

March 2019

## Modeling and Optimizing Routing Decisions for Travelers and On-demand Service Providers

Xinlian Yu

Follow this and additional works at: [https://scholarworks.umass.edu/dissertations\\_2](https://scholarworks.umass.edu/dissertations_2)



Part of the [Transportation Engineering Commons](#)

---

### Recommended Citation

Yu, Xinlian, "Modeling and Optimizing Routing Decisions for Travelers and On-demand Service Providers" (2019). *Doctoral Dissertations*. 1502.

[https://scholarworks.umass.edu/dissertations\\_2/1502](https://scholarworks.umass.edu/dissertations_2/1502)

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

**MODELING AND OPTIMIZING ROUTING DECISIONS  
FOR TRAVELERS AND ON-DEMAND SERVICE  
PROVIDERS**

A Dissertation Presented

by

XINLIAN YU

Submitted to the Graduate School of the  
University of Massachusetts Amherst in partial fulfillment  
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2019

Civil and Environmental Engineering

© Copyright by Xinlian Yu 2019

All Rights Reserved

# MODELING AND OPTIMIZING ROUTING DECISIONS FOR TRAVELERS AND ON-DEMAND SERVICE PROVIDERS

A Dissertation Presented

by

XINLIAN YU

Approved as to style and content by:

---

Song Gao, Chair

---

Eric Gonzales, Member

---

Ahmed Ghoniem, Member

---

Richard Palmer, Department Chair  
Civil and Environmental Engineering

## DEDICATION

*To my family, for their endless love and support.*

## **ACKNOWLEDGMENTS**

First of all, I would like to thank my advisor, Prof. Song Gao, for being a terrific teacher and a great advisor for me. She patiently taught me to see through problems from the right angles, and gave me valuable suggestions. She has been a brilliant, knowledgeable, and supportive adviser for me and I am deeply thankful to her for her time, energy, and support during these years. I want to thank my professors at UMass Amherst, and especially my committee members, Prof. Eric Gonzales and Prof. Ahmed Ghoniem, for their valuable guidance and comments. I also want to thank all the Transportation Engineering faculty members for instructing me in developing the skills that are required in this thesis. Finally, I would like to thank the Administrative Manager of UMass Transportation Center, Kris Stetsonthe and Academic Assistant of CEE department, Jodi Ozdarski, who always give me valuable help during my study. Last but not least, I want to thank my colleagues for making my doctoral study much more pleasurable.

## **ABSTRACT**

# **MODELING AND OPTIMIZING ROUTING DECISIONS FOR TRAVELERS AND ON-DEMAND SERVICE PROVIDERS**

FEBRUARY 2019

XINLIAN YU

B.S., SOUTHEAST UNIVERSITY

M.S., NANJING UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Song Gao

This thesis investigates the dynamic routing decisions for individual travelers and on-demand service providers (e.g., regular taxis, Uber, Lyft, etc).

For individual travelers, this thesis focuses on modeling and predicting route choice decision at two time scales: the day-to-day and within-day. For day-to-day route choice behavior, methodological development and empirical evidences are presented to understand the roles of learning, inertia and real-time travel information on route choices in a

highly disrupted network based on data from a laboratory competitive route choice game. A learning model based on the power law of forgetting and reinforcement is applied. The learning of routing policies instead of simple paths is modeled when real-time travel information is available, where a routing policy is defined as a contingency plan that maps realized traffic conditions to path choices. Using data from a competitive laboratory experiment, model parameter estimates are obtained from maximizing the likelihood of making the observed choices on the current day based on choices from all previous days. Prediction performance is then measured in terms of both one-step and full trajectory predictions. Traditionally, the routing policy model within each day is estimated with non-recursive model which requires prior choice set generation. In practice, sampling choice sets of routing policies is computationally costly and it does not scale well with the size of the network. In this thesis, a recursive logit model for route choice is formulated in a stochastic time-dependent (STD) network where the choice of path corresponds is formulated as a sequence of link choices, without sampling any choice sets. A decomposition algorithm is proposed for solving the value functions that relies on matrix operations so that the model can be estimated in reasonable time. Estimation and prediction results of the proposed model are presented using a data set collected from a subnetwork of Stockholm, Sweden.

Taxis and ride-sourcing vehicles play an important role in providing on-demand mobility in an urban transportation system. Unlike individual travelers, they do not have a specific destination when there's no passenger on board. The optimal routing of a vacant taxi is formulated as a Markov Decision Process (MDP) problem to account for long-term profit over the full working period at the highest level of spatial resolution. Two approaches are proposed to solve the problem. One is the model-based approach where a

model of the state transitions of the environment is obtained from queuing-theory based passenger arrival and competing taxi distribution processes. An enhanced value iteration for solving the MDP problem is then proposed making use of efficient matrix operations. The other is the model-free learning approach, which learns state-action values (and from that, the best policy) directly from observed trajectory data. This method is model-free, in that no transition models are needed and the system dynamics are embedded in the observed trajectories. Batch-model RL algorithm is applied to make more efficient use of collected data by the separation of learning and exploration steps. Both approaches are implemented and tested in a mega city transportation network with reasonable running time, and a systematic comparison of the model-based and model-free approaches is also provided.

## TABLE OF CONTENTS

	Page
<b>ACKNOWLEDGMENTS</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>LIST OF TABLES</b> .....	<b>xiv</b>
<b>LIST OF FIGURES</b> .....	<b>xvi</b>
 <b>CHAPTER</b>	
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 Background and Motivation .....	1
1.2 Thesis scope .....	2
1.2.1 Modeling and Predicting Individual Travelers' Route Choice Decision .....	3
1.2.2 Optimizing Vacant Taxi Routing Decisions .....	4
1.3 Thesis Contributions .....	5
1.4 Thesis Organization .....	9
 <b>2. LEARNING ROUTING POLICIES IN A DISRUPTED, CONGESTIBLE     NETWORK WITH REAL-TIME INFORMATION: AN     EXPERIMENTAL APPROACH</b> .....	<b>11</b>
2.1 Literature Review .....	13
2.1.1 Routing Policy Choice Models without Learning .....	13

2.1.2	Learning Model: Paths vs. Routing Policies . . . . .	14
2.1.3	Impacts of Exogenous Information on Learning . . . . .	16
2.2	Experiment Design and Descriptive Data Analysis . . . . .	18
2.2.1	Experimental Design . . . . .	18
2.2.2	Descriptive Data Analysis . . . . .	22
2.2.2.1	Trip Travel Time . . . . .	22
2.2.2.2	Risk Attitude Variations over Time . . . . .	25
2.2.2.3	Responses to Real-Time Information . . . . .	26
2.3	Modeling Approaches . . . . .	27
2.3.1	Routing Policy Learning and Choice Model Considering Overlapping . . . . .	27
2.3.2	Estimation Approach . . . . .	33
2.3.3	Systematic Utility Specification . . . . .	35
2.4	Results . . . . .	37
2.4.1	Estimation results . . . . .	39
2.4.2	Prediction Evaluation . . . . .	41
2.5	Summary . . . . .	43
<b>3.</b>	<b>A LINK-BASED RECURSIVE ROUTE CHOICE MODEL FOR STOCHASTIC AND TIME DEPENDENT NETWORKS . . . . .</b>	<b>45</b>
3.1	Literature review . . . . .	46
3.1.1	Route choice in an STD network: reaction to information . . . . .	46
3.1.2	Computation challenge: choice set generation and estimation . . . . .	47
3.2	Recursive Logit for Stochastic and Time-Dependent Networks . . . . .	49
3.2.1	Network Settings . . . . .	49
3.2.2	Recursive logit model . . . . .	51
3.3	Decomposition Method for the Maximum Likelihood Estimation . . . . .	55

3.3.1	Computation of the Value Functions .....	56
3.3.2	First Order Derivative .....	57
3.3.3	Decomposition-based Estimation Algorithm .....	60
3.4	Numerical Results .....	65
3.4.1	Network and Data .....	65
3.4.2	Model Specifications .....	66
3.4.3	Estimation results .....	67
3.4.4	Computational time results .....	70
3.5	Summary .....	71
<b>4.</b>	<b>A MARKOV DECISION PROCESS APPROACH TO VACANT TAXI ROUTING WITH E-HAILING .....</b>	<b>72</b>
4.1	Literature Review .....	74
4.1.1	Single Vacant Taxi Routing Problem .....	74
4.1.2	Other Related Work .....	79
4.2	Formulation of the Non-myopic Optimal Taxi Routing Problem .....	82
4.2.1	States and Actions .....	82
4.2.2	Passenger Arrival and Matching Probability on a Link .....	84
4.2.3	Passenger Destination Probabilities .....	86
4.2.4	State Transition Probabilities .....	87
4.2.5	Immediate Profit .....	88
4.2.6	The Bellman Equation .....	89
4.3	Solving the Bellman equation .....	90
4.4	Computational Tests .....	93
4.4.1	The Network, GPS Data and Experiment Setup .....	94
4.4.1.1	The Study Area and Network .....	94
4.4.1.2	GPS Data .....	94
4.4.1.3	Experiment Setup .....	98
4.4.2	Computational Performance .....	99

4.4.3	Evaluation .....	101
4.4.3.1	Heuristics .....	102
4.4.3.2	Trajectory Simulation .....	104
4.4.3.3	Results .....	107
4.5	Summary .....	115
<b>5.</b>	<b>OPTIMIZING VACANT TAXIS' ROUTING DECISIONS: A MODEL-FREE REINFORCEMENT LEARNING FRAMEWORK</b>	<b>117</b>
5.1	Problem Formulation .....	118
5.1.1	Model-free Reinforcement Learning without Transition Model .....	119
5.1.2	Discussion on Solving Time-dependent MDP .....	123
5.1.3	Discussion on Data Efficiency .....	124
5.2	Case Study .....	125
5.2.1	Data and Network .....	125
5.2.2	Data Pre-processing .....	125
5.2.3	Experiment setting .....	128
5.2.4	Computational Performance .....	128
5.2.5	Effectiveness Evaluation .....	130
5.2.5.1	Trajectory Simulation .....	130
5.2.5.2	Batch RL Performance against Sampling Data Size .....	131
5.2.5.3	Performance Comparison with Model-based Approach .....	133
5.3	Summary .....	136
<b>6.</b>	<b>CONCLUSIONS AND FUTURE DIRECTIONS</b> .....	<b>138</b>
6.1	Research Summary .....	138
6.2	Future Directions .....	140

<b>BIBLIOGRAPHY .....</b>	<b>145</b>
---------------------------	------------

## LIST OF TABLES

<b>Table</b>	<b>Page</b>
2.1 Average Trip Time Mean and Standard Deviation for All Sessions . . . . .	22
2.2 Trip travel time mean (min) and standard deviation (min) over time . . . . .	25
2.3 Instances used for perception update for each routing policy on each day for a hypothetical sequence of observed paths . . . . .	31
2.4 Model Specification . . . . .	35
2.5 Systematic Utility Specifications . . . . .	37
2.6 Estimation results of the main model . . . . .	38
2.7 Prediction Evaluation Results (One-step measured by $\bar{\rho}^2$ ; Full-trajectory measured by MSD) . . . . .	42
3.1 The joint probability distribution of all link travel times ( $p_1 = p_2 = p_3 = \frac{1}{3}$ ) . . . . .	51
3.2 The scheme of the event collection . . . . .	52
3.3 Estimation results . . . . .	68
3.4 Number of simulated paths with cycles . . . . .	70
4.1 An overview of single vacant taxi routing studies with network optimization . . . . .	78

4.2	Average unit profit and occupancy rate in the morning time intervals . . . . .	107
4.3	Difference between distributions of visited locations . . . . .	112
4.4	Summary statistics on spatial distributions of visited locations ( $X, Y$ are in meter) . . . . .	115

## LIST OF FIGURES

Figure	Page
2.1 Screen Shot of the Experiment Interface . . . . .	19
2.2 Histogram of trip travel time (in min) . . . . .	24
2.3 Average proportion of choosing Highway . . . . .	25
2.4 Observed number of “days” when the “bad routing policy” is taken in the Information scenario . . . . .	27
3.1 An illustrative small network . . . . .	51
3.2 Final log-likelihood values with discount factors from 0.50 to 0.98 . . . . .	69
4.1 Illustration of the passenger matching process on a link . . . . .	83
4.2 Distribution of observed unit profit and occupancy rate 5:30-11:30 am . . . . .	96
4.3 Pick-up and drop-off density 5:30-11:30 am on a weekday in April, 2015 (count per $km^2$ ) . . . . .	97
4.4 Convergence of the Value Iteration Algorithm . . . . .	100
4.5 Vacant taxi density at 5:30am (count per $km^2$ ) . . . . .	106
4.6 Distribution of unit profit and occupancy rate for three strategies 5:30-11:30 am . . . . .	109

4.7	Differences in unit profit of optimal vs. heuristic routing by starting zones .....	111
4.8	Empirical probability density function of the visited location from a starting location (red star) .....	114
5.1	The idea of Q-learning: learn while interacting with environment .....	121
5.2	The three distinct phases of the batch reinforcement learning process: 1: Collecting transitions with an arbitrary sampling strategy. 2: Application of batch reinforcement learning algorithms in order to learn the best possible policy from the set of transitions. 3: Application of the learned policy .....	121
5.3	Histograms for trip time and searching time during 5:30-11:30 am .....	126
5.4	Value Function Difference vs. Iterations .....	129
5.5	Average unit profit against the sample size for model-free approach .....	132
5.6	Distribution of unit profit and occupancy rate for three strategies 5:30-11:30 am .....	134
5.7	MB vs MF: average unit profit against length of time interval .....	135

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 Background and Motivation**

Human choices are central to the performance of a transportation system, as individual travelers and transportation service providers. It is from the interaction of human choices with the urban physical infrastructure that traffic patterns emerge, with either good or bad consequences. By understanding and optimizing decision-making process, more efficient transportation systems can be delivered. On one hand, humans are not automatons and it is not possible to force their actions. It is therefore essential that we understand the factors that affect travel-related choices and anticipate the actions of travelers in order to effectively design and manage the system. On the other hand, an efficient transportation system requires a lot of quality decision making. Optimization techniques can be applied to search for better travel decisions to improve mobility and operation efficiency.

Route choice plays a central role in many transport applications, including the design and implementation of effective intelligent transport systems, on-board navigation systems, and traffic information broadcasting. Modeling and optimizing route choice decisions are essential to forecast route choice behavior, understand drivers' reaction and adaptation to changing environment and improve future traffic conditions in transportation networks. Modeling and optimizing routing decisions are also challenging given the

complexity of representing human behavior, the uncertainty and dynamism of network composition and the large societal scale with thousands of travelers and service providers to make routing decisions.

Recent advances in information and communication technologies have been influencing drivers' route choice behavior greatly. For example, the provision of real-time or even personalized route guidance information allows drivers to adjust route choice behavior en route; the use of e-hailing apps(e.g., Uber, Lyft and DidiChuxing) allows both passengers and taxi drivers to find each other more quickly through smart phones. At the same time, new data sources such as GPS devices, mobile phone data records, smart card data and geo-coded social media records have contributed to the availability of high-quality traffic data at an unprecedented scale. This massive generation of traffic data allows to observe and understand mobility behavior on an unprecedented level of detail, which provides new opportunities for obtaining valuable insights and further optimizing operational efficiency by extracting every ounce of information. Nowadays, the challenge is to figure out how to use all this data effectively to inform planning activity, improve operations, reduce costs, and better serve travelers.

## **1.2 Thesis scope**

This thesis investigates the dynamic routing decisions of both travelers and on-demand service providers (e.g., regular taxis, Uber, Lyft, etc).

### **1.2.1 Modeling and Predicting Individual Travelers' Route Choice Decision**

For individual travelers, this thesis focus on the modeling and predicting of route choice behavior. Two types of route choice models can be defined. One is a fixed sequence of road segments, termed a path, while the other is a strategy that specifies road segments or the full path to take contingent on information on the decision environment, which is termed as "routing policy" in this thesis. In this thesis, dynamic routing policy is studied at different time scales according to the divisions of planning period: the within-day and day-to-day model.

Day-to-day route choice model focus on adjustment behavior of travelers' repeated route choice on a daily base. The underlying network conditions is changing everyday. The traveler gradually learns and updates his/her perception as experiences accumulate in memory, where a learning model is usually required to capture the updating process. Long-term longitudinal data is required in order to estimate dynamic day-to-day models. Longitudinal data is commonly used in other fields (e.g., financial study, health and employment) but has been used to a limited extent in transportation research due to the difficulty in collecting such data. In this thesis, a learning-based routing policy model was developed to capture travelers' adaptive route choice behavior with real-time information in a day-to-day learning process. A competitive laboratory behavioral experiment was designed and conducted in a network with stochastic incidents to collect the data associated with travelers' route choice decisions for parameter estimation.

Within-day routing policy model adapts en-route choices dynamically based on realized traffic condition. The underlying network distribution is assumed the same over days. Ding et al. (2015) estimated such a routing policy model with non-recursive model which

requires prior choice set generation that involves repeated executions of the optimal routing policy algorithm. However, sampling choice sets of routing policies is computationally costly and it does not scale well with the size of the network. In this thesis, a recursive logit model for route choice is formulated in a stochastic time-dependent (STD) network where the choice of path corresponds to a sequence of link choices, without sampling any choice sets of routing policies. A decomposition algorithm is proposed for solving the value functions that relies on matrix operations so that the model can be estimated in reasonable time.

### **1.2.2 Optimizing Vacant Taxi Routing Decisions**

Taxis and ride-sourcing vehicles play an important role in providing on-demand mobility services due to its great accessibility and convenience in urban areas. By the end of 2014, there were over 13,000 yellow cabs in New York City (NYC), serving more than 450,000 passengers daily (New York City Taxi, & Limousine Commission. (2014)).

Unlike individual travelers, taxis and ride-sourcing vehicles do not have a clear destination when they do not have a passenger on board. Excessive cruising of empty taxis not only leads to waste of fuel and time, but also generates additional traffic. The design and implementation of efficient intelligent routing and participation algorithms for taxis and ride-sourcing drivers is the key to improve taxi utilization and service quality to passengers, as well as reduce traffic congestion and energy consumption.

Most studies addressing the taxi routing problem focus on extracting passenger demand pattern from historical GPS trajectories and recommending a location or a sequence of potential pick-up points for taxi drivers (Powell et al., 2011; Hu, Gao, Chiu and Lin, 2012; Yuan, Zheng, Xie and Sun, 2013; Qu et al., 2014; Hwang et al., 2015). In these

studies, the routing process ends once the taxi finds a passenger. They do not optimize routing decisions continuously and adaptively to take into account downstream impacts.

In this thesis, the vacant taxi routing problem is formulated as a Markov Decision Process (MDP) so that long-term objectives can be taken into account instead of the immediate one of meeting the next customer. Two approaches are proposed to solve the problem. One is the model-based approach where a model of the state transitions of the environment is obtained from queuing-theory based passenger arrival and competing taxi distribution processes. The other is the model-free learning approach, which learns action values (and from that, the best policy) directly from observed trajectory data. This method is model-free, in that no transition models are needed and the system dynamics are embedded in the observed trajectories. Both approaches are implemented and tested in a large-scale network of Shanghai, China, and a systematic comparison of the model-based and model-free algorithms are also provided.

### **1.3 Thesis Contributions**

The contributions of the thesis to the knowledge base of modeling and optimizing routing decisions are summarized as follows:

#### **1. Learning Routing Policies in a Disrupted, Congestible Network with Real-Time Information: an Experimental Approach**

An econometric model of learning and choice of routing policies is developed, estimated and evaluated using data from a laboratory experiment of route choice game in a network disrupted by incidents following a pre-specified distribution. The contributions are two-fold.

- Methodologically, a general routing policy learning model based on the power law of memory decay and reinforcement is developed to account for overlapping of alternatives, extending Path Size Logit to a dynamic context.
- Empirically, the relative importance of learning against inertia is studied systematically in a more realistic setting compared to most laboratory experiments in the literature, where the network is disrupted by incidents, congestible, and has real-time information on incident occurrence. The comparison of one-step and full-trajectory predictions highlights the different role of learning in short- and long-term predictions, where one-step prediction entails predicting the next day's choice, while full-trajectory prediction entails predicting the next  $K$  days' choices, both of which are based on observed choices up to today.

## 2. A Link-based Recursive Route Choice Model for Stochastic and Time Dependent Networks

A recursive logit model for route choice decision is formulated in a stochastic time-dependent (STD) network, without sampling any choice set. A decomposition method is proposed to estimate the model efficiently. Estimation and prediction results are presented using data from a network situated in the Stockholm, Sweden.

- Methodologically, a recursive logit model for route choice decision is formulated in a stochastic time-dependent (STD) network where the choice of path corresponds to a sequence of link choices. A decomposition algorithm is proposed for solving the value functions that relies on matrix operations so that the model can be estimated in reasonable time.

- Empirically, estimation and prediction results of the proposed model are presented using a data set collected from a subnetwork of Stockholm, Sweden. Results show that the model can be estimated efficiently, and gives reasonable results for prediction.

### 3. A Markov Decision Process Approach to Vacant Taxi Routing with E-hailing

The objective is to develop a methodology for the vacant taxi routing optimization problem to achieve better optimality, practicality and computationally efficiency. Towards this end, contributions in modeling, problem formulation and solution algorithm design are made, detailed as follows.

- **Modeling** A queueing theory-based model for matching taxis and passengers is proposed to account for competition from other taxis and use of e-hailing apps. The routing decisions are based on the physical road network in contrast to cell/zone based, which enables more practical implementations including the generation of turn-by-turn guidance.
- **Problem formulation** The MDP formulation optimizes long-term expected profit over the complete working period, accounting fully for the impact of current decisions on future return over multiple pickups and drop-offs, and thus is able to integrate the array of factors (e.g., searching distance, searching time, pick-up probability, competition from other taxis, revenue from the next passenger) considered by other studies (Yuan, Zheng, Zhang and Xie, 2013; Hwang et al., 2015) in a single, theoretically appealing formulation. In the Shanghai case study, the MDP formula-

tion improves unit profit up to 27% and 8% over the random walk and local hotspot heuristic respectively; and improve occupancy rate up to 27% and 15% respectively.

- Efficient implementation of the solution algorithm    An enhanced value iteration for solving the MDP problem is proposed making use of efficient matrix operations, and its efficiency is tested in a mega city transportation network with reasonable running time. Existing studies handle computational efficiency by either adopting a cell/zone-based approach, or limiting the search to a local range if the physical road network is used. Computational efficiency is achieved for the complete network of a mega city at the highest level of spatial resolution through a recursive problem formulation and efficient implementation of the solution algorithm.

#### 4. Optimizing Vacant Taxis' Routing Decisions: a Model-free Reinforcement Learning Framework

The application of Reinforcement Learning (RL) is examined as a model-free approach to solve the vacant taxi routing problem. The contributions are detailed as follows.

- A model-free RL algorithm is applied to solve the empty taxi routing problem, in that no transition models are needed and the system dynamics are embedded in the observed trajectories. Batch RL algorithm is applied to make more efficient use of collected transition samples.
- The algorithm is implemented and tested in a real road network of Shanghai, China, and a systematic comparison of the model-based and model-free algorithms are also provided. Results show that batch RL is a sample efficient algorithm for vacant taxi routing so as to avoid extra modeling assumptions. It could still learn better

performance policies even from a small sample size. Overall, the performance of the learned policy increases with sample size.

- Comparison of the model-based and model-free algorithms show that both policies perform better than random walk despite not having any priori knowledge. Model-based method is more effective when the model perfectly matches the true dynamics but often at the cost of larger bias when the dynamics are not modeled accurately; while model-free method are less efficient but could achieve good asymptotic performance especially where the true dynamics cannot be modeled accurately.

## 1.4 Thesis Organization

The thesis is organized as follows. Chapter 2 and 3 investigate individual route choice behavior at different time scales. Chapter 4 and 5 optimize vacant taxi routing decisions.

In Chapter 2, an econometric model of learning and choice of routing policies is developed, estimated and evaluated using data from a laboratory experiment of route choice game in a network disrupted by incidents following a pre-specified distribution.

Chapter 3 proposes a recursive logit model for policy choice in STD networks. A decomposition algorithm for solving the value functions that relies on matrix operations is proposed and the model can be solved in reasonable time. Estimation and prediction results for a network situated in the Stockholm region, Sweden are then presented.

Chapter 4 formulates the vacant taxi routing problem as a Markov Decision Process (MDP) so that long-term objectives can be taken into account instead of the immediate one of meeting the next customer. A queueing theory-based model for matching taxis and passengers is proposed to account for competition from other taxis and use of e-hailing

apps. An enhanced value iteration for solving the MDP problem is proposed making use of efficient matrix operations, and its efficiency is tested in a mega city transportation network with reasonable running time.

Chapter 5 examines the application of a Reinforcement Learning method to solve the vacant taxi routing problem. This method is model-free, in that no transition models are needed and the system dynamics are embedded in the observed trajectories. The algorithm is implemented and tested in a mega city transportation network, and a systematic comparison of the model-based and model-free algorithms are also provided.

A summary of the thesis work and discussions on future directions are given in in Chapter 6.

## **CHAPTER 2**

### **LEARNING ROUTING POLICIES IN A DISRUPTED, CONGESTIBLE NETWORK WITH REAL-TIME INFORMATION: AN EXPERIMENTAL APPROACH**

Travelers make route choice decisions in an inherently uncertain environment, due to incidents, adverse weather, special events, and other travelers' behaviors. As a result, such decisions depend on travelers' evolving perceptions of the environment, which are usually formed by integrating two sources of information: personal experience, and exogenous travel information. Personal experience has been the primary source, stored in travelers' declarative memory and can be retrieved later to form perceptions before decisions are made. In contrast to experience, exogenous travel information provides descriptions of relevant aspects of the decision environment in text or graphically, e.g., "incident between Exit 10 and 11" on a variable message sign (VMS), and color-coded maps by Google Traffic. Such information has become increasingly available through radio, smartphone apps, and in-vehicle navigation systems. This study is concerned with learning for route choice decisions based on both personal experience and exogenous travel information provided in real-time.

