Center for Advanced Multimodal Mobility Solutions and Education

Project ID: 2017 Project 01

ESTIMATION OF ORIGIN-DESTINATION MATRIX AND IDENTIFICATION OF USER ACTIVITIES USING PUBLIC TRANSIT SMART CARD DATA

Final Report

by

Wei Fan (ORCID ID: https://orcid.org/0000-0001-9815-710X)
Zhen Chen (ORCID ID: https://orcid.org/0000-0001-6785-2652)

Wei Fan, Ph.D., P.E.
Director, USDOT CAMMSE University Transportation Center
Professor, Department of Civil and Environmental Engineering
The University of North Carolina at Charlotte
EPIC Building, Room 3261, 9201 University City Blvd, Charlotte, NC 28223
Phone: 1-704-687-1222; Email: wfan7@uncc.edu

for

Center for Advanced Multimodal Mobility Solutions and Education
(CAMMSE @ UNC Charlotte)
The University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC 28223

August 2018

ACKNOWLEDGEMENTS

This project was funded by the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE @ UNC Charlotte), one of the Tier I University Transportation Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT), under the FAST Act. The authors are also very grateful for all of the time and effort spent by DOT and industry professionals to provide project information that was critical for the successful completion of this study. The authors also would like to thank the www.openits.cn website for providing the data for this research.

DISCLAIMER

Table of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

The smart card (SC)-based automated fare collection (AFC) system has become the main method for collecting urban bus and rail transit (UBRT) fares in many cities worldwide. Such smart card technologies provide new opportunities for transportation data collection since the transaction data obtained through AFC system contains a significant amount of archived information which can be gathered and leveraged to help estimate public transit Origin-Destination (OD) matrices. Boarding location detection is an important step particularly when there is no automatic vehicle location (AVL) system or GPS information gathered in the database in some cases. With the analysis of raw data without AVL information in such research, an algorithm for trip direction detection needs to be built so that the direction for any bus in operation can be estimated. The transaction intervals between adjacent records will also need to be analyzed to detect the boarding clusters for all trips in sequence. Boarding stops can then be located with the help of route information and operation schedule. Alighting stop information can then be extracted with the analysis of relationships between records. Finally, the feasibility and practicality of such methodology will need to be tested.

The goal of this project is to develop a systematic transit passengers' origin-destination estimation methodology that utilizes data with only limited information to enhance the trip-chain-based OD estimation algorithms using the bus transit smart card data collected in Guangzhou City, China in this research. Specific objectives are to: 1) Detect the direction information of the records, 2) Detect the boarding cluster information of the records, 3) Extract boarding stop information, 4) Extract alighting location information, and 5) Generate OD matrix and analyze transit users' travel patterns. This report focuses on the algorithms that are developed and used to extract OD information step by step and to present data analysis results based on the extracted information to help the transit planners make informed decisions.

Chapter 1. Introduction

1.1 Problem Statement

The smart card (SC)-based automated fare collection (AFC) system has become the main method for collecting urban bus and rail transit (UBRT) fares in many cities worldwide. Such smart card technologies provide new opportunities for transportation data collection since the transaction data obtained through AFC system contains a significant amount of archived information which can be gathered and leveraged to help estimate public transit Origin-Destination (OD) matrices. Boarding location detection is an important step particularly when there is no automatic vehicle location (AVL) system or GPS information gathered in the database in some cases. With the analysis of raw data without AVL information in such research, an algorithm for trip direction detection needs to be built so that the direction for any bus in operation can be identified. The transaction intervals between adjacent records will also need to be analyzed to detect the boarding clusters for all trips in sequence. Boarding stops can then be located with the help of route information and operation schedule. Alighting stop information can then be extracted with the analysis of relationships between records. Finally, the feasibility and practicality of such methodology will need to be tested.

1.2 Motivation of Study

The purpose of this project is to develop a systematic approach to illustrating how passenger journey information gathered from the smart card data can be extracted and used for integrated network planning. Practical solution algorithms shall be developed for the robust origin-destination matrix estimation for public transit using smart card fare data. Case studies shall be conducted to present new information about transit temporal usage pattern, spatial usage pattern, and travel habits of different user groups that could be useful to transportation planners. The newly generated robust origin-destination matrix for the bus transit system can greatly help the decision makers plan, design, operate, and manage a more efficient public transit system.

1.3 Objectives of Study

The goal of this project is to develop a systematic transit passengers' origin-destination estimation methodology that utilizes data with only limited information to enhance the trip-chain-based OD estimation algorithms using the bus transit smart card data collected in Guangzhou City, China in this research. Specific objectives are to: 1) Detect the direction information of the records, 2) Detect the boarding cluster information of the records, 3) Extract boarding stop information, 4) Extract alighting stop information, and 5) Generate OD matrix and analyze transit users' travel patterns. This report focuses on the algorithms that are developed and used to extract OD information step by step and to present data analysis results based on the extracted information to help the transit planners make informed decisions.

1.4 Report Overview

The remainder of this report is organized as follows: Chapter 2 provides the general information about smart card technology, including the discussions about the history of smart card and relevant privacy concerns. This chapter also reviews the previous research on inferring bus transit OD information with smart card data (SCD). Since the SCD structures may be unique in different cases, the major topics include SCD based boarding location detection, SCD-based OD inference and SCD-based travel pattern analysis. Chapter 3 presents the raw data preparation steps. It starts with a primary analysis of the bus transit smart card data collected along four routes in Guangzhou, China. The route information, the structure and potential issues of the smart card data are also discussed. Chapter 4 presents the methodology developed for the transit users' boarding information identification. The bus travel time information identification method is discussed first. This is followed by the boarding direction detection and clustering of the boarding activities. The boarding stop identification algorithm and the results are also discussed. Chapter 5 discusses the alighting information identification. Different scenarios of the records are discussed. The transaction record matching algorithm is also described, and the alighting information identification methodology is then presented. Chapter 6 describes the further applications of this methodology and the inferred OD matrices. The data analysis results based on the extracted information are also demonstrated. Those results include the analysis of transit temporal usage pattern, spatial usage pattern, and travel habits of different user groups. Chapter 7 concludes the report with a summary of the findings and some discussions about possible improvements to enhance current practices. Future research directions are also given.

Chapter 2. Literature Review

2.1 Introduction

This chapter provides a comprehensive review and synthesis of the state-of-the-art and state-of-the-practice related to the SCD study, including SC technology, SC data structure, existing SCD based OD inference methodology and other applications of SCD. This should give a clear picture of SCD based OD estimation methodology, and SCD based trip characteristics as well as relevant data mining.

The following sections are organized as follows. Section 2.2 describes the history, the privacy concerns about SC and current SC utilization status within the major cities in U.S. Section 2.3 provides a comprehensive review of existing methodologies which focused on the detection of the boarding location with the help of smart card data. This section also presents a comprehensive review of existing methodologies which focused on the detection of the alighting location and estimation of OD matrix with the help of smart card data and other support databases such as AVL data. A review of relevant studies which examined transit users' travel pattern with the combination of the smart card data and other data sources is also performed. Finally, section 2.4 concludes this chapter with a summary.

2.2 Smart Card Technology Review

2.2.1 Smart Card Technology

As the smart card fare system has been widely implemented in the world today, SCD also plays an important role in the regional transportation system management. SCD can act as the data source to replace traditional travel survey in many respects.

Lessard (1993) conducted a research about the use of computer technology for urban transit operations. The smart card system was referred to as "a computerized fare collection and monitoring system." The new technology could act as a new component to improve transit management operations and to provide the best possible level of service.

Cunningham (1993) introduced the multi-modal access and payment system (MAPS) which could provide a common mechanism for the payment of transit fares, parking fees, and other transportation-related items. As transit users usually must deal with different schedules, fare options, and payment instruments when their trips involve more than one operator or mode shift, they need to have exact fare or pre-purchased tickets ready for each trip legs. The MAPS can provide a self-supporting, smart card-based fare collection system with transaction recording capabilities. In addition, the information extracted from the cards can also be used for future transit system management.

Stromberg (1995) presented an "electronic transit fare card system" in his patent to deal with the inconvenience of traditional fare collection method. For the new smart card technology, he defined his design as "a card system and method for accessing a public transit system that employs a card pre-encoded with trip data permitting a set number of trips on the transit system and a card processing unit to receive and process a card each time access to the transit system is desired”.

FHWA (1998) defined the smart card system as “electronic information systems that use plastic cards (similar to credit or debit cards) to store and process information. Used in fare-payment and parking applications” in a study on the intelligent transportation system and devices. Several other technologies could also be related to the smart card system (e.g., Automated Vehicle Location)

2.2.2 Smart Card Standards and Implementations

Standards are important to the smart card industry. Currently, the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) is “one of the worldwide standard-setting bodies for technology, including plastic cards.” The primary standards for smart cards are ISO/IEC 7816, ISO/IEC 14443, ISO/IEC 15693 and ISO/IEC 7501. Table 2.1 illustrates the main smart cards standards.

Table 2.1: Smart Card Standards

Technology	Level	Application
ISO/IEC 7816	International	Define the various aspects of the card and its interfaces, including the card’s physical dimensions, the electrical interface and the communications protocols
ISO/IEC 14443	International	Defines the interfaces to a “close proximity” contactless smart card
ISO/IEC 15693	International	Describes standards for “vicinity” cards
ISO/IEC 7501	International	Describes standards for machine-readable travel documents and has made a clear recommendation on smart card topology
ISO/IEC 18092	International	Defines communication modes for Near Field Communication Interface and Protocol (NFCIP-1)
FIPS 201	Federal	Provides the specifications for a standard Federal smart ID card
FIPS 197	Federal	Specifies a FIPS-approved cryptographic algorithm that can be used to protect electronic data
FIPS 140	Federal	Security Requirements for Cryptographic Modules

Source: http://www.smartcardalliance.org/smart-cards-intro-standards

Today, nearly all transit fare payment systems use contactless smart cards as the primary ticket medium. Major deployments are already operational in major cities around the world.

Table 2.2: U.S. Transit Smart Card Projects and Implementations

Location	Smart Card Projects
Atlanta	MARTA Breeze Card
Baltimore	MTA CharmCard
Boston	MBTA Charlie Card
Chicago	CTA Ventra
Houston	METRO Q Card
Los Angeles	LACMTA TAP Card
Miami	MDT EASY Card
Minneapolis-St. Paul	Metro Transit Go-To Card
New Jersey	NJ TRANSIT Tap>Ride
Newark	PANYNJ (SmartLink)
Philadelphia	PATCO FREEDOM Card
Salt Lake City	UTA EFC Card
San Diego	MTDB Compass Card
San Francisco	MTC (Clipper Card)
Seattle-Puget Sound	KC Metro ORCA Card
Washington, DC	Washington Metropolitan Transportation Authority SmarTrip

Source: http://www.smartcardalliance.org/smart-cards-applications-transportation/#smart-cards-and-transit

2.2.3 Smart Card Concerns

Several privacy concerns have been raised since the SCD has been used to track an individual's movements, conversations, etc. via smart card data.

Attoh-Okine (1995) summarized security concerns about the smart card technology. The violations that should be prevented were listed as follows: "Cloning: Every card should have unique identifiers and be assigned to a particular application; Card forgery: The cards should be personalized with a specific PIN (Personal Identification Number); Card tampering: Read and write access to the card should be designed to protect against accidental or intentional accidents; Repudiation: designed to stipulate specific transactions."

O'Connor (1994) discussed the privacy issues with a smart card in his research. There were several main privacy implications listed in his study including:

- *"The collection limitation principle"*
- *"The data quality principle"*
- *"The purpose specification principle"*
- *"The use limitation principle"*
- *"The openness principle"*
- *"The individual participation principle"*
- *"The accountability principle"*

Heydt-Benjamin (2006) discussed some interesting features of the transit ticketing as a problem domain and provided an architecture which is sufficient to meet the needs of a typical metropolitan transit system.

Dempsey (2015) pointed out that aggregating and correlating SCD can allow others to learn quite a lot about the behavioral patterns of individual SCD users. Dempsey concluded that “The greater the data base, and the more extensively SCD database correlated with other data bases, the less privacy the SCD user enjoys.”

Cottrill (2009) summarized the smart card privacy related act in her study. The author also introduced several technological approaches to privacy preservation, including “Separation of communication and authentication; Utilization of trusted third parties and Adaptive privacy preservation.”

Mayes (2008) described various attacks and countermeasures that could be applied to secure smart card applications in the book entitled “Smart Cards, Tokens, Security, and Applications.” The attacks can be divided into different categories including “invasive attacks, semi-invasive attacks, and non-invasive attacks.” The author also described how to conduct a careful evaluation of the implementation of commands on a smart card.

2.3 Smart Card Data Study

2.3.1 SCD Study Literature Review

2.3.1.1 Pelletier et al.’s research work

Pelletier et al. (2011) presented a literature review of the studies about SCD, which include the technologies of smart card auto fare system operation, privacy legal issues related to the smart card data usage, the management of smart card data system and several research avenues. This review firstly focused on the history of smart card implementation, and secondly introduced typical transit smart card system in Gatineau, Quebec. Then, the authors analyzed characteristics of the smart card system and the advantages/disadvantages by comparing it with other fare collection systems. After a review of the general information about the smart card system, the authors grouped the smart card data studies into three categories including strategic-level studies, tactical-level studies, and operational-level studies. Lastly, the authors introduced two different commercialization approaches: Smart card marketing experiments for both fare and non-fare policies in a public transit context.

2.3.1.2 Long et al.’s research work

Long et al. (2015) presented a comprehensive review of the studies about SCD and divided relevant literature into four categories: 1) Data processing and origin-destination inference: Since OD matrix inference is an important part of urban transportation planning and management, the majority of studies focused on how to utilize smart card data to conduct OD inference. In this part, the authors discussed several important assumptions of OD relationship and methodology and improvement to these assumptions as well as supplement of missing data and data mining methods. 2) Transit system operation and management: Researches on the SCD and transit system mainly focused on the relationship between SCD and long term planning in urban areas. Such studies discussed how SCD can be combined with or take the place of traditional traffic survey, and how SCD can be mined to forecast travel demand and achieve system optimization. 3) Spatial structure of cities: With the support of commuters' demographics information and mobile app data, SCD can help to recognize the spatial changes of city land use. 4) Mobility behavior and social networks: There are also studies about mining the characteristics of smart card users. Such research topics included estimating the travel probability of different groups of people, and investigating different behavior of transit users during special period, etc.

2.3.2 SCD Pre-Processing Methodology Review

As the SCD related transit system research efforts are highly dependent on whether the smartcard system correctly aggregates individual rides into trips, the pre-processing of SCD is helpful to improve the data quality and accuracy.

2.3.2.1 Robinson et al.'s research work

Robinson et al. (2014) conducted a research study to identify the causes as to why SCD collection systems may not correctly record the information. Various reasons were listed to illustrate why smartcard systems may generate erroneous data, which included: software problem, data problem, hardware problem and user problem. To ensure the quality of smart card data, the authors developed a methodology to filter out smartcard records which could distort the findings of analyses such as OD estimation and travel pattern classification.

Figure 2.1 presents Robinson et al.'s approach to pro-processing SCD. The SCD pre-processing procedures could be divided as follows: "Basic checks on the raw ride data; Identify erroneous tap-in and tap-out data; Aggregate rides into trips; Identify faulty data."

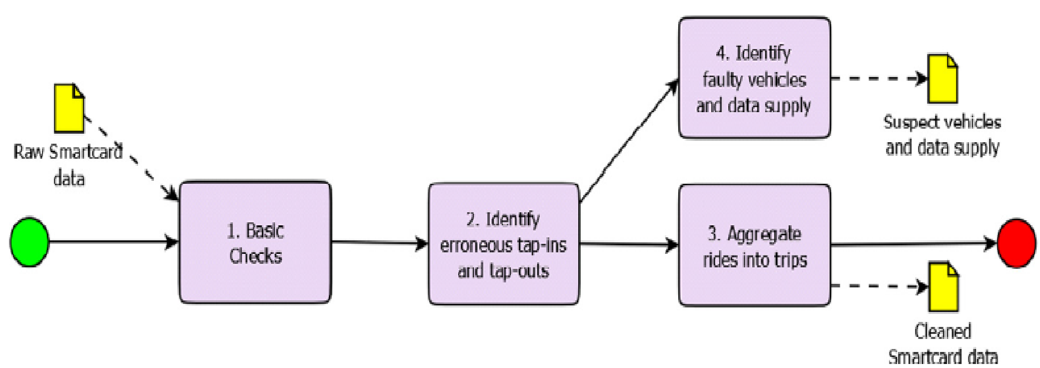

Figure 2.1: Robinson et al.'s Approach to Pre-processing SCD
Source: Robinson et al. (2014)

2.3.2.2 Qian et al.'s research work

Qian et al. (2015) used the transit system in Chengdu, Sichuan, China as an example and presented some methods of data cleaning to resolve the data quality issues in IC card and GPS data. Such data issues in the authors' case study included IC card replicated data, GPS replicated data and abnormal IC card data. With the help of data cleaning process, such records were deleted to guarantee the accuracy of further study.

Figure 2.2 presents Qian et al.'s approach to pre-processing SCD.

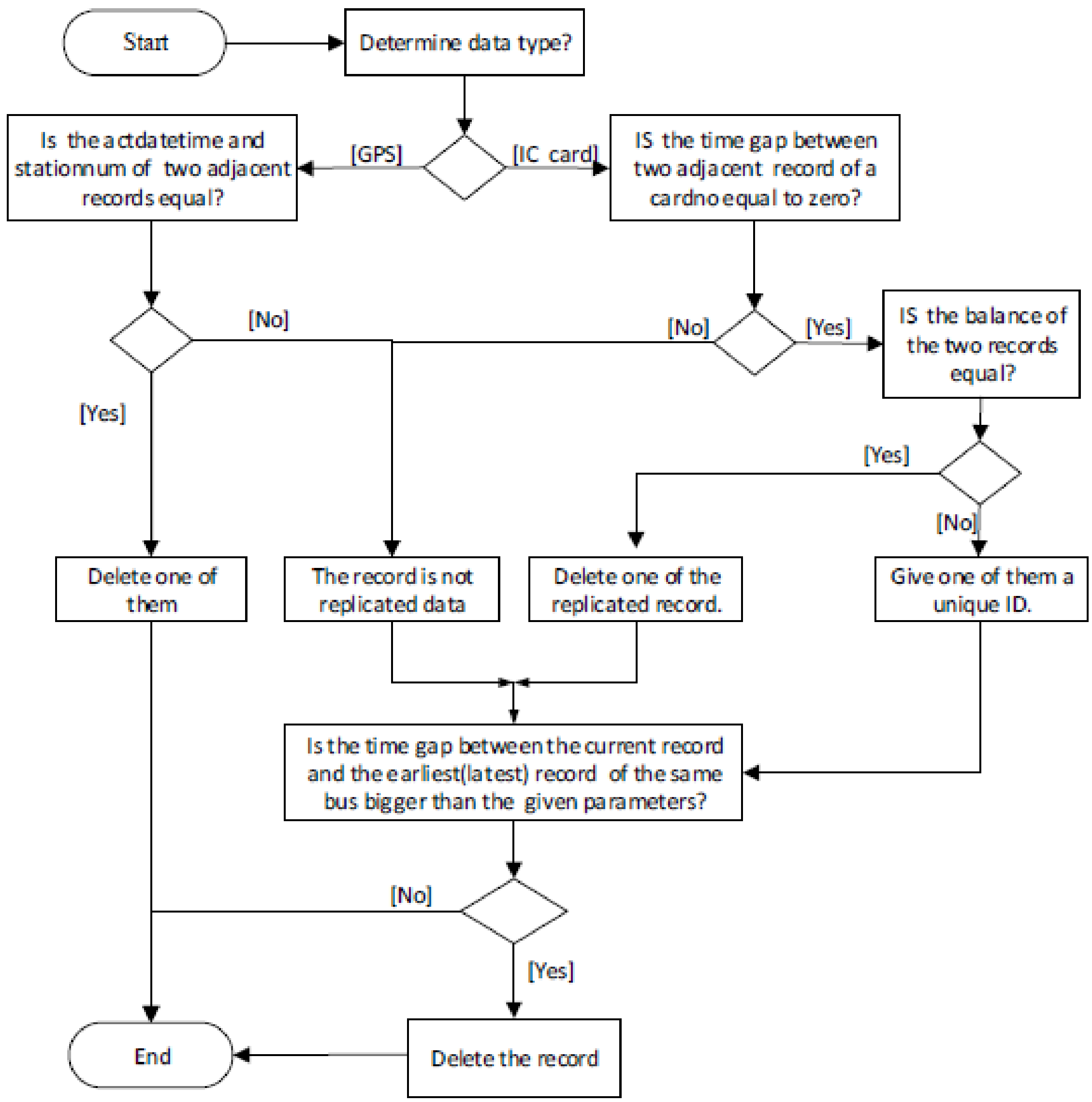

Figure 2.2: Qian et al.'s Approach to Pre-Processing SCD
Source: Qin et al. (2015)

2.3.3 SCD-Based Boarding Information Extraction Review

Identification of boarding location is usually the first step to the SCD based OD inference. However, one may need to deal with different cases since the data structure and support database are different under different cases. In particular, it is still very challenging to identify boarding stop information without other support data sources.

2.3.3.1 Barry et al.'s research work

Barry et al. (2008) developed an approach to identifying the specific boarding stops utilizing scheduled run time to estimate the location of a bus along its route at the time of the AFC transaction to solve the origin inference problem without the help of an AVL system. The challenge faced during the location estimation process in their study was that the transaction times of MetroCard system were truncated to 6-min intervals. To detect the approximate bus boarding locations, the scheduled run times between stops along the route were used to adjust the results with the transfer information obtained from SCD.

2.3.3.2 Wang et al.'s research work

Wang et al. (2011) inferred the trip origin with the help of both transaction record timestamp and AVL data. The AVL system can record the time when the bus doors open or close at each bus stop for each bus run. The boarding location of each record can be then inferred and finally some results were achieved with high accuracy. This research approach developed a very important boarding location identification methodology. Other support data sources also include on-bus survey data, general transit feed specification (GTFS) data, automatic passenger counting (APC) data, etc.

