P., Farrington, J., 1998. Bus and coach deregulation and privatization in Great Britain, with particular reference to Scotland. Journal of Transport Geography 6 (2), 135–141.

Wirasinghe, S. C. (1980). Nearly optimal parameters for a rail/feeder-bus system on a rectangular grid. Transportation Research Part A: General, 14(1), 33-40.

Wren, A., Rousseau, J.M., 1993. Bus driver scheduling – an overview. In: Daduna, J.R., Branco, I., Paixao, J.M.P. (Eds.), Computer-Aided Transit Scheduling.Springer, Berlin, Germany, pp. 173–187.

Wong, R.C.W., Leung, J.M.Y., 2004. Timetable synchronization for mass transit railway. In: Proceedings of the Ninth International Conference on Computer-Aided Scheduling of Public Transport (CASPT), San Diego, CA.

Xiong, J., Guan, W., Song, L., Huang, A., & Shao, C. (2013). Optimal routing design of a community shuttle for metro stations. Journal of Transportation Engineering, 139(12), 1211-1223.

Yeh, A. G. O., & Chow, M. H. (1996). An integrated GIS and location-allocation approach to public facilities planning—An example of open space planning. Computers, Environment and Urban Systems, 20(4-5), 339-350.

Farokhi Sadabadi, K., Hamedi, M., & Haghani, A. (2010). Evaluating moving average techniques in short-term travel time prediction using an AVI data set (No. 10-3144).

Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. IET Intelligent Transport Systems, 11(2), 68-75.

CURRICULUM VITAE

Yaojun Wang

Education

B.A., Shanghai University, May 2012

M.Sc., Shanghai University, May 2015

Ph.D., University of Wisconsin-Milwaukee, May 2018

Dissertation Title

Data-Driven Optimization Models for Feeder Bus Network Design

Work Experience

02/2012-05/2013 Mitsubishi Corp., P.R. China

Position: Quality Engineer

06/2017-Present A. O. Smith Corp. U.S.

Position: Senior Data Scientist

Awards

Second Place in CEAS Poster Competition	(2017)
Chancellor's Award in University of Wisconsin Milwaukee	(2017)
Chancellor's Award in University of Wisconsin Milwaukee	(2016)
Chancellor's Award in University of Wisconsin Milwaukee	(2015)
Outstanding TA Award	(2016)

Publication

Wang, Y., Yu, J.*, A Data-Mining Approach for Feeder Bus Network Design and Scheduling with Cellphone Data and Open GIS Tool. *Journal of Transport Geography.* [Under Review]

Wang, Y., Liu, Y.*, Li, X., "Data-driven design of urban rail feeder bus network with aggregated cellular data and open GIS tool," *Journal of Transportation Engineering*. [Under Review]

Li, X., Yu, J., Shaw, J., & **Wang, Y.** (2017). Route-Level Transit Operational-Efficiency Assessment with a Bootstrap Super-Data-Envelopment Analysis Model. J*ournal of Urban Planning and Development*, 143(3), 04017007.

Li, X., Liu, Y., **Wang, Y.**, & Gao, Z. (2016). Evaluating transit operator efficiency: An enhanced DEA model with constrained fuzzy-AHP cones. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(3), 215-225.

Li, X., Liu, Y., **Wang, Y.**, Liu, D., & Gao, Z. (2015). Multidimensional Assessment of Developing an Urban Public Transit Metropolis in China. *Journal of Urban Planning and Development,* 142(3), 04015021.