Travelers choose from a set of route alternatives based on perceived attributes of alternatives such as travel time, where perceptions evolve due to learning. Two types of route alternatives can be defined. One is a fixed sequence of road segments, termed a path,

while the other is a strategy that specifies road segments or the full path to take contingent on information on the decision environment. Various learning models of fixed paths are present in the literature, including Tang et al. (2017) based on human memory's power law of forgetting and reinforcement. Learning models of strategies, however, are not as prevalent in the literature. In recognition that travelers might be guessing others' decisions in a competitive route choice game, Selten et al. (2007b) defines strategies where the path to take depends on how the latest experienced travel time compare with the average travel time from all past experiences. When real-time traffic conditions are reported, such as in Lu et al. (2014), strategies are defined where the road segment to take depends on whether an incident on a downstream road segment is reported by a variable message sign. Learning of decision strategies in general has been studied in the Psychology literature, for example, Erev and Barron (2005b). The term "routing policy" is used in this thesis to refer to a strategy applied in a route choice context, and to differentiate from a general strategy.

This study builds upon the work in Lu et al. (2014), with three enhancements. First, in Lu et al. (2014), the learning and choice of routing policies are simplified to ignore overlapping of routing policies and the resulting correlations among choice alternatives, which could result in unrealistic route choice predictions. Secondly, model parameters in Lu et al. (2014) are estimated by matching aggregate, predicted route shared with observed shares, and thus precluding rigorous statistical results available from an econometric model estimated based on disaggregate data. Thirdly, the learning process in Lu et al. (2014) is based on exponential decay of memory, a commonly used method in the transportation community, where the perceived travel time for the current day is a convex combination

of the previously perceived travel time and the latest experienced travel time. Tang et al. (2017) uses the power law of memory decay and reinforcement, which is arguably more psychologically sound, and this study applies this law to the learning of routing policies instead of paths.

The remainder of this chapter is organized as follows. Section 2.1 provides a literature review. Next in Section 2.2, the route choice game is described, followed by a descriptive data analysis. Section 2.3 presents the modeling approach and specification of the utility function. Section 2.4 presents the estimation and prediction results. Conclusions and future directions are given in Section 2.5.

## **2.1 Literature Review**

The literature review has three major parts and focus on empirical studies. Routing policy choice models based on cross-sectional data are first reviewed, since they provide building blocks for the proposed learning model based on longitudinal data. Next, learning of fixed paths vs. routing policies are reviewed, recognizing that the majority of the literature has focused on path learning. Lastly, the literature on impacts of exogenous information on learning is reviewed, under different settings of the experiments (competitive vs. non-competitive) and using different evaluation methods (one-step vs. full-trajectory predictions).

### **2.1.1 Routing Policy Choice Models without Learning**

A number of studies extend the conventional path choice models based on cross-sectional data to networks with stochastic travel times. A static travel time distribution

is either specified by experimenters in a laboratory setting, or estimated using surveillance data from the field. The dynamic relationship between the current choice and past days' experience is not modeled.

Gao (2005b) develops a routing policy choice model where the user may update his/her route choice at any node of the road network depending on traffic conditions, and imbeds the model in a dynamic traffic assignment model. Empirical studies of the routing policy choice model have been carried out using cross-sectional data from both laboratory experiments (Razo and Gao, 2010; Razo and Gao, 2013a) and in-vehicle tracking and monitoring devices in real-life urban networks (Ding-Mastera et al., 2015). Using stated preference (SP) data, Razo and Gao (2010) shows the existence of strategic route choice behavior, where a simple hypothetical network is implemented as an interactive graphical map. Later on, Razo and Gao (2013a) estimates a rank-dependent utility model for explaining the choice among risky routing policies. Ding-Mastera et al. (2015) estimates a latent-class routing policy choice model using a taxi data set collected from a subnetwork of Stockholm, Sweden.

### **2.1.2 Learning Model: Paths vs. Routing Policies**

Almost all learning models for route choice have dealt with fixed paths, instead of routing policies. Please refer to Bogers et al. (2007) for a detailed description of different path learning models. A key question in a learning model is how past experiences are integrated to form a perception of path attributes. The most commonly used method in the transportation community is to treat the perceived travel time at time  $t$  as a convex combination of the perceived travel time and experienced travel time at time  $t - 1$  (see, e.g., Nakayama et al., 2001; Bogers et al., 2005; Lu et al., 2011b), which is equivalent to

assuming an exponential decay of memory. Tang et al. (2017) develops an instance-based learning (IBL) model based on the power law of forgetting and practice, an arguably more sound theory (Anderson and Schooler, 1991), which has been shown as a robust learning process triggered in a wide range of tasks from the simple repeated choice tasks to highly dynamic ones (Lejarraga et al., 2012). The endogeneity problem of a learning model due to the missing initial observations is studied in a follow-up paper (Guevara et al., Forthcoming). Some other studies have employed more complicated learning mechanism, e.g., Bayesian updating (Kaysi, 1991; Jha et al., 1998; Chorus et al., 2009). However, there is little empirical support that travelers are capable of carrying out Bayesian updating, and estimation problems of such learning models remain an open question.

A few studies have addressed routing policy learning where route choices are contingent on either experience or real-time information. Selten et al. (2007b) introduces a route choice game experiment in a two-route network, and proposes two response modes based on previous experiences: a direct one in which a traveler switches route following a bad payoff compared to the average payoff from all previous experiences, and a contrary one with the opposite reaction. The rationale for the latter is that in a route choice game where travelers are competing for good routes, one might gauge other travelers' responses and deliberately make a choice opposite to a popular one to avoid congestion. Klein et al. (2018) develops an agent-based model to study the emergent day-to-day traffic states in a simulation, assuming that agents are able to learn to comply with the daily route recommendations delivered by an advanced traveler information systems (ATIS) route recommendations. The agents' routing decisions are dependent on previous experiences. The estimation problem, however, is not addressed.

Lu et al. (2014) uses a dataset collected from a route choice game in a network subject to random disruptions, and develops a routing policy learning model where routing policies are contingent on real-time information instead of past experiences. The same dataset is used in the current study, and the learning model is enhanced in three aspects as detailed in the paragraph right before the organization of this chapter.

### **2.1.3 Impacts of Exogenous Information on Learning**

A large number of empirical studies have investigated the relationships between information, learning and level of uncertainty in route choice decisions. A detailed review can be found in Bifulco et al. (2014). Some experiments focus on the effects of feedback information in competitive environment (Avineri and Prashker, 2005; Bogers et al., 2007; Qi et al., 2018). Avineri and Prashker (2005) shows the existence of the payoff variability effect reinforced by travel time feedback: high payoff variability moves choice behavior toward random choice, which is related to the observation that variance in outcomes inhibits learning. Bogers et al. (2007) further demonstrates that enriching feedback information on foregone alternatives greatly expedites the learning process compared to a treatment without such information. Consistently, Qi et al. (2018) finds that enriching feedback information on foregone payoffs tend to reduce the proportion of people who firmly commit themselves to a unique route.

The effects of exogenous travel information have been studied in different experimental settings. Some are carried out in a competitive setting, where multiple human subjects make route choices simultaneously in a network, and travel times are determined by their collective choices based on an underlying performance function that links travel time with flow. Rapoport et al. (2014) finds that when pre-trip information about route conditions

(either good or bad) is provided, subjects switch more frequently between two routes in the treatment with four possible states than with two states. Klein and Ben-Elia (2017) shows that coupling punishments and rewards with recommendation significantly pushes the network closer to system optimal (SO). Other studies, which are the majority, are carried out in a non-competitive setting, where the travel time distribution is pre-determined by the experimenter (see, e.g., Avineri and Prashker, 2006; Ben-Elia et al., 2008; Ben-Elia and Shiftan, 2010; Shiftan et al., 2011; Mak et al., 2015; Ma and Di Pace, 2017). Avineri and Prashker (2006) finds that providing static information on the mean travel time on risky and fast routes makes subjects more likely to choose another route. In contrast, Ben-Elia et al. (2008) and Ben-Elia and Shiftan (2010) find that respondents learn faster and exhibit risk seeking behavior with information describing the ranges of travel times. In line with the above results, Shiftan et al. (2011) shows that when the most up-to-date travel time on both routes are provided pre-trip, individuals tend to prefer a riskier route if they had more experience. Ma and Di Pace (2017) shows that payoff variability effect is more prominent with lower information accuracy. As an exception, Mak et al. (2015) find that subjects' learning behavior is similar with or without en-route information.

Most of the empirical studies reviewed above evaluate route learning and choice models based on one-step prediction. A small number of studies have used full-trajectory prediction, including Selten et al. (2007b), Lu et al. (2014), and Zhang, Liu, Huang and Chen (2018). However, all these studies estimate model parameters using aggregate data as mentioned in 2.1.2.

## **2.2 Experiment Design and Descriptive Data Analysis**

The route choice experiment was designed based on the following principles: First, the traffic network is disrupted in that an incident could happen on one of the road segments with a certain probability, and may result in significant congestion. Second, real-time information on whether the incident occurred or not was provided en-route and a detour was available to avoid the incident if it did occur. Third, interactions among travelers and the collective effect of traveler choices on congestion were accounted for through link cost functions. The more travelers on the link, the higher the link travel time.

The experiment was carried out in a computer laboratory at the University of Massachusetts Amherst.

### **2.2.1 Experimental Design**

The laboratory experiment simulated a simplified network with random incidents, and human subjects acted as commuters traveling through this network on weekdays from work to home. Fig. 2.1 shows a screen shot of the experimental network containing three possible paths from the origin to destination. Three paths are defined:

- Path 1 (the upper path): Park Avenue;
- Path 2 (the middle path): Detour (Local 1 followed by Local 2);
- Path 3 (the lower path): Highway (Local 1 followed by I-99, an incident could occur on I-99 with a probability of 0.25).

Two scenarios, Information and No-Information, were designed and the difference was that, real-time information on whether an incident had occurred (incident indicator) was

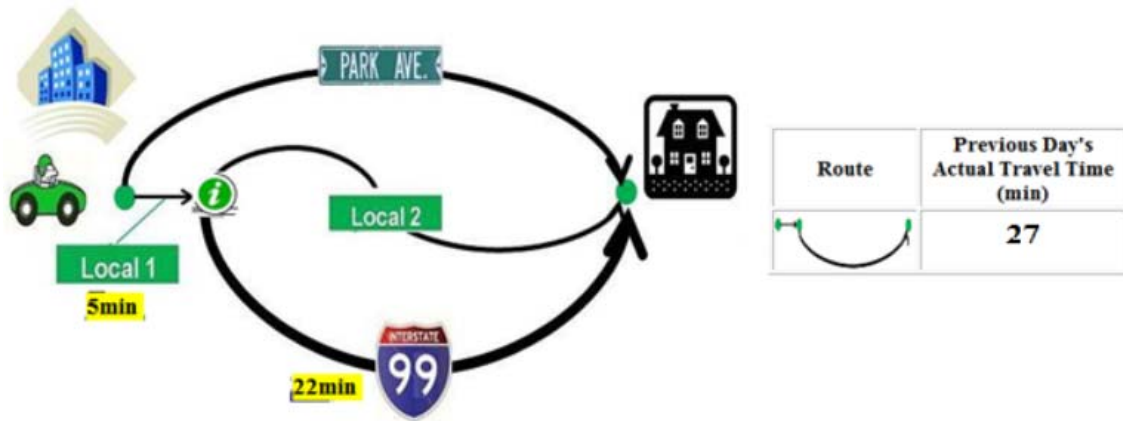


Figure 2.1: Screen Shot of the Experiment Interface

only provided in an Information scenario at the bifurcation towards Local 2 and I-99. Recruited participants were students and staff members at the University of Massachusetts Amherst. Each had to be at least eighteen years old and hold a valid U.S. driver's license with at least one year of driving experience. Future studies involving more diversified population groups are needed if conclusions from this exploratory phase are to be generalized to the general population. A total of 128 participants were divided randomly into 8 groups (sessions) of 16 members each, and no personal interactions were allowed among them throughout the experiment. Four groups participated in the Information scenario and the other 4 groups in the No-Information scenario.

Before each experiment session, participants were instructed about the nature of the experiment. They were notified that an incident could happen on I-99 with a chance of 1 out of 4, which could lead to a significant congestion. They were told that the travel time of a road will increase with the number of users on that road due to congestion effect. The specific link cost function was not revealed to them. Participants in an Information sce-

nario were told that they would receive information on whether an incident had occurred on I-99 at the bifurcation towards Local 2 and I-99 where the symbol “i” is located.

The task was described as making a series of work-to-home route choices within groups of 16 players who independently choose route on each round (or “day”) of play by clicking the appropriate radio button on the screen. Communication between the participants was not allowed. At the end of each “day”, travel times were calculated based on actual numbers of users on the links (link flow), from cost functions that participants did not know. The “day” on which an incident would occur was randomly generated in each 4-day block. To allow for comparison, the same incident profile was used for each paired sessions in the No-Information and Information scenario. At the beginning of each “day”, each participant received feedback on actual travel time and the incident indicator (if applicable) on the experienced path on the previous “day”. Thus the participant was able to learn the travel time distribution from his/her own travel experiences. The table on the screen in Fig. 2.1 provided participants with the previous day’s actual travel time on the experienced path. The numbers in the yellow boxes further showed actual travel times on the links along the experienced path.

The travel time (in minutes) on a given link is a function of the link flow specified in Eq. (2.1):

$$\begin{aligned}
T_{ParkAve}(x_{ParkAve}) &= 33.5 + 2x_{ParkAve} \\
T_{Local1}(x_{Local1}) &= 0.5x_{Local1} \\
T_{Local2}(x_{Local2}) &= 36.5 + 3.82x_{Local2} \\
T_{I-99}(x_{I-99}) &= \begin{cases} 20 + 0.5x_{I-99}, & \text{with probability 0.75 (normal condition)} \\ 20 + 27.5x_{I-99}, & \text{with probability 0.25 (incident condition)} \end{cases}
\end{aligned} \tag{2.1}$$

where  $x_a$  is the flow on link  $a$ , and  $T_a(x_a)$  the link travel time as a function of flow on link  $a$ . The coefficient to  $x_{I-99}$  is much larger under incident than normal condition, suggesting that a very high travel time on I-99 could emerge due to the incident.

In each session, participants completed route choices for a total of 120 “days”. The participants were not notified of the total number of “days” in advance, but only a rough estimate of the duration of the experiment, in order to reduce the likelihood that the participants would make “rushed” choices during final “days”.

In some route choice experiment, participants are paid based on their performance in the experiment, with better performance resulting in higher monetary rewards. This is a direct application of the payment scheme commonly used in experimental economics where the subjects’ tasks are usually directly related to monetary payoffs, such as choosing a lottery. It is argued that caution should be exercised in applying the same payment scheme to a travel choice task, which in the real world involves no monetary rewards. If performance-based incentives are given, it is implicitly assumed that the same value of time applies to every participant, which is not always the case and might bring unnecessary complications. Moreover, there is no solid proof that people would have the same risk attitude towards monetary gains (or losses) and travel time savings (or losses) and it was

advisable to be cautious in equating these two. In the current experiment, each participant was paid a flat payment of \$30. Participants were not distracted by any other tasks as in real life and the provision of feedback travel time on chosen route prompted participants to pay attention to and thus minimize travel times.

## 2.2.2 Descriptive Data Analysis

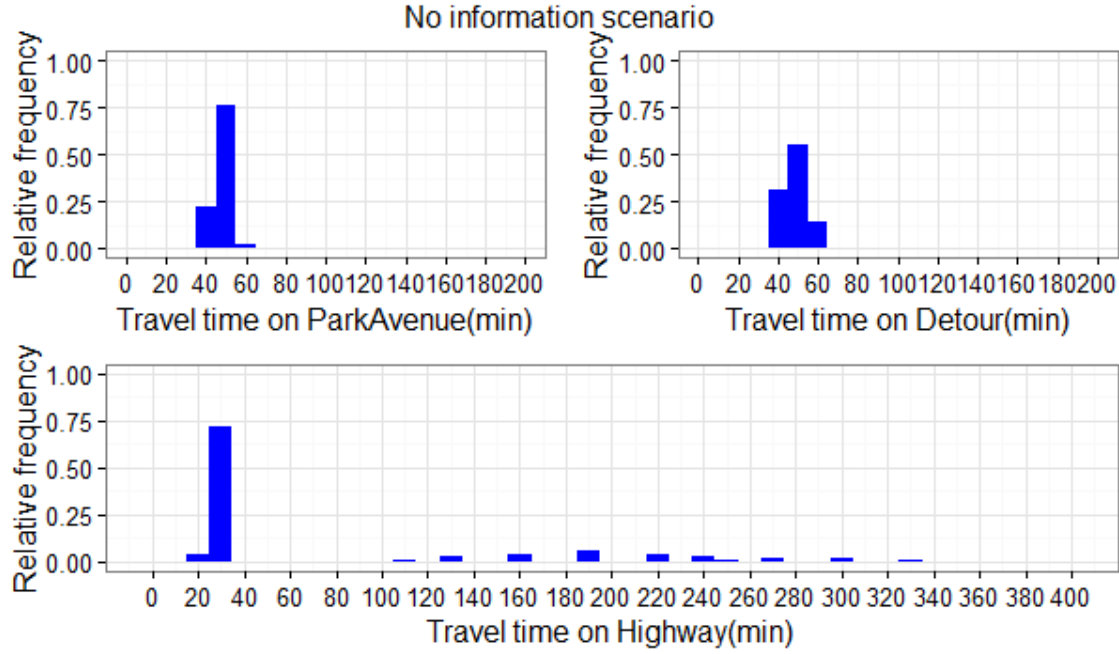
### 2.2.2.1 Trip Travel Time

Table 2.1: Average Trip Time Mean and Standard Deviation for All Sessions

Session	Mean travel time (min)		Standard deviation	
	No information	Information	No information	Information
1	58.17	42.16	39.23	18.10
2	58.77	40.92	39.76	19.31
3	64.66	42.36	51.99	17.86
4	57.52	43.08	34.62	18.85
Average	59.78	42.13	41.40	18.53
$H_0$ : Equality between Scenarios (5% one-sided)	Rejected		Rejected	

Table 2.1 shows the average trip times for all sessions. The No-Information sessions had an average length of 60 minutes, while the Information sessions had an average length of 41 minutes. According to the results of Wilcoxon-Mann-Whitney tests, Lu et al. (2011b) have shown that both the average trip travel time and the standard deviation in the No-Information scenario are significantly higher than that of the Information scenario.

Fig. 2.2 presents the histogram of trip travel time for each scenario. In the No-Information scenario, the travel times on Park Avenue and Detour mostly fall between 40 to 60 min and are slightly skewed. However, the distribution of travel time on Highway exhibits a



strong positive skew with a very long tail for extremely high travel times during an incident. In the Information scenario, the distribution of travel time on Park Avenue is similar to that in the No-Information scenario. The travel time on Detour, however, is slightly increased compared to that in the No-Information scenario. This suggests that some travelers avoided incidents by using Detour with real-time information.

Park Avenue and the Detour are denoted as safe routes and Highway as the risky route, given the former's much smaller travel time variability compared to the latter one. Table 2.2 presents the trip travel time mean and standard deviation of safe routes and risky route in blocks of 20 "days". Not surprisingly, both travel time mean and standard deviation of risky route are higher than those on safe routes in all blocks in each scenario due to the significant disruption on Highway. Moreover, the travel time mean and standard de-

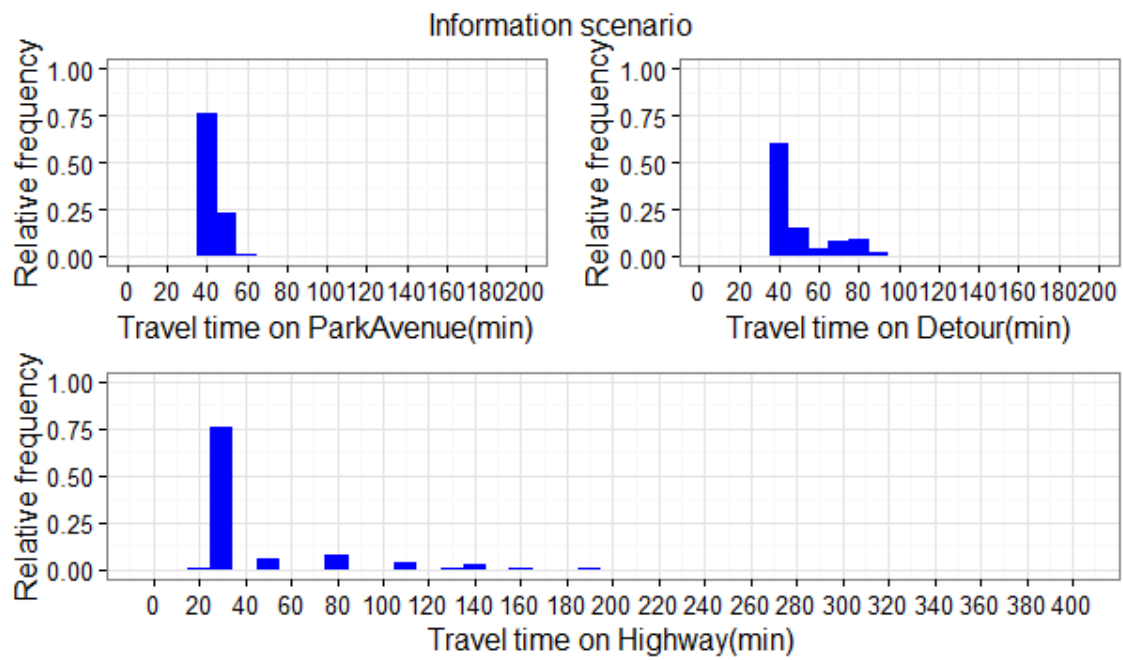


Figure 2.2: Histogram of trip travel time (in min)

violation of safe routes are almost stable over time, while the travel time mean and standard deviation for the risky route decreases over time in general in each scenario.

Table 2.2: Trip travel time mean (min) and standard deviation (min) over time

<b>No-Information scenario</b>						
	1-20 trials	20-40 trials	40-60 trials	60-80 trials	80-100 trials	100-120 trials
Mean_safe	48.16	48.24	47.75	48.47	48.46	47.69
Mean_risky	86.12	73.56	75.63	68.54	67.25	67.23
SD_safe	5.91	4.5	4.21	4.43	4.14	4.26
SD_risky	106.01	81.84	86.55	75.49	73.41	72.97
<b>Information scenario</b>						
	1-20 trials	20-40 trials	40-60 trials	60-80 trials	80-100 trials	100-120 trials
Mean_safe	46.88	46.57	46.57	46.51	45.92	46.43
Mean_risky	56.08	44.88	45.36	43.58	44.15	41.89
SD_safe	12.08	11.62	10.61	10.99	10.83	11.06
SD_risky	49.89	29.95	27.54	26.78	24.13	22.19

#### 2.2.2.2 Risk Attitude Variations over Time

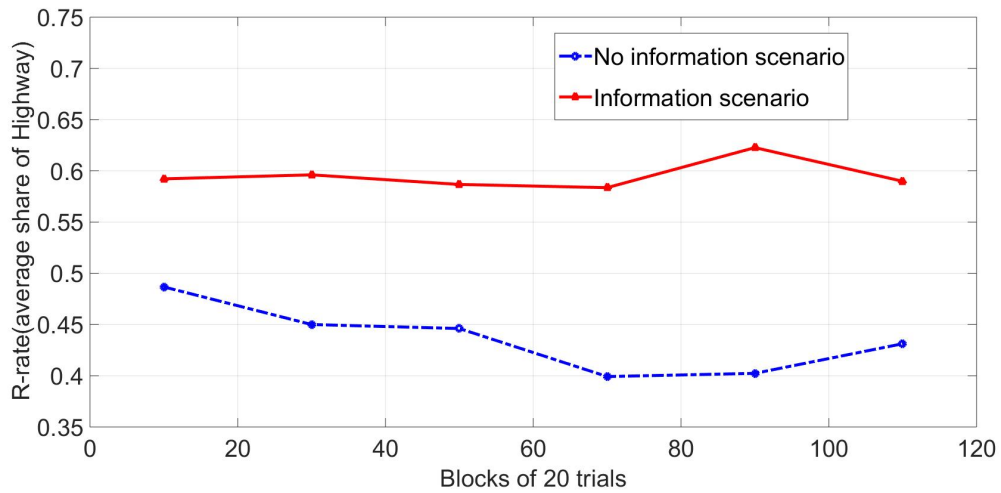


Figure 2.3: Average proportion of choosing Highway

Fig. 2.3 presents the average proportion of risky choices (R-rate) in blocks of 20 “days”. In the No-Information scenario, the R-rate is lower than 0.5 in all blocks and decreases in the first five blocks and then increases a little bit in the last block. This increased preference to the safe routes over time, despite the decreased travel time variability on the risky route and almost constant travel time variability on the safe routes, suggests that travelers appear more risk averse over time. This might be due to decreased exploration of the environment over time, and the “hot stove” effect where a very bad experience on the risky route prevents further exploration of that route. Note that the travel time variability observed by the modeler is likely different from that perceived by the travelers as they form perceptions based on personal experience only. Therefore it is not possible to give a definitive answer as to whether travelers true risk attitudes do change over “days” in the experiment.

The R-rate in the Information scenario is higher than that in the No-Information scenario as shown in Fig 2.3, not surprisingly given the significantly reduced travel time mean and variability on the risky route in the Information scenario.