2.3.3.3 Ma et al.'s research work

Ma et al. (2012) developed a Markov chain-based Bayesian decision tree algorithm to extract passengers' origin data from the Beijing flat-bus AFC database. This study introduced how to cluster transaction data with the following steps: "Sorting transaction timestamp in ascending order; including the record as another cluster if the time difference between itself and the previous record is larger than 60s; using 30 mins as the time threshold for a different trip". The transit stops were recognized via a Bayesian decision tree algorithm in which transmission matrix can be constructed as:

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{(n-1)1} & p_{(n-1)2} & \cdots & p_{(n-1)n} \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} = \begin{pmatrix} 1-\sum_{i=2}^{n} p_{1i} & p_{12} & \cdots & p_{1n} \\ 0 & 1-\sum_{i=2}^{n} p_{2i} & \cdots & p_{2n} \\ 0 & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & p_{(n-1)n} \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

where:

p_{ij} = the possibility of the next boarding stop being stop *j* conditioned on the previous boarding stop being *i*.

$$p_{ij} = \Pr(S_{k+1} = j \mid S_k = i) = \int_{z_{ij}-\Delta}^{z_{ij}+\Delta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \approx \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{{z_{ij}}^2}{2}\right) \cdot 2\Delta$$

where:

z_{ij} = the standardized travel speed between stops j and i.

However, the irregularity of the sample database and the time threshold configuration was a big issue with this method.

2.3.3.4 Song's research work

Song (2016) developed an algorithm to obtain the bus transit boarding information using raw smart card data based on the station labeling method. The research used 0.54 billion transaction records in China as a sample, which covered the information about 237 routes in five months. The author utilized a fixed time threshold with an adjustment element to distinguish the boarding records at different stops using the time difference between adjacent records. This research work represented a basic idea of boarding stops location detection.

2.3.3.5 Yu et al.'s research work

With only the transaction time, Yu et al. (2009) designed a hierarchical clustering method to classify the data source based on the time interval between two adjacent transaction records. The reference bus operation time was based on the fixed schedule and the criterion used was based on the difference between timestamp of the cluster and the arrival time of each stop on the schedule. However, it was very difficult to resolve all issues since the clusters could not be built without direction information.

Based on the studies mentioned above, different methodologies and techniques have been developed and used to study the SCD-based boarding information extraction. However, with the consideration of data issues such as data quality and sample size, specific boarding information identification methodology should be developed for a specific study site. Table 2.3 provides a summary of SCD-based boarding information extraction methods.

Table 2.3: Summary of SCD-Based Boarding Information Extraction Methods

Authors	Year	Data Type	Data Source	Case Study Location	Methodology	Achievements
Barry et al.	2009	Card ID; Bus ID; Boarding transaction time	SCD, Schedule data	NYC, US	Matching records with schedule Trip chain method	Extracted passengers' boarding stop information
Yu et al.	2009	Card ID; Route ID; Boarding transaction time; Bus ID; Bus operation data;	SCD	NA	Hierarchical clustering method	Extracted boarding stop information
Wang et al.	2011	Card ID; Route ID; Boarding transaction time; Bus ID; GPS data; Bus operation data;	SCD, AVL data	London, UK	Matching records with schedule and GPS data.	Extracted boarding stop information
Ma et al.	2012	Card ID; Boarding transaction time; Bus ID; Route ID	SCD	Beijing, China	Transaction data clustering; Markov chain-based Bayesian decision tree algorithm	Extracted transit passengers' origin
Song	2016	Card ID; Route ID; Boarding transaction time; Bus ID	SCD, Schedule data	Chengdu, China	Time labeling method; Greedy algorithm	Extracted boarding stop information

2.3.4 OD Inference without AVL and APC Data Review

Specifically, the transit rider OD matrix can potentially be extracted from the SC transaction database. Passengers' OD matrix estimation is a major part of transportation planning study. With the consideration of inconvenience involved in traditional travel surveys, SCD-based research has been becoming more and more popular recently. Historical SC transaction time data are one of the major data sources for transit users' OD information inference and have been widely used by both researchers and practitioners.

2.3.4.1 Barry et al.'s research work

Barry et al. (2002) presented one of the first research efforts to estimate OD matrix based on the NYC MetroCard data. This study also validated two important assumptions: "Most riders would like to return to the destination station of their last trip to begin next trip; the destination of high percentage riders at the end of the day is the origin stop where they begin their first trip of the day." In order to test these assumptions, the first step was to collect data from a travel diary survey conducted by New York metropolitan transportation council (NYMTC) and then to divide the data into two different groups: two trips (per day) group and more than two trips (per day) group. After implementing the methodology, results of comparing estimated OD patterns with counts data showed that the forecasting result was acceptable.

2.3.4.2 Trepanier et al.'s research work

Trepanier et al. (2007) conducted a SCD-based destination estimation research with the help of 2003 Gatineau, Quebec smart automatic fare collection (SCAFC) data. To resolve the issues of limited transaction information about the 'single trip', in which only one transaction record of a day existed, the authors tried to find a similar boarding which occurred in previous days. For example, if there was only one transaction record at day $n + 1$, and if at day $n - 1$, there was a boarding made almost at the same time of the day on the same route, then the estimated alighting stops for day $n - 1$ was used for the day $n + 1$. Figure 2.3 presents a sample scenario of similar boarding activities including the transaction records in three days.

The numerical result showed that 66% of all possible trip destinations were estimated. However, some of them were not found because the data were not suitable for developing the model or because the trip was 'single' and no destination information was available in other days' records. Although the 34% non-success rate appears high, this should not affect the results for the regular routes at peak hours. Incidentally, most of the rejected trips were non-regular and took place during off-peak hours.

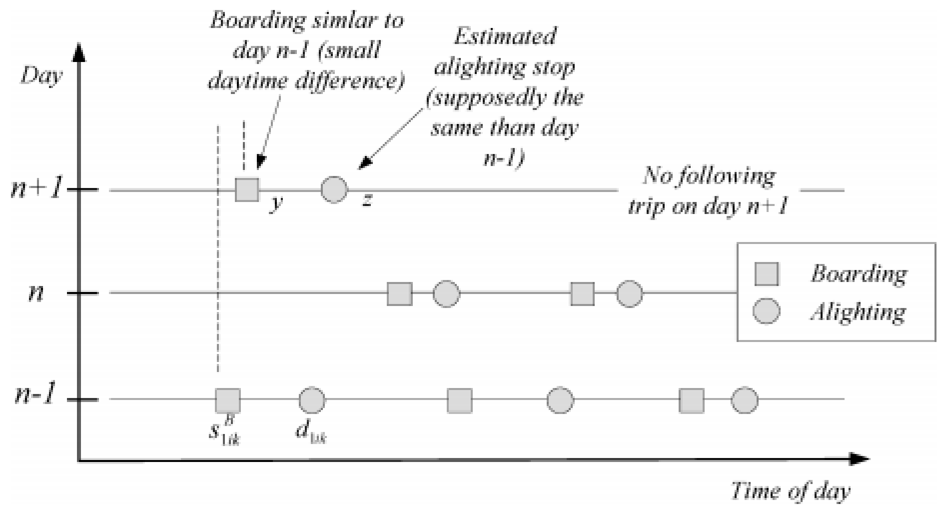

Figure 2.3: Sample of Similar Boarding Activities Scenario
Source: Trepenier et al. (2007)

2.3.4.3 Munizaga and Palma's research work

For the bus alighting location estimation, Munizaga and Palma (2011) utilized the trip chaining methodology to build up the OD matrix. Several assumptions were made in this research, including: (1) "if two transactions in a row are made in the Metro or along the same bus route, a destination is assumed between the two, regardless of the time interval, because it is very unlikely that someone would go out of the Metro network during a trip, unless he/she has something to do at that location." (2) "If the time elapsed to the next transaction is over two hours, then this next transaction is the first segment of a new trip."

2.3.4.4 Alsger et al.'s research work

Based on Barry et al.'s (2002) two assumptions, Alsger et al. (2015) presented and tested the trip chaining methodology with the consideration of Inter-transaction Time (ITT). This method assumed that the passenger starts his trip at B1 to first alighting stop and walks to next boarding stop B2 with no alighting information recorded. If there is a stop A1 located in the buffer zone of the second boarding stop B2, this stop can be used as the alighting stop of the last trip. For the buffer zone range calibration, different allowable transfer times have been assumed, ranging from 30 minutes to 90 minutes (Bagchi and White, 2004; Hofmann and O'Mahony, 2005; Nassir et al., 2011; Ma et al., 2013). After the estimation of different ITT, a sensitivity analysis was also conducted for different parameters and the result showed that 60-minute was the best time threshold in this study.

2.3.4.5 Nassir et al.'s research work

Nassir et al. (2011) developed two models to estimate alighting stop information without the help of AVL data. The first method was utilizing the trip chaining method and typical assumptions to infer alighting information. The second method was verifying the outputs by joining inferred data with APC system data. The authors pointed out that the reason of mismatching may be that other modes have been used during the transfer process. In order to resolve the issue of wrong inference in boarding location detection process, an alternative sub-model, which relaxed the search again in the other directions among the transaction locations before the record was excluded, was developed. This sub-model would increase the sizes of the output and enhance the accuracy.

2.3.4.6 Zhang et al.'s research work

Zhang et al. conducted a research study to infer boarding stop information without AVL data. To avoid the errors from being produced while using bus operation schedule only, the data collected by the on-bus survey in Changchun, Jilin, China was utilized together with SCD to infer the boarding and alighting information. The trip generation and attraction information of each traffic zone were then analyzed to help create OD matrix.

2.3.5 OD Inference with APC and AVL Data Review

Over the past several decades, many automatic systems have been introduced in the public transit area, including the smart card related automatic fare collection (AFC), automatic vehicle location (AVL), and automatic passenger counting (APC). APC and AVL data have also been becoming more and more important recently.

2.3.5.1 Cui's research work

Cui (2006) developed an algorithm to estimate a Bus Passenger Trip OD Matrix based on the AFC data, APC data, and AVL data. For the single route OD estimation, the Iterative Proportional Fitting (IPF) and Maximum Likelihood Estimation (MLE) models were developed based on the constructed seed matrix and marginal values. At the network level, the IPF methodology was utilized with the help of transfer flow estimation. Data from the Chicago Transit Authority (CTA) were used for testing while achieving the OD matrix estimation results.

2.3.5.2 Zhao's research work

Zhao (2006) proposed a method to combine the data from AFC and AVL systems to infer boarding and alighting information. These two systems were integrated by matching the vehicle location information from the AVL against the passenger trip information from the AFC to infer the individual passenger's boarding stops. A software tool was developed to facilitate the method implementation and the results of the application were reported by CTA. Figure 2.4 and 2.5 show the basic concepts of combining and processing AFC with AVL data, respectively.

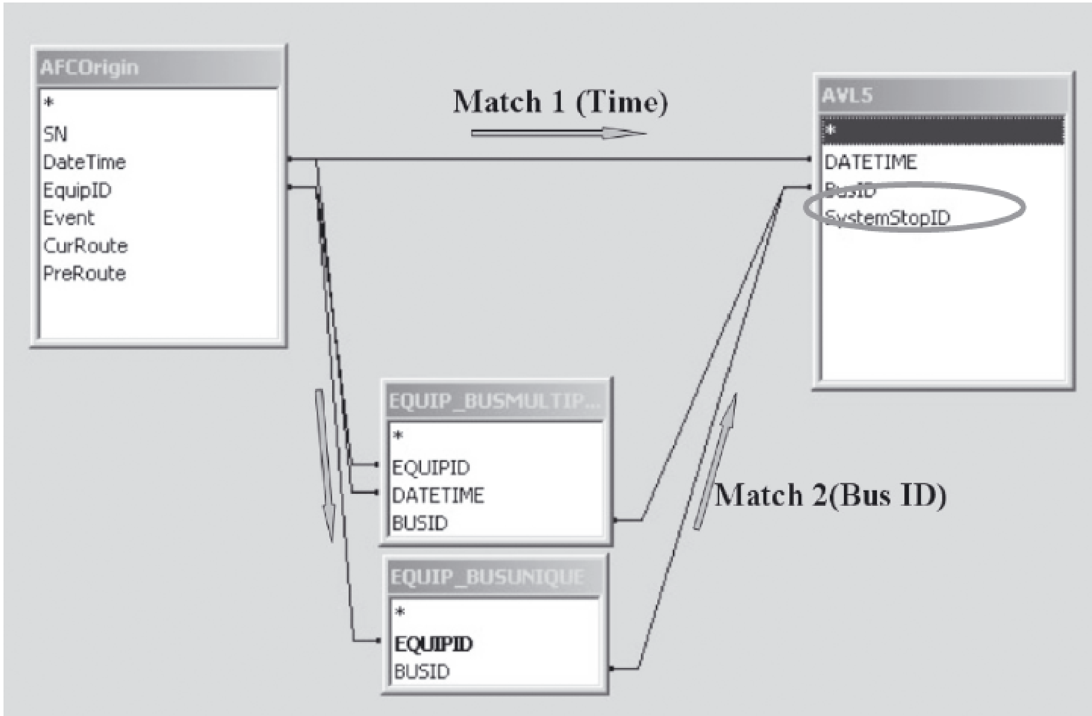

Figure 2.4: Combining AFC with AVL Database
Source: Zhao (2006)

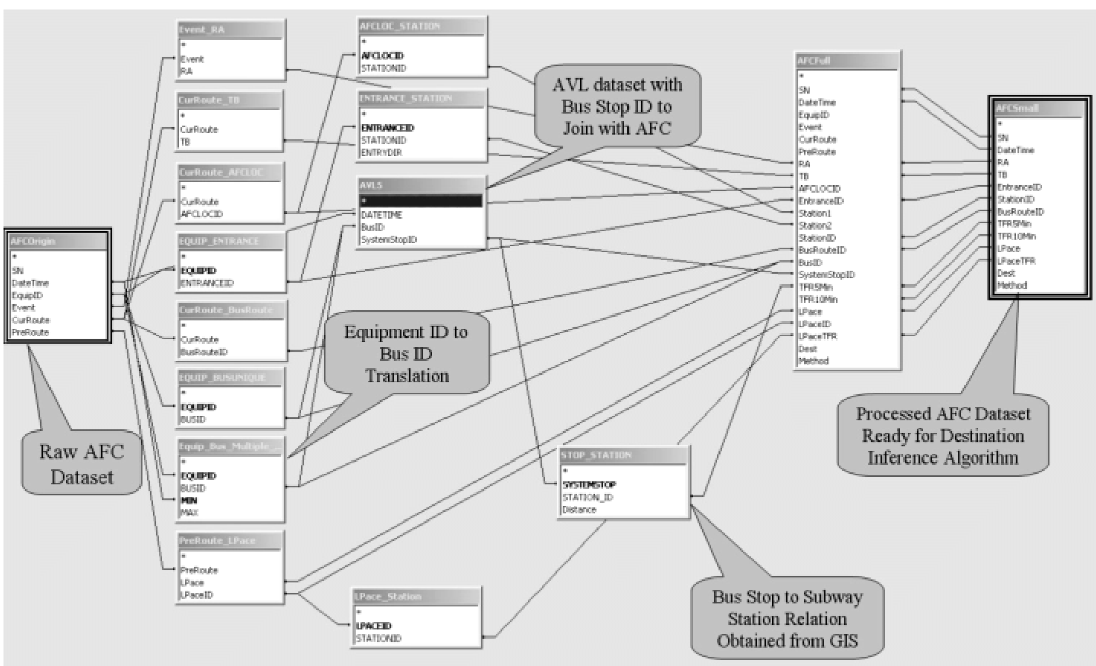

Figure 2.5: AFC and AVL Data Processing
Source: Zhao (2006)

2.3.5.3 Gordon et al.'s research work

Gordon et al. (2013) developed a method to infer all riders' OD information in a large public transit network. The methodology could be divided into four steps, including bus passenger origin inference; bus passenger alighting times and locations inference; bus passengers' interchange (transfer) information inference; and linked-journey flow metrics construction. During the bus passenger origin inference steps, passenger boarding locations were typically inferred through matching each AFC record with an AVL record.

Table 2.4 provides a summary of the SCD Based OD Estimation Methods.

2.3.6 SCD-based Travel Pattern Analysis Review

2.3.6.1 Jang's research work

Jang (2010) examined the possibilities of using data from closed systems for transportation planning applications in the city of Seoul. One specific feature of the Seoul AFC system is that different from the AFC system in many other cities, the AFC system in Seoul recorded each trip's entry and exit information, including passengers' boarding and alighting times and locations, as well as their transfer information. Based on the AFC data, Jang analyzed transfer trip patterns and identified interchange points that need improvements by examining the points with specifications such as: 1) interchange demand exceeding 5,000 per day; 2) the average interchange time exceeding 10 minutes; 3) the transfer rates larger than 50%; and 4) combining all these conditions above.

2.3.6.2 Devillaine et al.'s research work

Devillaine et al. (2012) developed a methodology to categorize trips according to different purposes after the detection and estimation of the location, time, duration and purpose of activities undertaken by public transit users by utilizing smart card databases. Data of Gatineau, Canada and Santiago, Chile were utilized to detect transit users' activities. With the activity identification module, the trips were labeled as transfer or non-transfer related. Four trip purposes were assigned to each activity based on the available dataset and estimated activity attributes along with differentiated criteria. With these constrained conditions set before, the authors gave the results of estimated trips and average trip times based on the number of stages in both cities. The study also revealed the different behavior activity patterns with the consideration of sociological, geo-political and culture differences.

2.3.6.3 Seaborn et al.'s research work

Seaborn et al. (2009) conducted basic data analysis based on the SCD collected from both bus and underground transit in London. Different groups of transfer activities were analyzed (including underground-to-bus and bus-to-underground and bus-to-bus). The transfer times were also categorized into three groups: low range, middle range, and high range. The analysis results demonstrated that passengers' transfer activities could be quantified with SCD.

Table 2.4: Summary of SCD-Based OD Estimation Methods

Authors	Year	Data Type	Data Source	Case Study Location	Methodology	Achievements
Barry et al.	2002	Card ID; Bus ID; Boarding transaction time; Exist counts.	SCD, Survey data	NYC, US	Trip chain built up by sequence.	Estimated OD matrix
Trépanier et al.	2004	Card ID; Boarding transaction time; Boarding stops; Card type	SCD, Survey data	Santiago, Chile	Trip chaining method; Comparison with household survey data	Estimated OD matrix
Cui	2006	Card ID; Route ID; Boarding transaction time; Last route ID; Boarding and alighting counts at stops	SCD, AVL	Chicago, US	Iterative Proportional Fitting (IPF) and Maximum Likelihood Estimation (MLE)	Estimated OD matrix at the segment level (3-5 combined stops)
Zhang et al.	2007	Card ID; Boarding transaction time; Driver ID; Route ID; On board survey data;	SCD, Survey data	Changchun, China	Comparison of records with on bus survey data	Estimated OD matrix
Wang	2010	Card ID; Route ID; Boarding transaction time; Boarding stops; Alighting transaction time; Alighting stops; Scheduled departure time; Actual departure time; GIS data	SCD, AVL	London, UK	Matching records with AVL data; Trip chaining method compared the results with survey data.	Estimated OD matrix
Nassir et al.	2011	Card ID; Card type; Boarding transaction time; Boarding stops; User type; Bus ID	SCD, AVL	Minneapolis, US	Matching records with AVL data; Sensitivity analyses	Estimated alighting stop information
Munizaga and Palma	2012	Card ID; Boarding transaction time; Bus GPS information	SCD, GPS	Santiago, Chile	Trip chaining method	Estimated OD matrix
Gordon et al.	2013	Card ID; Route ID; Boarding transaction time; Bus location	SCD, AVL	London, UK	Matching records with AVL data;	Estimated OD matrix and interchange information
Alsger et al.	2014	Card ID; Route ID; Boarding transaction time; Boarding stops; Alighting transaction time; Alighting stops	SCD	South East Queensland, Australia	Trip chaining method; ITT (Inter Transaction Time) analysis	Identified best ITT time threshold

2.3.6.4 Ma et al.'s research work

Ma et al. (2013) developed an effective data-mining methodology utilizing the Beijing transit data. In their research, boarding stop information was estimated using a Markov chain-based Bayesian decision tree algorithm. This algorithm can be used to extract changes in boarding volume with time between two transaction records, calculate the probabilities for all potential stops using speed profile from GPS data, and assume the stops with maximum probability as the boarding stops. The trip chains were identified based on a fixed temporal threshold. Density-based spatial clustering of applications with noise (DBSCAN) algorithm was also developed to analyze and detect transit riders' historical travel patterns. In order to detect individual travel pattern regularities, K-means clustering algorithm and rough-set theory were utilized. The comparison between different algorithms showed that rough set theory achieved more accurate and efficient results. This approach would be helpful in conducting travel behavior research, transit market analysis and transit OD estimation.

2.3.6.5 Ma and Wang's research work

Ma and Wang (2014) also conducted a research study to build a platform to combine GPS data and SCD together. The origin inference was also based on Bayesian inference theory to estimate the likelihood at each stop. The destination inference included spatiotemporal transfer activity identification, daily trip chain analysis and historical travel pattern integration. A light-weight transportation GIS data model was developed utilizing GTFS from google. This platform not only served as a data visualization platform to monitor transit network performance for planning and operations, it also intended to take advantage of e-science initiative for data-driven transportation research and applications. The data revealing functions included: "Transit Network-Level Speed Map; Transit Stop-Level Ridership Analysis; Transit Stop-Level Headway Distribution; and Transit Route-Level Travel Time Reliability."

2.3.6.6 Agard et al.'s research work

Agard et al. (2007) developed a K-means-clusters-based methodology to analysis public transport user behavior. The data contained card number, card type, and transaction period which were categorized into 20 groups. Four passenger pattern groups were generated after the application of the methodology. The results showed the proportion of each card type category in different travel pattern clusters. This study demonstrated a feasible data mining technique and could help transit agency to understand passenger behaviors better.

2.3.6.7 Kusakabe and Asakura's research work

Kusakabe and Askura (2014) developed a data fusion methodology to estimate absent behavioral attributes in smart card data by using survey-based data. Although the information on the boarding and alighting stops and time were utilizable in these two datasets, the precision of the survey information was not as accurate as SCD and SCD were not always identical to the individuals. However, if the trips of both SCD database and personal trip survey database had the same conditional probability distribution rate, the different attributes of two datasets can be estimated with probability model under high penetration rate. The probability function was:

"$p(c, g) = \sum_{F \in S} p(c|F) P_s(F, g)$"

In this equation built by the authors, the datasets can be categorized as:

c: The attributes that were only observed in the person trip survey data, such as trip purpose, origin, and destination.

g: Behavioral attribute only from smart card data, such as trip frequency.

F: Commonly behavioral attributes that were included in both datasets, such as boarding stations and times.

Naive Bayes classifier was also utilized by applying the probability distribution of behavior attributes to naive Bayes classifier. The attributes of trip purposes ('commuting to work'; 'commuting to school'; 'leisure'; 'business'; and 'returning home') were added to each trip in the smart card data.