### **2.2.2.3 Responses to Real-Time Information**

In a non-congested network, a traveler would stay away from Highway if they know for sure an incident occurred on it. In a congestible network, however, travelers are competing with others on the choice of the optimal route, where outcomes depend on the joint actions taken by all travelers involved. Fig. 2.4 presents the observed number of “days” for each participant when a seemingly bad routing policy is taken, that is, “take Highway with incident, and take Detour when no incident on Highway”. The figure shows a non-negligible number of instances of this routing policy, suggesting that some travelers could

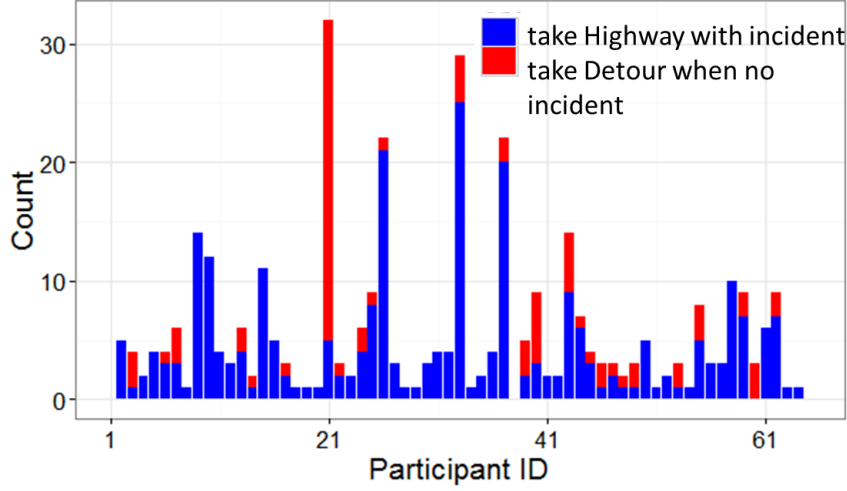


Figure 2.4: Observed number of “days” when the “bad routing policy” is taken in the Information scenario

be guessing other traveler’s responses to the incident indicator and trying to avoid road that they think others might use. In fact, an individual could still end up with an acceptable travel time by taking a seemingly bad routing policy if few or no other travelers choose it.

## 2.3 Modeling Approaches

### 2.3.1 Routing Policy Learning and Choice Model Considering Overlapping

With real-time information, a traveler is able to plan ahead for traffic information that he/she will receive in the future (Gao and Chabini, 2006b). Routing policy represents a traveler’s ability to incorporate real-time information not yet available at the time of decision. A fixed path is a special routing policy where any action is independent of traffic conditions. In this section, a routing policy learning and choice model is developed to capture travelers’ adaptive behavior with real-time information in a day-to-day context.

In the experimental network, the three fixed paths are also three routing policies 1 through 3. Two additional adaptive policies exist only in Information scenario: Routing Policy 4: “First take Local 1, and if the incident has occurred, take Local 2, otherwise take the risky route I-99”, which is referred to as “Avoid incident policy”; Routing Policy 5: “Take the detour when no incident is present, otherwise choose the risky route”, which is referred to as “Ignore incident policy”. A routing policy is not observable as it is a plan in traveler’s mind, and only the result of the plan execution, i.e., the realized path, can be observed. Given a particular day (situation), a routing policy is realized as only one fixed path. There are two possible traffic situations in the experimental network, i.e.,  $S = \{incident, normal\}$ . Routing Policy 4 (Avoid Incident policy) is realized as Detour in *incident* situation, and Highway in *normal* situation. And routing Policy 5 (Ignore Incident policy) is realized as Highway in *incident* situation, and Detour in *normal* situation.

We model travelers’ learning from experience using the instance-based learning theory (IBLT), which is originally proposed to describe decisions from experience in complex dynamic tasks (Gonzalez et al., 2003). It presents a process in which decisions are made from stored and retrieved experiences (called instances), based upon small samples and recently and frequently experienced outcomes. An instance is broadly defined by the context, decision and outcome of a previous choice that is encoded in the declarative memory. Learning resides in the activation mechanism that relies on the frequency and recency of past choices, i.e., more recent and frequent instances are more active in memory. Based on their levels of activation, instances that are relevant to the current decision context are retrieved and blended to produce perceptions of options. Memory decay is captured by the

power law of forgetting that is found in a number of psychological studies, see e.g., Anderson and Schooler (1991), Newell and Rosenbloom (1981), Rubin and Wenzel (1996) and Estes (2014).

Tang et al. (2017) developed the instance-based learning (IBL) model for fixed paths and test it in a two-path experiment network. Based on the similar idea, we propose an IBL model for routing policies. An instance is defined here as a past experience of link  $a$  on day  $t'$  and its associated outcome (realized link travel time),  $X_a(t')$ . An instance is stored in the declarative memory of the traveler, and its activation decays over time following a power law. Specifically, on day  $t$ , its activation is  $(t - t')^{-d}$ , where the decay parameter  $d$  captures the rate of forgetting, in that a smaller  $d$  value translates into higher activation in memory and  $(t - t')$  measures the recency of the experienced travel times. Eq. (2.2) shows the weight function of an experienced travel time from a past day  $t'$  for traveler  $n$  on day  $t$ , where the denominator is a summation of activations over all past experiences on link  $a$ .

Eq. (2.3) shows the perceived travel time of routing policy  $\mu$  for individual  $n$  on day  $t$ , which is the sum of perceived travel time on all links involved in policy  $\mu$ , and the perceived travel time on link  $a$  involved in policy  $\mu$  is the weighted average of realized travel time of link  $a$  on all previous days when link  $a$  is observed and is on the realized path of policy  $\mu$ .  $\Delta(a|\mu, t')$  is an indicator of the latent (routing policy) choice. It is a binary variable that equals 1 if link  $a$  is on the realized path of policy  $\mu$  on day  $t'$ , and 0 otherwise. This formulation also applies to a fixed path model. When  $\mu$  is a fixed path,  $\Delta(a|\mu, t')$  equals 1 when link  $a$  is on  $\mu$  in any traffic situations and 0 otherwise. In other words, the formulation collapse to the path-based model proposed in Tang et al. (2017). Note that the formulation ensures that only relevant link experiences are used to

form perceptions of a routing policy (or path), and thus perceptions are idiosyncratic. In contrast, day-to-day learning models as used in some studies, particularly those focusing on theoretical analysis of convergence, generally assume a shared memory of all travelers, in that the experience on a link from any individual is used to form a single perception for that link used by any path passing that link (see, e.g. Horowitz, 1984; Cantarella and Cascetta, 1995; Yang and Zhang, 2009).

$$W_{na}(t', t) = \frac{(t - t')^{-d}}{\sum_{\tau \in H_{na}(t)} (t - \tau)^{-d}} \quad (2.2)$$

$$T_{n\mu}(t) = \sum_a \left( \sum_{t' \in H_{na}(t)} \left( W_{na}(t', t) X_a(t') \right) \Delta(a|\mu, t') \right) \quad (2.3)$$

where

$t$ : current day;

$t'$ : a previous day when link  $a$  is observed,  $t' < t$ ;

$d$ : decay parameter that captures the rate of forgetting,  $d > 0$ ;

$H_{na}(t)$ : the set of indices of all days before day  $t$  when link  $a$  is observed for traveler  $n$ ;

$W_{na}(t', t)$ : weight of the experienced travel time on day  $t'$  for the perceived travel time on day  $t$  for traveler  $n$  and link  $a$ ;

$T_{n\mu}(t)$ : perceived travel time of policy  $\mu$  on day  $t$  for individual  $n$ ;

$X_a(t')$ : realized travel time of link  $a$  on day  $t'$ .

Table 2.3 illustrates the instances used to calculate the perceived travel time for each routing policy for an individual. The perceived link travel time corresponding to a routing policy is updated if the link is observed and is on the realized path of the policy on a given day. For example, on day 4, Detour, consists of Local 1 and Local 2, is observed with an incident on Highway. Since both Detour and Avoid Incident policy are realized as path Detour on day 4, the perceived link travel times on Local1 and Local2 corresponding to Detour and Avoid Incident policies are updated on day 4. The perceived link travel time on Local1 corresponding to Highway and Ignore Incident policies is also updated on day 4.

Table 2.3: Instances used for perception update for each routing policy on each day for a hypothetical sequence of observed paths

Day	Observed path	Observed links	Routing Policy					Incident
			Park Avenue	Detour	Highway	Avoid Incident policy	Ignore Incident policy	
1	Park Avenue	Park Avenue	$X_{ParkAve}(1)$	-	-	-	-	N/A
2	Highway	Local1	-	$X_{Local1}(2)$	$X_{Local1}(2)$	$X_{Local1}(2)$	$X_{Local1}(2)$	No
		I-99	-	-	$X_{I-99}(2)$	$X_{I-99}(2)$	-	
3	Detour	Local1	-	$X_{Local1}(3)$	$X_{Local1}(3)$	$X_{Local1}(3)$	$X_{Local1}(3)$	No
		Local2	-	$X_{Local2}(3)$	-	-	$X_{Local2}(3)$	
4	Detour	Local1	-	$X_{Local1}(4)$	$X_{Local1}(4)$	$X_{Local1}(4)$	$X_{Local1}(4)$	Yes
		Local2	-	$X_{Local2}(4)$	-	$X_{Local2}(4)$	-	
5	Highway	Local1	-	$X_{Local1}(5)$	$X_{Local1}(5)$	$X_{Local1}(5)$	$X_{Local1}(5)$	Yes
		I-99	-	-	$X_{I-99}(2)$	-	$X_{I-99}(2)$	

To account for overlapping routing policies in a Logit model, a link-based Policy Size (PoS) correction is added to the systematic utility function. As shown in Eq. (2.4),  $PoS_{n\mu}$ , the PoS for individual  $n$  and routing policy  $\mu$ , is the weighted average of Path Size (PS) over all situations  $s \in S$ . In each situation  $s$ , routing policy is realized as a single path, represented by the binary variable,  $\Delta(i|\mu, s)$  that equals 1 if policy  $\mu$  is realized as path  $i$  in situation  $s$  and 0 otherwise. The PS in situation  $s$  extends the original PS formulation (Ben-Akiva and Bierlaire, 1999b), by using  $T_{na}^s(t)$ , the perceived travel time on link  $a$  in situation  $s$  on day  $t$  by individual  $n$ , to calculate the weight of the size of link  $a$ , and the link size is the inverse of the number of paths using link  $a$  among the realized paths for individual  $n$  in situation  $s$ . Further details of the realized path set  $C_n^s$  is provided below.

$$PoS_{n\mu}(t) = \sum_{s \in S} \left( \sum_{i \in C_n} \left( \sum_{a \in \Gamma_i^s} \frac{T_{na}^s(t)}{\sum_{a \in \Gamma_i^s} T_{na}^s(t)} \frac{1}{\sum_{j \in C_n^s} \delta_{aj}} \right) \Delta(i|\mu, s) \right) P(s) \quad (2.4)$$

where

$S$ : the set of situations.

$P(s)$ : the objective probability of situation  $s$ .

$C_n$ : the universal path choice (without duplicate paths).

$\Gamma_i^s$ : the set of links on path  $i$  in situation  $s$ .

$T_{na}^s(t)$ : perceived travel time of link  $a$  on day  $t$  for individual  $n$  in situation  $s$ . Only experiences corresponding to situation  $s$  are extracted as instances to calculate  $T_{na}^s(t)$  as shown in Eq. (2.5), where  $H_{na}^s(t)$  is the set of indices up to day  $t$  in situation  $s$  when link  $a$  is experienced by individual  $n$ .

$$T_{na}^s(t) = \sum_{t' \in H_{na}^s(t)} \left( \frac{(t - t')^{-d}}{\sum_{\tau \in H_{na}^s(t)} (t - \tau)^{-d}} \right) X_a(t') \quad (2.5)$$

$C_n^s$ : realized path set for individual  $n$  in situation  $s$ . It varies with situation  $s$  and could include duplicate paths in each situation since a routing policy can be realized as different paths in different situations and different routing policies can be realized as the same path in situation  $s$ . For example,  $\tilde{C}_n^{incident} = \{path1, path2, path3, path2, path3\}$  and  $\tilde{C}_n^{normal} = \{path1, path2, path3, path3, path2\}$ . Note that the duplication of paths is corrected using Path Size.

$\delta_{aj}$ : link-path incidence variable, which equals 1 if link  $a$  is on realized path  $j$  and 0 otherwise.

The decay parameter  $d$  in an IBL model captures the rate of forgetting. The higher the value of the  $d$  parameter, the faster the decay in memory. Gonzalez and Dutt (2011) tests the IBL model with two datasets, each including a repeated-choice paradigm and a sampling choice paradigm. The results indicate similarity of the  $d$  values for the sampling and repeated-choice paradigms within each dataset. This supports the similarity of the cognitive processes involved in IBL. To simplify the estimation process, the decay parameter is fixed to a value previously estimated on a route-choice dataset (Tang et al., 2017), that is,  $d = 0.8$ . The  $d$  value 0.8 is also within the range of default or common values in the ACT-R architecture in Gonzalez and Dutt (2011) and Wong et al. (2010).

### 2.3.2 Estimation Approach

A routing policy is a plan in a traveler's mind and not observable. The routing policy model must be estimated based on path observations and thus a latent-choice specification

is needed, in that all routing policies whose realized paths on a given day match the path observation should be considered. Eq. (2.6) describes the likelihood of observing path  $i$  for individual  $n$  on day  $t$  as the sum of the likelihood of choosing policies from the routing policy choice set  $\tilde{C}_n$  that are realized as path  $i$  on day  $t$ , where  $P_{n\mu}(t)$  is the probability of choosing routing policy  $\mu$  for individual  $n$  on day  $t$  and  $\Delta(i|\mu, t)$  is a binary variable that equals 1 if policy  $\mu$  is realized as path  $i$  on day  $t$  and 0 otherwise.

$$P_{ni}(t) = \sum_{\mu \in \tilde{C}_n} P_{n\mu}(t) \Delta(i|\mu, t) \quad (2.6)$$

To account for repeated route choice observations from the same traveler, a Mixed Logit model is applied, where random parameter for the alternative specific constants (ASCs),  $\eta_{n\mu}$ , are included in the utility function to model the panel effect.  $L(i_{n1}, \dots, i_{nt}, \dots, i_{nT})$  is the likelihood of the sequence of observed paths for individual  $n$ ,

$$L_n(i_{n1}, \dots, i_{nt}, \dots, i_{nT}) = \int_{\eta} \prod_{t=1}^T \left( \sum_{\mu \in \tilde{C}_n} \frac{\exp(V_{n\mu}(t) + \eta_{n\mu} + \ln PoS_{n\mu}(t))}{\sum_{\omega \in \tilde{C}_n} \exp(V_{n\omega}(t) + \eta_{n\omega} + \ln PoS_{n\omega}(t))} \Delta(i_{nt}|\mu, t) \right) f(\eta) d\eta, \quad (2.7)$$

where  $i_{nt}$  is the observed path for individual  $n$  on day  $t$ ,  $V_{n\mu}(t)$  the systematic utility without the overlapping correction.  $\eta_{n\mu}$  varies over individual and is constant over choice alternatives for the same individual, and follows a normal distribution  $N(\alpha_{\mu}, \sigma_{\mu}^2)$ , where  $\alpha_{\mu}$  is the mean and  $\sigma_{\mu}$  the standard deviations (s.d). Note that  $\eta_{n\mu}$  for one alternative is normalized to 0.  $f(\eta)$  denotes the probability density function of  $\eta$ . The Mixed Logit probability is the weighted average of the Logit function evaluated at different values of  $\eta$ , with the weights given by the density function  $f(\eta)$ .

The Mixed Logit model is estimated by maximizing the likelihood of observed paths over a panel dataset with  $N$  individuals,  $\sum_{n=1}^N L_n$ . As the integral over a normal distribution has no close form expression, simulated maximum likelihood estimation with Halton draws is carried out (Bhat, 2001; Train, 2000).

### 2.3.3 Systematic Utility Specification

Two types of route choice models are considered for the Information scenario: path and routing policy choice model. The path choice model assumes that a traveler chooses a fixed path at the origin and follows it till the end, ignoring any real-time information provided en-route. While the routing policy choice model assumes that a traveler chooses a routing policy at the origin, suggesting that s/he is able to plan ahead for traffic information that will be received in the future.

The models are linear in the parameters, applied to both No Information and Information scenarios. The explanatory variables are given in Table 2.4, and discussions follow.

Table 2.4: Model Specification

Variable	Definition
ASC	alternative-specific constant, set to 0 for Park Avenue; random parameter following a normal distribution over individuals.
Perceived Travel Time	generic variable (in hour).
Stickiness	specific to three fixed paths; 1 if the path was observed on the previous day; 0 otherwise.
Post-incident Response	specific to Highway; 1 if Highway was observed and incident happened on Highway on the previous day; 0 otherwise.

- **Learning** “Perceived Travel Time” calculated by Eq. (2.3) in Section 2.3.1 represents learning based on experience.

- **Inertia** It has been argued that during routine activities, people tend to “place higher value on an opportunity if it is associated with the status quo” (Samuelson and Zeckhauser, 1988), as it can provide significant energy saving to cognitive thinking. Empirical evidence exists that inertia plays a significant role in people’s behavior (see, e.g. Bamberg and Schmidt, 2003). Inertia in route choice has been extensively studied and a review of empirical evidence can be found in Srinivasan and Mahmassani (2000). The impact of inertia in the presence of ATIS has been modeled recently in the joint day-to-day evolution of departure time and mode choices by Liu et al. (2017).

A variable, Stickiness, is defined as a lagged dependent variables (LDV) for each path, to model inertia. It equals 1 on day  $t$ , if the path was chosen on day  $t - 1$ , and 0 otherwise. As noted in Train (2002), the only problem is the necessity to assume initial conditions. Here Stickiness is calculated from the second day since the first observation has no preceding choice.

- **Immediate Response to an Incident** A sudden and large change in the decision environment, such as an incident in the highly disrupted network, might make past experiences immediately less informative for decisions in the near future.

This immediate adaption by is modeled by a dummy variable, “Post-incident Response”, specific to Highway. It is an LDV, and equals equals 1 on day  $t$  if the traveler experienced an incident on day  $t - 1$  and 0 otherwise.

The main model specification above examines the combined effects of inertia, post-incident response and learning. Three variants to the main model are also studied. They

are obtained by including some of the attributes from the main model. Table 2.5 summaries all the model specifications.

Table 2.5: Systematic Utility Specifications

	Attributes			
	ASC	Stickiness	Post-incident response	Instance-based learning(IBM)
<b>Main model</b>	✓	✓	✓	✓
<b>Variant # 1</b>		✓		
<b>Variant # 2</b>	✓	✓	✓	
<b>Variant # 3</b>	✓		✓	✓

## 2.4 Results

Table 2.6: Estimation results of the main model

Coef.	Description	No-Information		Information			
		Path Model		Path Model		Routing Policy Model	
		Value	t value	Value	t value	Value	t value
$\beta_{TT}$	Perceived Travel Time (hour)	-0.224	-2.66	-0.76	-5.54	-4.2	-8.31
$\beta_{STICK}$	Stickiness	27.1	93.19	24.2	71.45	24.8	3.78
$\beta_{POSRES}$	Post-incident Response	-12.9	-62.17	-11.5	-50.39	-8.34	-6.51
$\alpha_{Detour}$	mean of $ASC_{Detour}$	-0.656	-6.53	0.459	0.37	-3.75	-2.08
$\alpha_{Highway}$	mean of $ASC_{Highway}$	-0.23	3.47	0.548	7.73	-1.96	-5.65
$\alpha_{AvoideIncPolicy}$	mean of $ASC_{AvoideIncPolicy}$					1.46	15.08
$\alpha_{IgnoreIncPolicy}$	mean of $ASC_{IgnoreIncPolicy}$					-2.32	-3.71
$\sigma_{Detour}$	s.d. $ASC_{Detour}$	0.601	9.18	-0.529	-3.97	1.39	2.61
$\sigma_{Highway}$	s.d. $ASC_{Highway}$	0.296	0.067	0.177	1.43	0.685	3.53
$\sigma_{AvoideIncPolicy}$	s.d. $ASC_{AvoideIncPolicy}$					0.627	0.54
$\sigma_{IgnoreIncPolicy}$	s.d. $ASC_{IgnoreIncPolicy}$					1.3	4.52
No. of draws		500		500		500	
No. of observations		7680		7680		7680	
No. of parameters		7		7		11	
No. of individuals		64		64		64	
Null log-likelihood		-8437.342		-8437.342		-8437.342	
Final log-likelihood		-3083.71		-3898.78		-2969.23	
Rho-squared		0.634		0.544		0.648	
Adjusted Rho-squared		0.633		0.543		0.647	

### 2.4.1 Estimation results

Models are estimated using Python BIOGEME 2.4 (Bierlaire and Fetharison, 2009). The estimation results of the main model in both the No-Information and Information scenarios are presented in Table 2.6, where both a path model and routing policy model are estimated in the Information scenario, while only a path model is applicable in the No-Information scenario given the lack of real-time information.

The coefficient of Stickiness ( $\beta_{STICK}$ ) is significant at the 5% level and positive for all models in both scenarios. It suggests that travelers tend to repeat the choice from the previous day when everything else is equal. The coefficient of Perceived Travel Time ( $\beta_{TT}$ ) calculated from the IBL model is significant (at the 5% level) across models, suggesting that travelers learn from experience. The coefficient of Post-incident Response ( $\beta_{POSRES}$ ), capturing travelers' immediate response to an experienced incident on Highway, is significant at the 5% level across models. The negative sign indicates that travelers on average respond negatively immediately after the incident, which is on top of the negative impact already captured by the perceived travel time where the weight for the most recent outcome is the highest. A more in-depth look into the data reveals that a small number of travelers stay on Highway immediately after encountering the incident, which might be due to a misconception of negative correlation of incidents over days, that is, an incident on day  $t$  means no incident on day  $t + 1$ , while objectively incidents on different days are independent.

In the Information scenario, the routing policy model ( $\bar{\rho}^2 = 0.647$ ) fits the data much better than the path model ( $\bar{\rho}^2 = 0.543$ ). This suggests that travelers are able to look ahead for real-time information that will be available in the future and account for it in the current

decision. Specifically, the incident indicator at the bifurcation towards Highway and Local 2 allows the travelers to respond depending on the outcome of the incident indicator. Such flexibility makes the branch that contains Highway and Detour more attractive than either one of them as a standalone path alternative. In later discussions, the routing policy model will be used as the default model for the Information scenario.

The ratio of  $\beta_{TT}$  to  $\beta_{STICK}$  indicates the relative importance of learning compared with inertia, as Perceived Travel Time blends past experiences on any given alternative, while Stickiness simply entails repeating the previous choice, the opposite of learning. The ratio is numerically small, and a closer look at the data reveals that a dominating number of travelers keep staying on Park Avenue, an alternative with low variability and acceptable travel time. It is intuitive that travelers find sticking with such an alternative attractive, as it saves cognitive cost (Gao et al., 2011) yet provides reasonable travel time. Across scenarios, the ratio is larger in the Information scenario than in the No-Information scenario. As shown in Table 2.1, travel time variability is smaller in the Information scenario. It follows that learning plays a larger role in a less uncertain, but not deterministic (in which case there is no learning), environment. These findings corroborate with results in the literature (Ben-Elia et al., 2008; Ben-Elia and Shiftan, 2010; Tang et al., 2017), and suggest that providing real-time information promotes learning and more optimal behaviors and as an extension might potentially improve the performance of the system.

Table 2.7 shows the goodness-of-fit (measured by  $\bar{\rho}^2$ ) for the three variants to the main model. Not surprisingly, the main model has the best fit (highest  $\bar{\rho}^2$ ) among the four models for each scenario. Interestingly, Variant # 1, although a simple model with Stickiness only, fits the data quite well for each scenario, which suggests that inertia plays an impor-

tant role in predicting the next day's choice based on all previous days' choices. Variant # 2 is as good as the main model in the No-Information scenario, while not as much in the Information scenario, suggesting that without real-time information, immediate response to incident is the dominating behavior for travelers taking the risky branch, which together with the inertia on the safe route explain much of the observed choices. When real-time information is available, Variant # 2 is worse than the main model, suggesting that real-time information alleviates the myopic response to an incident. Variant # 3 has the lowest  $\bar{\rho}^2$  in either scenario, suggesting that in a highly disrupted network, long-term learning might be a poor predictor of the immediate response, as the high level of variability might lead people to behave less rationally.

#### 2.4.2 Prediction Evaluation

In Section 2.4.1, model parameter estimates are obtained by maximizing the likelihood of predicting choice on day  $t$  given all previous days' choices up to day  $t - 1$ , which can be referred to one-step prediction, measured by adjusted Rho-square  $\bar{\rho}^2$ . Such a prediction is suitable for short-term operational applications as the data keeps coming in, e.g., providing personalized route guidance to a traveler as his/her behaviors are continuously monitored by, e.g., a smartphone.

However, a full-trajectory prediction is needed for long-term planning applications and policy evaluations, i.e., predicting choices on day  $t$  through  $t + K$  given previous days' choices up to day  $t - 1$ . Therefore, models are also compared in terms of full-trajectory prediction.

The estimated models are used to generate a sequence of 120 days' predicted route choices for each of the 64 participants, where initial route choices were randomly gener-

ated based on the observed proportion of path share on the first day. The accuracy of the prediction is measured by the Mean Square Distance (MSD) of the model's predicted proportion of risky routes (Highway and Detour combined) compared with the observed proportion of risky routes. The larger the MSD value, the worse the full-trajectory prediction. Specifically, the 120 days are divided into blocks of  $m$  trials and let  $NumBlock = 120/m$  denote the number of blocks. The MSD is then calculated as

$$MSD = \frac{1}{R} \sum_{r=1}^R \frac{\sum_{j=1}^{NumBlock} (P_j^{pred} - P_j^{obser})}{NumBlock} \quad (2.8)$$

where  $R$  is the number of simulations,  $P_j^{pred}$  and  $P_j^{obser}$  the average of predicted and observed proportion of risky route choices across all 64 participants in the  $j^{th}$  block for each simulation, respectively. The MSDs are averaged across blocks of 10, 20 and 30 days to verify the robustness of the comparison results, at a given value of  $R$  at 200. Table 2.7 presents the MSDs of all models for each scenario with  $R = 200$  and  $m = 5$ .

Table 2.7: Prediction Evaluation Results (One-step measured by  $\bar{\rho}^2$ ; Full-trajectory measured by MSD)

	<b>No-Information</b>		<b>Information</b>			
	Path Model		Path Model		Routing Policy Model	
	$\bar{\rho}^2$	MSD	$\bar{\rho}^2$	MSD	$\bar{\rho}^2$	MSD
<b>Main Model</b>	0.633	0.0187	0.543	0.0203461	0.647	0.00912
<b>Variant # 1</b>	0.608	0.0523	0.534	0.0502146	0.625	0.0503
<b>Variant # 2</b>	0.633	0.0402	0.543	0.0365489	0.623	0.0331
<b>Variant # 3</b>	0.273	0.0304	0.240	0.e0306	0.413	0.0298

The main model performs best in both one-step and full-trajectory prediction in both scenarios with the largest  $\bar{\rho}^2$  and smallest MSD. Variant #1, with a good one-step fit, has the largest MSD value, suggesting that inertia cannot explain long-term behavioral adjust-

ments and learning must be included. In other words, a model that predicts that you simply repeat yesterday's choice does well in the short term, due to the updated observations before each prediction. It however does not do well in the long term as it cannot capture the learning from experience over time. Variants#2 and #3 differ by the inclusion/exclusion of inertia and learning. Variant#2 performs better in one-step prediction than Variant #3 does, but worse in full-trajectory prediction. The reversed prediction performance suggests that learning is a critical factor in long-term prediction.

## 2.5 Summary

In this chapter, methodological development and empirical evidences are presented to understand the roles of learning, inertia and real-time travel information on route choices in a highly disrupted network based on data from a laboratory competitive route choice game. A learning model based on the power law of forgetting and reinforcement is applied. The learning of routing policies instead of simple paths is modeled when real-time travel information is available, where a routing policy is defined as a contingency plan that maps realized traffic conditions to path choices. A deterministic correction to the Logit model in a learning context is developed, generalizing the cross-sectional counterpart to account for overlapping routing policies.

Model parameter estimates are obtained from maximizing the likelihood of making the observed choices on the current day based on choices from all previous days. Prediction performance is then measured in terms of both one-step and full-trajectory predictions. Based on choices up to today, one-step prediction entails predicting the next day's choice, while full-trajectory prediction entails predicting the next  $K$  days' choices. Three ma-

jor conclusions are drawn. First, the routing policy learning model can capture travelers' learning and choice behavior better than a path-based model under real-time travel information, as it accounts for travelers' forward-looking capabilities. Secondly, inertia exists where travelers stick to previously chosen routes and do not necessarily minimize travel time. Inertia plays a dominant role in one-step prediction, and a less important role in full-trajectory prediction, suggesting that learning is more important in longer term prediction. Third, relative importance of learning compared with inertia is more prominent in a less uncertain, but not close to deterministic, environment. Therefore decreasing uncertainty by providing real-time information could encourage learning and potentially more optimal decisions for individuals and the system.

A major direction for future research is to estimate and evaluate the proposed routing policy model using real world data. Such longitudinal data has become increasingly available with the prevalence of smartphone-based tracking. A major challenge in extending the study from the laboratory to real world is the generation of the choice set of routing policies, which is perceivably a very large set. Choice set generation methods as studied in Ding-Mastera et al. (2014) can be applied to generate static choice sets that do not change from day to day. Furthermore, a learning process can be operationalized to explicitly model the formation and adjustment of choice set (see, e.g., Han et al., 2011). The methods developed in Guevara et al. (Forthcoming) would be applied to correct for the endogeneity problem due to missing initial observations, a common problem in a longitudinal dataset.

## **CHAPTER 3**

### **A LINK-BASED RECURSIVE ROUTE CHOICE MODEL FOR STOCHASTIC AND TIME DEPENDENT NETWORKS**

This is a collaborative work with Dr.Tien Mai from Université de Montréal and CIR-RELT, Canada.

Predicting path choices in road networks is of importance in many transport applications. Discrete choice models based on the random utility maximization framework are often used for this purpose and the model parameters can be estimated by maximum likelihood using GPS trajectory data. The route choice modelling literature mainly focuses on networks where link or path attributes are static and deterministic even though most models can be extended to a deterministic and time-dependent case. Even in this simple network setting it is challenging to design models that can be consistently estimated and that can produce accurate forecasts in short computational time, in particular for large-scale networks. However, the assumption of static and deterministic networks is restrictive since many transport systems are inherently uncertain, for instance, due to variations in demand, incidents and weather.

In this Chapter, a recursive logit model is proposed in stochastic and time-dependent (STD) networks (i.e. link travel times are random variables with time-dependent distributions) which is considerably more challenging than the static and deterministic case

(Fosgerau et al., 2013b) . Estimation and prediction results are presented for a real case study in the city of Stockholm.

### **3.1 Literature review**

#### **3.1.1 Route choice in an STD network: reaction to information**

Real-time information is now common in traffic information systems and it allows travelers to adapt to actual traffic conditions and potentially mitigate the adverse effect of uncertainty. Interested readers may refer to (Ben-Elia et al., 2010; Ben-Elia et al., 2013; Ben-Elia and Avineri, 2015b; Li et al., 2017; Delle Site, 2018).

Modeling route choice in an STD network can be approached in two ways: as a path or as a routing policy. A path is a pre-specified set of concatenated links. Travelers who follow a path make decisions a priori and take a fixed set of links, regardless of the network conditions revealed during their trips. In contrast, travelers who follow a routing policy make decisions en route, depending on actual network conditions. A large number of models have been proposed, including the Multinomial Logit (MNL), and its corrections to deal with the overlapping problem, e.g., the C-Logit model (Cascetta et al., 1996) and Path Size Logit model (Ben-Akiva and Bierlaire, 1999a). More complicated models continue to be developed such as Multinomial Probit (Bolduc and Ben-Akiva, 1991), Error Component model (Frejinger and Bierlaire, 2007a), latent route choice model with network-free data (Bierlaire and Frejinger, 2008), model assuming a universal choice set estimated based on a sampling approach (Frejinger et al., 2009a), model assuming a universal choice set estimated through repeated link choices based on a dynamic discrete

choice approach (Fosgerau et al., 2013b), and cross-nested logit based on sampling of choice sets (Lai and Bierlaire, 2015b).

In contrast, a routing policy choice model was first developed in (Gao, 2005a) where the user may update his/her route choice at any node of the road network depending on traffic conditions, and imbedded in a dynamic traffic assignment model. In contrast, the least expected travel time path ignores the possible information collected during the trip, and thus is generally less effective than an optimal routing policy. Indeed, a path is a special routing policy, in which the next link for a given node is chosen a priori. (Gao et al., 2008b) studied two types of models that account for travelers' adaptation to real-time information: an adaptive path model and a routing policy choice model based on synthetic data and a simplified network. Empirical studies of the routing policy choice to this date have only been carried out with SP data (Razo and Gao, 2010; Razo and Gao, 2013a). Traffic prediction models where routing policy choices are assumed for travelers have also been studied using simplified networks either in an equilibrium context (Gao, 2012a) or a disequilibrium context (Boyer et al., 2015).

### **3.1.2 Computation challenge: choice set generation and estimation**

Based on the characteristics of route choice studies, we categorize our problem as a study on adaptive routing policy with real-time information in stochastic and time-dependent network. Ding-Mastera et al. (2015) estimated such a routing policy model with non-recursive model which requires prior choice set generation that involves repeated executions of the optimal routing policy algorithm. Even if the process is done only once for multiple simulations, the computer memory required to store generated routing policies could be a concern in very large networks.

Due to the stochasticity, multiple states could be reached given a current state and thus the expected utility needs to sum over all possible next states in the Bellman equation. The estimation of the RL model requires solving a system of non-linear equations within a dynamic programming framework, which is considerably more complex than solving linear equations than in the deterministic case.

While the non-recursive model can accommodate a wide range of systematic utility specifications, including non-additive attributes, the recursive model however by design can only accommodate additive attributes, which make extra efforts needed to include important attributes such as travel time variability (generally not additive when correlation among link travel times exist) and to employ non-linear utility function such as the prospect theory (Tversky and Kahneman, 1992). This property largely increases the difficulty to solve the policy model.

Other than the above computation complexity, the time-dependency in link travel time also adds complexity to algorithmic development for applications in large-scale networks. In order to make the node-based state transition happen between intervals, time needs to be discretized into small enough intervals which greatly expands the state space.

In this chapter, the above estimation challenge is addressed by developing a two-level states space representation and propose a decomposition method that allows to reduce the number of linear systems to be solved when tackling the DP problem of the RL model.

## 3.2 Recursive Logit for Stochastic and Time-Dependent Networks

### 3.2.1 Network Settings

The whole period of interests within a day is divided into discrete time intervals. Let  $t$  denote a time interval which is represented as a non-negative integer,  $t = \lfloor x/\delta \rfloor$ , where  $x$  is the original wall-clock time and  $\delta$  is the length of a time interval which is equal to the shortest link travel time.

Let  $\mathcal{G} = (\mathcal{N}, \mathcal{A}, \mathcal{T}, \mathcal{P})$  be an stochastic time-dependent (STD) network, where  $\mathcal{N}$  is the set of nodes,  $\mathcal{A}$  is the set of links,  $\mathcal{T}$  is the set of time intervals over the period of interest  $\{0, 1, \dots, |\mathcal{T}| - 1\}$ , and  $\mathcal{P}$  is the probabilistic description of link travel times. Beyond time period  $|\mathcal{T}| - 1$ , travel times are assumed to be static and deterministic. Each node  $k \in \mathcal{N}$  has a set of outgoing nodes  $A(k)$ . Similarly to the RL model in deterministic networks (Fosgerau et al., 2013b), an absorbing state is associated with destinations by adding a dummy node  $d$  to the network, and denote  $\tilde{\mathcal{N}} = \mathcal{N} \cup \{d\}$ .

Along the line of previous work (Gao and Chabini, 2006b), the probabilistic description of link travel times are represented by a joint distribution of time-dependent random variables. Define a support point as a distinctive value that a discrete random variable can take, or a distinctive vector of values that a discrete random vector can take depending on the context. The joint probability distribution  $\mathcal{P}$  of all link travel times is characterized by a set of support points:  $\mathcal{P} = \{v_1, v_2, \dots, v_R\}$ , where  $v_r$  is a vector with a dimension  $|\mathcal{T}| \times |\mathcal{A}|$ ,  $r = 1, 2, \dots, R$ , and  $R$  is the number of support points. The  $r^{\text{th}}$  support point has a probability of  $p_r$  and  $\sum_{r=1}^R p_r = 1$ . This choice is motivated by the fact that observed travel times are usually aggregated and stored in a discrete number  $\mathcal{T}$  of time segments for practical purposes (Rilett and Park, 2001).

In this study, the information is composed of link travel times and are revealed during the execution of the routes. We formulate the route choice problem under perfect online information (POI), i.e., the current-information includes travel times on all links up to the current decision time (Gao and Huang, 2012b).

Let  $\hat{\tau}_{kj}(t)$  denote the random travel time on link  $(k, j)$  at time  $t$ . At time  $t$ , the traveler has knowledge of the realizations of  $\hat{\tau}_{kj}(t)(\forall (k, j) \in \mathcal{A})$  until  $t$ . Each joint realization of the  $(t + 1) * |\mathcal{A}|$  random variables,  $\hat{\tau}_{kj}(t')(\forall (k, j) \in \mathcal{A}, t' \leq t)$ , corresponds to a unique subset of  $\{v_1, v_2, \dots, v_R\}$ . To represent the concept of information, event collection  $\mathbf{q}$  is introduced as a subset of support points that are compatible with the realized travel times,  $\mathbf{q} = \{v_r | \hat{\tau}_{kj}(t) = \pi_{kj}(t), \forall (k, j) \in \mathcal{A}, t' \leq t \text{ for a certain } t\}$ , where  $\pi_{kj}(t)$  is the realization of  $\hat{\tau}_{kj}(t)$ . As more information (i.e.  $t$  increases) is collected,  $q$  can split into multiple event collections. When  $q$  becomes a singleton, the network collapse to a deterministic network. Let  $\mathbf{q}(t)$  be the set of all possible event collections at time  $t$ , and the element of  $\mathbf{q}(t)$  is an event collection. Specifically,  $q(|\mathcal{T}| - 1) = \{\{v_1\}, \{v_2\}, \dots, \{v_R\}\}$ .

An illustrative example is shown in Figure 3.1 and Table 3.1 with three nodes, three links and three time periods ( $t=0,1,2$ ). Table 3.2 shows the scheme of event collections. Each cell in the represents a event collection. At time 0, there is only one possible event collection  $\mathbf{q}(0) = \{v_1, v_2, v_3\}$ , as travel times on all links are the same across the two support points at time 0. At time 1, , when more link travel time realizations are available, it splits into two event collections,  $\mathbf{q}(1) = \{v_1, v_2, v_3\}$ . At time 2, similarly,  $\{v_2, v_3\}$  further splits into two event collections,  $\{v_2\}$  and  $\{v_3\}$  and there are three event collections in total. Each event collection contains a single support point, implying that the network becomes deterministic beyond time period 2.

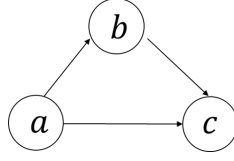


Figure 3.1: An illustrative small network

Table 3.1: The joint probability distribution of all link travel times ( $p_1 = p_2 = p_3 = \frac{1}{3}$ )

Time	Link	$v_1$	$v_2$	$v_3$
0	(a, b)	1	1	1
0	(a, c)	2	2	2
0	(b, c)	1	1	1
1	(a, b)	1	2	2
1	(a, c)	3	1	1
1	(b, c)	2	2	2
2	(a, b)	1	2	1
2	(a, c)	3	1	2
2	(b, c)	2	2	1

### 3.2.2 Recursive logit model

Define the set of states by  $S$  and a state in  $S$  by  $s = (k, t, \mathbf{q})$ , where  $k$  denotes current node,  $t$  denotes current time and  $\mathbf{q}$  denotes current information set.

Consider an individual traveler travelling from an origin to a destination  $d$ . The traveler starts from a state  $(k, t, \mathbf{q})$  and reached the next state by choosing an action  $a$  (next node) from the set of outgoing nodes  $A(k)$ . The traveler will also have a potentially different event collection at node  $a$ , which accounts for realized link travel times between  $t$  and the arrival time at the end of link  $(k, a)$ . The arrival time  $t'$  at node  $a$  is uniquely determined by  $\mathbf{q}'$ , and vice versa. The traveler continues the routing decision process based on dynamically involved event collections.

Table 3.2: The scheme of the event collection

t=0	$v_1$	$v_2$	$v_3$
t=1	$v_1$	$v_2$	$v_3$
t=2	$v_1$	$v_2$	$v_3$

At each state  $s = (k, t, \mathbf{q}) \in S$ , an instantaneous utility,  $u(a|k, t, \mathbf{q}) = v(a|k, t, \mathbf{q}) + \mu\epsilon(a)$ , is associated with the action of choosing node  $a \in A(k)$ , where the random terms  $\epsilon(a)$  are assumed i.i.d. extreme value type I (also known as the Gumbel distribution) with zero mean and they are independent of everything else in the model with scale parameter  $\mu$ . The traveler aims at maximizing the sum of instantaneous utility  $u(a|k, t, \mathbf{q})$  and expected downstream utilities, which is defined by taking the continuation of this process into account via the Bellman's equation (Bellman, 1957).

$$V(k, t, \mathbf{q}) = \mathbb{E} \left[ \max_{a \in A(k)} \left\{ v(a|k, t, \mathbf{q}) + \rho \sum_{\mathbf{q}' \in \mathbf{q}(t')} V(a, t', \mathbf{q}') P_r(\mathbf{q}' | \mathbf{q}) + \mu\epsilon(a) \right\} \right], \quad (3.1)$$

where  $\rho > 0$  is a discount factor,  $t'$  is the arrival time at the chosen node  $a$ , and  $t' = t + \tau(a|k, t, \mathbf{q})$ , where  $\tau(a|k, t, \mathbf{q})$  is the travel time between node  $k$  and node  $a$  conditional on state  $(k, t, \mathbf{q})$ , in which  $\tau(a|k, t, \mathbf{q})$  is deterministically specified based on the POI. Moreover,  $\mathbf{q}'$  is one of the possible event collection at  $t'$  and  $\mathbf{q}(t')$  is the set of all possible event collections at  $t'$ . Note that at the destination  $V(d, t, \mathbf{q}) = 0, \forall t, \mathbf{q}$ , where  $d$  is the destination of the trip.

The probability of transforming to  $\mathbf{q}'$  from  $\mathbf{q}$ ,  $P(\mathbf{q}' | \mathbf{q})$ , is computed as

$$P(\mathbf{q}' | \mathbf{q}) = \frac{\sum_{v_r | v_r \in \mathbf{q}' \cap \mathbf{q}} p_r}{\sum_{v_r | v_r \in \mathbf{q}} p_r}, \quad (3.2)$$

where  $p_r$  is the probability of support point  $v_r$ .

The value functions are then given recursively by a logsum according to the Gumbel distribution property, as shown in Eq.(3.3).

$$V(k, t, \mathbf{q}) = \begin{cases} \mu \ln \left( \sum_{a \in A(k)} \exp \left( \frac{1}{\mu} (v(a|k, t, \mathbf{q}) + \rho \sum_{\mathbf{q}' \in \mathbf{q}(t')} V(a, t', \mathbf{q}') P_r(\mathbf{q}' | \mathbf{q})) \right) \right) & \forall (k, t, \mathbf{q}) \in S, k \neq d \\ 0 & k = d \end{cases} \quad (3.3)$$

The transition probability between two states  $(a, t', \mathbf{q}')$  and  $(k, t, \mathbf{q})$  is the product of the probability of transforming to  $\mathbf{q}'$  from  $\mathbf{q}$  and the probability that the traveler chooses node  $a$  from state  $(k, t, \mathbf{q})$ . The probability that the traveler chooses node  $a$  from state  $(k, t, \mathbf{q})$  is given by a multinomial logit model. And we can further simplify the denominator, and represent it by the value function according to Eq.(3.3).

$$\begin{aligned} P(a, t', \mathbf{q}' | k, t, \mathbf{q}) &= \frac{\exp \left( \frac{1}{\mu} (v(a|k, t, \mathbf{q}) + \rho \sum_{\mathbf{q}' \in \mathbf{q}(t')} V(a, t', \mathbf{q}') P_r(\mathbf{q}' | \mathbf{q})) \right)}{\sum_{a' \in A(k)} \exp \left( \frac{1}{\mu} (v(a'|k, t, \mathbf{q}) + \rho \sum_{\mathbf{q}' \in \mathbf{q}(t')} V(a', t', \mathbf{q}') P_r(\mathbf{q}' | \mathbf{q})) \right)} P_r(\mathbf{q}' | \mathbf{q}) \\ &= \exp \left[ \frac{1}{\mu} \left( v(a|k, t, \mathbf{q}) + \rho \sum_{\mathbf{q}' \in \mathbf{q}(t')} V(a, t', \mathbf{q}') P_r(\mathbf{q}' | \mathbf{q}) - V(k, t, \mathbf{q}) \right) \right] P_r(\mathbf{q}' | \mathbf{q}). \end{aligned} \quad (3.4)$$

Given an observations  $\sigma = \{(k_1, t_1, \mathbf{q}_1), \dots, (k_I, t_I, \mathbf{q}_I)\}$ , the probability of  $\sigma$  is

$$P(\sigma) = \prod_{i=1}^{I-1} P(k_{i+1}, t_{i+1}, \mathbf{q}_{i+1} | k_i, t_i, \mathbf{q}_i).$$

It is important to note that, in the deterministic RL model presented in (Fosgerau et al., 2013a), the value functions in the path probability cancel out and the IIA property holds. However in stochastic RL model, when we take the log of the state probability in (3.4), the value functions do not cancel out, so the IIA does not hold.

The log-likelihood function could be derived as follows

$$LL(\beta) = \frac{1}{N_{obs}} \sum_{n=1}^{N_{obs}} \sum_{i=0}^{I_n-1} \ln P(k_{i+1}^n, t_{i+1}^n, \mathbf{q}_{i+1}^n | k_i^n, t_i^n, \mathbf{q}_i^n). \quad (3.5)$$

The first order derivatives are required for an efficient optimization algorithm for estimating the model parameters. They can be obtained by taking the first derivative of (3.5) and computed via the first derivatives of the probabilities  $P(k_{i+1}^n, t_{i+1}^n, \mathbf{q}_{i+1}^n | k_i^n, t_i^n, \mathbf{q}_i^n)$ . The gradient of  $P(a, t', \mathbf{q}' | k, t, \mathbf{q})$  with respect to a parameter  $\beta$  is

$$\begin{aligned} & \frac{\partial \ln P(a, t', \mathbf{q}' | k, t, \mathbf{q})}{\partial \beta} \\ &= \frac{1}{\mu} \left( -\frac{\partial V(k, t, \mathbf{q})}{\partial \beta} + \frac{\partial v(a | k, t, \mathbf{q})}{\partial \beta} + \rho \sum_{\mathbf{q}' \in \mathbf{q}(t')} \frac{\partial V(a, t', \mathbf{q}')}{\partial \beta} P_r(\mathbf{q}' | \mathbf{q}) \right). \end{aligned} \quad (3.6)$$

Thus, the value  $V(k, t, \mathbf{q})$  and  $\frac{\partial V(k, t, \mathbf{q})}{\partial \beta}$  are needed to compute the log-likelihood and its gradients. We discuss the computation of the value functions as well as their gradients in the following section.

The resulting model is expensive to estimate and apply because the state space is large and the destination and state specific value functions need to be solved. In the following

a decomposition algorithm is presented in which the state space is decomposed into two sub-set of states, one describing the physical network and the deterministic attributes and another one describing the probabilistic transitions between states (event collection and travel time distributions). This allows to efficiently reformulate and simplify the computation of the value functions as well as the state choice probabilities.

### **3.3 Decomposition Method for the Maximum Likelihood Estimation**

The nested fixed point algorithm proposed by (Rust, 1987) is used to solve the recursive model (RL). The general idea is to combine an outer iterative nonlinear optimization algorithm for searching over the parameter space with an inner algorithm for solving the expected maximum utilities (or the value functions). As mentioned, compared to other recursive route choice models proposed previously (Fosgerau et al., 2013b; Mai et al., 2015c; Mai, 2015), the most challenging issue when estimating the RL in STD networks comes from the stochasticity of the network. More precisely, Bellman equation in (3.1) becomes complicated to be solved. In the following, a way to decompose the computation of the value functions into several simple steps is proposed, so that the Bellman equation can be solved efficiently.

The computations of the value functions and their gradients is first formulated, then these computations is decomposed into sequences of simple matrix operations, which allows to quickly compute the log-likelihood function and estimate the RL model.

### 3.3.1 Computation of the Value Functions

Consider the stochastic network with the set of states  $S = \{(k, t, \mathbf{q}) \mid k \in \mathcal{N}, t \in \mathbf{T}, \mathbf{q} \in \mathbf{Q}\}$ . Let also denote by  $S^d$  the set  $\{(d, t, \mathbf{q}) \mid \forall t, \mathbf{q}\}$ , set of destination states. For each node  $s = (k, t, \mathbf{q}) \in S$ , define mappings  $n(s) = k \in \mathcal{N}$ ,  $\pi(s) = t$ , and  $i(s) = (\mathbf{q})$ , i.e.,  $n(s)$  refers to the physical node in the transportation network associated with state  $s$ ,  $\pi(s)$  refers to the time, and  $i(s)$  refers to the travel information  $\mathbf{q} \in \mathbf{Q}$ . With these mapping, define the set of arcs in the stochastic network as  $\mathcal{A}^S = \{(s, s') \mid n(s') \in A(n(s)), P_r(i(s') \mid i(s)) > 0\}$ . In other words, two nodes  $s$  and  $s'$  in  $S$  are connected if their corresponding physical nodes are connected, and the transforming probability between  $i(s)$  and  $i(s')$  is greater than zero. Moreover, define the transforming probabilities between two states  $s, s' \in S$  as

$$P_r(s' \mid s) = \begin{cases} P_r(i(s) \mid i(s')) & \text{if } (s, s') \in \mathcal{A}^S \\ 0 & \text{otherwise .} \end{cases}$$

The Bellman equation can be written as

$$V(s) = \begin{cases} \mu \ln \left( \sum_{\substack{a \in A(k) \\ k=t(s)}} \exp \left( \frac{1}{\mu} v(a \mid s) + \frac{\rho}{\mu} \sum_{\substack{s' \in S \\ n(s')=a}} V(s') P_r(s' \mid s) \right) \right) & s \in S \setminus S^d \\ 0 & s \in S^d. \end{cases}$$

If define a vector  $Z$  of size  $|S|$  with element  $Z_s = \exp(\frac{V(s)}{\mu})$ ,  $\forall s' \in S$ , the value of  $Z$  can be computed by solving the following recursive equation

$$Z_s = \begin{cases} \sum_{\substack{a \in A(k) \\ k=t(s)}} \exp \left( \frac{1}{\mu} v(a \mid s) \right) \prod_{\substack{s' \in S \\ n(s')=a}} Z_{s'}^{\rho P_r(s' \mid s)} & \text{if } s \in S \setminus S^d \\ 1 & \text{if } s \in S^d. \end{cases} \quad (3.7)$$

Now define matrices  $M$  and  $U$  of size  $|S| \times |\tilde{\mathcal{N}}|$  with entries

$$M_{sa} = \exp\left(\frac{1}{\mu}v(a|s)\right), \quad U_{sa} = \prod_{\substack{s' \in S \\ n(s')=a}} Z_{s'}^{\rho P_r(s'|s)}, \quad \forall s \in S, a \in \tilde{\mathcal{N}}, \quad (3.8)$$

and a vector  $b$  of size  $|S|$  with zero values for states  $s \in S \setminus S^d$  and one values for destination ones, i.e.,  $b_s = 1$  if  $s \in S^d$ . The recursive formulation in (3.7) can be written as

$$Z = (M \circ U)e + b, \quad (3.9)$$

where  $e$  is a vector of size  $(|\tilde{\mathcal{N}}|)$  with unit entries, and  $\circ$  is the element-by-element product. Based on this recursive equation, the value of  $Z$  can be estimated by iteratively perform (3.9) until we get a fixed point solution. More precisely, at the beginning (iteration  $l = 0$ ), we start with an initial value  $Z^0$ . At iteration  $l$ , suppose that the current value of vector  $Z$  is  $Z^{(l)}$ , we update the next value  $Z^{l+1}$  by performing the two following steps.