2.3.6.8 Long et al.'s research work

Long et al. (2014) used SCD to identify the spatiotemporal travel patterns of transit users. The users were categorized into four groups, including: "transit users who travel significantly earlier than average riders; users who ride in unusual late hours; users who commute in excessively long distance; and users who make significantly more trips per day."

In addition, household survey data were used as the support database to identify the users' socioeconomic background. The research findings were helpful for the transit management agency to better understand transit users' travel patterns.

2.3.6.9 Jánošíková et al.'s research work

Jánošíková et al. (2014) described how to use the SCD in a logit model of route. Four attributes were taken into consideration including in-vehicle travel time, walking time to transfer, the number of transfers, and headway of the line. A case study was conducted in the city of Žilina in Slovak Republic, and 115,007 transaction records were collected. The relevant importance of these attributes was estimated for both morning peak hour and off-peak hour. The original estimation results showed that the in-vehicle travel time played the most significant role. In order to correct correlation in the unobserved utility across repeated choices, the bootstrap method was utilized for resampling in the research. The aggregate elasticities for these 4 attributes were calculated and the relative ratios table was built. The research findings would be helpful in analyzing and enhancing the service quality.

2.3.6.10 Kieu et al.'s research work

Kieu et al. (2013) conducted a study on mining spatial temporal characteristics of transit users. Queensland transit SCD were used in this study. The trip chaining method was built using 60 min as the time threshold. A DBSCAN approach was utilized to retrieve travel regularity of each frequent transit user. The first step of the methodology was analyzing the whole travel itineraries of the passenger. The second step was examining the first boarding points of each cluster in step 1. The last step was finding repeated time patterns for each regular trip. Finally, an example of using travel regularity passenger classifications and personal travel time variability estimation was given.

2.3.6.11 Chang and Zhao's research work

Chang and Zhao (2016) presented a comprehensive data mining method to extract individual transit users' travel patterns from the dataset with incomplete information. The trip chains of each transit user were generated based on individual trip records. A DBSCAN approach was utilized to mine the transit users' travel patterns based on their historical trip chains. The travel pattern clustering was conducted under three categories, respectively, including: "Spatial-temporal regular", "Spatial regular", and "Temporal regular". This study can help transit agency to understand transit users' travel behavior variability more easily.

Table 2.5: Summary of SCD-Based Travel Patterns Analysis

Authors	Year	Data Type	Data Source	Case Study Location	Methodology	Achievements
Agard et al.	2007	Card ID; Route ID; Boarding transaction time; Boarding stops; Trip direction	SCD, GPS data	Gatineau, Canada	K-means clustering;	Revealed travelers' behavior
Seaborn et al.	2009	Card ID; Mode; Journey stage sequence number; Start information; End information; Date	SCD	London, UK	Trip chaining method	Estimated route connectivity information
Jang	2010	Bus Transfer information	SCD, GPS data	Seoul, Korea	NA	Revealed SC holders' historical travel patterns
Kieu et al.	2013	Card ID; Route ID; Boarding transaction time; Boarding stops; Alighting transaction time; Alighting stops; Trip direction	SCD, Survey data	Brisbane city, Australia	Travel itineraries reconstruction; DBSCAN algorithm	Revealed SC holders' historical travel patterns
Devillaine et al.	2013	Card ID; Card Type; Boarding transaction time; Boarding stops; Bus ID	SCD	Gatineau, Canada	Trip chaining method;	Identified transit users' trip characteristics
Ma et al.	2013	Card ID; Route ID; Boarding transaction time; Boarding stops; Alighting transaction time; Alighting stops	SCD	Beijing, China	Trip chaining method; DBSCAN algorithm; K-means clustering.	Revealed SC holders' historical travel patterns
Ma and Wang	2014	Card ID; Route ID; Boarding transaction time; Boarding stops; Alighting transaction time; Alighting stops; Transit speed	SCD, AVL	Beijing, China	Transit path–finding algorithm	Constructed a web-based platform analyzing transit performance
Kusakabe and Asakura	2014	Trip ID; Boarding information; Alighting information; Trip purpose; Card ID	SCD, Survey data	NA	Naive Bayes probabilistic model	Enhanced understanding of travelers' behavior
Jánošíková et al.	2014	In-vehicle travel time; Walking time; Number of transfers; Headway	SCD	Žilina, Slovak	Multinomial logit model	Revealed travelers' behavior
Long et al.	2016	Boarding transaction time; Boarding stops; Alighting transaction time; Alighting stops	SCD, Survey data	Beijing, China	NA	Enhanced understanding of travelers' behavior
Chang and Zhao	2016	Card ID; Boarding transaction time; Boarding stops; Route ID; Bus ID	SCD	Guangzhou, China	Trip chaining method; DBSCAN clustering	Revealed SC holders' travel patterns

2.4 Summary

A comprehensive review and synthesis of the past research and development based on the SCD are conducted. Both theoretical and practical applications have been discussed in this chapter. This review is intended to provide a solid reference and assist in the OD estimation process. It also gives a clear picture of the current situation in SCD utilization and the directions that the SCD-related studies may take in the near future.

Chapter 3. Case Study

3.1 Introduction

The chapter presents the raw database obtained and utilized in this study. The following sections are organized as follows. Section 3.2 shows the basic database information. Section 3.3 presents the basic information about each route. Finally, section 3.4 concludes this chapter with a summary.

3.2 Basic Database Information

The study utilized Guangzhou, China transit system smart card database for transit OD estimation and trip purpose analysis. A transaction record is generated each time a passenger boards. Data for 4 bus routes in 5 weekdays that are used in this study include 100,000 transactions.

The fare collection system used boarding-only controls and the information was gained from the smart card reader and then transferred to the data management center. Since the system was not integrated with the AVL system, the data recorded did not have the stop ID showing where the passengers paid their fares. Instead, the data had a transaction time record and general information on the transit route was linked to the equipment on the bus. The alighting information was not recorded.

Table 3.1: Smart Card Raw Data (Sample)

Route	Card ID	Card Type	Bus ID	Transaction Time
Route A	'9999995338410746	General	'60750014	'20150101011816
Route A	'9999990069526128	General	'60750014	'20150101002722
Route A	'9999998626055540	Student	'60750014	'20150101002658

A sample of transactions records of the smart card system is shown in Table 3.1 which contains the following information:

Card ID: The card ID is a 16-digit number that uniquely identifies a smart card. This field is critical as it allows selection of transit trips by a particular passenger. It is assumed in this study that the system allows multiple users to share the same card. In that case, additional processing is required to eliminate the redundant records.

Card Type: The card type in this study includes general card, student card, disable card and senior card. Different categories of cards can help detect transit users' activity characteristics more easily.

Route Number: This field indicates which route the transit record belongs to.

Bus ID: This field indicates on which bus the transaction takes place. The bus ID information can help to infer the direction of the transit.

Transaction Time: Transaction time field contains timestamp information that is accurate to the second. This field is necessary to sort the transit trips (transactions) of a passenger in sequential order.

3.3 Basic Information about Each Route

The basic information about 4 bus transit routes is introduced in this section. Each route has two operational directions.

Figure 3.1 presents the maps of each transit route.

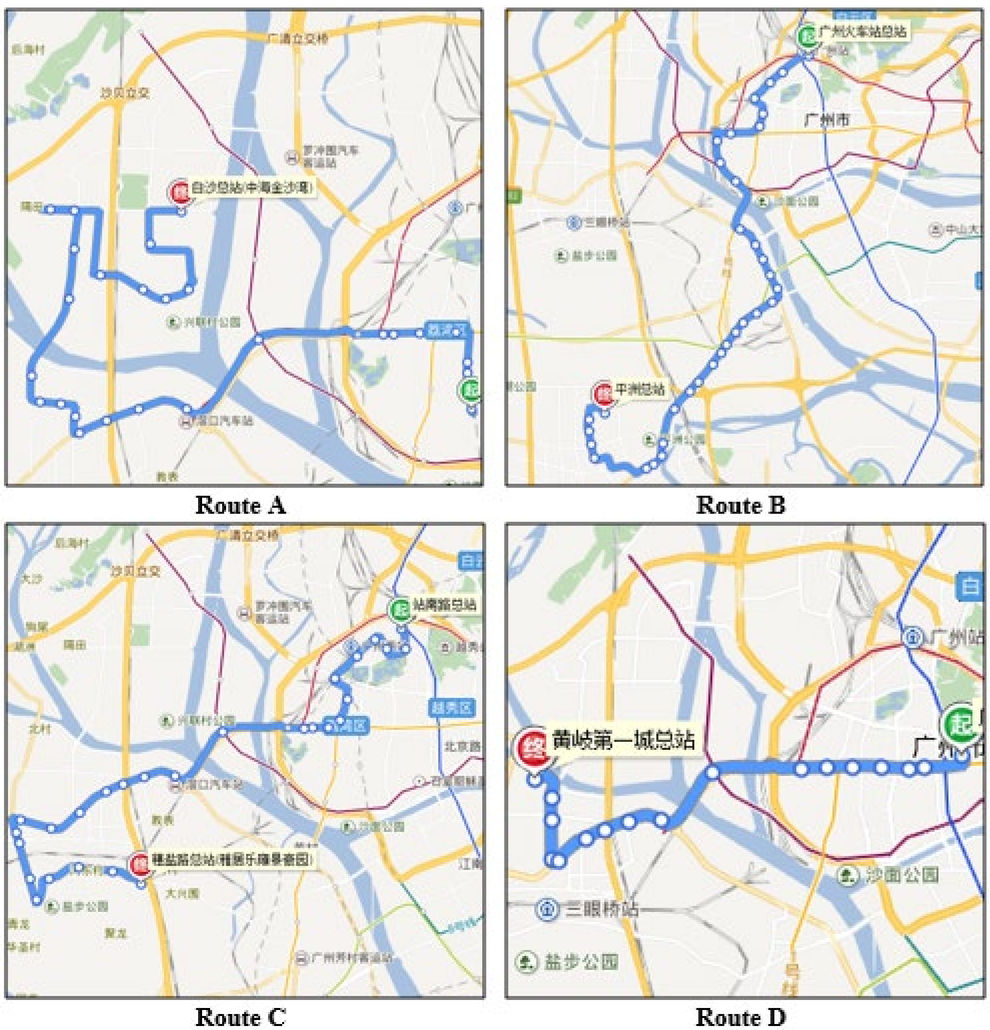

Figure 3.1: Maps of Each Transit Route

Route A upstream: Route A upstream goes through a bustling commercial district and the passenger volume is very high. There are 26 stops in total on the route, and the average operation time from the starting station to the last stop is 72 minutes and the total distance is 17.7 km. The distance between each station can be detected via the GPS data of the route.

Table 3.2 presents the distance information between each stop along route A upstream direction.

Table 3.2: Route A Upstream Basic Information

Station Number	Station Name	Zone Number	Travel Distance (m)
1	Kangwang Lu (Shangxiajiu) Bus Terminal	1	0
2	Hualin Temple	2	676
3	Kangwang Zhonglu	3	161
4	Chen Clan Academy	4	645
5	Liwan Lu Entrance	5	508
6	Shiluji	6	379
7	Zhongshan 8 Lu	7	158
8	Qiaozhong	8	1800
9	Guangfo Lu Entrance (Jiaokou Coach Station)	9	1800
10	Guangfo Lu East	10	730
11	Huangqi Downtown	11	505
12	Hongwei Lu Entrance	12	415
13	Hongwei Lu	13	324
14	Poyang Lu (Jingxiutang Pharmaceutical Factory)	14	400
15	Huangqi Administration Service Center	15	290
16	Shimen Middle School Entrance	16	466
17	Shamian Xincheng	17	1300
18	Jiaoyu Lu Entrance	18	729
19	Jianshe Dadao	19	623
20	Michong	20	344
21	Shaxi	21	2000
22	Baisha Town Hall	22	652
23	Xinglian	23	372
24	Baisha	24	375
25	Sanjun Market	25	1200
26	Baisha (Zhonghai Jingshanwan) Bus Terminal	26	906
Total			17758

Route A downstream: Route A downstream goes through a bustling commercial district and the passenger volume is very high. There are 26 stops in total on the route, and the average operation time from the starting station to the last stop is 75 minutes and the total distance is 18.7 km. The distance between each station can be detected via the GPS data of the route.

Table 3.3 presents the distance information between each stop along route A downstream direction.

Table 3.3: Route A Downstream Basic Information

Station Number	Station Name	Zone Number	Travel Distance (m)
1	Baisha (Zhonghai Jingshanwan) Bus Terminal	26	0
2	Sanjun Market	25	1700
3	Baisha	24	1300
4	Xinglian	23	362
5	Baisha Town Hall	22	415
6	Shaxi	21	575
7	Michong	20	2000
8	Jianshe Dadao	19	550
9	Jiaoyu Lu Entrance	18	427
10	Shamian Xincheng	17	851
11	Shimen Middle School Entrance	16	1100
12	Huangqi Administration Service Center	15	500
13	Poyang Lu (Jingxiutang Pharmaceutical Factory)	14	237
14	Hongwei Lu	13	471
15	Hongwei Lu Entrance	12	338
16	Huangqi Downtown	11	356
17	Guangfo Lu East	10	471
18	Guangfo Lu Entrance (Jiaokou Coach Station)	9	882
19	Qiaozhong	8	1900
20	Zhongshan 8 Lu	7	1800
21	Shiluji	6	327
22	Liwan Lu Entrance	5	254
23	Chen Clan Academy	4	554
24	Kangwang Zhonglu	3	478
25	Hualin Temple	2	332
26	Kangwang Lu (Shangxiajiu) Bus Terminal	1	534
Total			18714

Route B upstream: Route B upstream goes through the research region from north to south via several major roads and could represent the characteristics of north-south transit well. There are 34 stops in total on the route, and the average operation time from the starting station to the last stop is 90 minutes and the total distance is 22.8 km. The distance between each station can be detected via the GPS data of the route.

Table 3.4 presents the distance information between each stop along route B upstream direction.

Table 3.4: Route B Upstream Basic Information

Station Number	Station Name	Zone Number	Travel Distance (m)
1	Bus Terminal at Guangzhou Railway Station	1	0
2	Zhanqian Lu	2	1200
3	Xizhan	3	511
4	Xihua Luwei	4	722
5	Caihongqiao	6	685
6	Liwan Lu	7	422
7	Zhongshan 8 Lu	8	1000
8	Liwanhu Park West Entrance	9	519
9	Ruyifang	10	733
10	Huangsha Dadao (Stomatological Hospital of Guangzhou Medical University)	11	799
11	Fangcun	12	1900
12	Xiafangcun	13	742
13	Dachongkou	14	513
14	Hedong Xincun	15	534
15	Guangzhong Pier	16	488
16	True Light High School (Fusheng Garden)	17	578
17	Guangzhou Peiying Middle School	17	630
18	Baihedong	18	537
19	Hedong Lu West	19	357
20	Xilang	20	500
21	Maicun	21	335
22	Yuwei Bridge East	22	416
23	Huadi Dadao (Hainancun)	23	737
24	Huadi Dadao South	24	415
25	Shawei Bridge	25	750
26	Pingzhou Interchange	26	1200
27	Taiping Lu Entrance	27	406
28	Guicheng Elderly Center	28	417

Station Number	Station Name	Zone Number	Travel Distance (m)
29	Pingzhou Hotel	29	267
30	Pingxi Lu Entrance	30	2200
31	Guicheng Traffic Management Bureau	31	454
32	Nanhai No. 2 People's Hospital	32	423
33	Kongxi	34	394
34	Pingzhou Bus Terminal	35	1000
Total			22784

Route B downstream: Route B downstream goes through the research region from south to north via several major roads and could represent the characteristics of north-south transit well. There are 36 stops in total on the route, and the average operation time from the starting station to the last stop is 98 minutes and the total distance is 25 km. The distance between each station can be detected via the GPS data of the route.

Table 3.5 presents the distance information between each stop along route B downstream direction.

Table 3.5: Route B Downstream Basic Information

Station Number	Station Name	Zone Number	Travel Distance (m)
1	Pingzhou Bus Terminal	35	0
2	Nanhai No. 2 People's Hospital	34	1000
3	Guicheng Traffic Management Bureau	33	433
4	Pingxi Lu Entrance	32	284
5	Bengchong	31	393
6	Guangdong General Hospital Pingzhou Branch	30	2200
7	Pingzhou Jade Street	29	309
8	Pingzhou Hotel	29	529
9	Guicheng Elderly Center	28	444
10	Taiping Lu Entrance	27	507
11	Pingzhou Interchange	26	188
12	Shawei Bridge	25	1100
13	Huadi Dadao South	24	509
14	Huadi Dadao (Hainancun)	23	649
15	Yuwei Bridge East	22	624
16	Maicun	21	339
17	Xilang	20	281
18	Hedong Lu West	19	689
19	Baihedong	18	437

Station Number	Station Name	Zone Number	Travel Distance (m)
20	Guangzhou Peiying Middle School	17	731
21	Guangzhong Pier	16	885
22	Hedong Xincun	15	650
23	Dachongkou	14	486
24	Xiafangcun	13	541
25	Fangcun	12	820
26	Huangsha Dadao (Stomatological Hospital of Guangzhou Medical University)	11	2100
27	Ruyifang	10	610
28	Liwanhu Park West Entrance	9	489
29	Zhongshan 8 Lu	8	837
30	Liwan Lu	7	989
31	Caihongqiao	6	477
32	Dongfeng Xilu	5	1000
33	Xiyuan	4	484
34	Xizhan	3	402
35	Zhanqian Lu	2	867
36	Bus Terminal at Guangzhou Railway Station	1	1600
Total			24883

Route C upstream: Route C upstream goes through the research region from east to west via several major roads and could represent the characteristics of east-west transit well. There are 24 stops in total on the route, and the average operation time from the starting station to the last stop is 71 minutes and the total distance is 17.5 km. The distance between each station can be detected via the GPS data of the route.

Table 3.6 presents the distance information between each stop along route C upstream direction.

Table 3.6: Route C Upstream Basic Information

Station Number	Station Name	Zone Number	Travel Distance (m)
1	Suiyan Lu (Majestic Garden) Bus Terminal	1	0
2	Yaju Lanwan	2	780
3	Sanyanqiao (Guangdong Grain & Oil Market)	3	744
4	Hubang Market	4	695
5	Yanbu Hospital	5	1000

Station Number	Station Name	Zone Number	Travel Distance (m)
6	Hubang Lu Entrance	6	568
7	Lian'an	7	878
8	Yanbu Dadao North	8	223
9	Yanbu Entrance	9	512
10	Dazhuanwan (Huangqi Kaimin Tea Market)	10	490
11	Jiazhou Plaza	11	986
12	Hongwei Lu Entrance	12	554
13	Huangqi Downtown	13	344
14	Guangfo Lu East	14	489
15	Guangfo Lu Entrance (Jiaokou Coach Station)	15	884
16	Qiaozhong	16	1900
17	Zhongshan 8 Lu	17	1800
18	Shiluji	18	158
19	Liwan Lu	20	784
20	Caihongqiao	21	471
21	Xiyuan	22	1500
22	Xizhan	23	413
23	Zhanqian Lu	24	863
24	Zhannan Lu Bus Terminal	26	468
Total			17504

Route C downstream: Route C downstream goes through the research region from west to east via several major roads and could represent the characteristics of east-west transit well. There are 27 stops in total on the route, and the average operation time from the starting station to the last stop is 72 minutes and the total distance is 17.4 km. The distance between each station can be detected via the GPS data of the route.

Table 3.7 presents the distance information between each stop along route C downstream direction.

Table 3.7: Route C Downstream Basic Information

Station Number	Station Name	Zone Number	Travel Distance (m)
1	Zhannan Lu Bus Terminal	26	0
2	Military Hospital	26	506
3	Zhanqian Henglu	25	590
4	Zhanqian Lu	24	458
5	Xizhan	23	530

Station Number	Station Name	Zone Number	Travel Distance (m)
6	Xihua Luwei	22	680
7	Caihongqiao	21	680
8	Liwan Lu	20	498
9	Liwan Lu Entrance	19	325
10	Shiluji	18	355
11	Zhongshan 8 Lu	17	175
12	Qiaozhong	16	1800
13	Guangfo Lu Entrance (Jiaokou Coach Station)	15	1800
14	Guangfo Lu East	14	730
15	Huangqi Downtown	13	505
16	Hongwei Lu Entrance	12	415
17	Jiazhou Plaza	11	683
18	Dazhuanwan (Huangqi Kaimin Tea Market)	10	833
19	Yanbu Entrance	9	701
20	Yanbu Dadao North	8	533
21	Lian'an	7	230
22	Hubang Lu Entrance	6	815
23	Yanbu Hospital	5	477
24	Hubang Market	4	834
25	Sanyanqiao (Guangdong Grain & Oil Market)	3	726
26	Yaju Lanwan	2	738
27	Suiyan Lu (Majestic Garden) Bus Terminal	1	778
Total			17395

Route D upstream: Route D is similar to Route C. Route D upstream goes through the research region from east to west via several major roads and could represent the characteristics of east-west transit well. There are 17 stops in total on the route, and the average operation time from the starting station to the last stop is 51 minutes and the total distance is 11.7 km. The distance between each station can be detected via the GPS data of the route.

Table 3.8 presents the distance information between each stop along route D upstream direction.

Table 3.8: **Route D Upstream Basic Information**

Station Number	Station Name	Zone Number	Travel Distance (m)
1	Guangwei Lu Bus Terminal	1	0
2	Zhongshan 6 Lu	2	1100
3	Ximenkou (Zhongshan 6 Lu)	3	303
4	Zhongshan 7 Lu	4	554
5	Chen Clan Academy	5	605
6	Liwan Lu Entrance	6	554
7	Zhongshan 8 Lu	6	477
8	Qiaozhong	7	1800
9	Guangfo Lu Entrance (Jiaokou Coach Station)	8	1900
10	Guangfo Lu East	9	730
11	Huangqi Downtown	10	505
12	Hongwei Lu Entrance	11	415
13	Jiazhou Plaza	12	748
14	Huanghai Lu	13	202
15	Huangqi Department Building	14	806
16	Shimen Middle School Entrance	15	434
17	Huangqi Diyicheng Bus Terminal	16	879
Total			12012

Route D downstream: Route D is similar to Route C. Route D downstream goes through the research region from west to east via several major roads and could represent the characteristics of east-west transit well. There are 17 stops in total on the route, and the average operation time from the starting station to the last stop is 53 minutes and the total distance is 12.1 km. The distance between each station can be detected via the GPS data of the route.

Table 3.9 presents the distance information between each stop along route D downstream direction and the zone number of each stop.