1. Compute  $U^{(l)}$  by (3.8), using  $Z^{(l)}$ ,
2.  $Z^{(l+1)} \leftarrow (M \circ U^{(l)})e + b$ .

The iterative process is stopped when a fixed point solution is found, i.e.,  $\|Z^{(l+1)} - Z^{(l)}\| \leq \gamma$  for a given threshold  $\gamma > 0$ . It can be shown that if the discount factor  $\rho < 1$ , then the method converges to an fixed point solution after a finite number of iterations (Rust, 1987).

### 3.3.2 First Order Derivative

The first-order derivation of the log-likelihood function is critical for the maximum likelihood estimation. In this section, the gradients of the value functions as well as the

log-likelihood function is derived. In particular, similar to the computation of the value functions, the computation of the gradients is formulated as matrix operations, which then can be decomposed into a sequence of simpler steps.

First the analytical formula of  $\partial Z_s / \partial \beta$ ,  $s \in S$  is derived, with respect to the model parameters  $\beta$ . From (3.9)

$$Z_s = b_s + \sum_{\substack{a \in A(k) \\ k=t(s)}} e^{\frac{1}{\mu} v(a|s)} \prod_{\substack{s' \in S \\ n(s')=a}} Z_{s'}^{\rho P_r(s'|s)}, \quad \forall s \in S. \quad (3.10)$$

Define a matrix  $K(|S| \times |\tilde{\mathcal{N}}|)$  with entries

$$K_{sa} = e^{\frac{1}{\mu} v(a|s)} \prod_{\substack{s' \in S \\ n(s')=a}} Z_{s'}^{\rho P_r(s'|s)}, \quad \forall s \in S, a \in \tilde{\mathcal{N}}.$$

Then

$$\ln K_{sa} = \frac{1}{\mu} v(a|s) + \rho \sum_{\substack{s' \in S \\ n(s')=a}} P_r(s'|s) \ln Z_{s'}, \quad \forall s \in S, a \in \tilde{\mathcal{N}}. \quad (3.11)$$

Taking the first derivative of (3.11) with respect to a parameter  $\beta_q$  we obtain

$$\frac{\partial K_{sa}}{\partial \beta_q} = K_{sa} \left( \frac{1}{\mu} \frac{\partial v(a|s)}{\partial \beta_q} + \sum_{\substack{s' \in S \\ n(s')=a}} \frac{\rho P_r(s'|s)}{Z_{s'}} \frac{\partial Z_{s'}}{\partial \beta_q} \right). \quad (3.12)$$

Note that

$$Z_s = b_s + \sum_{\substack{a \in A(k) \\ k=t(s)}} K_{sa}, \quad \forall s \in S. \quad (3.13)$$

So if take the derivative of (3.13) and substitute to (3.12), then

$$\begin{aligned}\frac{\partial Z_s}{\partial \beta_q} &= \sum_{\substack{a \in A(k) \\ k=t(s)}} K_{sa} \left( \frac{1}{\mu} \frac{\partial v(a|s)}{\partial \beta_q} + \sum_{\substack{s' \in S \\ n(s')=a}} \frac{\rho P_r(s'|s)}{Z_{s'}} \frac{\partial Z_{s'}}{\partial \beta_q} \right) \\ &= \sum_{\substack{a \in A(k) \\ k=t(s)}} \frac{1}{\mu} K_{sa} \frac{\partial v(a|s)}{\partial \beta_q} + \sum_{s' \in S} \left( \frac{\rho K_{sa} P_r(s'|s)}{Z_{s'}} \right) \frac{\partial Z_{s'}}{\partial \beta_q}.\end{aligned}\tag{3.14}$$

So, if define two matrices  $J(|S| \times |\tilde{\mathcal{N}}|)$ ,  $D(|S| \times |\tilde{\mathcal{N}}|)$  with entries

$$J_{sa} = \frac{1}{\mu} K_{sa} \frac{\partial v(a|s)}{\partial \beta} \text{ and } D_{ss'} = \left( \frac{\rho K_{sa} P_r(s'|s)}{Z_{s'}} \right), \forall s, s' \in S, a \in \tilde{\mathcal{N}}\tag{3.15}$$

then (3.14) can be written in a matrix form as

$$\frac{\partial Z}{\partial \beta_q} = J e + D \frac{\partial Z}{\partial \beta_q},\tag{3.16}$$

or equivalently,

$$\frac{\partial Z}{\partial \beta_q} = (I - D)^{-1} J e\tag{3.17}$$

where  $I$  is the identity matrix. So the gradients of  $Z$  can be obtained by solving the system of linear equations in (3.17). Note that

$$\frac{\partial V(s)}{\partial \beta_q} = \frac{\mu}{Z_s} \frac{\partial Z_s}{\partial \beta_q}, \forall s \in S.$$

This allows to compute the gradient of the log-likelihood function through (3.5) and (3.6).

### 3.3.3 Decomposition-based Estimation Algorithm

The value functions and their gradients can be obtained by performing (3.9) and (3.16). However, these matrix operations are costly to perform, due to the large number of states. In this section, a way to decompose these operations into simpler ones is presented in order to accelerate the computations.

In order to perform the value iteration method and compute the gradients presented in the previous sections we need to compute several matrices. Some of them are easy to obtained, e.g., matrix  $M$ , but some are not straightforward to compute quickly, e.g.  $U$ ,  $K$ . The process to simplify these operations is given in the following. First, define a matrix  $F(|S| \times |S|)$  with entries

$$F_{ss'} = P_r(s'|s), \forall s, s' \in S,$$

and note that

$$\ln U_{sa} = \sum_{\substack{s' \in S \\ n(s')=a}} \rho P_r(s'|s) \ln Z_{s'}, \forall s \in S, a \in \tilde{\mathcal{N}}. \quad (3.18)$$

Moreover, to deal with the sum in (3.18) we define a matrix  $T$  of size  $|S| \times |\tilde{\mathcal{N}}|$  where its entries are

$$T_{sk} = \begin{cases} 1 & \text{if } k = n(s) \\ 0 & \text{otherwise} \end{cases} \quad \forall s \in S$$

and the other elements equals zeros. Three propositions are introduced to support the computations

**Proposition 1.** *Given a matrix  $X$  of size  $|S| \times |S|$ , and a matrix  $U(|S| \times |\tilde{\mathcal{N}}|)$  with entries*

$$U_{sa} = \sum_{\substack{s' \in S \\ n(s')=a}} X_{ss'}, \forall s \in S, a \in \tilde{\mathcal{N}}, \text{ then } U = X \times T.$$

*Proof.* Indeed, each element  $sa$  of matrix  $X \times T$  can be computed as

$$(X \times T)_{sa} = \sum_{w \in S} X_{sw} T_{wa}.$$

Moreover, according the the definition of  $T$  we have, given  $w \in S$ ,  $T_{wa} = 1$  if and only if  $n(w) = a$ , so, we have the following result

$$(X \times T)_{sa} = \sum_{w \in S, n(w)=a} X_{sw} = U_{sa}.$$

□

**Proposition 2.** *Given matrices  $X, Y$  of size  $|S| \times |S|$ , and a matrix  $U(|S| \times |\tilde{N}|)$  with entries  $X_{ss'} = U_{sa} Y_{ss'}$ ,  $\forall s, s' \in S$ ,  $a = n(s')$ , then  $X = (U \times T^T) \circ Y$ , where  $^T$  is the transpose operator.*

*Proof.* Each element  $ss'$  of  $(U \times T^T) \circ Y$  can be computed as

$$[(U \times T^T) \circ Y]_{ss'} = Y_{ss'} \sum_{a \in \tilde{N}} U_{sa} T_{as'}^T = Y_{ss'} U_{sa} T_{s'a} = Y_{ss'} U_{sa},$$

where  $a = n(s')$ .

□

**Proposition 3.** *Given matrices  $X, Y$  of size  $|S| \times |S|$ , and a vector  $b(|S|)$  with entries  $X_{ss'} = Y_{ss'} b_{s'}$ ,  $\forall s, s' \in S$ , then  $X = Y \times \text{diag}(b)$ , where  $\text{diag}(b)$  a square diagonal matrix with the elements of vector  $b$  on the main diagonal.*

*Proof.* Each element  $ss'$  of  $Y \times \text{diag}(b)$  can be computed as

$$[Y \times \text{diag}(b)]_{ss'} = \sum_{w \in S} Y_{sw} \text{diag}(b)_{ws'} = Y_{ss'} \text{diag}(b)_{s's'} = Y_{ss'} b_{s'}.$$

The result is obtained.  $\square$

Now, for the sake of simplicity, given a matrix  $X$  of size  $|S| \times |S|$ , denote by  $\ln(X)$  and  $\exp(X)$  two matrices of the same size with  $X$  with elements

$$[\exp(X)]_{ss'} = \begin{cases} \exp(X_{ss'}) & \text{if } (s, s') \in \mathcal{A}^S \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad [\ln(X)]_{ss'} = \begin{cases} \ln(X_{ss'}) & \text{if } (s, s') \in \mathcal{A}^S \\ 0 & \text{otherwise,} \end{cases}$$

and if  $X$  is a matrix of size  $|S| \times |\tilde{\mathcal{N}}|$ ,  $\ln(X)$  and  $\exp(X)$  are defined as

$$[\exp(X)]_{sa} = \begin{cases} \exp(X_{sa}) & \text{if } a \in A(n(s)) \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad [\ln(X)]_{sa} = \begin{cases} \ln(X_{sa}) & \text{if } a \in A(n(s)) \\ 0 & \text{otherwise,} \end{cases}$$

and if  $X$  is a vector of size  $|S|$ , then  $\ln(Z)$  is also of size  $|S|$  with entries  $[\ln(Z)]_s = \ln(Z_s)$ ,  $\forall s \in S$ . According to Propositions 1 and 3, the matrix  $\ln(U)$  defined in (3.18) can be computed as

$$\ln(U) = [F \times \text{diag}(\ln(Z))] \times T.$$

This equation allows to compute  $U$  as well as perform the value iteration quickly.

Now let's turn our attention to the computation of the gradients. Indeed, the matrices  $J$  and  $D$  are needed in order to obtain  $\frac{\partial Z}{\partial \beta_q}$ . Note that,  $K = M \circ U$ , so  $J$  can be written as

$$J = K \circ V_q = V_q \circ (M \circ U),$$

where  $V_q$  is a matrix of size  $|S| \times |\tilde{\mathcal{N}}|$  with entries  $\frac{\partial v(a|s)}{\partial \beta_q}$ . Moreover, according to (3.15), matrix  $D$  can be computed based on matrix  $K$  (of size  $|S| \times |\tilde{\mathcal{N}}|$ ) and matrices  $F$  (of size  $|S| \times |S|$ ) and vector  $Z(|S|)$ . So, using Propositions 2 and 3, the matrix  $D$  can be obtained through  $K$ ,  $F$ ,  $Z$  and  $T$  as follows

$$D = [(K \times T^T) \circ F] \times \text{diag}(1/Z),$$

where  $1/Z$  is a vector of size  $|S|$  with entries  $1/Z_s$ ,  $\forall s \in S$ . So, finally, all the computations can be formulated via some simple matrix operations, which are summarized in Algorithms 1 and 2. Algorithm 1 contains steps to compute the value function, and Algorithm 2 concerns the computation of the Jacobian of the value function. The algorithms can be implemented by performing sequences of simple matrix operations. The matrices, even are large in terms of size, but sparse. It is important to note that the computational complexity of sparse operations is proportional to the number of non-zero elements and the row and column sizes, but should be independent of the number of all the elements in the matrix.

Finally, the outer algorithm of the nested fixed point algorithm is briefly discussed. This algorithm is based on a nonlinear optimization algorithm, such as a trust region or line search algorithm (Nocedal and Wright, 2006). Rust (1988) suggest using the BHHH

---

**Algorithm 1** Solving Bellman equation

---

*# 1. Initializing the value functions and the discount factor*

$Z = Z^0; \rho \leq 1;$

*# 2. Computing the Value iteration*

**do**

$Z_{prev} = Z$

$Y = [F \times \text{diag}(\ln(Z))] \times T;$

$U = \exp(Y);$

$Z = (M \circ U)e + b;$

**while**  $\|Z_{prev} - Z\| < \gamma;$

$V = \mu \ln(Z);$ 

---

---

**Algorithm 2** Computing gradients

---

$Y = [F \times \text{diag}(\ln(Z))] \times T;$

$K = M \circ \exp(Y);$

$J = V_q \circ (M \circ U);$

$D = [(K \times T^T) \circ F] \times \text{diag}(1/Z);$

**foreach**  $q$  **do**

$\frac{\partial Z}{\partial \beta_q} = (I - D)^{-1} J e;$

$\frac{\partial V}{\partial \beta_q} = \mu \left( \frac{\partial Z}{\partial \beta_q} \right) \circ (1/Z);$

**end**

---

approximation (Berndt et al., 1974) to obtain an efficient nonlinear optimization. Indeed, once the value functions as well as their first derivatives can be computed, the BHHH approximation can be easily obtained. However, note that the performance of the BHHH approach may be poor when close to the optimum solution in case of misspecification (Bastin et al., 2005; Mai et al., 2014). Structured quasi-Newton techniques then can be used to address this issue (Mai et al., 2014).

## 3.4 Numerical Results

### 3.4.1 Network and Data

A subnetwork of Stockholm, Sweden is considered which includes the Arlanda airport area, northeast inner city, and the connecting corridor. The network has 2772 nodes, 5447 links (including 619 stochastic links). The data consists of 500 observations and the study time period is from 6h30 to 9h A.M. There are 30 breaking points. The trace generation process is time-based with data from November 1, 2012 through January 18, 2013, covering the time intervals of Mondays through Fridays, resulting in 56 days (support points). They are matched to the road network using a 4-step map-matching method designed for sparse Floating Car Data (FCD), which is data collected from traced vehicles that “float” with the traffic (Rahmani and Koutsopoulos, 2013). The time period is discretized by 10-sec. The numbers of states in the STD network range from 33,476 to 445,315, and the numbers of links are between 65,364 and 844,285.

A non-parametric method is used to compute the link travel times per time interval using the map-matched GPS data. For each road segment between a pair of GPS coordinates, the observed travel time (i.e., the difference between the time stamps) is decomposed to the traversed links proportionally to their free-flow speeds and overlapping lengths. The weighted average, where the weight reflects the overlap with both the considered link and other links, over observations from different vehicles within the same time interval is the estimated link travel time. The travel time estimation is performed for each time interval separately for each day in the data set, producing an empirical, joint travel time distribution. Please refer to (Rahmani et al., 2015) for detailed evaluation of the method. With the available data, there are link-day-interval combinations for which the travel time can-

not be estimated due to lack of observations. These missing values are filled in through a sequence of inter/extrapolation steps. Furthermore, unreasonably high or low link travel times are removed to produce reliable estimates.

A link is treated as deterministic when there is not enough variation of travel time over time and day, or not enough observations to derive reliable travel time estimates. In this case, a single mean travel time is estimated across all days and time intervals.

### 3.4.2 Model Specifications

Four different attributes are used to specify the instantaneous utility function. First, travel time  $TT(a|k, t, \mathbf{q})$  from node  $k \in \mathcal{N}$  to node  $a \in \mathcal{N}$ , conditional on information  $(t, \mathbf{q})$ . Second, the standard deviation of travel time  $sTT(a|k)$  on link  $(k, a) \in \mathcal{A}$ . Third,  $mTT(a|k)$  is the averaged travel time of link  $(k, a)$ . And fourth, link constant  $LC(a|k)$  that is equal to 1 for every  $(k, a) \in \mathcal{A}$ . Some additional variables are also defined to model the behavior of travelers with respect to the length and variation of the travel time, namely, dummy-long-trip  $DL(n)$  that is equal to 1 if the travel time of trip  $n$  is greater than 15 minutes, and 0 otherwise, and dummy-high-variance-travel-time variables  $DHV(a|k)$  that is equal to 1 if  $\frac{sTT(a|k)}{mTT(a|k)} > 0.2$ , and equal to 0 otherwise. The RL model is estimated using the following utility function associated with arc  $(a, k) \in \mathcal{A}$ , observation  $n$ , and information  $(t, \mathbf{q})$ .

$$\begin{aligned}
 v^n(a|k, t, \mathbf{q}) = & DL(n) \left( \beta_{TT}^1 TT(a|k, t, \mathbf{q}) DHV(a|k) + \beta_{TT}^2 mTT(a|k) (1 - DHV(a|k)) \right) \\
 & + (1 - DL(n)) \beta_{mTT} mTT(a|k, t, \mathbf{q}) + \beta_{LC} LC(a|k) + \beta_{PS} PS(a|k),
 \end{aligned} \tag{3.19}$$

where  $PS(a|k)$  is an attribute designed to correct the utilities for overlapping policies in a STD network. The attribute is similar to the Link Size (LS) attribute proposed in (Fosgerau et al., 2013b) for deterministic networks. This attribute is computed through the policy choice probabilities calculated based on the above instantaneous utility function, in which the term  $\beta_{PS}PS(a|k)$  is excluded, and with parameters  $[\tilde{\beta}_{TT}^1, \tilde{\beta}_{TT}^2, \tilde{\beta}_{mTT}, \tilde{\beta}_{LC}] = [-1.8, -3.7, -3.6, -1.0]$ . Note that the use of the attribute LC refers to the number of crossings on paths. The intuition behind the utility function in (3.19) is as follows. It is reasonable to assume that the travelers only take into consideration additional information  $(t, \mathbf{q})$  for long trips, and for arcs whose the travel times have high variances.

### 3.4.3 Estimation results

In this section, estimation results for the RL model are presented using the above STD network. It is important to note that, in these experiments, the RL model without discount factor become difficult to estimate. More precisely, if  $\rho = 1$ , the value iteration method is often not able to return fixed-point solutions. Therefore, we therefore estimate the RL model with  $\rho \leq 0.98$ . Table 3.3 reports the estimation results with  $\rho = 0.98, 0.95, 0.90$  and 0.85. Clearly, all the parameter estimates are significant, and all have their expected signs. The  $\beta$  estimates given by different discount factors are similar. Specifically,  $\beta_{TT}^1$  is more negative than  $\beta_{TT}^2$ , which indicates that travelers are more sensitive to travel time conditional on real-time information than mean travel time for long trips. Besides,  $\beta_{PS} < 0$  shows that travelers do not like crossings in general. Moreover, there are remarkable improvements in final log-likelihood when decreasing the discount factor. In other words, we obtain better models in in-sample fit with lower discount factors.

Table 3.3: Estimation results

	Discount factors			
Parameters	0.98	0.95	0.90	0.85
$\beta_{TT}^1$	-3.20	-3.19	-3.14	-3.11
Rob. Std. Err.	0.28	0.27	0.28	0.28
Rob. t-test(0)	-11.43	-11.91	-11.26	-11.11
$\beta_{TT}^2$	-1.16	-1.14	-1.11	-1.10
Rob. Std. Err.	0.22	0.22	0.21	0.21
Rob. t-test(0)	-5.27	-5.15	-5.33	-5.24
$\beta_{mTT}$	-4.21	-4.18	-3.98	-3.98
Rob. Std. Err.	0.33	0.32	0.36	0.35
Rob. t-test(0)	-12.93	-11.00	-11.37	-11.37
$\beta_{LC}$	-1.03	-1.02	-0.98	-0.98
Rob. Std. Err.	0.04	0.04	0.05	0.05
Rob. t-test(0)	-25.75	-23.01	-20.95	-19.60
$\beta_{PS}$	-0.12	-0.12	-0.12	-0.12
Rob. Std. Err.	0.02	0.02	0.02	0.02
Rob. t-test(0)	-6.00	-5.58	-5.63	-6.00
Log-likelihood	-3385.01	-3354.68	-3329.42	-3315.14

In the context of a sequential route choice model, discount factor represents travelers' degree of spatial cognition of networks as a parameter. The value of a discount factor is assumed to be between zero and one. A large value of discount factor means that drivers evaluate the future expected utility with great weight. When discount factor equals one, travelers evaluate the expected downstream utility and the instantaneous utility of the next link with equal weights, suggesting that travelers are perfectly looking ahead. On the other hand, when discount factor equals zero, travelers choose the next link based only on its instantaneous utility, suggesting that travelers are completely myopic.

In order to illustrate the effect of the discount factor on the final log-likelihood values, the model is estimated with  $\rho \in \{0.98, 0.96, \dots, 0.5\}$ , with a remark that with  $\rho \leq 0.5$  several numerical issues are encountered when solving the Bellman equation. Figure 3.2 plots the final log-likelihood values given by discount factors in the interval  $[0.5, 0.98]$ . Interestingly, the figure shows that the model becomes better in in-sample fit with lower discount factors. The final log-likelihood values are stable with  $\rho \in [0.5, 0.85]$ , but decrease dramatically with  $\rho = 0.85$  to  $0.98$ .

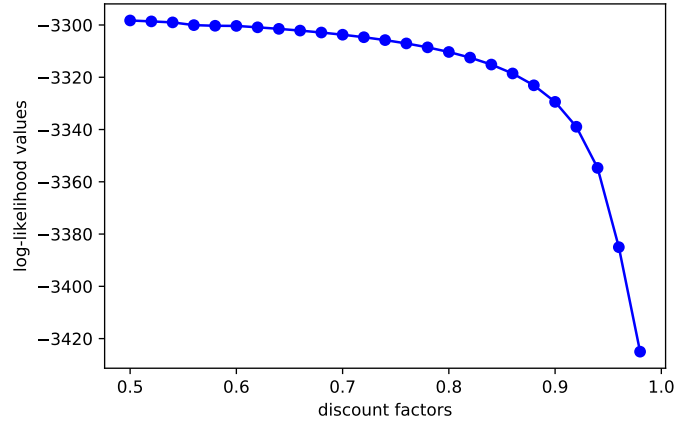


Figure 3.2: Final log-likelihood values with discount factors from 0.50 to 0.98

As discussed before, lower discount factor means that travelers are more myopic and place less weight on the the future downstream expected utilities. As a result, route choices could follow approximately random walk with extremely smaller discount factor, leading to circles during a trip. Next, whether the value of discount factor is reasonable or not is checked by looking at the number of cycles in the generated paths according to the RL model, as a measure of random walk.

The prediction performance of the RL model with two extreme discount factors, i.e.,  $\rho = 0.98$  and  $\rho = 0.5$  is studied. The model is estimated with the two discount factors, and the estimates are then used to simulate 200 paths per each observed origin-destination (OD) pair. In total, there are 100,000 paths are generated. Table 3.4 reports the number of simulated paths without and with cycles. Approximately 98% of the generated paths are cycle-free, for both  $\rho = 0.98$  and  $\rho = 0.50$ . This also means that there are only about 2% of the generated paths containing cycles. For  $\rho = 0.50$  there are a few paths of 4, 5 and 6 cycles are generated, and for  $\rho = 0.98$  there is no path of more than 2 cycles generated. The results indicate that, with a low discount factor of 0.5, the RL model still gives good simulation results.

Table 3.4: Number of simulated paths with cycles

Discount factor	N° simulated paths	N° paths with cycles						
		0	1	2	3	4	5	6
0.98	100,000	97,896	2082	21	1	0	0	0
0.50	100,000	98,020	1936	29	8	2	1	3

#### 3.4.4 Computational time results

The code is implemented in MATLAB 2015 and all computations are carried out on an Intel(R) 3.20 GHz machine with a x64-based processor. It is running with the Window 8 64-bit Operating System.

It requires approximately half an hour to compute a log-likelihood value, if the code is not parallelized. The optimization algorithm (i.e. the outer one) needs around 30 to 40 iterations to converge. So, it takes about 15 to 20 hours to estimate the model with

a non-parallelized code. It is worth mentioning that if we parallelize the code and use a machine of 8 CPUs to estimate the model, the estimation can be done in only 2 to 3 hours.

For solving the Bellman equation, the value iteration needs around 100 to 150 iterations to approximate fixed point solutions. The computational time needed to compute the value functions, as well as their gradients depends on the size of the STD network, and varies between 2 to 100 seconds. It is interesting to mention that if estimating the RL model (Fosgerau et al., 2013b) with the static and deterministic network, the estimation only requires about 20 seconds. The difference in computation time is, clearly, due to the fact that the size of the static network is smaller than the STD one, and solving the Bellman equation on the static RL can be done quickly by solving systems of linear equations, which is not the case with the stochastic RL model.

### **3.5 Summary**

This chapter investigates the routing policy choice problems in a stochastic time-dependent (STD) network. Firstly, a recursive logit model for STD networks is presented in which the probabilistic choice of the next link is modeled at each link, following the framework of dynamic discrete choice models. Next, an algorithm for solving the value functions that relies on matrix operations is proposed. Estimation and prediction results are then presented using data from a subnetwork situated in the Stockholm region, Sweden. Results show that the model can be estimated efficiently, and gives reasonable results for prediction.

## **CHAPTER 4**

### **A MARKOV DECISION PROCESS APPROACH TO VACANT TAXI ROUTING WITH E-HAILING**

Taxis play an important role in providing on-demand mobility. Compared to other forms of public transportation, the advantages of taxis include speediness, privacy, comfort, door-to-door service and longer operation hours with wide spatial coverage. Traditionally, vacant taxis cruise on roads searching for customers. In recent years, thanks to the proliferation of GPS-enabled smartphones, e-hailing applications (e.g., Uber, Lyft, and Didi Chuxing) are widely adopted by ride-sourcing drivers and in some cases, traditional taxi drivers (Didi Chuxing in China) to receive requests from nearby customers. The driver of a hired taxi usually aligns his/her routing objective with the passenger's, given the paramount importance of customer service (taking a detour to get higher fare is unprofessional and rare), and such a problem has been well studied. The interesting question is how to route vacant taxis. Taxis cruising on roads not only result in wasted gas and time for taxi drivers but also generate additional traffic in a city. Therefore, how to improve the utilization of taxis is of importance to both taxi drivers and the society.