Table 3.9: **Route D Downstream Basic Information**

Station Number	Station Name	Zone Number	Travel Distance (m)
1	Huangqi Diyicheng Bus Terminal	16	0
2	Shimen Middle School Entrance	15	879
3	Huangqi Department Building	14	503
4	Huanghai Lu	13	719
5	Jiazhou Plaza	12	535
6	Hongwei Lu Entrance	11	571
7	Huangqi Downtown	10	354

Station Number	Station Name	Zone Number	Travel Distance (m)
8	Guangfo Lu East	9	647
9	Guangfo Lu Entrance (Jiaokou Coach Station)	8	720
10	Qiaozhong	7	1800
11	Zhongshan 8 Lu	6	1800
12	Shiluji	6	324
13	Chen Clan Academy	5	815
14	Zhongshan 7 Lu	4	524
15	Ximenkou (Zhongshan 6 Lu)	3	663
16	Zhongshan 6 Lu	2	478
17	Guangwei Lu Bus Terminal	1	772
Total			12104

3.4 Summary

This chapter analyzes the data structure of the raw data based on which the methodology is to be developed in order to achieve the schedule information. Due to the limitation of the sample size and the variety of attributes of the raw data, the schedule information will be very important and will be developed later in the study.

Chapter 4. Boarding Information Identification

4.1 Introduction

The chapter introduces the boarding location information identification methodology developed and utilized in this study. The following sections are organized as follows. Section 4.2 presents the assumptions and constraints in this study. Section 4.3 describes how to identify the bus operation time. Section 4.4 describes how to identify the direction information. Section 4.5 explains how the boarding cluster information is generated for the further boarding stop information analysis. Section 4.6 shows the methodology of the boarding location identification. Finally, section 4.7 concludes this chapter with a summary.

4.2 Assumptions and Constraints

In the steps of cleaning the raw data and developing the prototype boarding and alighting location identification algorithms, the following assumptions have been made to minimize the potential errors:

- Only a limited amount of information about the route is available. Since there is no detailed bus schedule information, the operation time between two consecutive stops will be calculated based on the distance between stops and the roadway geometry condition along each section.
- The schedule information is calculated based on the distance and there is no accurate timestamp at a specific stop. Therefore, the first record of a trip is arranged to be at the first stop of the route.
- Boarding records in a short time period are categorized as one boarding cluster.
- Most riders would like to return to the destination station of their last trip to begin next trip. Therefore, the destination of one trip is labeled based on the boarding location of its next transfer trip.
- The destination of most riders at the end of day is the origin stop where they begin their first trip of the next day.

4.3 Bus Operation Time Identification

As there is no available published/fixed schedule of the buses, the operation times between stations are computed based on the operational length and number of intersections between bus stations, as well as the total distance of the route. The distance between stations can be detected via GIS data of the route. 25 km per hour (km/h) is assumed and utilized as the operation speed of all buses (Tantiyanugulchai and Bertini, 2003, Chen and Fan, 2018). The travel time between stops can be illustrated as:

$$T_{ij} = D_{ij}/V_{average} + (T - D_{total}/V_{average}) \cdot U_{ij}$$

D_{ij} stands for the distance between two adjacent stops i and j on the map; $V_{average}$ is the average operational speed of the bus (i.e., 25 km/h in this study); T is the total travel time of a whole bus trip on the schedule (i.e., from first station to the terminal station); and U_{ij} is the adjustment made to the travel time between stops, which can be different due to different factors including unsignalized intersections, U-turns, roundabouts and traffic signals involved.

Table 4.1 presents a sample of travel time between stops. The detailed information about each route could be found in Appendix A.

Table 4.1: Route A Upstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Kangwang Lu (Shangxiajiu) Bus Terminal	0	0
2	Hualin Temple	676	176
3	Kangwang Zhonglu	161	23
4	Chen Clan Academy	645	197
5	Liwan Lu Entrance	508	126
6	Shiluji	379	107
7	Zhongshan 8 Lu	158	75
8	Qiaozhong	1800	469
9	Guangfo Lu Entrance	1800	261
10	Guangfo Lu East	730	210
11	Huangqi Downtown	505	125
12	Hongwei Lu Entrance	415	60
13	Hongwei Lu	324	47
14	Poyang Lu	400	110
15	Huangqi Administration Service Center	290	68
16	Shimen Middle School Entrance	466	68
17	Shamian Xincheng	1300	188
18	Jiaoyu Lu Entrance	729	184
19	Jianshe Dadao	623	116
20	Michong	344	50
21	Shaxi	2000	394
22	Baisha Town Hall	652	94
23	Xinglian	372	80
24	Baisha	375	80
25	Sanjun Market	1200	200
26	Baisha (Zhonghai Jingshanwan) Bus Terminal	906	183
Total		17758	3692

4.4 Direction Information Identification

Since the transaction database only records the transaction time and contains no information about the bus direction, the boarding stops cannot be determined by matching transaction times with the estimated transit operation schedule directly. To infer trip direction information, many studies utilized a 30-minute time gap as the time threshold for direction identification (Ma et al., 2012, Chen and Fan, 2018). Note that such time gap is not the only way which can be used to detect the change in direction because the time spent by the buses at terminal station is unknown. For example, when the bus is operated during peak hours, it may leave the terminal after only a short break due to the need to serve high passenger demands.

To detect the bus trip direction, it is necessary to study the transaction records of each bus based on the time sequence and label the direction with the original operation time per trip on the schedule as a reference. In this regard, the direction of the first transaction period is marked. The direction labeling process will use the following criteria:

- The direction will be changed when the time gap between the current transaction and the last transaction is more than 30 minutes;
- The time difference between current record and first record of last transaction sequence is close to the operation time on schedule.
- The records with the transaction time before 6 a.m. on any day will be classified as belonging to the trip series of the previous day.

In short, the trip direction detection algorithm is presented below:

Step 1: Sort the data by Route ID/Bus ID/Transaction date/Transaction Time.

Figure 4.1 presents the detailed information about the record sorting process.

Column		Sort On	Order
Sort by	Route ID	Values	A to Z
Then by	Bus ID	Values	Smallest to Largest
Then by	Transaction Date	Values	Oldest to Newest
Then by	Transaction Time	Values	Smallest to Largest

Figure 4.1: Record Sorting Process

Step 2: Label the direction as 1 (upstream) to the first transaction record of the day after 6 a.m.

Step 3: Read and label next transaction record: Record the time difference between current record j and previous record i as t_{ij} and also label the time difference between current record and the first record as t'_j.

Step 4: Label direction information: If t_{ij} < 15 minutes and t'_i < T (total operation time for a whole bus trip along this route on the schedule), then label the direction of current record the same as the previous one. If t_{ij} > 15 minutes, and the t'_i of the previous record < T, then label the status of current record as 'hold'. Continue reading the records. When another t_{ij} > 15 minutes and t'_i > T, determine the direction change status based on the value $|t'_i$ - T| and choose the record with the minimum (among all $|t'_i$ - T| values) as the direction change record.

Figure 4.2 illustrates the detail information on the trip direction identification algorithm.

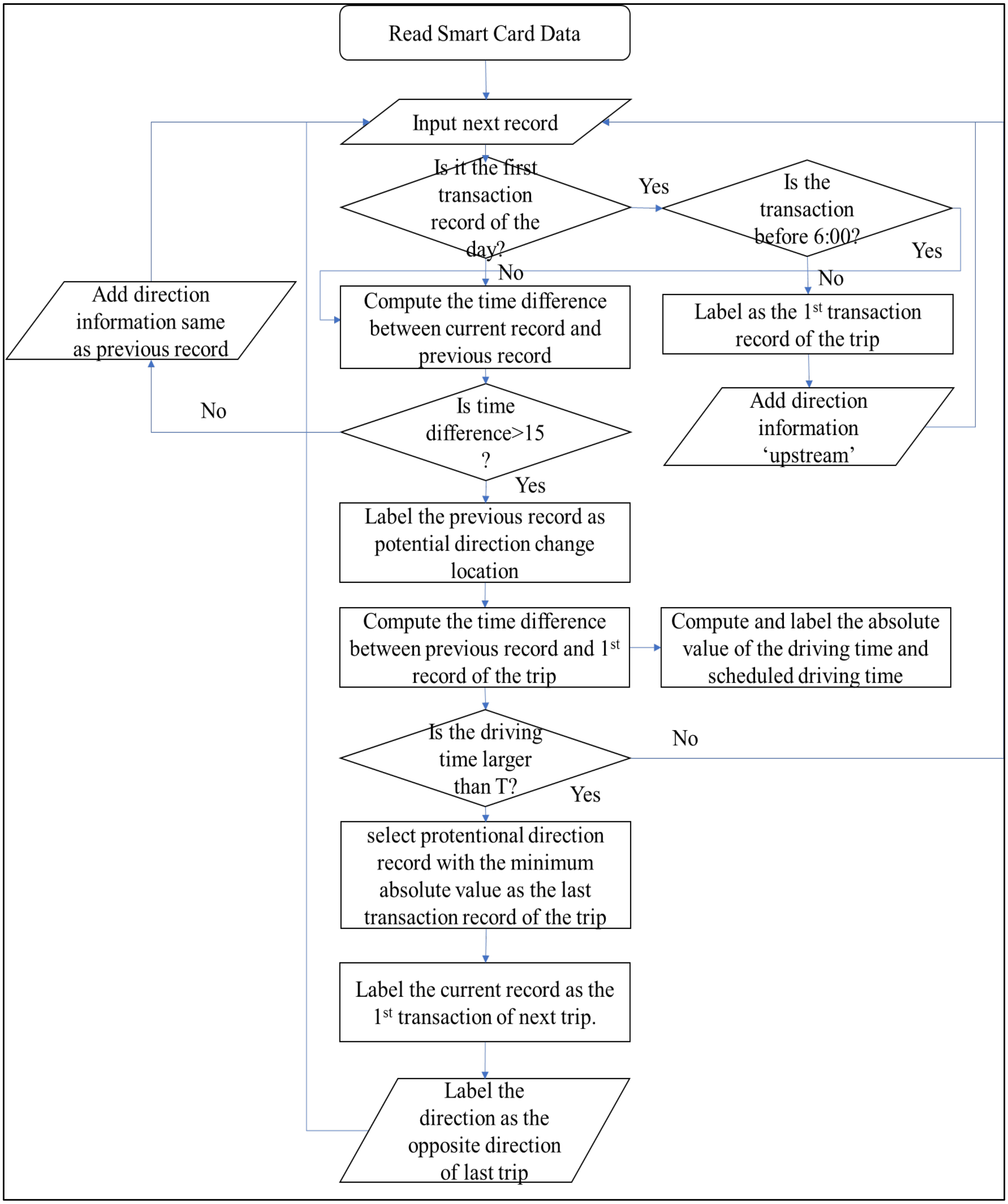

Figure 4.2: Trip Direction Identification Algorithm

4.5 Transaction Data Clustering

As known, many different scenarios could exist in the real world. For example, congestion can occur during peak hours, and there may be no boarding activities at the first and/or last stops. Therefore, it is impossible to guarantee that the buses will be operated according to the fixed schedule. Based on the transaction records, it will be convenient and also more accurate to group several transaction records together into a same boarding cluster. Because several passengers usually board in an intensive period of time, and therefore, multiple smart cards swiping activities will occur at one specific bus stop, the boarding clusters can be labeled based on the time interval between the transactions.

The process contains the following steps:

Step 1: The records are sorted by the sequence of route ID/bus ID/transaction date/transaction time; and

Step 2: The time interval threshold for two consecutive records is 60 seconds (Ma et al., 2012). If the interval is within 60 seconds, records are grouped into the same boarding cluster; otherwise, the boarding cluster will be different and changed.

Figure 4.3 presents the detailed information about boarding cluster identification algorithm.

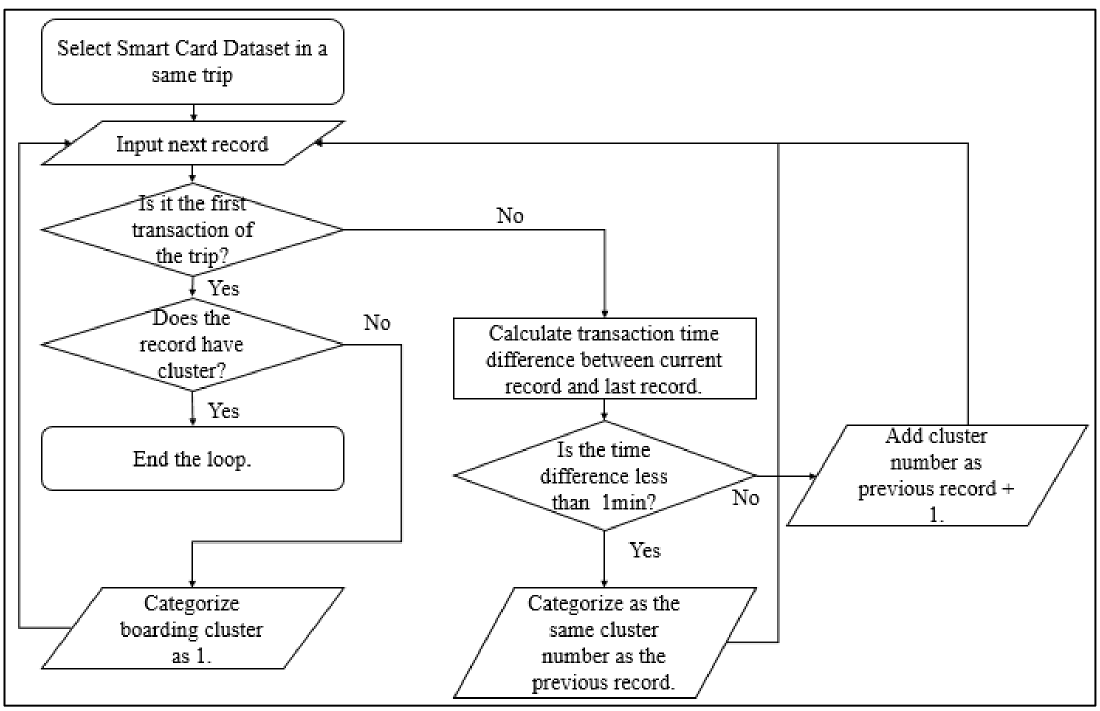

Figure 4.3: Boarding Cluster Identification Algorithm

Table 4.2 presents an example of the clustering results. Note that the number of clusters indicates the boarding activities during a whole trip along a route. It is obvious that there can be no boarding passengers at some stops. For each cluster, timestamps will be added to both the first and last record for the convenience of future analyses.

Table 4.2: Boarding Cluster Identification Results (Sample)

Route ID	Transaction ID	Card ID	Card Type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Cluster
RouteA	10061	90009698490	Regular Card	75	2015/1/1	0:15:05	B	1
RouteA	10062	95208436630	Regular Card	75	2015/1/1	0:16:54	B	2
RouteA	10063	99857026060	Regular Card	75	2015/1/1	0:16:55	B	2
RouteA	2411	91005456310	Regular Card	75	2015/1/1	0:17:31	B	2
RouteA	2412	99455451040	Regular Card	75	2015/1/1	0:17:33	B	2
RouteA	2451	99460138890	Senior Card	75	2015/1/1	0:27:04	B	3

4.6 Boarding Stop Information Extraction

After the identification of boarding clusters, the specific boarding stops can be inferred based on the difference in timestamps between adjacent boarding clusters. The first boarding cluster is assumed to belong to the first stop (i.e., origin) of the route.

Based on the results of direction identification, the bus operation times of most trips are shorter than the scheduled operation time of a whole bus trip (Chen and Fan, 2018). By examining the data, this can be attributed to few passengers' boarding activities occurring during several intermediate (particularly the last few) bus stops of the trip. In addition, traffic delays should also be considered under the situation of congestion during peak hours. Hence, the boarding stop information extraction will use the following rules, which will ensure that different clusters be assigned to different bus stops:

The boarding stop of first boarding cluster is labeled as "stop 1".

The difference between the timestamps of the first record of boarding cluster n+1 and last record of boarding cluster n is Δt_{n+1}; The boarding stop ID of boarding cluster n is i; and the operation time between station i and i+1 is $a_{i\,(i+1)}$ on the schedule.

If $\Delta t_{n+1} < a_{i\,(i+1)}$, then label the boarding stop as i+1 and to the records with boarding cluster n+1.

If $\Delta t_{n+1} > a_{i\,(i+1)}$, compare Δt_{n+1} and the operation time $a_{i\,(i+2)}$ between station i and i+2 on the schedule……until $a_{i\,(i+k)} > \Delta t_{n+1}$, then label the boarding stop as i+k-1 and to the records with boarding cluster n+1.

Figure 4.4 presents the rules used in the boarding information extraction process.

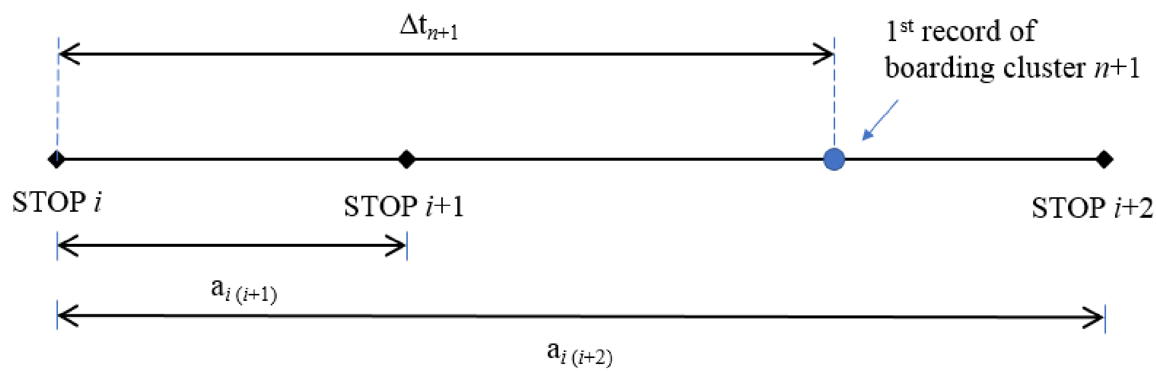

Figure 4.4: **Boarding Information Extraction**

By applying the algorithm developed above to process the smart card data collected from Guangzhou, China, the boarding information is successfully identified. The information gathered out of this chapter includes:

Direction information. After the direction information is labeled to each record, the potential relationships between records can be further analyzed.

Boarding cluster information. With the help of boarding cluster identification, the transaction records for a whole trip along a route are divided into different cluster groups in order to identify the transactions occurred at different stops. The results of cluster identification can greatly reduce potential errors involved in the boarding stop identification and boarding passenger count estimation processes.

Boarding stop information. Finally, the results show the estimated boarding location of each transaction records. Additional information about passenger boarding counts at each stop could be mined by analyzing the SCD with estimated boarding locations.

Table 4.3 presents the detailed results of the boarding stop information extraction. Three new columns, which represent the boarding direction, boarding cluster and boarding stop ID of each record, are added.

Table 4.3: Boarding Information Identification Results (Sample)

Route ID	Transaction ID	Card ID	Card Type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Cluster	Boarding Stop ID
RouteA	10044	99852761030	Regular Card	75	2015/1/1	0:14:44	B	1	1
RouteA	10061	90009698490	Regular Card	75	2015/1/1	0:15:05	B	1	1
RouteA	10062	95208436630	Regular Card	75	2015/1/1	0:16:54	B	2	2
RouteA	10063	99857026060	Regular Card	75	2015/1/1	0:16:55	B	2	2
RouteA	2411	91005456310	Regular Card	75	2015/1/1	0:17:31	B	2	2
RouteA	2412	99455451040	Regular Card	75	2015/1/1	0:17:33	B	2	2
RouteA	2451	99460138890	Senior Card	75	2015/1/1	0:27:04	B	3	5
RouteA	4350	90770525860	Regular Card	75	2015/1/1	7:14:02	A	1	1
RouteA	4351	98449075430	Regular Card	75	2015/1/1	7:16:31	A	2	2
RouteA	4352	96219065390	Regular Card	75	2015/1/1	7:16:36	A	2	2
RouteA	4353	98232009250	Student Card	75	2015/1/1	7:17:36	A	3	3
RouteA	3726	92256025680	Student Card	75	2015/1/1	7:17:38	A	3	3
RouteA	4439	97860198820	Senior Card	75	2015/1/1	7:17:56	A	3	3
RouteA	4440	97035705710	Regular Card	75	2015/1/1	7:19:06	A	4	4
RouteA	4441	96438204420	Regular Card	75	2015/1/1	7:19:09	A	4	4
RouteA	5073	91446095450	Regular Card	75	2015/1/1	7:20:01	A	4	4
RouteA	3830	97046638410	Regular Card	75	2015/1/1	7:20:02	A	4	4
RouteA	4403	94055185710	Regular Card	75	2015/1/1	7:21:20	A	5	5
RouteA	7891	92511324600	Student Card	75	2015/1/1	7:21:27	A	5	5
RouteA	7892	90073109240	Regular Card	75	2015/1/1	7:21:29	A	5	5
RouteA	7893	99733607940	Regular Card	75	2015/1/1	7:21:31	A	5	5
RouteA	3589	97057255310	Student Card	75	2015/1/1	7:21:35	A	5	5
RouteA	3590	93450025200	Regular Card	75	2015/1/1	7:21:36	A	5	5
RouteA	5369	91138055650	Student Card	75	2015/1/1	7:21:39	A	5	5
RouteA	5370	94907428630	Regular Card	75	2015/1/1	7:21:42	A	5	5
RouteA	3082	99133334000	Regular Card	75	2015/1/1	7:21:45	A	5	5

4.7 Summary

This chapter describes the schedule based boarding location identification methodology step by step. All activities conducted in the sections of this chapter aim to minimize the potential errors. The developed boarding location identification methodology is very general and can be used to detect origin information based on the SC transaction data, both with and without other support databases.

Chapter 5. Alighting Information Identification

5.1 Introduction

The chapter presents the alighting information identification methodology developed and utilized in this study. The following sections are organized as follows. Section 5.2 describes different scenarios of the transaction records. Section 5.3 explains different scenarios of the transaction records with the same ID. Section 5.4 shows the analysis results of the case study. Finally, section 5.5 concludes this chapter with a summary.

5.2 Scenario Analyses

Based on the literature review and the analysis of raw data, the alighting information for only parts of the records can be detected. Therefore, the transaction records for each card ID are categorized into different scenarios based on the number of the records, timestamp of the records, etc. This step aims to help better understand the alighting information identification methodology.

5.2.1 Scenario 1: A Single Trip

Under this scenario, there is only one transaction record of that specific card ID. Based on the boarding information identification methodology described in Chapter 4, the trip origin (boarding stop) of this record can be inferred. However, the destination information is not recorded, and it seems not possible to detect the alighting information for such records. Therefore, the records under this scenario are categorized as a single trip scenario and no alighting information can be identified.