In an earlier study by the co-authors (Hu, Gao, Chiu and Lin, 2012), a dynamic programming model of routing vacant taxis was proposed to depict the decisions at intersections according to the passenger arrival rate. However, the expected search time is only minimized for the next customer, which might be inefficient in the long run. For

example, driving to the airport might not minimize the search time for the next customer, but it brings in a higher chance of a long trip for the next customer and thus the profit might be higher overall. For this reason, experienced taxi drivers would not simply make their customer-search decisions depending on the current searching time/profit, but would also consider the subsequent possible states that could be encountered. The majority of vacant taxi routing studies in the literature, in fact, only considers an optimization problem until meeting the next customer (Zhang et al., 2015; Qu et al., 2014; Dong et al., 2014; Hwang et al., 2015; Huang et al., 2015), and in some cases, the revenue from the next customer (Yuan, Zheng, Zhang and Xie, 2013).

This study formulates the vacant taxi routing problem as a Markov Decision Process (MDP) so that long-term objectives can be taken into account instead of the immediate one of meeting the next customer. Some studies apply reinforcement learning (RL) (Sutton and Barto, 1998c) and adopt MDP formulations (Han et al., 2016; Verma et al., 2017), yet their treatments are usually not fully developed, in that important modeling issues such as e-hailing are ignored and space is highly aggregated. A typical RL algorithm also is purely data driven without taking advantage of the understanding of the underlying physical process, that is, no state transition probabilities are derived.

The remainder of this chapter is organized as follows. Section 4.1 provides a review of the literature. Section 5.1 formulates an MDP problem for the vacant taxi routing problem, defining states, actions and transition probabilities, followed by the presentation of an efficient solution algorithm in Section 4.3. Section 4.4 presents the computational experiments using GPS data from Shanghai, China to evaluate the merits of the proposed

methodology. Finally, Section 4.5 concludes the study and discusses potential directions for future work.

## **4.1 Literature Review**

### **4.1.1 Single Vacant Taxi Routing Problem**

The single vacant taxi routing problem is also known as the taxi recommender problem in the literature. The emergence of GPS tracking has facilitated the study of taxi routing problems, and massive taxi GPS datasets have attracted the attention of researchers from various fields with expected cross-fertilization of methods from transportation engineering, operations research, and computer science. The general problem statement is that given the location of a vacant taxi in an urban area, find the optimal decisions regarding its spatial movements.

Some studies recommend one or multiple pick-up locations without solving a network-based optimization problem. The attractiveness of a location with respect to the current location of the taxi is calculated using a number of factors, such as the distance between the current location and the pick-up location, expected revenue of trips from the pick-up location, waiting time for the next passenger at the pick-up location, and the probability of getting matched with a passenger on the way from the current location to the pick-up location. Some or all of the factors are manipulated to generate a single metric of attractiveness, and locations are ranked accordingly (Powell et al., 2011; Hwang et al., 2015; Zhang et al., 2016). The approach abstracts away the taxi cruising process on roads, and thus does not have a physically meaningful objective in the problem formulation, such

as maximizing profit or minimizing search time, although the evaluation of the methods is usually based on such measures.

Table 4.1 presents a list of studies (including the current one) where some form of network-based optimization is carried out. Note that the network can be either the physical road network, cell/zone-based or a number of locations connected by abstract links (see the “Network Representation” field).

There are three broad categories of taxi routing decisions (see the “Decisions” field):

- A route is defined as a sequence of connected links (physical road segments or abstract connections between locations) without metrics attached to the end node (labeled “Route” in the table). Different from a regular commuter routing problem, a vacant taxi does not have a definitive destination, and thus constraints are added to ensure that the optimal route does not become unrealistic, such as within a certain distance or time (Yuan, Zheng, Zhang and Xie, 2013; Dong et al., 2014), with a fixed number of segments (Qu et al., 2014; Huang et al., 2015), and with at least one expected passenger (Zhang et al., 2015).
- A route with metrics attached to the end node, the so-called “parking” place where taxies queue for passengers (hotels, transportation hubs) as in Yuan, Zheng, Zhang and Xie (2013). Metrics attached at the end node (such as revenue miles from the next passenger) are included in the optimization together with those from the route.
- A routing policy which maps any state (node) to an action (link) (Hu, Gao, Chiu and Lin, 2012; Han et al., 2016; Verma et al., 2017). An MDP formulation generates policies and in this paper, they are called routing policies. Note that the MDP

framework allows natural extension to include other important features of the transportation network that might affect the optimal routing, such as time-of-day, traffic condition, special events, by expanding the state space. All the MDP problems in the literature, as well as this study, are with infinite horizon, and thus either a terminal state is defined, such as finding a passenger (Hu, Gao, Chiu and Lin, 2012), or a discount factor of less than 1 is applied to returns from the future so that a convergent policy can be found (Han et al., 2016).

The "Planning Horizon" field in the table summarizes how far into the future a study considers the routing problem. Define a cycle as the vacant taxi trip from a drop-off location to the pick-up location for the next passenger, succeeded by the hired trip to the passengers' destination. Some studies consider only the first half cycle ("Half cycle"), that is, until a passenger is picked up, while a few consider the revenue from the next passenger ("One cycle"). Some studies do not have a clear probabilistic description of the cruising process ("N/A"). This study adopts a multi-cycle approach to account for the long-term profit, similar to studies based on RL (Han et al., 2016; Verma et al., 2017).

The matching of a vacant taxi and a passenger (see the "Matching Probability" field) is usually set simply as the observed fraction of matched taxis over all taxis present at a link/cell/zone, with the exception of Hu, Gao, Chiu and Lin (2012). Clustering sometimes is carried out to resolve the issue of data sparsity. Competition from other vacant taxis is modeled in Zhang et al. (2015) by accounting for relative locations of multiple taxis on the same link in calculating the matching probability, while Yuan, Zheng, Zhang and Xie (2013) use the queue length at the "parking" place as an indirect measure of competition, in that it is not factored in matching probability but used in optimization as an either objective

or constraint. This study uses a space Poisson distribution for vacant taxis and factors it directly in calculating matching probabilities.

The objective of the optimization problems also vary in the literature. Closely related to the planning horizon, when a half-cycle is considered, the objective function is usually the search travel time or distance, while when a full cycle is considered, revenue from the next passenger can be included. Studies without a clear model of the matching process (Qu et al., 2014; Dong et al., 2014) have objectives that are not well defined, despite that sometimes the name suggests otherwise. A planning horizon of multiple cycles allows for long-term objective to be included, and resolves the issue of the inconsistency between the optimization objective and evaluation criteria. Often times, a half-cycle or one cycle is adopted so that optimization can be done quickly, yet in evaluation, taxis are simulated for multiple cycles. The MDP formulation ensures that the correct objective is optimized.

Table 4.1: An overview of single vacant taxi routing studies with network optimization

	<b>Network Representation</b>	<b>Decisions</b>	<b>Planning Horizon</b>	<b>Matching Probability</b>	<b>Objective</b>
Hu, Gao, Chiu and Lin (2012)	Physical	Routing policy	Half cycle	Queuing theory; no competition	Minimal searching
Zhang et al. (2015)	Physical	Route	Half cycle	Empirical frequency; with competition	Shortest path with at least one expected passenger
Yuan, Zheng, Zhang and Xie (2013)	Physical	"Parking" place and the route to it	One cycle	Empirical frequency; competition by queue length	Combinations of 1) revenue miles per unit searching time 2) searching time, 3) pick-up probability 4) queue length at "parking" place
Qu et al. (2014)	Physical	Route	N/A	Empirical frequency	Maximal "profit" a fixed number of
Dong et al. (2014)	Physical	Route	N/A	N/A	Maximal score (related to revenue with distance cons
Huang et al. (2015)	Locations and their connections	A sequence of fixed number of pick-up points	Half cycle	Empirical frequency	Minimal searching
Verma et al. (2017)	Cell-based	Routing policy	Multi-cycle	Empirical frequency	Maximal revenue
<b>This Study</b>	Physical	Routing policy	Multi-cycle	Queuing theory; with competition	Maximal profit

#### 4.1.2 Other Related Work

In this section, related studies that are not directly addressing vacant taxi routing problem is reviewed. They either might provide insights into the formulation and solution of the taxi routing problem (e.g., the vehicle routing problem), or can provide support to a future extension of the current approach (taxi demand and destination prediction).

- **Vehicle Routing Problems** Models and algorithms developed for non-myopic vehicle routing problem (VRP) under uncertainty with look-ahead policies and rolling horizons (e.g., Mitrović-Minić et al., 2004; Thomas and White, 2004; Ferrucci et al., 2013) might provide insights for taxi routing problems in terms of accounting for future unknown demand and efficient solution algorithms. Thomas and White (2004) formulated a Markov Decision Process (MDP) in which known customers may ask for service with a known probability. Mitrović-Minić et al. (2004) included double-horizon heuristic that minimizes route distance for customers served in the near-term. Ferrucci et al. (2013) presented a tabu search approach for the delivery of newspapers and applied temporal and spatial clustering of future requests, assumed to be known as a time-space Poisson distribution, which guides vehicles into request-likely areas. It is however recognized that the taxi problem is different. In a typical VRP, the service of a customer does not bring the vehicle to another location, while a taxi does and the destination is not known until the request is taken. This significantly increases the geographic spread of taxi movements. In addition, a taxi (without carpooling service) can serve only one quest at one time and a new request does not come up until the old request is finished (unless a dispatcher is sending request during the previous ride).

- **Taxi Demand and Destination Prediction** Accounting for future states in taxi searching behavior requires sound models of geographic and temporal distributions of taxi demand and destination prediction. Several approaches have been proposed to predict taxi demand distribution which could be combined with the optimal taxi routing model including the time-series forecasting techniques such as the time-varying Poisson model and the autoregressive integrated moving average (ARIMA) (Li et al., 2012; Moreira-Matias et al., 2013), the multi-level clustering technique where demand over neighboring cells are aggregated, and the neural network based algorithms (e.g., Ke et al., 2017; Xu, Rahmatizadeh, Bölöni and Turgut, 2017). For instance, Zhao et al. (2016) implemented and compared three models, i.e., the Markov algorithm, Lempel-Ziv-Welch algorithm, and neural network. The results showed that neural network performed better with the lower theoretical maximum predictability while the Markov predictor had better performance with the higher theoretical maximum predictability. Some socio-demographical and built-environment variables have also been in use for predicting taxi passenger demand (Qian and Ukkusuri, 2015).
- **Taxi Driver Search Behavior** Another category of related work aims to model accurately the actually observed customer-search behavior of vacant taxi drivers. These studies provide comprehensive and quantitative insight into factors affecting taxi drivers' incomes and assist in developing effective optimization algorithms for taxi operations. Yang and Wong (1998) developed a model to determine the taxi movements on a given road network. Their study was further improved to capture congestion effects (Wong et al., 2001; Yang et al., 2005b), multiple user classes (Wong

et al., 2008), stochastic searching processes (Wong et al., 2005; Yang et al., 2010), day-to-day learning processes (Kim et al., 2005) and search frictions between vacant taxis and taxi customers (Yang and Yang, 2011; Yang et al., 2014). Wong, Szeto, Wong and Yang (2014) formulated and validated multinomial logit (MNL) models to predict vacant taxi drivers' zone choices for customer searching in both peak and off-peak hours. Szeto et al. (2013) further extended the consideration to every hour in a day. Wong, Szeto and Wong (2014) formulated a cell-based network and modeled the local customer-search movement of vacant taxi drivers based on the probability of successfully meeting the next taxi customers. Wong et al. (2015a) proposed a sequential logit-based vacant taxi behavior model predicting searching paths as a sequence of choices of adjacent zones while heading to their designated zones as compared with the model of Wong, Szeto and Wong (2014). Wong et al. (2015b) further combined the two proposed models to a two-stage modeling approach, cell-based model for local (within zone) search decisions and ESL for zonal decisions to predict vacant taxi movements in searching for customers. Qin et al. (2017) categorize taxi drivers into three levels according to their revenue and develop a generalized multilevel ordered logit (GMOL) model to find the significant factors that influence revenue.

- **Others** A large number of studies have been conducted to better understand and improve the taxi market, focusing an array of topics, such as taxi equilibrium assignment analysis (e.g., Yang et al., 2005a; Yang et al., 2010; Yang and Yang, 2011; Long et al., 2017), taxi fleet dispatching systems (e.g., Seow et al., 2010; Hou et al., 2013; Lowalekar et al., 2016), ride-sharing/carpooling problems (e.g., Hosni

et al., 2014; Lee and Savelsbergh, 2015; Qian et al., 2017; Masoud and Jayakrishnan, 2017b), pricing in taxi/ride-sourcing market(e.g., Qian and Ukkusuri, 2017b; Zha et al., 2017), validating user equilibrium with taxi trajectory data (Xie et al., 2017) and route planning through taxi trajectory mining (Yang, Kwan, Pan, Wan and Zhou, 2017).

## 4.2 Formulation of the Non-myopic Optimal Taxi Routing Problem

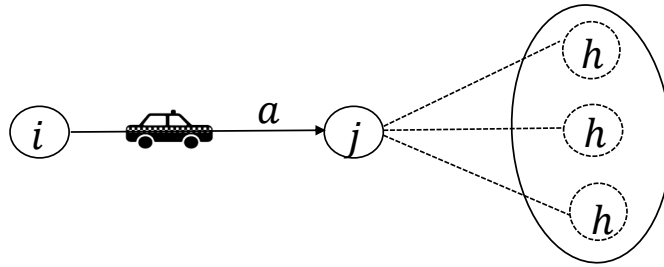
The defining characteristics and assumptions of the optimal taxi routing problem are:

- The vehicle routing problem is applicable for a single taxi;
- When a taxi is hired, the routing problem is reduced to a fastest path problem from the passenger's origin to destination and is not studied explicitly in this paper;
- Passenger arrivals and vacant taxi distribution are assumed independent;
- Passenger arrivals at different nodes are assumed independent;
- A passenger is matched with the nearest vacant taxi.

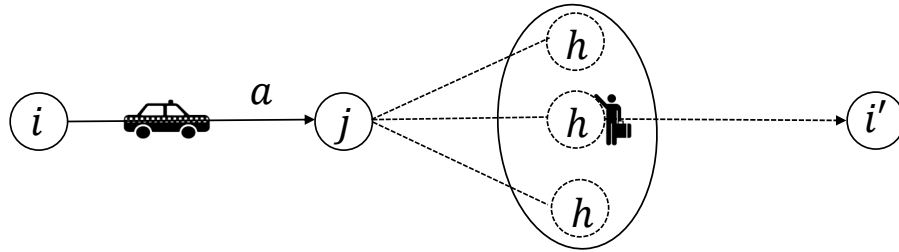
Accordingly, an MDP formulation for the taxi routing problem is presented.âĀĬ

### 4.2.1 States and Actions

A taxi driver's routing decisions over a time horizon on a given day is modeled as an MDP. A taxi travels in a traffic network  $G = (N, A)$ .  $N$  is the set of nodes and  $A$  the set of links. There is at most one directional link,  $a$ , from the source node  $i$  to sink node



(a) A vacant taxi starting from node  $i$  taking action  $a$  is not matched with any passenger while traversing link  $a$ . The next state is node  $j$ .



(b) A vacant taxi starting from node  $i$  taking action  $a$  is matched with a passenger at node  $h$  with a destination  $i'$  while traversing link  $a$ . The next state is node  $i'$ .

Figure 4.1: Illustration of the passenger matching process on a link

$j$ .  $A(i)$  is the set of downstream links of  $i$ . The taxi is actively searching for, or carrying passengers during a planning horizon.

When a taxi is hired, the routing problem is reduced to a fastest path problem from the passenger's origin to destination, and is not studied explicitly in this paper. The state of a vacant taxi is defined as the current node  $i \in N$ . The action set for state  $i$  is the set of outgoing links  $A(i)$ . For a given state  $i$  and action  $a \in A(i)$ , two types of transition to a new state  $i'$  could happen (see Fig. 4.1). a) The taxi is not matched with any passenger when traversing link  $a = (i, j)$ , and the next state is  $j$ , the sink node of  $a$ . b) The taxi is matched with a passenger at  $h$  (not necessarily the same as  $j$  due to e-hailing) when traversing link  $a$ , and the next state is the destination node of the passenger  $i'$ . To calculate state transition probabilities, the passenger matching probability on a link (Section 4.2.2) and passenger destination probabilities (Section 4.2.3) are needed.

#### 4.2.2 Passenger Arrival and Matching Probability on a Link

Passengers arrive at link  $a$  following a one-dimensional space-time Poisson process with rate  $\lambda_a$  per hour per mile. For modeling convenience, these are simplified as homogeneous time Poisson processes at each node with a constant average passenger demand rate, and the arrival rate at node  $j$  (per hour),  $\lambda_j = \sum_{a \in B(j)} \lambda_a l_a$ , where  $l_a$  is the length of link  $a$  and  $B(j)$  the set of incoming links. The combined process over all nodes is also a Poisson process with arrival rate  $\lambda = \sum_{j \in N} \lambda_j$ . In practice, demand rate  $\lambda_j$  is often approximated by observed met demand rate. Statistical analysis can be carried out to build a predictive model for the demand rate as a function of built environment variables (e.g., residential density, and employment by business type such as hotel and nightclub),

time of day, and weather condition (e.g., Phithakkitnukoon et al., 2010; Moreira-Matias et al., 2012) .

When e-hailing is used, it is assumed that the nearest vacant taxi to a passenger gets matched to the passenger. Vacant taxis around node  $n$  at any given point of time follow a two-dimensional spatial Poisson distribution with density  $\gamma_n$ . For a given node  $n$ , the probability of a vacant taxi  $r$  miles away (based on right-angle travel) being the nearest vacant taxi is the probability of no vacant taxi in a square (denoted by  $S$ ) rotated at 45 degree centered at node  $n$  with area equal to  $2r^2$  (Larson and Odoni, 1981), namely,

$$P_n(r) \{X(S) = 0\} = \exp(-2\gamma_n r^2). \quad (4.1)$$

where  $X(S)$  denotes the number of empty taxis contained in the square  $S$ .

The set of potential pick-up nodes is limited to those that are within a certain distance to the vacant taxi. Let  $N(j)$  denotes the sets of nodes within a certain matching distance  $R$  to node  $j$ . That is,  $R$  is the farthest distance between a passenger and a taxi where a request can go through. The combined process over all nodes in  $N(j)$  is also a Poisson process with arrival rate  $\lambda_{N(j)} = \sum_{j \in N} \lambda_j$ . The matching distance can be different for different areas, and it probably changes as a function of time as well. For example, during slow hours, drivers are willing to pick up a passenger who is far away. During busy hours, drivers are less likely to accept trips with long pick-up time. In this study, the rates are assumed to be static within a study period (say, 2 hours), and future research will address time-varying rates.

Consider a vacant taxi with e-hailing traversing link  $a$ . It gets matched with a passenger at node  $h$  when the following conditions are all satisfied.

- A passenger arrives at node  $h$  during the traversal time  $\tau_a$ ,
- The arrival at node  $h$  is earlier than arrivals on all other nodes in  $N(j)$ ,
- The taxi in question is the nearest vacant taxi to node  $h$ .

The probability of having at least one arrival from any node in  $N(j)$  during  $\tau_a$  is  $1 - \exp(-\lambda_{N(j)}\tau_a)$ . The probability that an arrival from node  $h$  is earlier than all other nodes is  $\frac{\lambda_h}{\lambda_{N(j)}}$  (Larson and Odoni, 1981). The product of the two probabilities is the probability that the earliest arrival during  $\tau_a$  happens at node  $h$ . The matching probability,  $p_{a,h}$ , is the product of the probability that the earliest arrival during  $\tau_a$  happens at node  $h$  and the probability that the taxi in question is the nearest vacant taxi to node  $h$ , namely,

$$p_{a,h} = \begin{cases} \frac{\lambda_h}{\lambda_{N(j)}} (1 - \exp(-\lambda_{N(j)}\tau_a)) \exp(-2\gamma_h \mathcal{L}_{a \rightarrow h}^2), & \text{if } h \in N(j) \\ 0, & \text{otherwise} \end{cases}, \quad (4.2)$$

where  $\mathcal{L}_{a \rightarrow h}$  is the right-angle distance from link  $a$  to node  $h$ , which can be approximated as the distance from the middle point of link  $a$ .

For those taxis that pick up passengers along the roads without e-hailing, it usually requires that the taxi and passenger to be no more than 1 block away from each other, thus the pick-up node set without e-hailing,  $N(j)$ , is a subset of the pick-up nodes with e-hailing.

#### 4.2.3 Passenger Destination Probabilities

The probability of a passenger picked up at node  $h$  having node  $k$  as the destination,  $p_{h \rightarrow k}$ , can be approximated by the observed fraction of passengers picked up at node  $h$

going to  $k$ . When no passenger pick-up is observed at node  $h$ , the probability is undefined. To resolve this issue, the study area is divided into zones such that any zone has strictly positive number of pick-ups. Let node  $h$  be in zone  $\mathcal{H}$  and node  $k$  in zone  $\mathcal{K}$ . Assume each node in zone  $\mathcal{K}$  has equal probability of being the destination node, and the destination probability is

$$p_{h \rightarrow k} = \begin{cases} \frac{p_{\mathcal{H} \rightarrow \mathcal{K}}}{m_{\mathcal{K}}}, & \forall \mathcal{H} \neq \mathcal{K} \\ \frac{p_{\mathcal{H} \rightarrow \mathcal{K}}}{m_{\mathcal{K}} - 1}, & \forall \mathcal{H} = \mathcal{K}, \forall h \neq k \\ 0, & \text{if } h = k \end{cases}, \quad (4.3)$$

where  $p_{\mathcal{H} \rightarrow \mathcal{K}}$  is the probability of a passenger picked up in zone  $\mathcal{H}$  having zone  $\mathcal{K}$  as the destination zone, and  $m_{\mathcal{K}}$  is the number of nodes in zone  $\mathcal{K}$ . The equal probability assumption can be easily relaxed.

The proposed modeling methodology could be applied to any study area using different sizes of zones. The sizes of the zones should be designed carefully based on the required level of modeling accuracy and the information available to the modeler. An unnecessarily large zone would mask traffic pattern differences that might be important for taxis finding customers. If the sizes were too small, the relevant data collected would be statistically unreliable, and the number of samples in each zone would be insufficient to provide representative means on the model parameters. In practice, it is advised to use the traffic analysis zones (TAZs) in a regional planning model as the basis for calculating passenger destination probabilities.

#### 4.2.4 State Transition Probabilities

For a given state  $i$  and action  $a \in A(i)$  with a sink node  $j$ , the transition probability,  $p_{ii'|a}$  is defined as follows:

$$p_{ii'|a} = \begin{cases} 1 - \sum_{h \in N(j)} p_{a,h} + \sum_{h \in N(j)} p_{a,h} p_{h \rightarrow j}, & \text{if } i' = j \\ \sum_{h \in N(j)} p_{a,h} p_{h \rightarrow i'}, & \text{if } i' \neq j \end{cases} \quad (4.4)$$

In the first case, the next state of the taxi is the sink node  $j$ . The probability of arriving at node  $j$  is the sum of the probability of arriving at  $j$  without getting matched and the probability of picking up a passenger from node  $h$  with destination  $j$ . In the second case, the next state is not the sink node  $j$ . In this case, a passenger from node  $h$  with destination  $i'$  ( $i' \neq j$ ) is matched, and the taxi arrives at node  $i'$  after picking up the passenger from node  $h$  and carrying the passenger from  $h$  to  $i'$ , both following shortest paths. The probabilities are summed over all possible  $h$ . The taxi continues the routing process after dropping off the passenger.

#### 4.2.5 Immediate Profit

It follows that the immediate profit of going from state  $i$  to  $i'$  given action  $a$  can be written as follows:

$$g_{ii'|a} = \begin{cases} \frac{-\alpha\tau_a(1 - \sum_{h \in N(j)} p_{a,h}) + \sum_{h \in N(j)} [F(d_{h \rightarrow j}) - \alpha(\tau_a + \mathcal{T}_{j \rightarrow h} + \mathcal{T}_{h \rightarrow j})] p_{a,h} p_{h \rightarrow j}}{1 - \sum_{h \in N(j)} p_{a,h} + \sum_{h \in N(j)} p_{a,h}}, & \text{if } i' = j \\ \frac{\sum_{h \in N(j)} [F(d_{h \rightarrow i'}) - \alpha(\tau_a + \mathcal{T}_{j \rightarrow h} + \mathcal{T}_{h \rightarrow i'})] p_{a,h} p_{h \rightarrow i'}}{\sum_{h \in N(j)} p_{a,h} p_{h \rightarrow i'}}, & \text{if } i' \neq j \end{cases} \quad (4.5)$$

where  $\alpha$  is the taxi operating cost per unit time, and  $\mathcal{T}_{j \rightarrow h}$  ( $\mathcal{T}_{h \rightarrow i'}$ ) is the fastest path travel time from node  $j$  to  $h$  ( $h$  to  $i'$ ).  $F(d_{h \rightarrow i'})$  is the taxi fare of an occupied trip from pick-up node  $h$  to destination node  $i'$ , where  $d_{h \rightarrow i'}$  is the occupied travel distance from node  $h$  to  $i'$ .  $F(d_{h \rightarrow i'})$  can be calculated by Eq. (4.6), a piecewise linear fare structure used in the study area of Shanghai in this study and can be adapted to other forms that depends on distance and/or travel time.

$$F(d_{h \rightarrow i'}) = \begin{cases} f_0, & \text{if } d_{h \rightarrow i'} \leq d_0 \\ f_0 + \beta(d_{h \rightarrow i'} - d_0), & \text{if } d_0 \leq d_{h \rightarrow i'} \leq d_1 \\ f_0 + \beta(d_1 - d_0) + \gamma(d_{h \rightarrow i'} - d_1), & \text{if } d_{h \rightarrow i'} \geq d_1 \end{cases}, \quad (4.6)$$

Similar to the state transition equation, the expected payoff is calculated for two different cases. In the case where the next state is  $j$ , the payoff is either the negative of the operating cost of traversing link  $a$ , which is  $-\alpha\tau_a$ , with the probability of not matched with a passenger,  $\sum_{h \in N(j)} p_{a,h}$ , or the taxi fare of going from  $h$  to  $j$  minus the operating cost of traversing  $a$ , going from  $j$  to  $h$  and from  $h$  to  $j$ , which is  $F(d_{h \rightarrow j}) - \alpha(\tau_a + \mathcal{T}_{j \rightarrow h} + \mathcal{T}_{h \rightarrow j})$ , with the probability of getting matched with a passenger whose destination is  $j$ . Note that the probabilities are normalized. In the second case where the taxi is matched with a passenger whose destination is  $i' \neq j$ , the same calculation of fare minus operating cost is carried out, with normalized probabilities.

#### 4.2.6 The Bellman Equation

Let  $V^*(i)$  denote the optimal expected payoff starting from state  $i$ . The taxi driver chooses the action at each state  $i$  to maximize the expected payoff that is the sum of the

expected immediate payoff and the expected downstream payoff, which is the expectation of the payoff over all possible next state  $i'$ . The optimal expected payoff is obtained by solving the Bellman equation (Bellman, 1957) as follows:

$$V^*(i) = \max_{a \in A(i)} \sum_{i' \in N} [g_{ii'|a} + \rho V^*(i')] p_{ii'|a}, \forall i \in N. \quad (4.7)$$

where  $\rho$  is discount factor, and  $0 \leq \rho \leq 1$ . Since this is an infinite horizon problem,  $\rho$  is set to be a number slightly smaller than 1 to ensure the existence of finite optimal expected payoff.

The optimal routing policy is then written as follows:

$$\mu^*(i) = \arg \max_{a \in A(i)} \sum_{i' \in N} [g_{ii'|a} + \rho V^*(i')] p_{ii'|a}, \forall i \in N. \quad (4.8)$$

### 4.3 Solving the Bellman equation

The Bellman equation can be solved by value iteration (Bellman, 1957).  $V^*(i)$  is termed the value function, and at each iteration, the value function at each state is updated by Eq. (4.7) where the value function estimates from the previous iteration are substituted into the right-hand side of the equation to obtain new estimates at the left-hand side. The time complexity is  $O(|A| \cdot |N|^2)$  per iteration, with  $|A|$  actions and  $|N|$  states.

Vectorizing batch operations avoid expensive for-loops and significantly improves computational performance (van der Walt et al., 2011). The Bellman equation is thus reformulated as a series of matrix operations.

The transition probability matrix,  $P(|NA| \times |N|)$ , and immediate payoff matrix,  $G(|NA| \times |N|)$ , are defined as follows:

$$P_{ia,i'} = \begin{cases} p_{ii'|a}, & \text{if } a \in A(i) \\ 0, & \text{otherwise} \end{cases}, \quad \forall i, i' \in N, a \in A \quad (4.9)$$

and

$$G_{ia,i'} = \begin{cases} g_{ii'|a}, & \text{if } a \in A(i) \\ 0, & \text{otherwise} \end{cases}, \quad \forall i, i' \in N, a \in A \quad (4.10)$$

Define an expected payoff vector  $V(|N| \times 1)$  with entry  $V_i$  as the expected payoff starting from state  $i, i \in N$ . According to Eq. (4.7), define

$$Y_{ia,1} = \begin{cases} \sum_{i'} g_{ii'|a} p_{ii'|a} + \rho \sum_{i'} V(i') p_{ii'|a}, & \text{if } a \in A(i) \\ 0, & \text{otherwise} \end{cases}, \quad \forall i, i' \in N, a \in A \quad (4.11)$$

To re-write Eq. (4.11) as matrix operations, define vector  $b(|N| \times 1)$  where  $b_{i,1} = 1$ ,  $\forall i \in N$ . Note that  $((G \circ P) \cdot b)_{ia,1} = \sum_{i'} g_{ii'|a} p_{ii'|a}$ , and  $(P \cdot V)_{ia,1} = \sum_{i'} V(i') p_{ii'|a}$ , where  $\circ$  is the element-by-element multiplication operator and  $\cdot$  the matrix multiplication operator. Eq. (4.11) then can be written as

$$Y = (G \circ P) \cdot b + \rho P \cdot V. \quad (4.12)$$

The max operator in Eq. (4.7) needs to operate on a matrix where a row represents a state and the columns represent expected payoffs from all feasible actions from that state

(all outgoing links from the node). Thus  $Y$  is reshaped into a matrix  $U(|N| \times |A|)$ , such that

$$U_{i,a} = Y_{ia,1} \quad (4.13)$$

Note that  $U_{i,a} = 0$ , if  $a \notin A(i)$ , that is,  $a$  is not an outgoing link of node  $i$ . The expected payoff of an outgoing link could be negative (taxi driver losing money), and if a max operator is directly applied to  $U_i$ , which takes the maximum over all columns, an infeasible action with an expected payoff of 0 could be chosen as the optimal. Therefore an exponential transformation is applied to  $U_{i,a}$ , if  $a \in A(i)$ . Define the operator  $\exp(U)$  such that

$$\exp(U)_{i,a} = \begin{cases} \exp(U_{i,a}), & \text{if } a \in A(i), \\ 0, & \text{otherwise.} \end{cases}, \quad \forall i \in N, a \in A, \quad (4.14)$$

and let  $W = \exp(U)$ . Eq. (4.7) can then be written as

$$V_i = \ln(\max_a(W_{i,a})), \forall i \in N, a \in A \quad (4.15)$$

Algorithm 3 shows the procedure for solving the Bellman equation via matrix operations. At each iteration, the expected payoff vector  $V$  is updated (Lines 6-9). The iteration is stopped when  $\|V_{prev}/V - 1\|_\infty < \epsilon$  for a given threshold  $\epsilon > 0$ , where  $\|\cdot\|_\infty$  is the maximum norm.

---

**Algorithm 3** Solving Bellman's equation

---

**Input:** transition probability matrix  $P$ , immediate payoff matrix  $G$ , all-ones vector  $b$  and convergence criteria  $\epsilon$

```
begin
  1. Initializing the value functions and the discount factor
   $V = 0$ ;  $0 \leq \rho \leq 1$ ;
  2. Value iteration
  do
     $V_{prev} = V$ ;
     $Y = (G \circ P) \cdot b + \rho P \cdot V$ ;
    Define matrix  $U(|N| \times |A|)$ , where  $U_{i,a} = Y_{ia}$ ,  $i \in N, a \in A$ ;
     $W = \exp(U)$ ;
     $V = \ln(\max_a W)$ ,  $\mu = \arg \max_a U$ ;
  while  $\|V_{prev}/V - 1\|_\infty < \epsilon$ ;
  return  $V, \mu$ 
end
```

---

## 4.4 Computational Tests

The objectives of the computational tests are three-fold:

- To demonstrate the efficiency of the solution algorithm in solving the proposed MDP problem in a real-life, large-scale network,
- To understand the differences between the proposed MDP and baseline heuristics in terms of unit profit and occupancy rate, and
- To understand the solution patterns of the proposed MDP problem, in relation to baseline heuristics.

#### **4.4.1 The Network, GPS Data and Experiment Setup**

##### **4.4.1.1 The Study Area and Network**

Shanghai is the most populated metropolitan area in China, with a land area of more than  $6300 \text{ km}^2$ , with a population of over 24 million. Shanghai urban area spreads broadly, including the city center landmark area (the Bund and People's square) and several other central areas. Shanghai roadway network is comprised of 13,531 nodes and 30,167 directed links, excluding connectors. The travel time on each link of the network is computed based on the speed limit by road type. Shortest paths and travel times between all nodes are pre-calculated and stored in a look-up table.

The study area is divided into 4,518 zones with smaller zones in populated urban areas and larger zones in suburban areas. The zones are small enough so that it is realistic to assume that all nodes within a destination zone have an equal probability of being the destination node, an assumption required in Eq. (4.3). Note that the states, actions and state transitions are network-based instead of zone-based. The purpose of the zones is to ensure a large enough sample size in calculating passenger destination probabilities.

##### **4.4.1.2 GPS Data**

GPS trajectories with 10-second gaps and indicators of hired vs. vacant status are available from one of the major taxi companies (market share of approximately 25%) in Shanghai. The market share is deemed sufficiently large to deduce the movements of taxis in general and provide adequate evaluation platform for the proposed methodology.

Data cleaning is carried out to remove obvious mistakes, for example, hired trips with exceptionally short travel distances over a long period, and very short travel times. The

mistakes could be due to the GPS device malfunctions, poor connectivity to satellites in the urban areas surrounded by highrise buildings, or human errors by taxi drivers who operated the devices. Occupied trips are eliminated if its distance is shorter than 500 *m*, and/or if its travel time is shorter than 1 min.

In Shanghai, taxi drivers work for one day and rest for one day. They usually shift between 5-6 am . Average daily working time is 14.8 hours (Lv et al., 2017). An inquiry (Qin et al., 2017) found that taxi drivers in Shanghai have a flexible meal schedule with varying length during the daytime (scattering across 11:00-14:00 and 16:00-19:30), it is thus very complicated to distinguish between the status waiting for passengers and the status taking a meal. We eventually select the period 5:30-11:30 am as our study time to ensure that it's continuous and long enough for making non-myopic routing decisions. It should be noted that the proposed methodology can be directly applied to longer study periods.

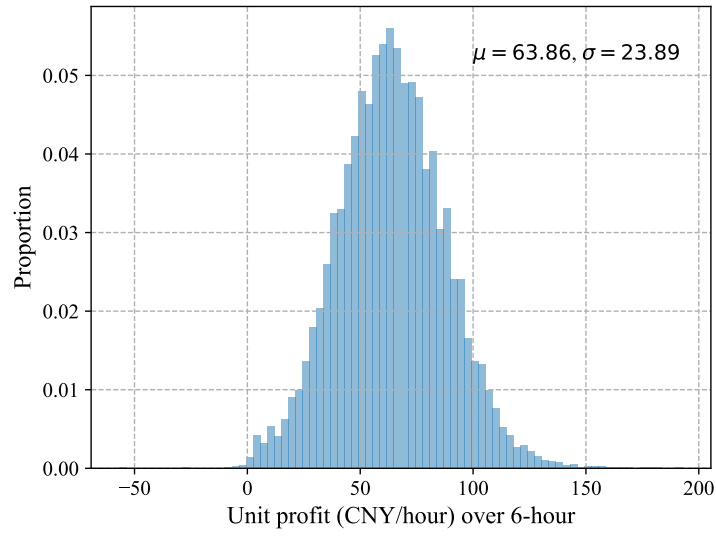
After data cleaning, trajectories from 12,017 taxis in 5:30-11:30 am on a representative weekday in April, 2015 are used for the case study. No special events or holidays which may introduce great trip variability were reported during the study period.

Taxi fares are charged based on distance traveled. The parameters for calculating taxi fare  $F(d_{h \rightarrow i'})$  in Eq. (4.6) are set as  $f_0 = 14$ ,  $d_0 = 3$ ,  $d_1 = 15$ ,  $\beta = 2.5$  and  $\gamma = 3.6$ .<sup>1</sup> The unit operating cost,  $\alpha$ , is assumed to be 0.5 CNY/min.

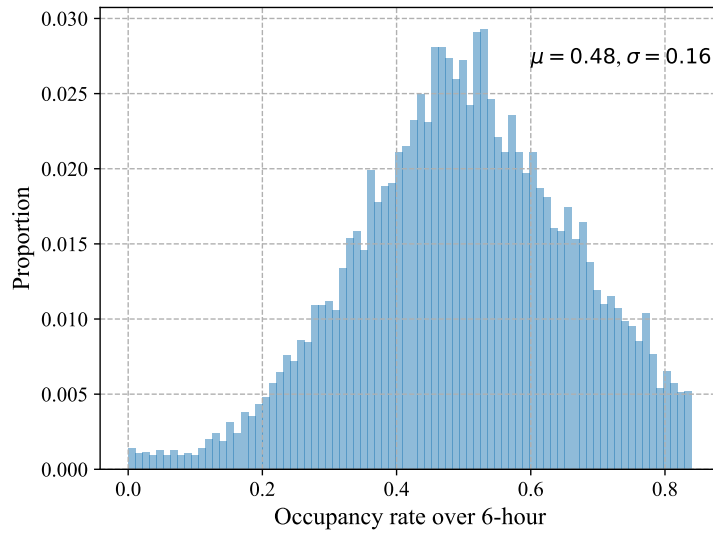
Fig. 4.2a shows the distribution of unit profit (CNY/hour) during the 6-hour study period with a mean of 63.86 CNY/hour and a standard deviation of 23.89 CNY/hour.

---

<sup>1</sup>source: <https://www.travelchinaguide.com/cityguides/shanghai/transportation/taxi.html>

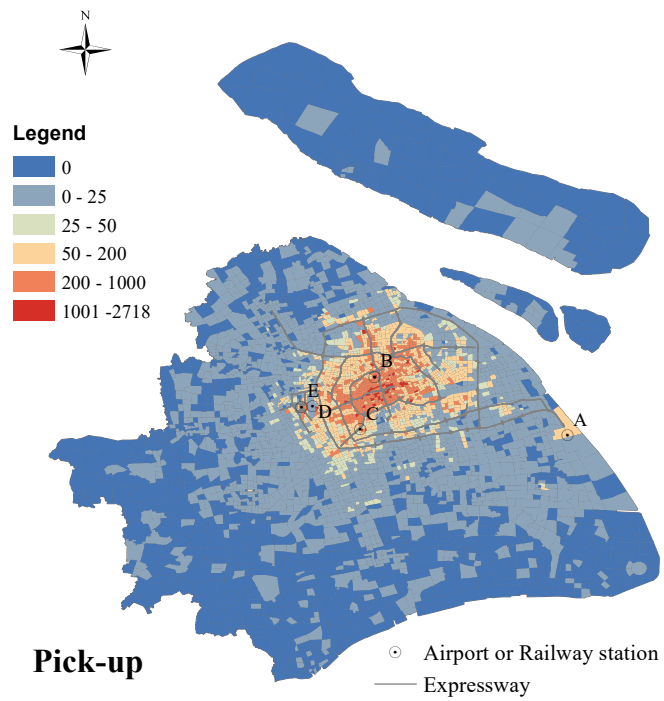


(a) Distribution of unit profit

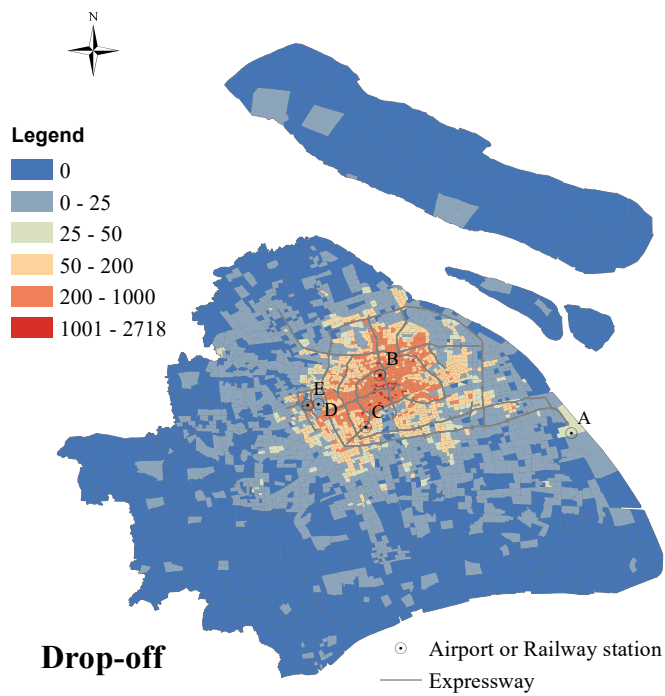


(b) Distribution of occupancy rate

Figure 4.2: Distribution of observed unit profit and occupancy rate 5:30-11:30 am



(a) Pick-up



(b) Drop-off

Figure 4.3: Pick-up and drop-off density 9:30-11:30 am on a weekday in April, 2015  
(count per  $km^2$ )

Profits are below 47.36 CNY for the lowest 25% of drivers, while above 80.03 CNY for highest 25%, indicating a significant profit difference across drivers. Fig. 4.2b shows the distribution of occupancy rate (the quotient between occupied time and the total working time) with a mean of 0.48 and a standard deviation of 0.16. Occupancy rates are below 0.39 for the lowest 25% of drivers, while above 0.61 for the highest 25%, showing a similar dispersion among drivers.

originate or end in regions containing airports or railway stations. There are very few trips on the islands and they are thus omitted in the remainder of the paper.

#### 4.4.1.3 Experiment Setup

The whole study period is divided into three 2-hour time intervals: 5:30-7:30 am, 7:30-9:30 am and 9:30-11:30 am.

For each interval, the passenger arrival rate  $\lambda(j)$  and vacant taxi density rate  $\gamma(j)$  at any node  $j$  are assumed time-invariant and calculated from the historical data. Passenger arrival rate at each node is set as the average count of pick-ups per hour during each time interval. The calculation of vacant taxi density rate is more involved as it is a time-wise average of spatial rate. A buffer is created for each node, which is a circle centered at the node with a radius of 500 m, approximately the 75th percentile of all link length. At any given time instance (say, 8:00am), the vacant taxi density for a node is the number of vacant taxis within the buffer at that time instance (a snapshot) divided by the area of the buffer. Multiple snapshots are taken every 15 minutes over each 2-hour time interval and the average over all snapshots is used as the vacant taxi density for the 2-hour period.

The matching node set  $N(j)$  is comprised of nodes within 1 km radius of node  $j$ , that is, the vacant taxi is eligible to be matched with passengers within 1 km crow-fly distance.

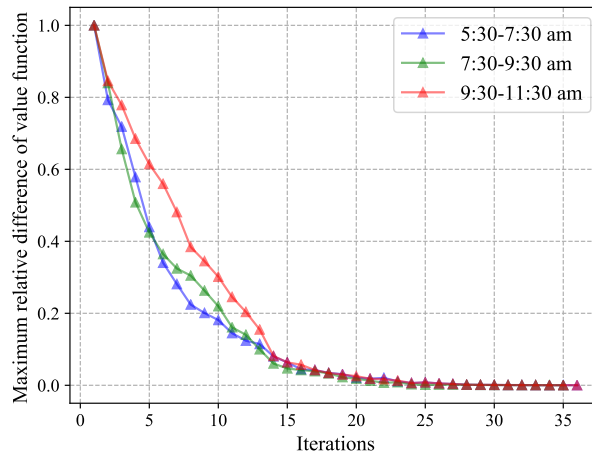
The matching radius is set such that both picking up along the roads and e-hailing at high demand density areas are accommodated. It is conceivably higher in a low demand area or period where drivers need to drive a relatively long distance to pick up passengers with e-hailing. Given that the data come from a traditional taxi company and the percentage of e-hailing is conceivably small, the radius is set to be relatively short.

Passenger destination probability is calculated directly based on Eq. (4.3) from the data for each 2-hour time interval.

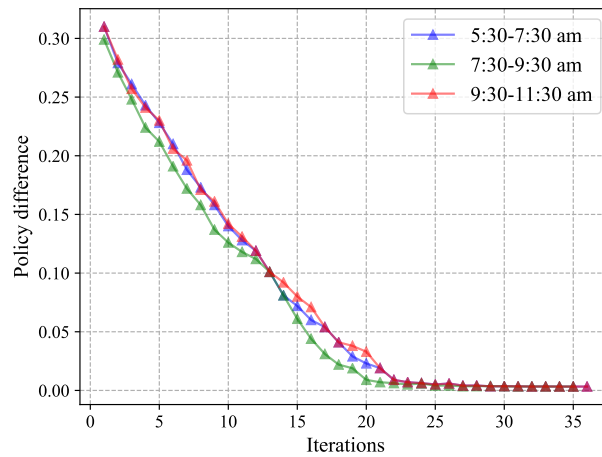
#### **4.4.2 Computational Performance**

The value iteration algorithm is coded in Python 3.5 with NumPy. All computations are carried out on a workstation with an eight-core 3.0GHz Xeon E5-1660 processor and 64GB RAM. The discount factor is set as 0.95. For each 2-hour time interval, the running time per iteration is about 16.15 min (969 sec) when for-loops are used, and is reduced to about 1.98 min (118 sec) with matrix operations (Section 4.3), an 8x speed-up. It is expected that the speed-up will be higher with more cores to process the matrix operations in parallel, and real-time efficiency can thus be achieved.

Fig. 4.4a shows the maximum relative difference of value function over successive iterations as a function of the number of iterations. The relative value function difference reaches below 0.01 at about the 25th iteration. Fig. 4.4b shows the relative change in routing policy as a function of the number of iterations. The policy change is calculated as the fraction of states whose optimal action changes from the last iteration. The relative policy difference reaches below 0.01 at about 22nd iteration. The convergence patterns are similar across time intervals.



(a) Value Function Difference vs. Iterations



(b) Policy Difference vs. Iterations

Figure 4.4: Convergence of the Value Iteration Algorithm

#### **4.4.3 Evaluation**

In this section, the optimal routing policy is compared with a number of baseline heuristics using two metrics: unit profit and occupancy rate. The metrics are generated by executing the optimal policy and heuristics in the network for the study period of 5:30 through 11:30 am, and the developed taxi/passenger matching model and passenger destination model are postulated as the true model based on which taxi and passenger matching is simulated. Comparisons between the optimal policy and heuristics provide insights into the value of the proposed method and are precursor to real world evaluation.

These comparisons aim to confirm the advantage of the proposed method over some commonly used heuristics, under the condition that perfect models of matching and passenger destination choice are available, which gives an indication of its potential performance in the real world, when such perfect models usually do not exist.

The comparison of the optimal routing policy with the observed taxi driver routing choices cannot be made until the optimal routing policy is implemented in the real world. Another way is to build a high-fidelity traffic simulation testbed where taxi driver behavioral models are calibrated using observed trajectory data, and then evaluate the optimal routing policy against the calibrated, simulated drivers in the simulation testbed. The comparison with real world drivers or simulated drivers calibrated against real world data is important in assessing the method's value. However, implementing the proposed method in the field or developing and calibrating a high-fidelity traffic simulation model is beyond the scope of the study, and left for future research.

#### 4.4.3.1 Heuristics

Three heuristics are defined as follows in increasing order of sophistication.

- **Random walk** A vacant taxi chooses an outgoing link randomly and once matched with a passenger, takes the fastest path to deliver the passenger. This is the simplest strategy.
- **Global hotspot** A vacant taxi heads toward the zone with the highest demand density (number of pick-ups per  $km^2$ ), following the fastest path to the central node of the zone (the node closet to the centroid of the zone), and do random walks in this zone until matched with a passenger. Once matched, it takes the fastest path to deliver the passenger. Note that the taxi could get matched on the way to the highest demand density zone.
- **Local hotspot** The global hotspot strategy can be inefficient when the taxi is far from the highest demand density zone. A more sensible strategy is to move to higher demand zones sequentially (see, e.g., taxi driver behavioral studies by Wong et al., 2015a; Wong et al., 2015b). This strategy can be viewed as a spatially aggregated, partially myopic approximation of the proposed optimal routing policy. Note that a majority of previous taxi routing studies (see, e.g., Yuan, Zheng, Zhang and Xie, 2013; Dong et al., 2014; Qu et al., 2014; Huang et al., 2015) recommend cruising routes with a maximum cruising distance or time, a constraint that is needed to avoid unrealistically long routes, necessitated by the myopic nature of their methods. The local hotspot heuristic is in the general family of partially myopic strategies.

The study area is divided into equal-sized square cells to represent roughly the range of a local area. The length of each cell is set at the 75<sup>th</sup> percentile of straight line distance from drop-off location to the next pick-up location (not the actual search distance), which equals 5 km approximately. A zone (which is smaller than a cell) is considered a member of a cell if its central node falls in the cell. The queuing-based model developed in Section 5.1 is applied to match a vacant taxi with passengers while the taxi moves following the Local hotspot strategy, and the Local hotspot strategy is implemented as follows:

Step 0: Calculate the demand density in each zone during each of the 2-hour period.

Step 1: For a vacant taxi start from node  $i$ , calculate the shortest path  $P_z$  (in terms of a sequence of nodes) from  $i$  to the centroid of the zone with the highest demand density in the current cell.

Step 2: Check if the taxi is matched with a passenger at each node while moving along path  $P_z$

Step 2-1: If the taxi is not matched with any passenger, the next state is the subsequent node on path  $P_z$ , and time  $t$  is updated.

Step 2-2: If the taxi is matched with a passenger during the movement, the next state is the destination of the passenger and time  $t$  is updated; go to Step 1.

Step 3: Let the taxi move randomly within the current zone for 15 min, and check if it is matched with a passenger at each node.

Step 3-1: If the taxi not matched with any passenger, the next state is the sink node of the link it takes, and time  $t$  is updated.

Step 3-2: If the taxi is matched with a passenger during the movement, the next state is the destination of the passenger and time  $t$  is updated; go to Step 1.

Step 4: Pick a neighboring cell  $C$  that contains the highest demand density zone  $Z$  among all zones in all neighboring cells. Calculate the shortest path  $P_c$  (in terms of a sequence of nodes) from current node to the centroid of zone  $Z$  with the highest demand density in the chosen cell  $C$ .

Step 5: Check if the taxi is matched with a passenger at each node while moving along path  $P_c$ .

Step 5-1: If the taxi is not matched with any passenger, the next state is the subsequent node on path  $P_c$ , and time  $t$  is updated.

Step 5-2: If the taxi is matched with a passenger, the next state is the destination of the passenger and time  $t$  is updated; go to Step 1.

Step 6: Steps 1-5 are repeated until  $t$  reaches the end of the whole study period.

#### **4.4.3.2 Trajectory Simulation**

A simulation of a single taxi's trajectory starting from various locations for the 6-hour study period is conducted according to each of the routing strategies. The parameters of the taxi/passenger matching model and the passenger destination probability model are the same as those used in generating the optimal routing policy (Section 4.4.1). The simulation of the optimal routing policy is needed because 1) the optimal value function obtained from solving the Bellman equation has a discount factor of 0.95 for convergence reason, yet the profit in real life should not be discounted given the relatively short time intervals, that is, one dollar earned now and one hour later should be treated as equal valued; 2) the occupancy rate is not available from solving the Bellman equation.