Table 5.1 presents a data sample of a single trip scenario.

Table 5.1: **Sample Record of a Single Trip**

Route ID	Transaction ID	Card ID	Card type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Stop
RouteA	74	97305414700	Regular Card	2400515	2015/1/1	19:49:50	A	2

Figure 5.1 presents an example of a single trip scenario. The alighting information on this record cannot be detected as no further information can be mined from the database.

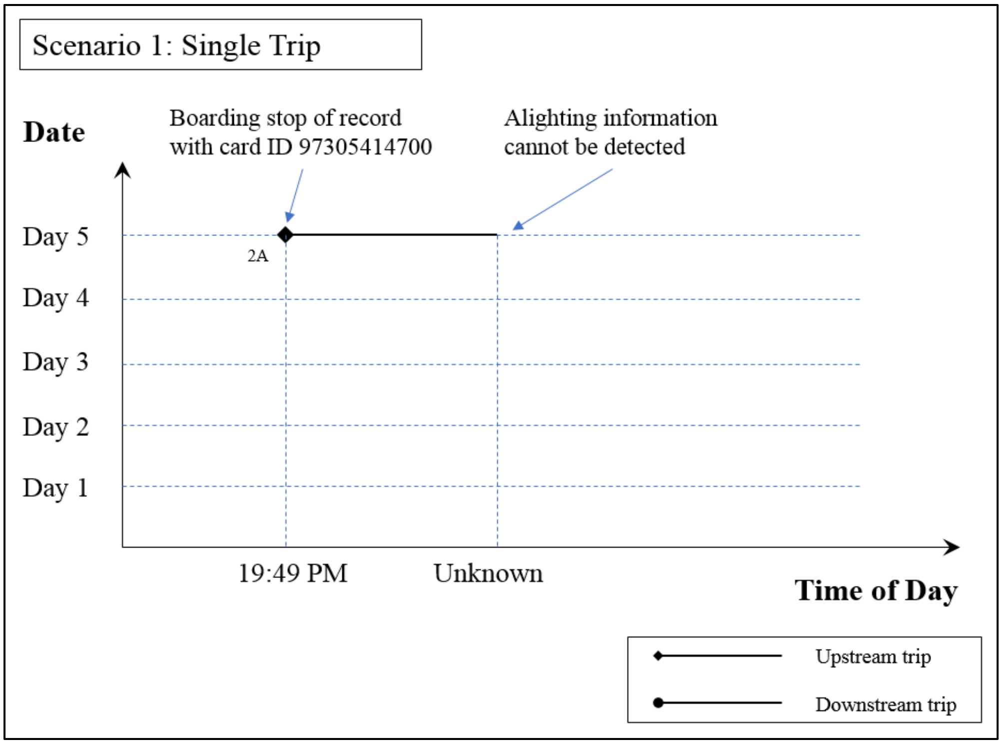

Figure 5.1: A Single Trip Scenario

5.2.2 Scenario 2: A Single Trip Chain

Under this scenario, there are multiple records of that specific card ID and the transaction date and period of all the records are same. The direction of these records is same. Based on the boarding information identification methodology described in Chapter 4, the trip origin (boarding stop) of these records can be inferred. Based on the trip chaining theory, the alighting information about the first few records could also be detected except the last record. For the last record of the trip chain, it seems not possible to detect the alighting information. Therefore, the records under this scenario are categorized as a single trip chain scenario

Table 5.2 presents a data sample of the single trip chain.

Table 5.2: Sample Records of a Single Trip Chain

Route ID	Transaction ID	Card ID	Card type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Stop
RouteC	70666	95930288860	Senior Card	90857004	2015/1/1	8:40:47	A	8
RouteC	55903	95930288860	Senior Card	30580050	2015/1/1	9:14:01	A	11

Figure 5.2 presents an example of a single trip chain scenario. The alighting information about the last record cannot be detected as no further information can be mined from the database.

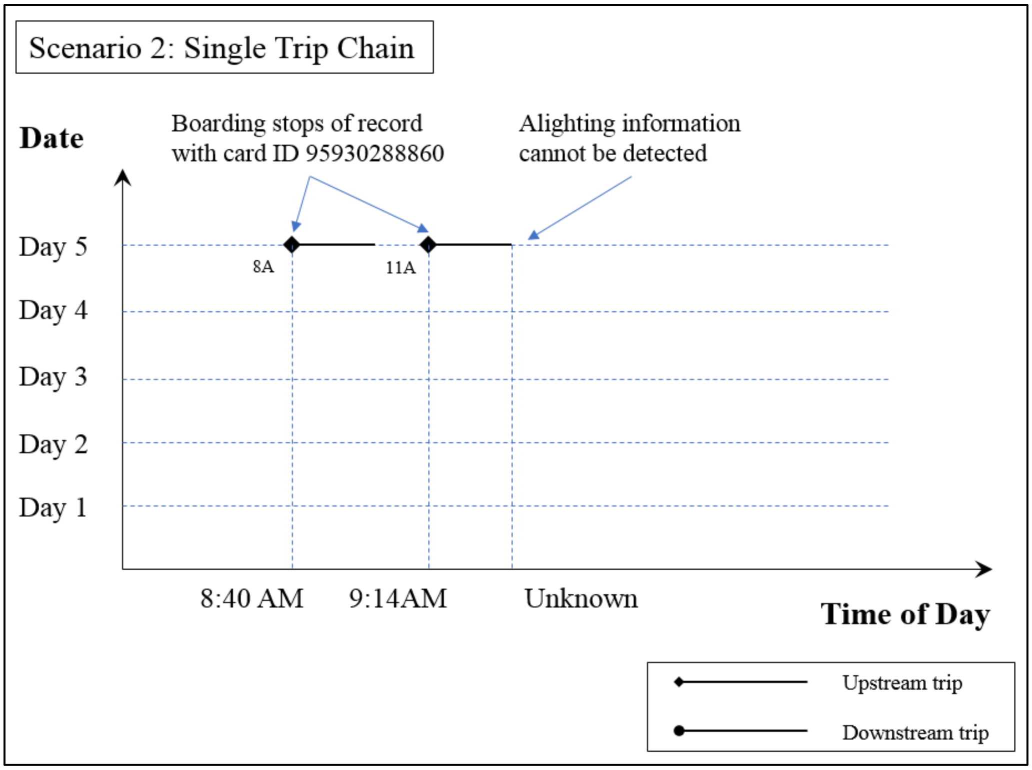

Figure 5.2: A Single Trip Chain Scenario

5.2.3 Scenario 3: Multiple Single Trips in Same Period

Under this scenario, there are multiple records of that specific card ID. Based on the boarding information identification methodology described in Chapter 4, the trip origin (boarding stop) of these records can be inferred. However, all the records occurred in a same time period (AM period or PM period) in different days and there is still no helpful information to use to detect the potential alighting information. The records under this scenario are categorized as multiple single trips in different days and no alighting information can be identified.

Table 5.3 presents a data sample of multiple single trips in different days.

Table 5.3: Sample Records of Multiple Single Trips in Different Days

Route ID	Transaction ID	Card ID	Card type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Stop
RouteA	23256	90000646530	Regular Card	60582050	2015/1/2	18:29:31	A	7
RouteA	12146	90000646530	Regular Card	5000005	2015/1/3	18:29:26	A	4

Figure 5.3 presents an example of the “multiple single trips in different days” scenario. The alighting information on these records cannot be detected as no further information can be mined from the database.

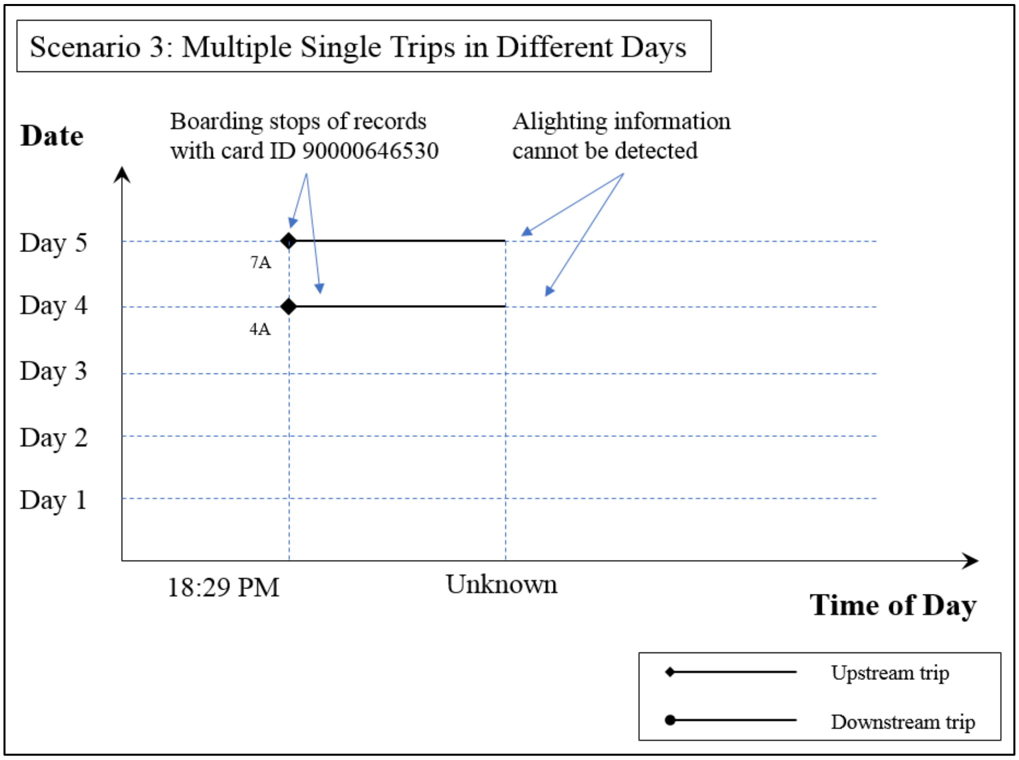

Figure 5.3: Multiple Single Trips in Different Days Scenario

5.2.4 Scenario 4: A Single Trip and Trip Chain in Same Period

Under this scenario, there are multiple records of that specific card ID. All the records happened in a same period (AM period or PM period). The directions of these records are same. Based on the boarding information identification methodology described in Chapter 4, the trip origin (boarding stop) of these records can be inferred. The alighting information on the first few records in the trip chain can be detected. However, there is still no helpful information to use to help detect the potential alighting information about other records. The records under this scenario are categorized as multiple single trip and trip chain in the same period.

Table 5.4 presents a data sample of a single trip and trip chain in the same period.

Table 5.4: Sample Records of a Single Trip and Trip Chain Scenario

Route ID	Transaction ID	Card ID	Card type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Stop
RouteD	87812	98537901310	Regular Card	60100304	2015/1/2	17:54:21	B	1
RouteD	91893	98537901310	Regular Card	4000715	2015/1/2	19:43:01	B	11
RouteD	89104	98537901310	Regular Card	7210197	2015/1/3	17:18:11	A	9

Figure 5.4 presents an example of a single trip and trip chain in the same period scenario.

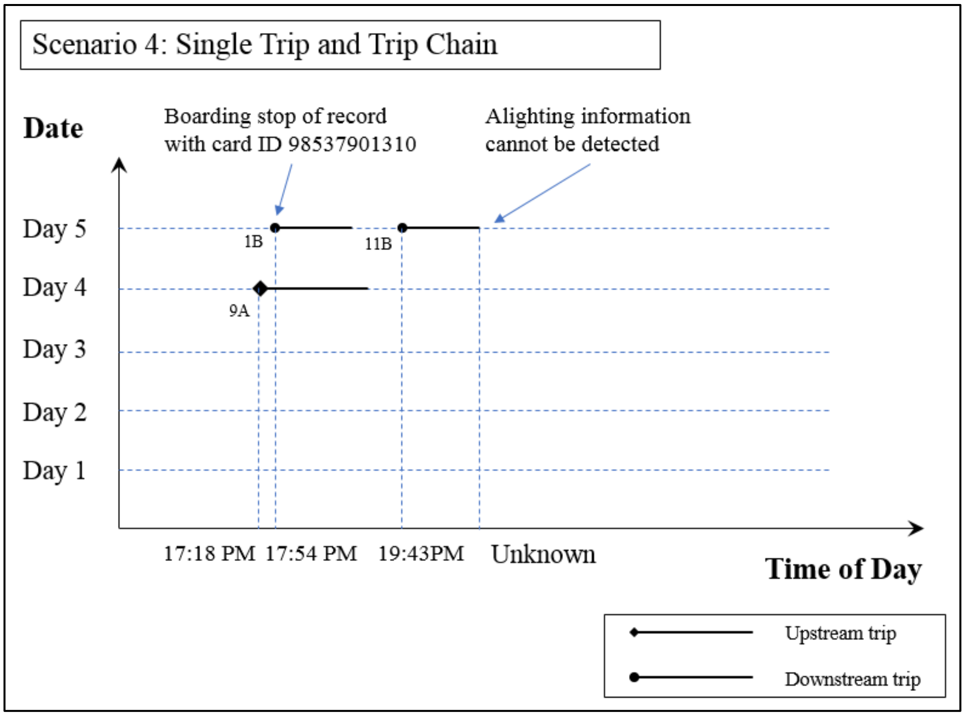

Figure 5.4: A Single Trip and Trip Chain Scenario

5.2.5 Scenario 5: Trips with Different Directions

Under this scenario, there are multiple records in same day of that specific card ID. The directions of these records are different. Based on the boarding information identification methodology described in Chapter 4, the trip origin (boarding stop) of these records can be inferred. The boarding stop is the potential alighting stop of another record.

Table 5.5 presents a data sample of trips with different directions.

Table 5.5: Sample Records of Trips with Different Directions Scenario

Route ID	Transaction ID	Card ID	Card Type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Stop
RouteB	44512	92046091410	Regular Card	40448004	2015/1/4	9:53:27	A	4
RouteB	49817	92046091410	Regular Card	78004510	2015/1/4	18:39:43	B	12

Figure 5.5 presents an example of trips with different directions scenario. The boarding information on one record can be the potential alighting stop of the matched record.

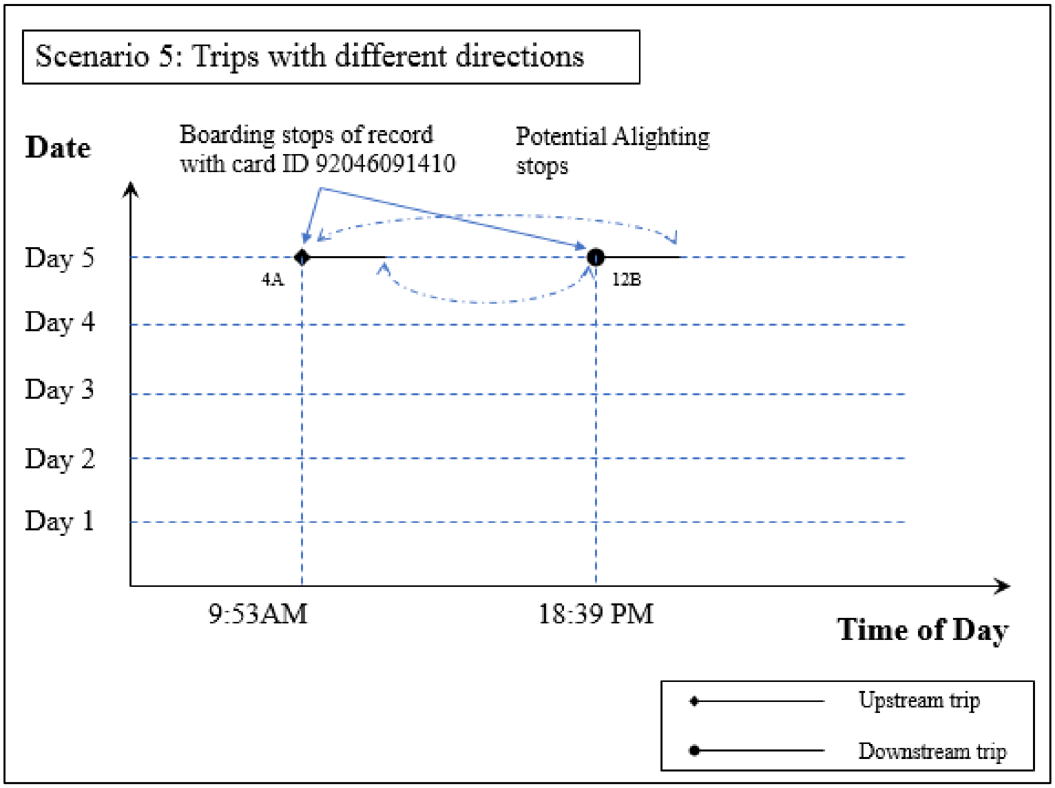

Figure 5.5: Trips with Different Directions Scenario

5.2.6 Scenario 6: Commute Trips in Different Days

Under this scenario, there are multiple records of that specific card ID occurred in different days during both AM and PM period. The direction of the records should also be different. Based on the boarding information identification methodology described in Chapter 4, the trip origin (boarding stop) of these records can be inferred. The potential alighting information on the records under this scenario could also be mined for the further analysis.

Table 5.6 presents a data sample of commute trips.

Table 5.6: Sample Records of Commute Trips Scenario

Route ID	Transaction ID	Card ID	Card type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Stop
RouteB	36559	93737492690	Regular Card	30547050	2015/1/2	22:16:59	A	7
RouteB	34173	93737492690	Regular Card	78004510	2015/1/3	10:39:26	B	4

Figure 5.6 presents an example of commute trips scenario. The boarding information about these records can be the potential alighting stop of the matched record.

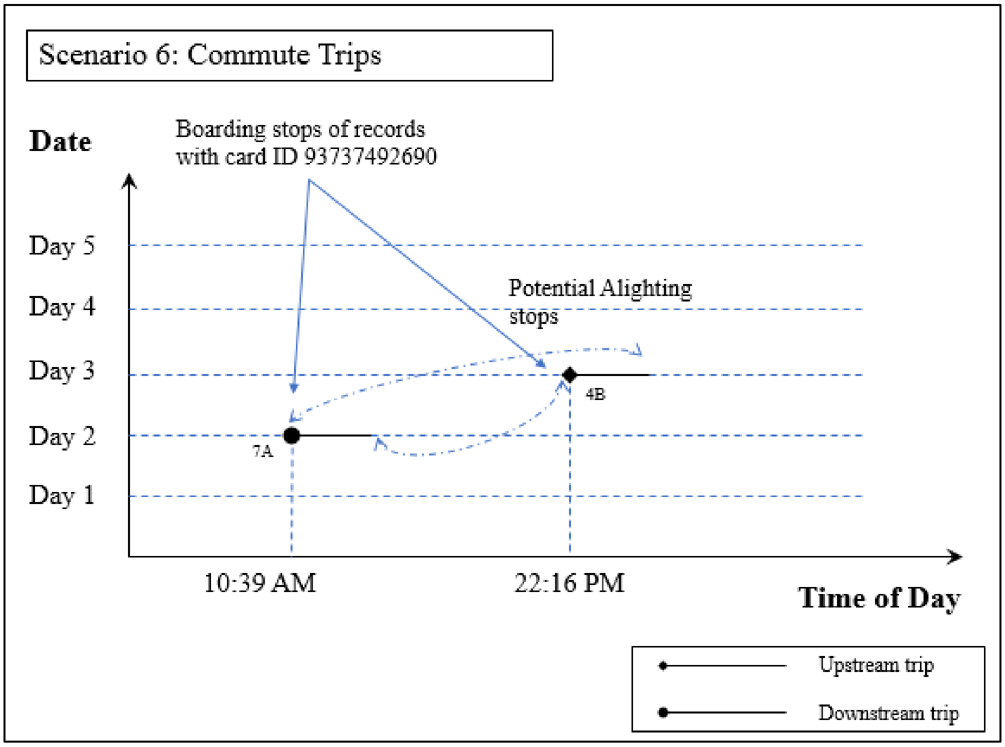

Figure 5.6: Commute Trips Scenario

5.2.7 Scenario 7: Commute Trips with Trip Chain

Under this scenario, there are multiple records of that specific card ID happened in different days. The direction of the records in different periods should also be different. Based on the boarding information identification methodology described in Chapter 4, the trip origin (boarding stop) of these records can be inferred. The potential alighting information on the records under this scenario could also be mined for the further analysis.

Table 5.7 presents a data sample of commute trips with trip chain.

Table 5.7: Sample Records of Commute Trips with Trip Chain Scenario

Route ID	Transaction ID	Card ID	Card type	Bus ID	Transaction Date	Transaction Time	Direction	Boarding Stop
RouteA	422	95092298010	Regular Card	2008035	2015/1/1	11:57:43	A	1
RouteA	532	95092298010	Regular Card	5000005	2015/1/1	12:17:19	A	8
RouteA	6277	95092298010	Regular Card	5500290	2015/1/1	14:55:24	B	3

Figure 5.7 presents an example of commute trips with trip chain scenario. The boarding information on these records could be the potential alighting stop of the matched record.

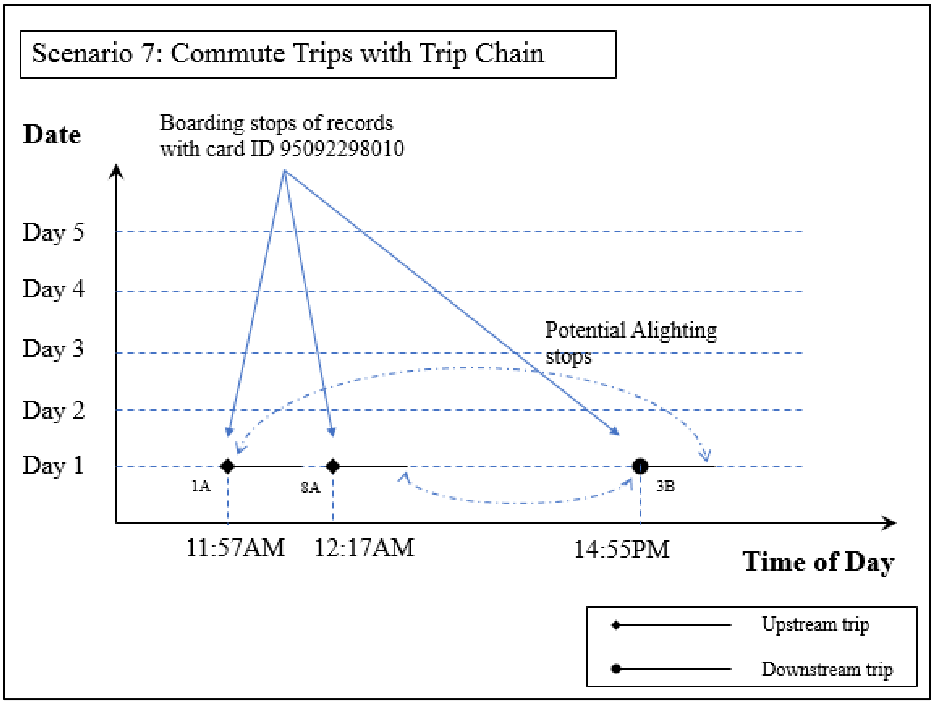

Figure 5.7: Commute Trips with Trip Chain Scenario

5.3 Transaction Record Matching

As mentioned before, the alighting information is necessary to detect transit riders' OD information. In order to identify alighting information, Barry's (Barry et al., 2002) assumptions are used here to help mine the potential information from the inadequate database. Based on the analysis of seven scenarios in this chapter, the records could be matched based on the following criteria:

The criteria for the records matched as 'Commute':

Same 'Card ID', same 'Transaction date', different 'Direction';

Same 'Card ID', different 'Transaction date', different 'Time period', and different 'Direction';

The criteria for the records matched as 'Transfer':

Same 'Card ID', same 'Transaction date', same 'Direction' or different 'Route', and same 'Time period'.

For the other situation, match the record as 'No matching'.

Table 5.8 shows the sample of the transaction record matching results. There are two new columns (Transfer Matching Results and Commute Matching Results) that indicate the transaction ID of matching records for the current record.

Table 5.8: Transaction Records Matching Results

Route ID	Transaction ID	Card ID	Bus ID	Transaction Date	Transaction Time	Period	Direction	Boarding Stop	Transfer Matching Results	Commute Matching Results
RouteD	1	90000080210	3008405	2015/1/1	9:40:38	AM	B	1B	N/A	N/A
RouteD	2	90000080210	1600615	2015/1/1	18:28:41	PM	B	6B	N/A	N/A
RouteB	3	90000128240	80215004	2015/1/2	23:06:50	PM	A	24A	N/A	N/A
RouteB	4	90000128240	78004510	2015/1/3	21:43:26	PM	B	1B	N/A	N/A
RouteC	5	90000207730	76001211	2015/1/1	19:15:03	PM	A	16A	N/A	N/A
RouteA	6	90000207730	37064090	2015/1/2	15:33:01	PM	B	1B	N/A	N/A
RouteB	7	90000306440	3000585	2015/1/1	16:26:32	PM	A	3A	N/A	N/A
RouteB	8	90000306440	3008947	2015/1/3	16:17:35	PM	A	7A	N/A	N/A
RouteD	9	90000312530	75001660	2015/1/2	15:08:47	PM	A	9A	N/A	N/A
RouteD	10	90000312530	75001660	2015/1/2	17:41:20	PM	B	1B	N/A	N/A
RouteD	11	90000414560	16064050	2015/1/4	18:52:44	PM	B	1B	N/A	N/A
RouteD	12	90000414560	4210117	2015/1/5	18:44:59	PM	B	2B	N/A	N/A
RouteD	13	90000418850	3008405	2015/1/4	7:38:39	AM	A	2A	N/A	14
RouteD	14	90000418850	4210117	2015/1/5	19:47:33	PM	B	N/A	N/A	13
RouteA	15	90000434960	5500490	2015/1/2	9:54:18	AM	A	4A	N/A	16
RouteA	16	90000434960	9500496	2015/1/2	15:39:25	PM	B	22B	N/A	15
RouteB	17	98497328700	9800193	2015/1/1	17:48:29	PM	B	4B	N/A	18,19,20
RouteB	18	98497328700	75063070	2015/1/2	6:53:30	AM	A	1A	N/A	17
RouteB	19	98497328700	75063070	2015/1/3	6:49:58	AM	A	1A	N/A	17
RouteB	20	98497328700	9800193	2015/1/4	6:50:57	AM	A	1A	N/A	17
RouteD	21	98509428970	70847004	2015/1/1	7:43:21	AM	A	2A	N/A	24
RouteD	22	98509428970	60100304	2015/1/2	9:18:52	AM	A	3A	N/A	24
RouteD	23	98509428970	4210117	2015/1/2	21:56:45	PM	A	16A	N/A	N/A
RouteD	24	98509428970	77008740	2015/1/4	22:07:06	PM	B	11B	N/A	21,22

5.4 Alighting Stop Information Identification

Based on the transaction record matching results and the scenarios discussed earlier in this chapter, the alighting stop information on each record could be mined. For the transfer matching records, the alighting stop identification follows the trip chaining theory as discussed in the literature review section. For the commute matching records, if a record has only one match, it can be inferred that the boarding stop of one record is the alighting stop of the other. For the record with multiple matches, the selection criteria are listed below:

The first level of selection is: Select the matched records with the transaction occurred in the same day. Based on the assumption mentioned before, such passengers would most likely go back to their origin of the first trip of the day.

The second level of selection is: If there is no matched record occurred in the same day, then select the record with the transaction occurred during peak hour since the activity during peak hour will likely exhibit the regular patterns.

The third level of selection is: If there is no matched record in the same day and during peak hour, then select the record with the boarding stop that is nearest to the terminal station.

Based on the record matching results and the matching record selection results discussed above, the alighting stop information identification method is applied to the study case.

Table 5.9 presents a sample of alighting stop identification results and note that a new column which represents the alighting stop identification results of each record is added.

Table 5.10: Alighting Stop Identification Results (Sample)

Route ID	Transaction ID	Card ID	Bus ID	Transaction Date	Transaction Time	Period	Direction	Boarding Stop	Alighting Stop
RouteD	1	90000080210	3008405	1/1/2015	9:40:38	AM	B	1B	N/A
RouteD	2	90000080210	1600615	1/1/2015	18:28:41	PM	B	6B	N/A
RouteB	3	90000128240	80215004	1/2/2015	23:06:50	PM	A	24A	N/A
RouteB	4	90000128240	78004510	1/3/2015	21:43:26	PM	B	1B	N/A
RouteC	5	90000207730	76001211	1/1/2015	19:15:03	PM	A	16A	N/A
RouteA	6	90000207730	37064090	1/2/2015	15:33:01	PM	B	1B	N/A
RouteB	7	90000306440	3000585	1/1/2015	16:26:32	PM	A	3A	N/A
RouteB	8	90000306440	3008947	1/3/2015	16:17:35	PM	A	7A	N/A
RouteD	9	90000312530	75001660	1/2/2015	15:08:47	PM	A	9A	N/A
RouteD	10	90000312530	75001660	1/2/2015	17:41:20	PM	B	1B	N/A
RouteD	11	90000414560	16064050	1/4/2015	18:52:44	PM	B	1B	N/A
RouteD	12	90000414560	4210117	1/5/2015	18:44:59	PM	B	2B	N/A
RouteD	13	90000418850	3008405	1/4/2015	7:38:39	AM	A	2A	N/A
RouteD	14	90000418850	4210117	1/5/2015	19:47:33	PM	B	N/A	2A
RouteA	15	90000434960	5500490	1/2/2015	9:54:18	AM	A	4A	22B
RouteA	16	90000434960	9500496	1/2/2015	15:39:25	PM	B	22B	4A
RouteB	17	98497328700	9800193	1/1/2015	17:48:29	PM	B	4B	1A
RouteB	18	98497328700	75063070	1/2/2015	6:53:30	AM	A	1A	4B
RouteB	19	98497328700	75063070	1/3/2015	6:49:58	AM	A	1A	4B
RouteB	20	98497328700	9800193	1/4/2015	6:50:57	AM	A	1A	4B
RouteD	21	98509428970	70847004	1/1/2015	7:43:21	AM	A	2A	11B
RouteD	22	98509428970	60100304	1/2/2015	9:18:52	AM	A	3A	11B
RouteD	23	98509428970	4210117	1/2/2015	21:56:45	PM	A	16A	N/A
RouteD	24	98509428970	77008740	1/4/2015	22:07:06	PM	B	11B	2A

5.5 Summary

This chapter describes the alighting information identification methodology step by step. All activities conducted in the sections of this chapter aim to minimize the potential errors. The developed alighting information identification methodology is very general and can be used to detect destination information based on the SC transaction data, both with and without other support databases.

Chapter 6. Numerical Results Analysis

6.1 Introduction

This chapter presents the numerical results of analyses based on the boarding location information and alighting location information extracted before. The following sections are organized as follows. Section 6.2 shows the OD matrix achieved by applying the integrated approach. Section 6.3 presents the analysis of transit temporal usage pattern. Section 6.4 discusses the analysis of transit spatial usage pattern. Section 6.5 describes the analysis of the trip characteristics of different user groups. Finally, section 6.6 concludes this chapter with a summary.

6.2 OD Matrices

With the analysis of the boarding and alighting information as discussed before, the OD matrices on each route can be built up to help analyze the transit network characteristics. The detailed OD matrices are shown in Appendix B.

6.3 Analysis of Temporal Usage Pattern

Based on the trip direction identification results, it is possible to calculate the average passenger counts and operation time of each bus trip during each time period. The results are excluded if there is only one boarding cluster in a whole trip since it will be impossible to calculate the operation time for such records. The passenger boarding counts during each period are also presented in the chart. The results of passenger counts indicate that the crest values of boarding activity occur during different peak hours. During the AM period, the highest volume occurs at 8–9 a.m. and the top three periods are 8–9, 10–11 and 9–10 a.m. During the PM period, the highest volume occurs at 5–6 p.m. and the top three periods are 5–6, 6–7, and 4–5 p.m. The reason behind this could be explained as follows: Most of government institutions and enterprises begin their work around 9 a.m. and finish their work around 6 p.m. Therefore, the travel patterns of citizens in Guangzhou, China follow exactly the same. This result is also consistent with previous studies (Liu et al. 2009, Tao et al., 2014).

Furthermore, the two charts in Figure 6.1 also indicate that the operation time will increase as the passenger boarding counts increase.

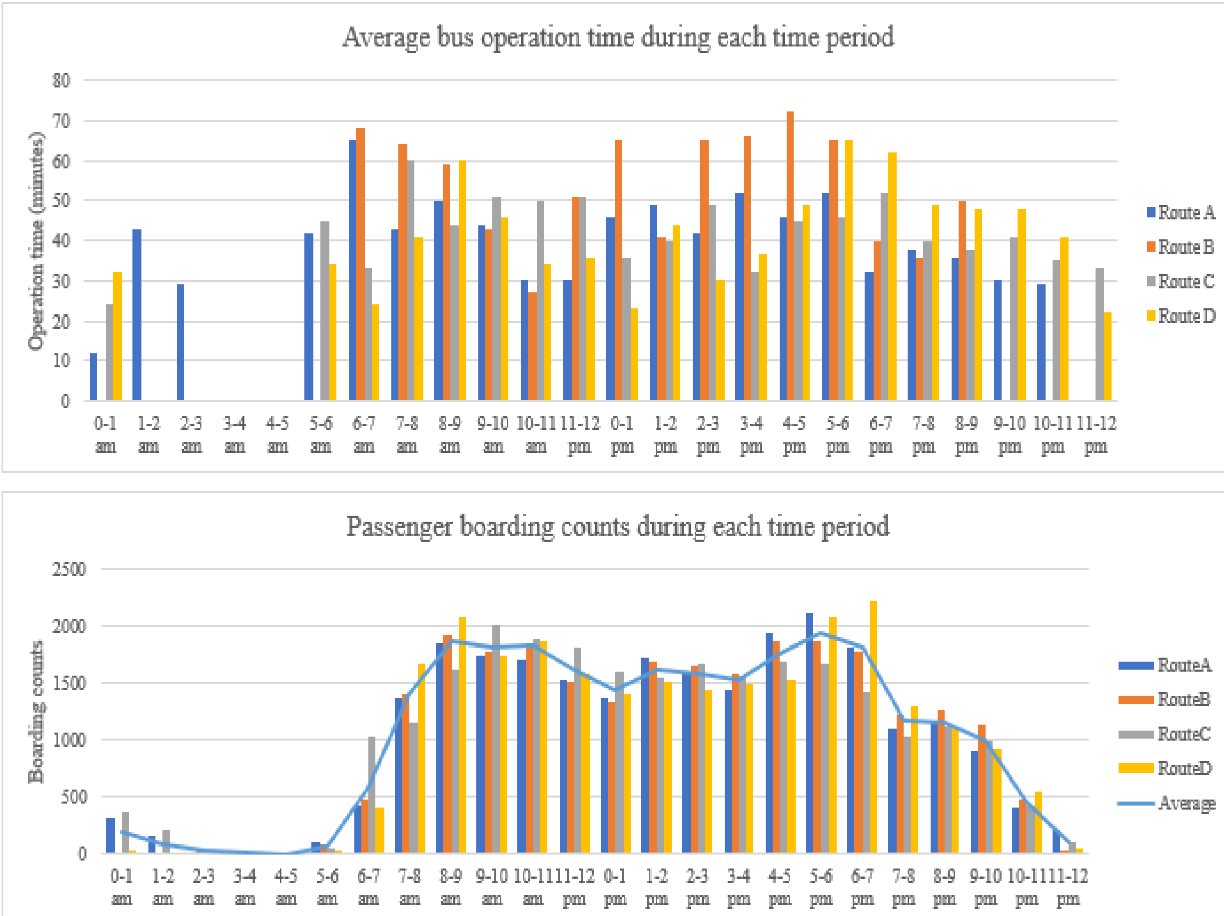

Figure 6.1: Boarding Counts and Operation Time during Each Time Period

6.4 Analysis of Spatial Usage Pattern

Based on the boarding stop identification results of each record, it is also possible to estimate the frequency of boarding activities at each stop. As the results show in Figure 6.2-6.5 below, most boarding activities occur at the first several stops, and the passenger boarding counts decrease as the bus stops get closer to the terminal. The passenger boarding activities rarely occur at last several stops. The reason for such high passenger volume occurrence at first stop could be explained as the result of the assumption of labeling the boarding stop of first boarding cluster as stop 1. The decreasing trend at last several stops is also consistent with previous studies (Liu et al., 2009). That clearly indicates both the travel habits of local passengers and the passenger boarding location characteristics from the traffic network point of view.

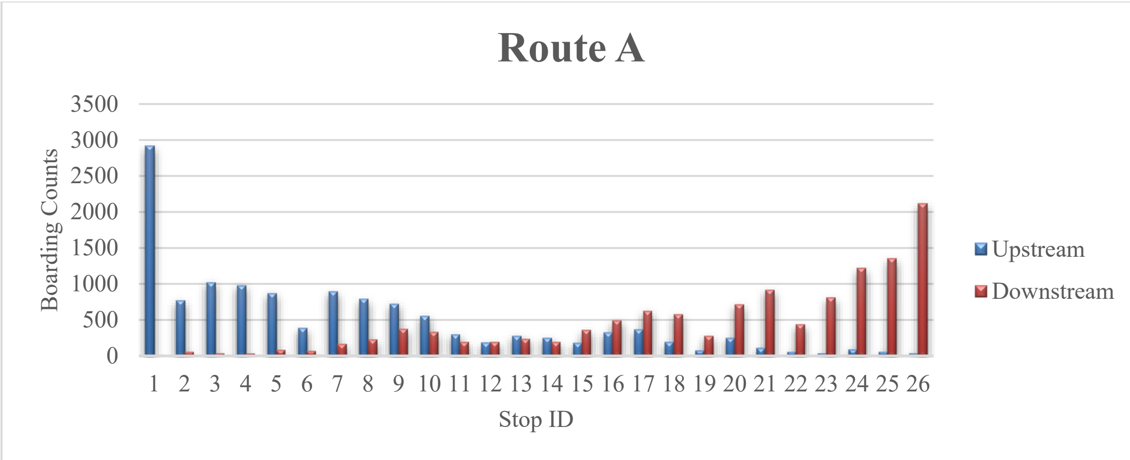

Figure 6.2: Frequency of Passengers' Boarding Activities on Route A

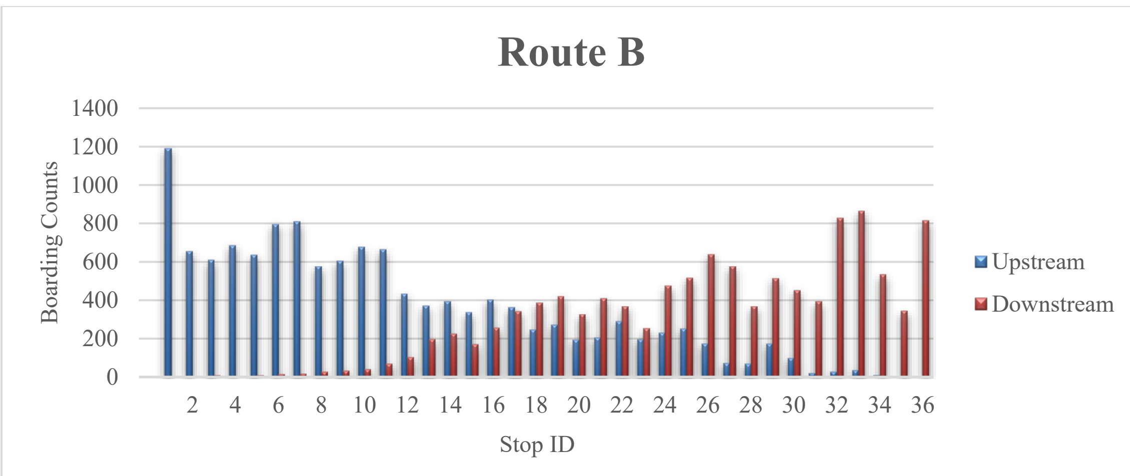

Figure 6.3: Frequency of Passengers' Boarding Activities on Route B

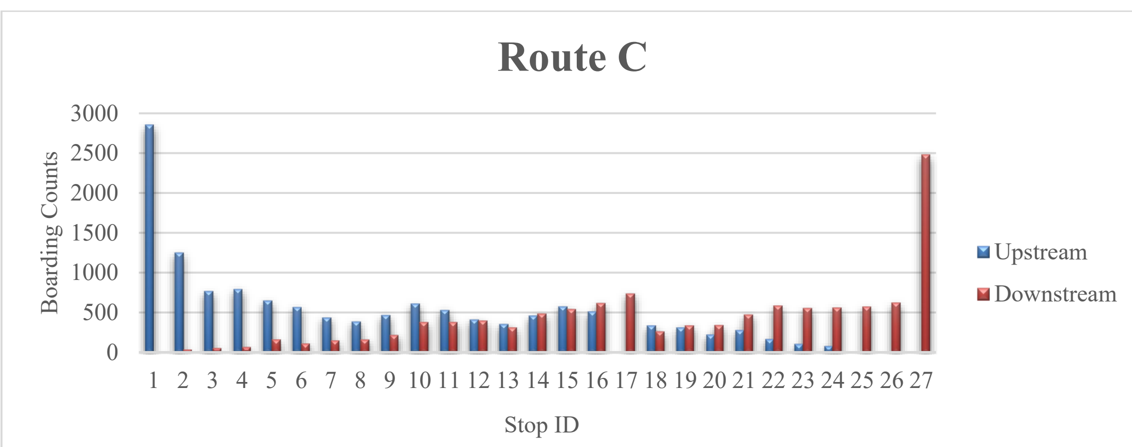

Figure 6.4: Frequency of Passengers' Boarding Activities on Route C

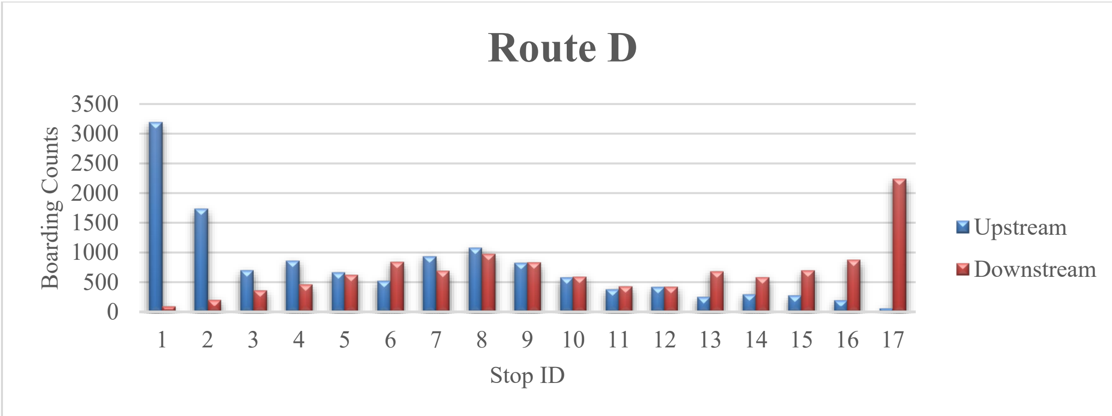

Figure 6.5: Frequency of Passengers' Boarding Activities on Route D

6.5 Analysis of Trip Characteristics of Different Group of Users

Based on the information on each record based on the card type, it is possible to analyze the trip characteristics of each group of users. The passenger boarding counts of each group during each period are presented in Figure 6.6-6.9.

6.5.1 Regular Card Users' Activity Pattern

Figure 6.6 presents the passenger boarding counts of the regular card users on each route. The frequency distribution reveals that regular card user group demonstrated notable peak-hour patterns. This result is also consistent with previous studies (Tao et al., 2014)

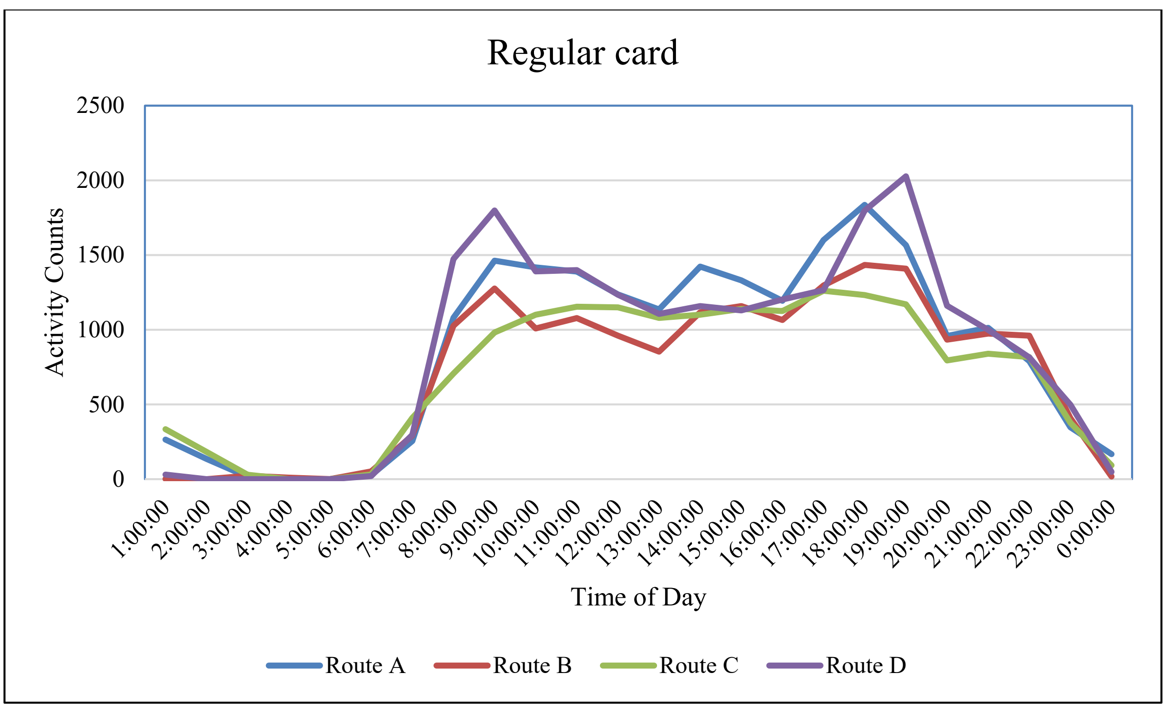

Figure 6.6: Frequency of Regular Card Users' Activities

6.5.2 Senior Card Users' Activity Pattern

Figure 6.7 presents the passenger boarding counts of the senior card users on each route. In comparison with the regular card users, the boarding pattern of seniors has no pronounced peak-hour patterns throughout the day. However, the peak period of the 9:00-10:00 AM in the figure is still consistent with previous studies (Tao et al., 2014).

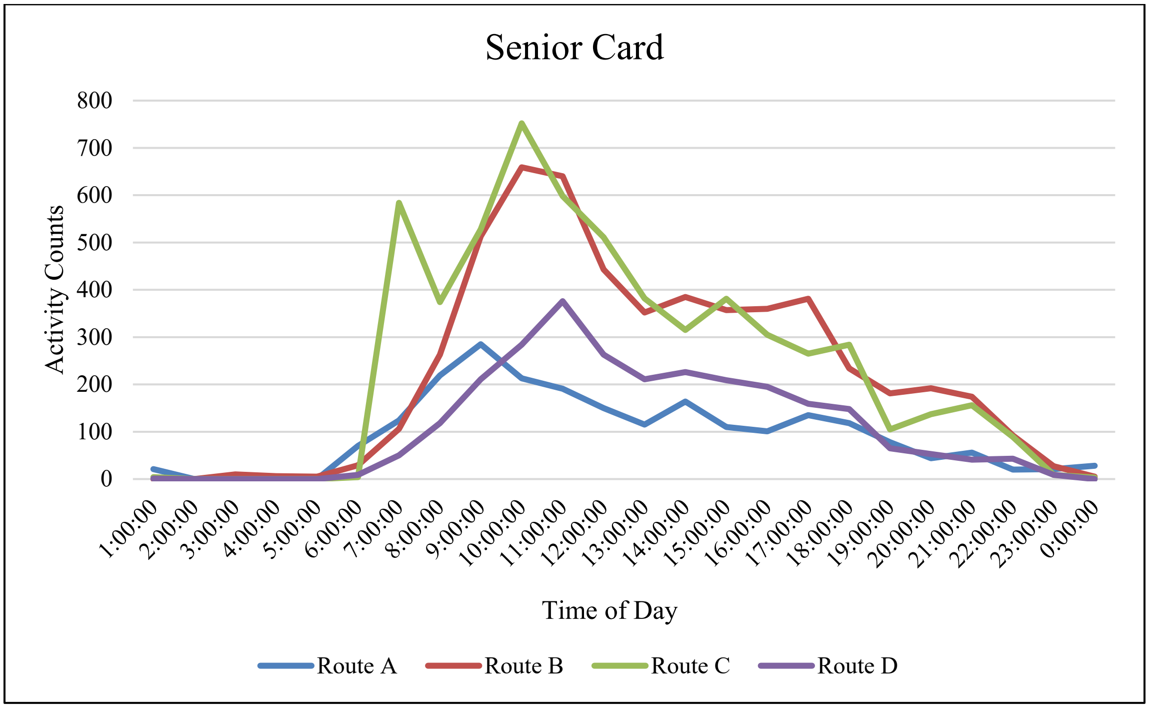

Figure 6.7: **Frequency of Senior Card Users' Activities**

6.5.3 Student Card Users' Activity Pattern

Figure 6.8 presents the passenger boarding counts of the senior card users on each route. In contrast with the regular card users, the boarding pattern of student has pronounced multi-peak period patterns throughout the day and the highest boarding frequency occurs during PM peak hour (which is 18:00-20:00). It could be explained as school students still have a tight daily schedule similar to the regular card user group. However, some of them have more travel activities at noon and require more mobility than the regular card users, which makes the proportion of their activity at noon higher than regular card users.

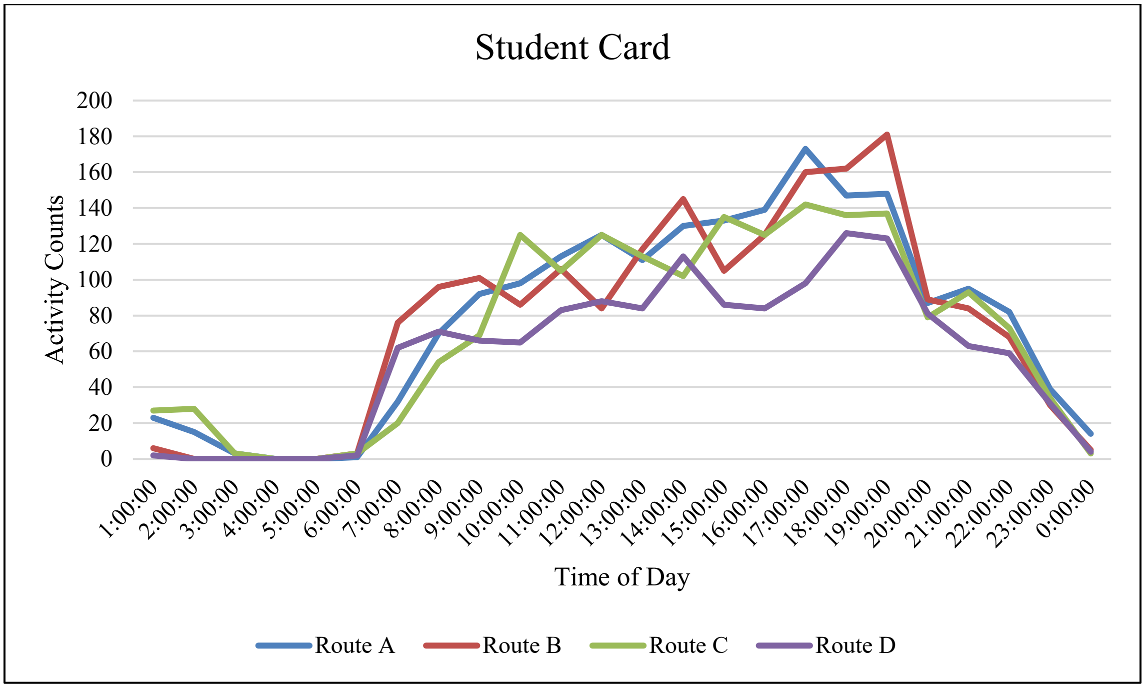

Figure 6.8: Frequency of Student Card Users' Activities

6.5.4 Disable Card Users' Activity Pattern

Figure 6.9 presents the passenger boarding counts of the disable card users on each route. In contrast with the regular card users, the boarding patterns of the disable card users have no pronounced peak-hour patterns throughout the day.

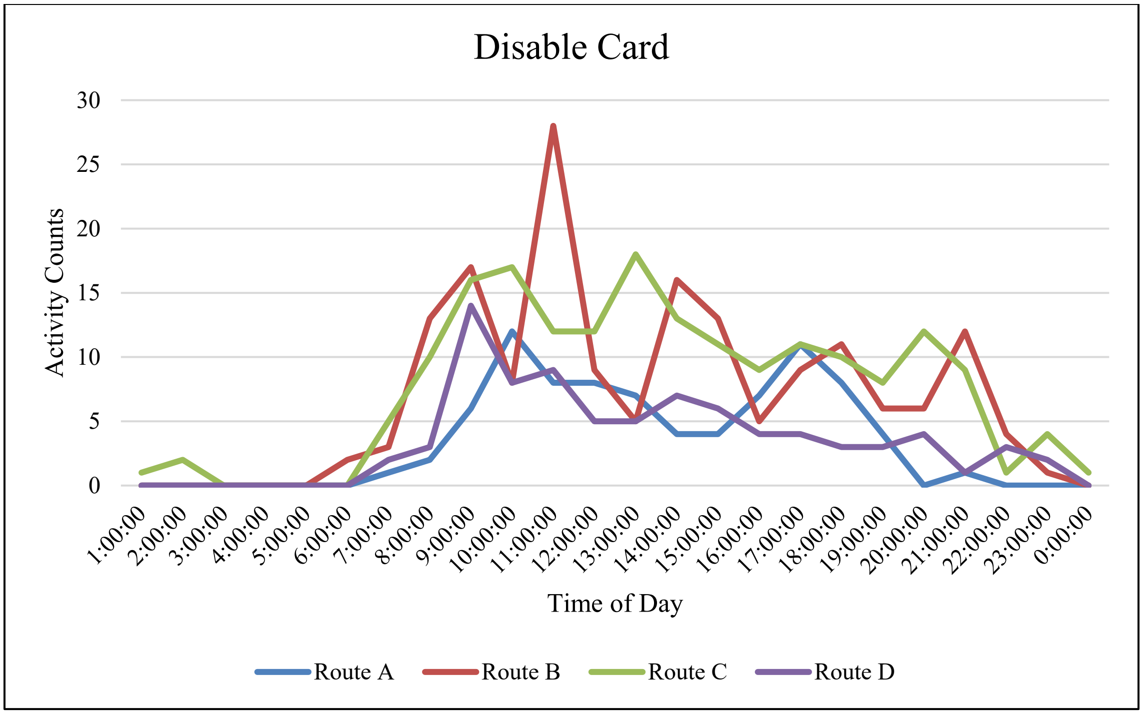

Figure 6.9: Frequency of Disable Card Users' Activities

6.6 Summary

This chapter summarizes the numerical results based on the approaches developed and presented in previous chapters, including the OD matrices on different routes, the average bus operation time and boarding activity counts during each time period, boarding activity counts at each stop on different routes, and the frequency of activity of each user group.

Chapter 7. Summary and Conclusions

7.1 Introduction

The chapter summarizes the results achieved from this study. The following sections are organized as follows. Section 7.2 describes the summary of the key results. Section 7.3 draws conclusions and also gives the future research directions.

7.2 Summary of Key Results

With the applications of the algorithms as developed and used to process the smart card data collected from Guangzhou, China, the boarding and alighting information are successfully identified in this study. The information gathered as results of this study can be summerized as follows:

Direction information. The direction information is important in mining the smart card data. With the help of bus direction information, the time difference between the first and last records can be calculated, which represents the actual travel time from the first stop to the last stop with boarding records/activities.

Boarding cluster information. With the help of boarding cluster identification, the transaction records for a whole trip along a route are divided to different cluster groups in order to identify the transactions occurred at different stops. The boarding cluster identification results can greatly reduce potential errors involved in the boarding stop identification and boarding passenger count estimation.

Boarding stop information. The boarding stop information is one of the key findings in this research. Additional information about passenger boarding counts at each stop can be mined and analyzed.

Alighting stop information. Finally, the results show the estimated alighting location of each transaction records. Based on the boarding and alighting information of each record, the OD matrix is built up for each route.

7.3 Conclusions and Future Research Directions

It is very challenging for one to conduct the OD estimation if there is no boarding location information recorded by the AFC system. This research aims to develop a methodology to extract the boarding and alighting location information using the available transaction records with only basic route information, transaction time and few transfer activities, but without GPS, passenger counts, and survey data. To reduce errors, the algorithms that are developed are used to process the data in the order of identification of boarding direction, boarding clusters, boarding stops and alighting stops. The boarding and alighting stop information are successfully extracted.

The methodology developed and results obtained in this study can be helpful for the OD estimation related work in the real world. However, with the limited amount of SCD, the network level OD matrix extraction framework is not discussed in this study. In the future, the transfer activities could be mined from the database if it contains information about more routes and for longer time periods. The cluster analysis can also be conducted to reveal passengers' travel patterns with the help of survey data, card type, and land use data. Furthermore, the boarding and alighting stop information can also be validated based on the real-world data if available.

References

1. Agard, B., Morency, C. and Trépanier, M. (2007), "Mining Public Transport User Behaviour from Smart Card Data." *IFAC Proceeding Volumes*, 39, pp. 399-404.
2. Alsger, A.A., Mesbah, M., Ferreira, L. and Safi, H. (2015), "Use of Smart Card Fare Data to Estimate Public Transport Origin–Destination Matrix." *Transportation Research Record: Journal of the Transportation Research Board*, 2535, pp. 88-96.
3. Attoh-Okine, N. O. and Shen, L. (1995), "Security Issues of Emerging Smart Cards Fare Collection Application in Mass Transit." Proceedings of In Vehicle Navigation and Information Systems Conference, July 30 – Aug 2, 1995, Seattle, WA.
4. Bagchi, M. and White, P. (2004), "What Role for Smart-Card Data from Bus Systems," *Municipal Engineer*, 157(1), pp.39-46.
5. Barry, J., Freimer, R. and Slavin, H. (2009), "Use of Entry-only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City." *Transportation Research Record: Journal of the Transportation Research Board*, 2112, pp. 53-61.
6. Barry, J., Newhouser, R., Rahbee, A. and Sayeda, S. (2002), "Origin and Destination Estimation in New York City with Automated Fare System Data." *Transportation Research Record: Journal of the Transportation Research Board*, 1817, pp. 183-187.
7. Chang, Y. and Zhao, C. (2016), "Travel Pattern Recognition Using Smart Card Data in Public Transit." *International Journal of Emerging Engineering Research and Technology*, 4(7), pp. 6-13.
8. Chen, Z. and Fan, W. (2018), "Extracting Bus Transit Boarding Stop Information Using Smart Card Transaction Data." *Journal of Modern Transportation*, 26(3), pp. 209-219.
9. Cottrill, C. (2009), "Approaches to Privacy Preservation in Intelligent Transportation Systems and Vehicle-Infrastructure Integration Initiative." *Transportation Research Record: Journal of the Transportation Research Board*, 2129, pp. 9-15.
10. Cui, A. (2006), Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems, Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.
11. Cunningham, R. F. (1993), Smart Card Applications in Integrated Transit Fare, Parking Fee and Automated Toll Payment Systems-The MAPS Concept. Conference Proceedings National Telesystems Conference, June 16-17, 1993, Atlanta, GA.
12. Dempsey, P. S. (2015), Privacy Issues with the Use of Smart Cards. Available at: http://dx.doi.org/10.2139/ssrn.2692652
13. Devillaine, F., Munizaga, M. and Trépanier, M. (2012), "Detection of Activities of Public Transport Users by Analyzing Smart Card Data." *Transportation Research Record: Journal of the Transportation Research Board*, 2276, pp. 48-55.
14. Gordon, J., Koutsopoulos, H., Wilson, N. and Attanucci, J. (2013), "Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data." *Transportation Research Record: Journal of the Transportation Research Board*, 2343, pp. 17-24.
15. Heydt-Benjamin, T., Chae, H. J., Defend, B. and Fu, K. (2006), *Privacy for Public Transportation. In Privacy Enhancing Technologies*. International Workshop on Privacy Enhancing Technologies, Springer, Berlin, Heidelberg.

16. Hofmann, M. and O'Mahony, M. (2005), "Transfer Journey Identification and Analyses from Electronic Fare Collection Data." Intelligent Transportation Swystems, September 16-16, 2005, Vienna, Austria.
17. Jang, W. (2010), "Travel Time and Transfer Analysis Using Transit Smart Card Data." *Transportation Research Record: Journal of the Transportation Research Board*, 2144, pp. 142-149.
18. Jánošíková, Ľ., Slavík, J. and Koháni, M. (2014), "Estimation of A Route Choice Model for Urban Public Transport Using Smart Card Data." *Transportation Planning and Technology*, 37(7), pp. 638-648.
19. Kieu, L., Bhaskar, A. and Chung, E. (2013), "Mining Temporal and Spatial Travel Regularity for Transit Planning." Proceeding of 36th Australasian Transport Research Forum , 2013, Brisbane, Queensland, Australia
20. Kusakabe, T. and Asakura, Y. (2014), "Behavioural Data Mining of Transit Smart Card Data: A Data Fusion Approach." *Transportation Research Part C: Emerging Technologies*, 46, pp. 179-191.
21. Lessard, R. (1993), "The Use of Computer for Urban Transit Operations." Vehicle Navigation and Information Systems Conference, October 12-15, 1993, Ottawa, Canada
22. Liu, L., Hou, A., Biderman, A., Ratti, C. and Chen, J. (2009), "Understanding Individual and Collective Mobility Patterns from Smart Card Records: A Case Study in Shenzhen." 12th International IEEE Conference on Intelligent Transportation Systems, October 4-7, 2009, St. Louis, MO, USA
23. Long, Y., Sun, L. and Sui, T. (2015), "A Review of Urban Studies Based on Transit Smart Card Data." In Urban Planning Forum, Vol. 3, No. 10, pp. 70-77.
24. Long, Y., Liu, X. and Zhou, J. (2016), "Early Birds, Night Owls, and Tireless/Recurring Itinerants: An Exploratory Analysis of Extreme Transit Behaviors in Beijing, China." Habitat International 57, pp. 223-232
25. Ma, X., Wang, Y., Chen, F., and Liu, J. (2012), "Transit Smart Card Data Mining for Passenger Origin Information Extraction." *Journal of Zhejiang University-Science C*, 13(10), pp.750-760.
26. Ma, X., Wu, Y., Wang, Y., Chen, F. and Liu, J. (2013), "Mining Smart Card Data for Transit Riders' Travel Patterns." *Transportation Research Part C: Emerging Technologies*, 36, pp. 1-12.
27. Ma, X. and Wang, Y. (2014), "Development of a Data-Driven Platform for Transit Performance Measures Using Smart Card and GPS Data." *Journal of Transportation Engineering*, 140(12), pp.04014063.
28. Mayes, K. E. and Markantonakis, K. (2008), *Smart Cards, Tokens, Security and Applications* Springer, New York.
29. Munizaga, M.A. and Palma, C. (2012), "Estimation of a Disaggregate Multimodal Public Transport Origin–Destination Matrix from Passive Smartcard Data from Santiago, Chile." *Transportation Research Part C: Emerging Technologies*, 24, pp.9-18.
30. Nassir, N., Khani, A., Lee, S., Noh, H. and Hickman, M. (2011). Transit Stop-Level Origin–Destination Estimation Through Use of Transit Schedule and Automated Data Collection System. *Transportation Research Record: Journal of the Transportation Research Board*, 2263, pp. 140–150.
31. O'Connor, K. (1994), "Smart Cards, Privacy Issues." *Journal of Law and Information Science*, 5, pp. 245.

32. Pelletier, M.P., Trépanier, M. and Morency, C. (2011), "Smart Card Data Use in Public Transit: A Literature Review." *Transportation Research Part C: Emerging Technologies*, 19(4), pp. 557-568.
33. Qian, Y., Xu, L. and Ye, P. (2015), "A Method of Data Cleaning Based on IC Card and GPS Data in Chengdu." Fifth International Conference on Transportation Engineering, September 26-27, 2015, Dalian, China
34. Robinson, S., Narayanan, B., Toh, N. and Pereira, F. (2014), "Methods for Pre-Processing Smartcard Data to Improve Data Quality." *Transportation Research Part C: Emerging Technologies*, 49, pp.43-58.
35. Seaborn, C., Attanucci, J. and Wilson, N. (2009), "Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data." *Transportation Research Record: Journal of the Transportation Research Board*, 2121, pp. 55-62.
36. Siwek, S. J. (1998), Transportation Planning and ITS: Putting the Pieces Together. FHWA. Available at: https://www.fhwa.dot.gov/publications/research/operations/its/transplanits.cfm.
37. Smart Card Alliance. *About Smart Cards: Applications: Transportation.* Available at: http://www.smartcardalliance.org/smart-cards-applications-transportation/#smart-cards-and-transit.
38. Smart Card Alliance. *Smart Card Standards and Specifications*. Available at: http://www.smartcardalliance.org/smart-cards-intro-standards/?redirect=http%3A%2F%2Fwww.smartcardalliance.org%2Fsmart-cards-intro-standards#isoiec-standards.
39. Song, Z. (2016), "Research on Large Scale OD Matrix Estimation Method Based on Bus IC Card Data." *Application Research of Computers*, 33(7), pp. 2007-2013.
40. Stromberg, R. E. (1995), U.S. Patent No. 5,450,051. Washington, DC: U.S. Patent and Trademark Office.
41. Tantiyanugulchai, S. and Bertini, R. (2003), "Arterial Performance Measurement Using Transit Buses as Probe Vehicles." Proceedings of the IEEE International Conference on Intelligent Transportation Systems, October 12-15, 2003, Shanghai, China
42. Tao, S., Rohde, D. and Corcoran, J. (2014), "Examining the Spatial–Temporal Dynamics Of Bus Passenger Travel Behaviour Using Smart Card Data and the Flow-Comap." *Journal of Transport Geography*, 41, pp. 21–36.
43. Trépanier, M., Tranchant, N. and Chapleau, R. (2007), "Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System." *Journal of Intelligent Transportation Systems*, 11(1), pp. 1-14.
44. Wang, W., Attanucci, J. P. and Wilson, N. H. (2011), "Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems." *Journal of Public Transportation*, 14(4), pp. 131-150.
45. Yu, Y., Deng, T. and Xiao, Y. (2009), "A Novel Method of Confirming the Boarding Station of Bus Holders." *Journal of Chongqing Jiaotong University*, 28(1), pp. 121-125.
46. Zhao, J., Wilson, N.H. and Rahbee, A. (2006), "Estimating Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection." *Computer-Aided Civil and Infrastructure Engineering*, 22, pp. 376-387.
47. Zhang, L., Zhao, S., Zhu, Y. and Zhu, Z. (2007), "Study on the Method of Constructing Bus Stops OD Matrix Based on IC Card Data." International Conference on Wireless Communications, Networking and Mobile Computing, September 21-25, 2007, Shanghai, China

Appendix A: Travel Time between Stops

Table A.1: Route A Upstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Kangwang Lu (Shangxiajiu) Bus Terminal	0	0
2	Hualin Temple	676	176
3	Kangwang Zhonglu	161	23
4	Chen Clan Academy	645	197
5	Liwan Lu Entrance	508	126
6	Shiluji	379	107
7	Zhongshan 8 Lu	158	75
8	Qiaozhong	1800	469
9	Guangfo Lu Entrance (Jiaokou Coach Station)	1800	261
10	Guangfo Lu East	730	210
11	Huangqi Downtown	505	125
12	Hongwei Lu Entrance	415	60
13	Hongwei Lu	324	47
14	Poyang Lu (Jingxiutang Pharmaceutical Factory)	400	110
15	Huangqi Administration Service Center	290	68
16	Shimen Middle School Entrance	466	68
17	Shamian Xincheng	1300	188
18	Jiaoyu Lu Entrance	729	184
19	Jianshe Dadao	623	116
20	Michong	344	50
21	Shaxi	2000	394
22	Baisha Town Hall	652	94
23	Xinglian	372	80
24	Baisha	375	80
25	Sanjun Market	1200	200
26	Baisha (Zhonghai Jingshanwan) Bus Terminal	906	183
Total		17758	3692

Table A.2: Route A Downstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Baisha (Zhonghai Jingshanwan) Bus Terminal	0	0
2	Sanjun Market	1700	374
3	Baisha	1300	239
4	Xinglian	362	78
5	Baisha Town Hall	415	86
6	Shaxi	575	83
7	Michong	2000	417
8	Jianshe Dadao	550	80
9	Jiaoyu Lu Entrance	427	62
10	Shamian Xincheng	851	200
11	Shimen Middle School Entrance	1100	159
12	Huangqi Administration Service Center	500	123
13	Poyang Lu (Jingxiutang Pharmaceutical Factory)	237	85
14	Hongwei Lu	471	68
15	Hongwei Lu Entrance	338	100
16	Huangqi Downtown	356	52
17	Guangfo Lu East	471	68
18	Guangfo Lu Entrance (Jiaokou Coach Station)	882	230
19	Qiaozhong	1900	301
20	Zhongshan 8 Lu	1800	312
21	Shiluji	327	149
22	Liwan Lu Entrance	254	88
23	Chen Clan Academy	554	131
24	Kangwang Zhonglu	478	120
25	Hualin Temple	332	99
26	Kangwang Lu (Shangxiajiu) Bus Terminal	534	128
Total		18714	3834

Table A.3: Route B Upstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Bus Terminal at Guangzhou Railway Station	0	0
2	Zhanqian Lu	1200	267
3	Xizhan	511	121
4	Xihua Luwei	722	151
5	Caihongqiao	685	192
6	Liwan Lu	422	61
7	Zhongshan 8 Lu	1000	238
8	Liwanhu Park West Entrance	519	75
9	Ruyifang	733	106
10	Huangsha Dadao	799	116
11	Fangcun	1900	322
12	Xiafangcun	742	154
13	Dachongkou	513	121
14	Hedong Xincun	534	124
15	Guangzhong Pier	488	71
16	True Light High School (Fusheng Garden)	578	130
17	Peiying Middle School	630	91
18	Baihedong	537	78
19	Hedong Lu West	357	52
20	Xilang	500	72
21	Maicun	335	95
22	Yuwei Bridge East	416	60
23	Huadi Dadao (Hainancun)	737	153
24	Huadi Dadao South	415	60
25	Shawei Bridge	750	155
26	Pingzhou Interchange	1200	174
27	Taiping Lu Entrance	406	105
28	Guicheng Elderly Center	417	107
29	Pingzhou Hotel	267	39
30	Pingxi Lu Entrance	2200	598
31	Guicheng Traffic Management Bureau	454	112
32	Nanhai No. 2 People's Hospital	423	61
33	Kongxi	394	104
34	Pingzhou Bus Terminal	1000	238
Total		22784	4604

Table A.4: Route B Downstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Pingzhou Bus Terminal	0	0
2	Nanhai No. 2 People's Hospital	1000	230
3	Guicheng Traffic Management Bureau	433	63
4	Pingxi Lu Entrance	284	84
5	Bengchong	393	100
6	Guangdong General Hospital	2200	425
7	Pingzhou Jade Street	309	45
8	Pingzhou Hotel	529	77
9	Guicheng Elderly Center	444	107
10	Taiping Lu Entrance	507	116
11	Pingzhou Interchange	188	70
12	Shawei Bridge	1100	159
13	Huadi Dadao South	509	116
14	Huadi Dadao (Hainancun)	649	137
15	Yuwei Bridge East	624	90
16	Maicun	339	49
17	Xilang	281	83
18	Hedong Lu West	689	100
19	Baihedong	437	63
20	Peiying Middle School	731	149
21	Guangzhong Pier	885	171
22	Hedong Xincun	650	94
23	Dachongkou	486	113
24	Xiafangcun	541	78
25	Fangcun	820	161
26	Huangsha Dadao	2100	304
27	Ruyifang	610	131
28	Liwanhu Park West Entrance	489	71
29	Zhongshan 8 Lu	837	121
30	Liwan Lu	989	271
31	Caihongqiao	477	112
32	Dongfeng Xilu	1000	230
33	Xiyuan	484	113
34	Xizhan	402	143
35	Zhanqian Lu	867	211
36	Bus Terminal at Guangzhou Railway Station	1600	338
Total		24883	4927

Table A.5: Route C Upstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Suiyan Lu (Majestic Garden) Bus Terminal	0	0
2	Yaju Lanwan	780	153
3	Sanyanqiao (Guangdong Grain & Oil Market)	744	187
4	Hubang Market	695	140
5	Yanbu Hospital	1000	224
6	Hubang Lu Entrance	568	122
7	Lian'an	878	167
8	Yanbu Dadao North	223	72
9	Yanbu Entrance	512	114
10	Dazhuanwan (Huangqi Kaimin Tea Market)	490	71
11	Jiazhou Plaza	986	183
12	Hongwei Lu Entrance	554	160
13	Huangqi Downtown	344	90
14	Guangfo Lu East	489	150
15	Guangfo Lu Entrance (Jiaokou Coach Station)	884	168
16	Qiaozhong	1900	275
17	Zhongshan 8 Lu	1800	261
18	Shiluji	158	63
19	Liwan Lu	784	193
20	Caihongqiao	471	108
21	Xiyuan	1500	336
22	Xizhan	413	100
23	Zhanqian Lu	863	204
24	Zhannan Lu Bus Terminal	468	108
Total		17504	3648

Table A.6: Route C Downstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Zhannan Lu Bus Terminal	0	0
2	Military Hospital	506	73
3	Zhanqian Henglu	590	133
4	Zhanqian Lu	458	114
5	Xizhan	530	124
6	Xihua Luwei	680	146
7	Caihongqiao	680	194
8	Liwan Lu	498	120
9	Liwan Lu Entrance	325	47
10	Shiluji	355	99
11	Zhongshan 8 Lu	175	73
12	Qiaozhong	1800	451
13	Guangfo Lu Entrance (Jiaokou Coach Station)	1800	261
14	Guangfo Lu East	730	201
15	Huangqi Downtown	505	121
16	Hongwei Lu Entrance	415	60
17	Jiazhou Plaza	683	194
18	Dazhuanwan (Huangqi Kaimin Tea Market)	833	216
19	Yanbu Entrance	701	149
20	Yanbu Dadao North	533	125
21	Lian'an	230	81
22	Hubang Lu Entrance	815	118
23	Yanbu Hospital	477	117
24	Hubang Market	834	168
25	Sanyanqiao (Guangdong Grain & Oil Market)	726	105
26	Yaju Lanwan	738	107
27	Suiyan Lu (Majestic Garden) Bus Terminal	778	113
Total		17395	3711

Table A.7: Route D Upstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Guangwei Lu Bus Terminal	0	0
2	Zhongshan 6 Lu	1100	364
3	Ximenkou (Zhongshan 6 Lu)	303	85
4	Zhongshan 7 Lu	554	162
5	Chen Clan Academy	605	129
6	Liwan Lu Entrance	554	80
7	Zhongshan 8 Lu	477	110
8	Qiaozhong	1800	261
9	Guangfo Lu Entrance (Jiaokou Coach Station)	1900	275
10	Guangfo Lu East	730	228
11	Huangqi Downtown	505	114
12	Hongwei Lu Entrance	415	142
13	Jiazhou Plaza	748	190
14	Huanghai Lu	202	29
15	Huangqi Department Building	806	199
16	Shimen Middle School Entrance	434	104
17	Huangqi Diyicheng Bus Terminal	879	168
Total		12012	2641

Table A.8: Route D Downstream Travel Time between Stops

Station Number	Station Name	Travel Distance (m)	Actual Travel Time (s)
1	Huangqi Diyicheng Bus Terminal	0	0
2	Shimen Middle School Entrance	879	364
3	Huangqi Department Building	503	85
4	Huanghai Lu	719	162
5	Jiazhou Plaza	535	129
6	Hongwei Lu Entrance	571	80
7	Huangqi Downtown	354	110
8	Guangfo Lu East	647	261
9	Guangfo Lu Entrance (Jiaokou Coach Station)	720	275
10	Qiaozhong	1800	228
11	Zhongshan 8 Lu	1800	114
12	Shiluji	324	142
13	Chen Clan Academy	815	190
14	Zhongshan 7 Lu	524	29
15	Ximenkou (Zhongshan 6 Lu)	663	199
16	Zhongshan 6 Lu	478	104
17	Guangwei Lu Bus Terminal	772	168
Total		12104	2641

Appendix B: OD Matrices

B.1.1 OD MATRIX FOR ROUTE A

	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
1	0	0	0	1	3	2	1	1	4	3	0	2	6	3	10	11	9	7	5	10	14	10	12	18	14	41
2	0	0	0	0	1	0	1	2	1	4	3	0	0	0	3	2	4	2	5	7	5	0	4	4	8	9
3	0	0	0	0	1	2	0	1	4	3	3	3	8	1	7	8	11	6	4	9	7	4	8	9	8	18
4	1	0	0	0	2	1	0	1	2	1	1	1	2	1	8	9	9	6	3	7	11	3	7	9	10	7
5	3	1	0	2	0	0	0	0	1	3	1	2	2	2	2	1	2	4	1	4	6	1	5	4	5	5
6	1	0	2	1	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1	3	2	2	2	6	1	3
7	1	1	0	0	0	0	0	1	0	1	0	1	0	2	2	4	5	1	1	6	5	3	6	14	7	9
8	1	2	2	1	0	0	1	0	1	1	0	0	1	1	0	1	7	2	3	9	2	1	8	14	13	12
9	4	1	3	2	1	1	0	1	0	0	1	0	1	0	2	2	2	2	0	3	3	0	10	13	9	11
10	3	4	3	1	4	0	2	1	0	0	0	0	1	0	0	1	3	1	1	2	3	1	4	7	6	13
11	0	3	3	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	2	5	5	3	3
12	1	0	3	1	2	0	1	0	0	0	0	0	0	1	0	1	1	0	0	1	0	2	2	1	5	4
13	4	0	8	2	2	0	0	1	0	1	1	0	0	0	0	3	1	0	0	1	1	3	5	5	2	1
14	3	0	1	1	2	0	2	1	0	0	1	1	0	0	0	0	0	1	0	0	2	2	5	7	0	0
15	10	3	9	7	2	1	0	0	1	0	0	0	0	0	0	0	3	0	0	2	0	0	4	3	6	1
16	11	3	8	8	1	1	4	1	2	1	0	1	3	0	0	0	1	0	1	1	2	1	3	6	3	4
17	9	4	12	10	2	1	5	7	1	3	0	1	1	0	3	1	0	1	0	2	2	3	1	11	2	4
18	7	3	7	5	4	1	1	2	2	2	1	0	0	1	0	0	1	0	0	0	1	0	0	2	1	1
19	7	2	4	2	1	1	1	3	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0
20	10	7	9	8	4	3	6	9	2	2	0	1	1	0	2	1	2	0	0	0	1	2	1	1	0	6
21	14	5	6	12	6	2	5	2	0	3	1	0	1	2	0	2	2	1	0	1	0	1	0	3	0	1
22	10	0	4	3	1	2	3	1	0	1	2	2	3	2	0	2	3	0	0	2	1	0	0	0	0	1
23	10	4	9	7	4	2	8	10	9	5	5	3	5	6	4	3	2	0	1	1	0	0	0	1	0	0
24	18	3	7	11	3	5	14	14	13	5	6	0	5	7	2	5	11	2	1	1	4	0	1	0	4	0
25	13	7	7	10	5	1	8	13	11	5	4	5	2	0	5	3	2	1	1	0	0	0	0	4	0	0
26	42	9	19	7	5	4	9	13	10	13	3	4	1	1	0	4	4	1	0	6	1	1	0	0	0	0

B.1.2 OD MATRIX FOR ROUTE B

	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
1	0	0	1	2	1	2	0	1	2	3	3	1	3	1	8	7	7	8	10	4	2	7	1	4	3	0	4	1	4	5	4	4	3	2	7
2	0	0	0	0	0	0	0	0	1	0	1	3	1	2	3	2	2	2	1	1	0	3	3	5	2	0	2	2	0	2	1	3	0	2	6
3	1	0	0	1	0	0	1	1	0	0	0	1	1	1	4	4	1	4	2	2	0	3	4	1	1	3	2	0	2	3	5	4	0	4	3
4	2	0	0	0	0	0	0	0	0	1	1	2	0	0	1	2	6	6	0	6	2	3	0	5	2	2	0	1	2	3	6	8	0	3	6
5	1	0	0	0	0	0	0	0	0	1	1	2	0	0	2	0	2	5	1	4	1	8	3	3	4	1	3	0	4	9	8	1	1	2	6
6	2	0	0	0	0	0	0	0	0	2	1	0	1	2	2	0	1	3	3	3	0	8	9	3	3	4	1	4	5	4	3	3	1	2	3
7	1	0	1	0	0	0	0	1	0	0	1	2	1	2	2	4	1	4	4	1	1	4	6	6	4	5	4	1	0	5	6	7	3	2	2
8	2	0	1	0	0	0	1	0	1	1	0	3	0	2	0	0	2	1	3	1	0	4	2	3	3	3	5	3	0	3	0	0	3	0	6
9	2	1	0	0	0	0	0	1	0	0	2	0	1	0	0	0	3	3	0	1	4	2	5	8	5	1	3	1	0	7	8	2	0	2	7
10	3	0	0	1	1	2	0	2	0	0	0	0	1	0	0	0	4	1	2	3	1	3	3	10	2	1	2	1	3	4	5	5	0	1	13
11	3	1	0	1	1	1	1	0	2	0	0	0	0	1	0	0	1	1	1	0	1	3	2	1	3	2	4	4	3	5	7	0	0	3	8
12	1	3	2	2	3	0	2	3	0	0	0	0	1	1	0	0	0	2	1	3	1	2	1	5	1	1	1	0	5	3	1	1	3	5	6
13	1	2	1	0	0	1	1	0	1	1	0	1	0	0	1	0	1	0	1	0	1	0	2	0	0	0	2	0	1	3	2	3	0	1	3
14	2	2	1	0	0	2	3	2	1	0	1	1	0	0	0	0	2	1	0	0	0	0	2	3	2	1	3	0	2	5	1	1	0	0	3
15	8	2	3	1	2	2	2	0	0	0	0	0	1	0	0	2	3	1	0	0	0	0	1	2	0	0	1	0	2	1	5	1	1	5	6
16	8	2	4	2	0	0	4	0	0	0	0	0	0	0	2	0	1	0	0	0	1	0	1	2	1	0	4	1	1	2	5	1	1	3	3
17	6	2	1	7	2	0	0	2	2	4	1	0	1	2	2	1	0	0	0	1	1	0	0	0	0	1	0	0	1	0	3	3	1	1	6
18	6	1	2	5	5	3	3	1	3	1	1	2	0	1	0	0	0	0	1	0	0	2	0	2	2	0	1	0	2	3	4	1	2	1	4
19	9	1	2	0	1	1	3	4	0	2	1	1	1	0	0	0	0	1	0	1	0	0	1	0	1	0	1	1	0	0	1	0	0	1	4
20	5	1	3	5	3	3	1	1	1	3	0	3	0	0	0	0	1	0	1	0	1	2	0	0	0	0	0	1	1	1	0	2	0	2	3
21	2	0	0	2	1	0	1	0	2	1	1	1	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	1	2	1
22	5	3	1	2	7	8	3	3	2	2	3	2	0	0	0	1	0	2	0	2	0	0	0	1	2	0	2	1	0	1	2	2	2	1	3
23	1	3	3	0	3	8	5	3	3	2	3	1	2	2	1	1	0	0	1	0	1	0	0	0	2	0	1	0	3	2	1	0	1	2	3
24	3	4	2	6	3	2	4	4	6	10	2	5	0	3	1	2	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	2	2	0	1
25	3	2	1	2	5	4	4	2	5	4	2	1	0	2	0	1	0	2	1	0	0	2	2	1	0	0	1	0	1	0	1	1	0	3	0
26	0	0	2	2	0	3	5	3	2	1	2	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	1	2	0	0	4	2
27	6	2	2	0	3	1	6	5	3	2	4	1	2	3	1	4	0	1	1	0	0	2	1	1	1	0	0	0	0	0	1	0	0	0	0
28	2	2	0	0	0	5	1	3	1	1	4	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	1	1	1	1
29	3	0	3	1	3	5	0	0	0	4	3	6	1	2	1	1	1	2	0	1	0	1	3	0	1	2	0	1	0	0	2	1	0	0	2
30	5	1	3	2	9	5	4	3	7	5	5	3	3	5	1	2	1	3	0	1	1	1	2	0	0	1	0	0	0	0	0	1	0	0	1
31	2	2	5	8	8	3	6	0	8	5	8	1	2	1	6	4	3	4	1	0	1	2	1	0	1	3	2	0	3	0	0	0	1	0	0
32	3	4	4	6	2	2	6	0	1	2	0	1	3	1	1	2	3	1	0	2	0	3	1	3	0	0	0	2	1	0	0	0	0	0	0
33	3	0	1	0	1	1	4	3	0	0	0	3	0	0	1	1	1	2	0	0	1	1	1	2	0	0	0	1	0	0	1	0	0	0	2
34	3	2	3	3	2	2	1	1	2	1	3	5	0	0	5	2	1	1	1	2	1	1	2	0	3	4	0	1	0	0	0	0	0	0	0
35	7	6	2	4	5	4	2	4	8	12	7	4	3	3	6	3	4	3	3	3	2	3	3	1	1	2	0	1	2	1	0	0	1	0	0

B.1.3 OD MATRIX FOR ROUTE C

	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
1	0	5	7	19	4	12	10	9	21	29	25	15	23	18	34	22	6	6	6	9	17	8	23	16	19	76
2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	7	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	13	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
7	10	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
8	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
9	25	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0
10	28	0	1	0	0	0	1	0	1	0	2	2	1	3	5	5	1	1	0	1	3	1	0	0	2	24
11	26	0	1	0	0	0	2	0	1	2	0	1	4	3	0	4	3	2	0	2	1	1	0	1	4	23
12	16	0	2	0	0	0	0	0	0	2	1	0	0	0	0	2	0	0	0	0	1	1	0	2	0	13
13	25	0	1	0	0	1	0	0	0	0	4	0	0	1	2	0	0	0	1	0	1	1	1	0	2	13
14	17	0	0	0	0	0	0	0	1	4	3	0	1	0	3	0	1	1	1	0	3	1	3	0	1	12
15	33	0	0	0	0	1	0	0	0	3	0	0	2	2	0	0	4	1	0	0	1	0	4	1	4	20
16	22	0	0	0	0	0	0	1	0	5	3	2	0	1	0	0	2	0	1	0	3	5	3	4	1	8
17	6	0	0	2	0	1	1	0	1	0	3	0	0	1	4	2	0	1	1	0	0	2	1	0	1	13
18	7	0	0	0	0	0	0	0	0	1	2	1	0	1	1	0	1	0	0	0	2	1	1	3	2	6
19	6	0	0	0	0	0	0	0	0	0	0	0	1	1	3	1	1	0	0	1	0	0	1	1	2	8
20	9	0	0	0	0	0	0	0	0	1	2	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
21	18	0	0	0	0	0	0	0	0	3	1	1	1	2	1	1	0	2	0	0	0	0	0	0	0	0
22	7	0	0	0	0	0	0	0	0	1	1	1	1	1	0	5	2	1	0	0	0	0	0	0	0	0
23	25	0	0	0	0	0	0	0	0	0	0	0	1	3	4	3	1	1	1	0	0	0	0	0	0	0
24	16	0	0	0	0	0	0	0	0	0	1	2	0	0	1	4	0	3	1	0	0	0	0	0	0	0
25	20	0	0	0	0	0	0	0	0	2	4	0	2	1	4	1	1	1	2	0	0	0	0	0	0	0
26	71	0	0	0	0	0	0	0	0	24	23	13	14	11	21	8	13	6	8	0	0	0	0	0	0	0

B.1.4 OD MATRIX FOR ROUTE D

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	68	78	99	121	92	89	63	23	22	22	37	19	20	41	94
2	63	0	43	55	79	48	69	41	23	17	19	16	13	15	20	49
3	73	44	0	24	22	17	24	18	8	5	7	13	6	7	10	16
4	90	57	19	0	37	16	26	12	8	4	11	10	11	13	10	25
5	107	79	20	37	0	23	14	12	12	7	16	17	11	6	6	35
6	92	40	20	14	21	0	14	9	6	10	7	7	17	6	9	30
7	91	59	24	22	15	15	0	20	16	13	13	16	10	10	17	54
8	57	39	19	13	11	8	22	0	27	17	13	24	17	21	26	45
9	26	20	9	9	12	4	12	31	0	14	9	15	19	18	27	39
10	23	17	5	3	7	11	11	21	12	0	5	6	9	17	16	39
11	22	13	6	8	13	7	13	15	8	5	0	7	7	11	14	18
12	37	10	9	9	17	7	14	26	16	8	5	0	10	8	15	26
13	18	12	7	10	9	18	12	25	17	9	10	9	0	7	12	14
14	22	18	6	11	6	3	9	24	21	16	8	7	7	0	15	18
15	37	24	13	12	6	9	17	29	24	13	14	18	12	15	0	26
16	106	50	18	24	30	28	52	46	34	39	15	23	13	18	20	0