A strategy  $\mu(i, H)$  specifies a probability vector associated with outgoing links, i.e.,  $\mu(i, H) \rightarrow (p_1, \dots, p_a, \dots, p_{|A(i)|})$ ,  $a \in A(i)$ , for each node  $i$  and routing history  $H$ . For the optimal routing policy, random walk and global hotspot heuristics,  $H$  is empty as the action does not depend on the routing history. For the local hotspot heuristic, the routing history  $H$  represents whether the cruising time in the current zone has reached 15 minutes. The optimal routing policy and the global hotspot heuristic are deterministic strategies, in that exactly one of the outgoing links at any node is assigned probability 1. The random walk and local hot spot heuristics are random strategies, in that outgoing links of certain nodes are assigned probabilities other than 0 or 1. A clock is advanced along the simulated trajectory to determine whether the simulation has reached the end of the study period.

The execution of a strategy from any given node  $i$  could result in multiple realizations of trajectories due to the random processes of passenger arrival, competition with other vacant taxis, and passenger destination choice. Following are the steps to simulate a single taxi trajectory for a specific strategy:

1. For a taxi at node  $i$ , an action  $a = (i, j)$  is chosen according to the specific strategy.
2. The location of a matched passenger,  $h$ , is sampled according to matching probability, as in Eq. (4.2).
3. If no matching happens (with probability  $1 - \sum_{h \in N(j)} p_{a,h}$ ), the taxi moves to node  $j$  and the clock is advanced by  $t_a$ . The sampling of the location of a matched passenger, i.e. Step 2, is then repeated.
4. If matching happens, the taxi moves to the passenger location, node  $h$ , and the clock is advanced by  $\tau_a + \mathcal{T}_{j \rightarrow h}$ .

5. The passenger destination  $i'$  is sampled according to Eq. (4.3). The taxi moves to  $i'$ , and the clock is advanced by  $\mathcal{T}_{h \rightarrow i'}$ .
6. Steps 1 through 5 are repeated until the clock reaches the end of the study period.

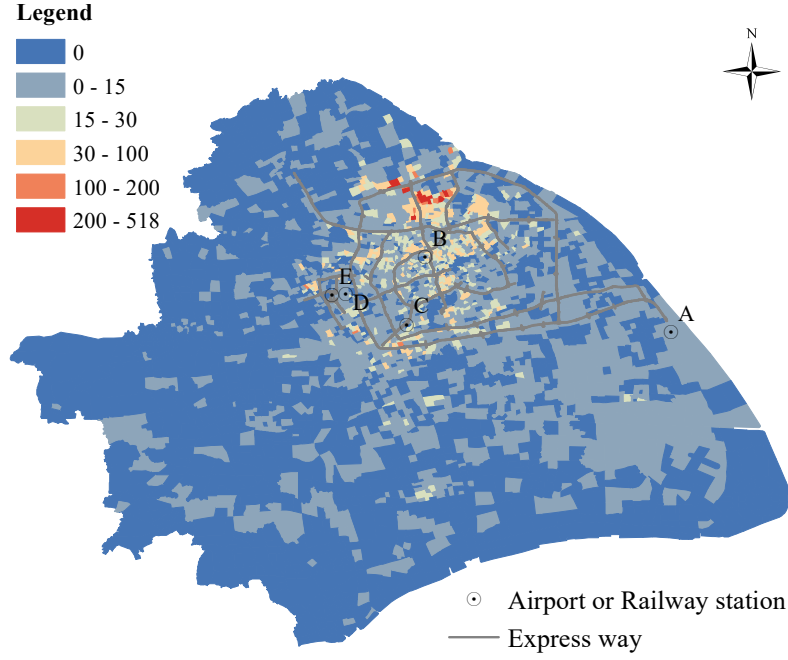


Figure 4.5: Vacant taxi density at 5:30am (count per  $km^2$ )

For any given strategy, 1200 trajectories are simulated from each initial state (node) where there is a positive number of observed vacant taxis at that node at 5:30 am. For each trajectory, the unit profit (CNY/hour), occupancy rate (percentage of time with a passenger onboard) and other relevant measures (such as time spent in each pre-defined cells) are calculated. The average unit profit and occupancy rate for each node is an estimate of the expected unit profit and occupancy rate. The sample size has been increased until the sample average at each node stabilizes.

#### 4.4.3.3 Results

Table 4.2: Average unit profit and occupancy rate in the morning time intervals

Strategy	Unit profit (CNY/hour)				Occupancy rate			
	5:30-7:30	7:30-9:30	9:30-11:30	All intervals	5:30-7:30	7:30-9:30	9:30-11:30	All intervals
Random walk	61	78	72	71	0.40	0.45	0.42	0.42
Global hotspot	68	77	77	74	0.44	0.46	0.47	0.45
Local hotspot	69	86	85	80	0.46	0.47	0.48	0.48
Optimal policy	72	93	92	87	0.47	0.54	0.54	0.52

First the average unit profit and occupancy rate over the network for each strategy is presented. Taxis do not start randomly over the network. Fig. 4.5 shows a heat map of the starting locations of vacant taxis at 5:30 am. The high density starting locations are in the city center and its surrounding areas. The most flourishing ones are indicated on Fig. 4.5: (1) Jiali Sleepless City in Zhabei District, the Central Ring Commercial Circle and Zhenru Commercial Center to the North; (2) Around Shanghai rail way station; (3) Around Hongqiao airport and Hongqiao railway station to the West; and (4) Meilongzhen Plaza, the Commercial Center in Minhang District to the South, with some university campuses. To account for the starting position distribution, the empirical distribution of average unit profit (or occupancy rate) over the network is weighted by the starting position distribution. Practically, each sample average for a given node is duplicated by the number of observed vacant taxis from that node at 5:30 am, and the resulting data points are used to plot the histogram in Fig. 5.6 and the average is taken over all data points to generate Table 4.2.

Table 4.2 shows that the optimal policy performs the best in each interval in terms of both average unit profit and occupancy rate. Specifically, over the 6 hours, the average unit profit for the optimal policy is 23.0% higher than the random walk, 17.0% higher than the global hotspot and 8.4% higher than the local hotspot strategy. This suggests that it is beneficial to take into account subsequent pick-ups and drop-offs beyond the immediate next customer. The increases are higher during higher-demand time intervals, probably due to more room for improvement.

It is noted that the global hotspot strategy generates smaller unit profit than the local hotspot strategy, suggesting that travel time/distance to the next potential customer is a very important factor. Given its inferior performance, the global hotspot strategy will be omitted in the more detailed analyses to follow. The random walk strategy is kept for later analysis as it provides bottom line performance.

While maximizing taxi utilization is not the optimization criterion in the proposed optimization problem, it is still observed that the optimal policy is able to increase the average occupancy rate by 23.8% over the random walk, 15.6% over the global hotspot and 8.3% over the local hotspot strategy. It is intuitive that these two metrics are highly positively correlated, as less time spent on searching for passengers suggests more time spent on making money.

For a given strategy, the average unit profit is the highest during the morning peak interval (7:30 - 9:30 am) due to higher demand, and it becomes only slightly lower for late morning interval (9:30 - 11:30 am), suggesting that taxi trips might have a less pronounced morning peak than regular commuter trips. The flat pattern is also present for the average occupancy rate.

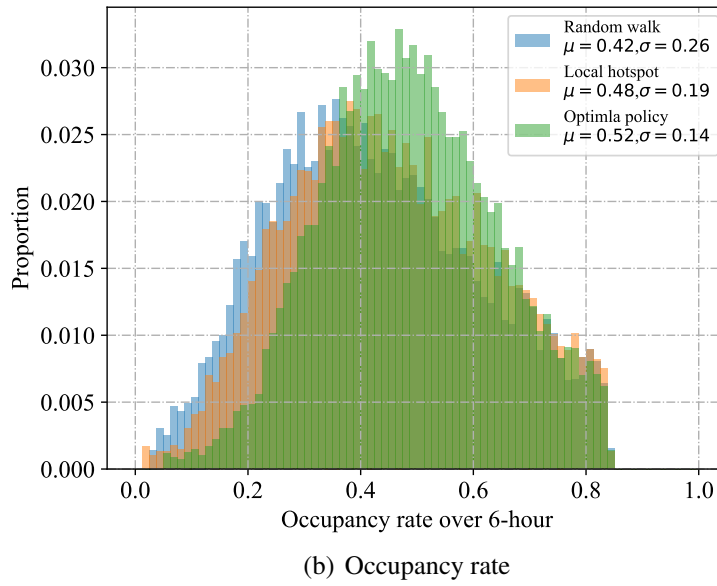
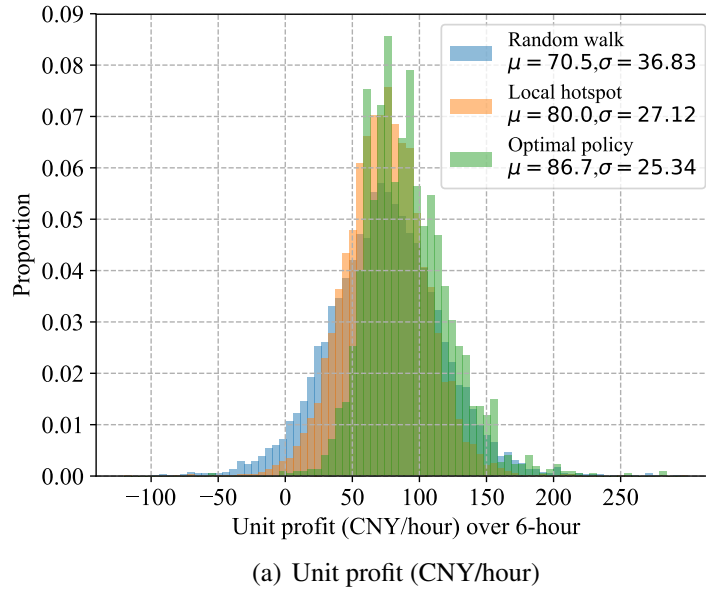
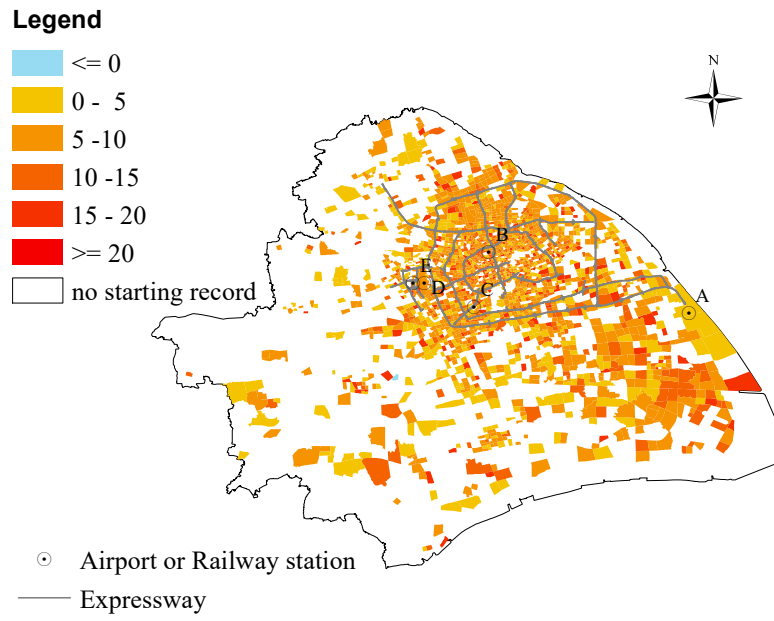


Figure 4.6: Distribution of unit profit and occupancy rate for three strategies 5:30-11:30 am

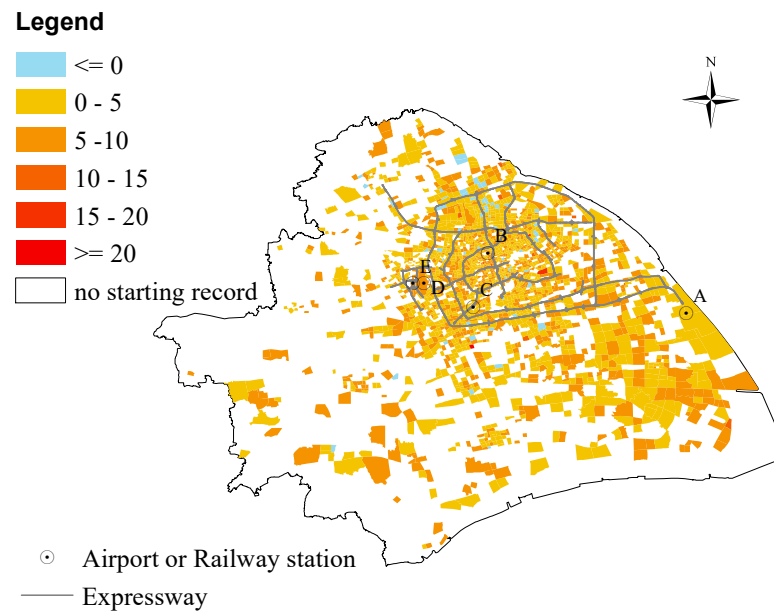
Fig. 5.6 shows the distribution of unit profit and occupancy rate for the 6-hour study period for each of the three remaining strategies: optimal policy (green), local hotspot (orange) and random walk (blue). It can be seen that the distributions for the optimal policy shift to the right, suggesting an across-the-board increase instead of isolated extremely large increases from certain locations.

To further understand the spatial pattern differences among the different strategies, analyses are done by starting locations. Fig. 4.7 (a) shows the the difference in average unit profit by starting location between the optimal policy and random walk strategy. Not surprisingly, a taxi following the optimal policy can make more profit than random walk no matter where it starts from (yellow through red). Fig. 4.7 (b) shows the the difference in average unit profit by location between the optimal policy and local hotspot strategy. From most starting locations, the average unit profit of taxis taking the optimal policy is higher. The relationship is reversed for some locations (blue), due to the fact that the local hotspot strategy has a state space larger than the optimal policy, expanded by including routing history in its state. The optimal policy is optimal among all policies defined based on the same state space, that is, the nodes, but is not necessarily so compared to a more flexible strategy.

The position of a taxi at any time starting from a given location following a given strategy can be described by the coordinates  $(X, Y)$  in a two-dimensional plane, where  $X$  and  $Y$  are the projected longitude and latitude (in meters) respectively (Maling, 2013).  $X$  and  $Y$  are continuous random variables given the underlying passenger matching and destination choice process. To obtain probability distributions and summary statistics of the two random variables, the study area is discretized into a grid where each square cell is



(a) Optimal Policy vs Random Walk



(b) Optimal Policy vs Local Hotspot

Figure 4.7: Differences in unit profit of optimal vs. heuristic routing by starting zones

3 *km* long, and the fraction of time the taxi spent in each cell over the 6-hour study period is the empirical probability of  $(X, Y)$  in that cell,  $p_{x,y}$ .

The Hellinger distance (Le Cam and Yang, 2012) is used to measures the difference between two probability distributions  $P$  and  $P^*$ , that is,

$$H(P, P^*) = \frac{1}{\sqrt{2}} \sqrt{\sum_{(x,y) \in G_{(x,y)}} (\sqrt{p_{x,y}} - \sqrt{p_{x,y}^*})^2}. \quad (4.16)$$

The Hellinger distance is between 0 and 1, and a larger value indicates a larger difference. Table 4.3 presents the Hellinger distance between the distribution of visited locations of the optimal policy and random walk, and between that of the optimal policy and local hotspot for four different starting locations, of which two are major transportation hubs and two are major commercial areas. As expected, the difference between the optimal policy and random walk is larger than the difference between the optimal policy and local hotspot.

Table 4.3: Difference between distributions of visited locations

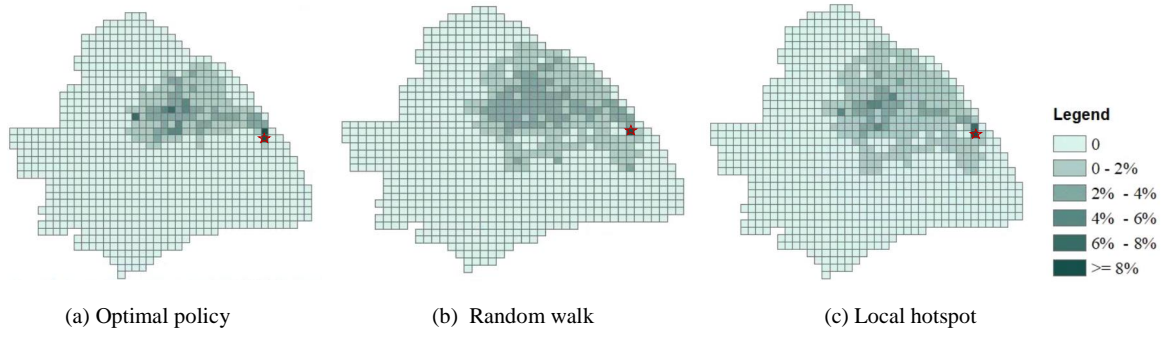
Starting location	Optimal vs Random walk	Optimal vs Local hotspot
Pudong Airport	0.4117	0.3023
Around Hongqiao Airport and Hongqiao Railway Station	0.3928	0.3653
Central Ring Commercial Circle and Zhenru Commercial Center	0.5424	0.4969
Meilongzhen Plaza, the South Commercial Center	0.5128	0.4879

Fig. 4.8 shows the empirical distributions of visited locations from four starting locations. While the movement pattern differs across starting locations, a similar trend can be observed, that is, the trajectories are more widely distributed for the two heuristic strategies. For a taxi starting from a transportation hub (Pudong Airport, around Hongqiao Airport or Hongqiao Railway Station), the trajectory following the optimal policy is mostly between the airports and railway station. The trajectory starting from either of two commercial areas is distributed more widely, probably due to more diversity in the origins and destinations of passengers in commercial areas compared to transportation hubs.

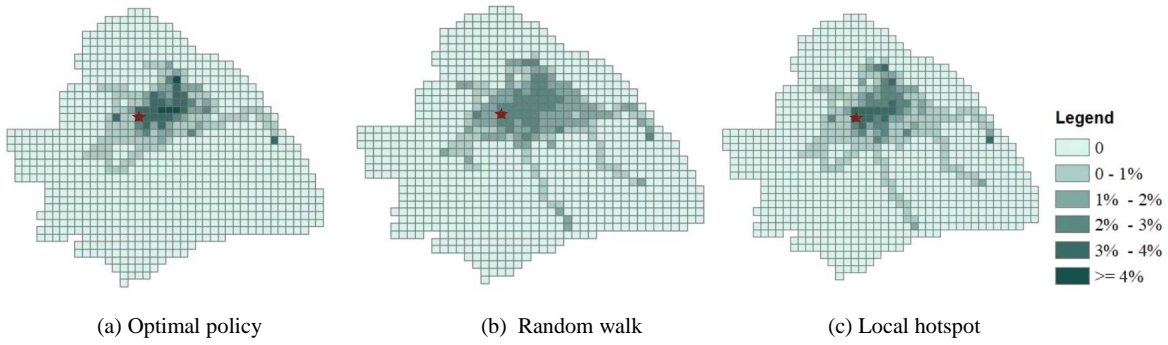
Table 4.4 shows summary statistics of visited locations for the three strategies starting from four different locations. A larger  $X$  means more east and a larger  $Y$  means more north. A positive covariance suggests the trajectory is more northeast/southwest than northwest/southeast, and a negative covariance suggests that the trajectory is more northwest/southeast than northeast/southwest.

For a taxi starting from one of the transportation hubs and the Meilongzhen Plaza, the expected location following the optimal policy is to the northwest of that following the random walk or local hotspot strategy (smaller  $E[X]$  and larger  $E[Y]$ ), while for a taxi starting from Central Ring Commercial Circle and Zhenru Commercial Center, the expected location following the optimal policy is to the northeast of that following the two heuristic strategies. Nor surprisingly, random walk has the largest variance among the three strategies for each of the four locations.

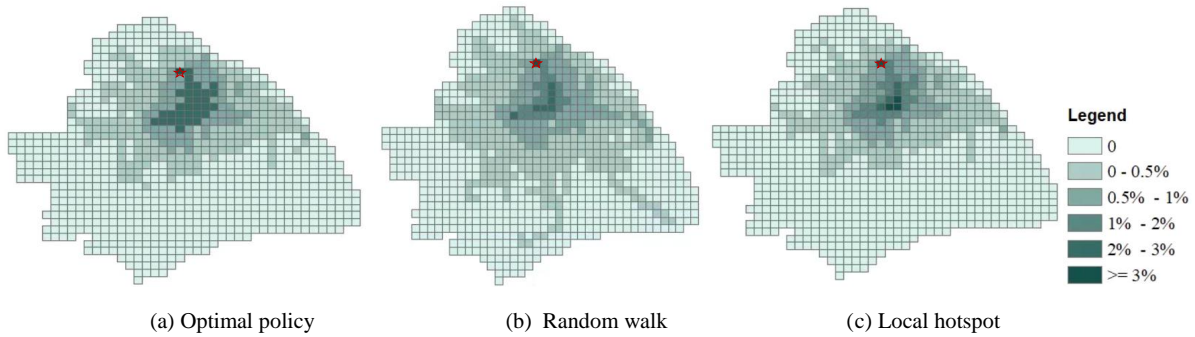
**(1) Pudong Airport**



**(2) Around Hongqiao Airport & Hongqiao Railway Station**



**(3) Jiali Sleepless City, the Central Ring Commercial Circle and Zhenru Commercial Center**



**(4) Meilongzhen Plaza, the South Commercial Center**

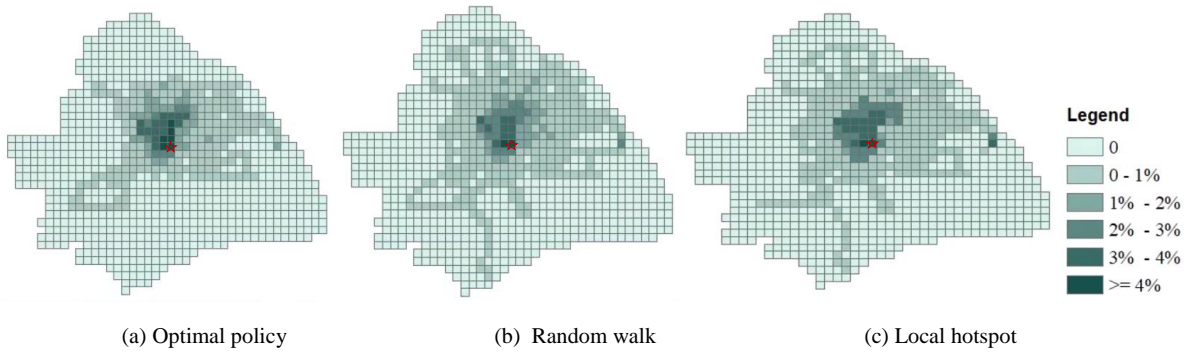


Figure 4.8: Empirical probability density function of the visited location from a starting location (red star)

Table 4.4: Summary statistics on spatial distributions of visited locations ( $X, Y$  are in meter)

Starting location	Strategy	$E[X]$	$E[Y]$	$Var[X]$	$Var[Y]$	$Cov(X, Y)$
Pudong Airport	Optimal policy	1353473	3634788	2144078	3990332	-1309082
	Random walk	1353955	3633768	4360890	5317421	-2126156
	Local hotspot	1353821	3633815	4016870	4931743	-1126138
Around Hongqiao Airport and Hongqiao Railway Station	Optimal policy	1351548	3638640	1219876	1780287	1005313
	Random walk	1352130	3638216	2454627	2245158	-1141663
	Local hotspot	1351999	3638527	2377582	1750020	-1541638
Central Ring Commercial Circle and Zhenru Commercial Center	Optimal policy	1352473	3634619	4141008	3190332	-2309082
	Random walk	1352004	3645415	5252598	4246513	-1396572
	Local hotspot	1352015	3645085	4065228	4005653	-1854668
Meilongzhen Plaza, the South Commercial Center	Optimal policy	1351575	3632361	3745147	4343224	1402910
	Random walk	1351861	3631450	5557328	4820842	1499701
	Local hotspot	1351703	3631959	4010645	5268561	2473367

## 4.5 Summary

In this paper, the single vacant taxi routing problem is investigated, which aims at maximizing long-term expected profit over the complete working period. Theoretical contributions in modeling and problem formulation as well as practical contributions in computational efficiency are provided, which builds the foundation for real-world implementations of taxi routing. A queueing theory-based model for matching taxis and passengers is proposed to account for competition from other taxis and use of e-hailing apps. The problem is formulated as a Markov decision process, taking into account the impact of current decisions on future return over multiple pickups and drop-offs. To improve computation efficiency, an enhanced value iteration algorithm for solving the MDP is proposed via matrix operations.

Numerical experiments in a mega city suggest that matrix operation helps to achieve 8x speed-up in computation. Simulation experiments are conducted to compare the performance of the proposed strategy with a number of baseline heuristics. The MDP formulation improves unit profit by 23.0% and 8.4% over the random walk and local hotspot heuristic respectively; and improve occupancy rate by 23.8% and 8.3% respectively. Empirical spatial distributions of taxi location from a few starting locations and following the various strategies are obtained from the simulation, and the heuristics are shown to have in general more spread-out spatial patterns than the optimal policy. Specifically the optimal policy concentrates between major transportation hubs if starting from one of them. The trajectory starting from either of two commercial areas is distributed more widely than that from a transportation hub.

## **CHAPTER 5**

### **OPTIMIZING VACANT TAXIS' ROUTING DECISIONS: A MODEL-FREE REINFORCEMENT LEARNING FRAMEWORK**

In chapter 4, a model-based approach is presented for the vacant taxi routing problem, where a model of the state transitions of the environment is obtained from queuing-theory based passenger arrival and competing taxi distribution processes. Dynamic Programming (DP) algorithm is then applied to solve the problem. While model-based DP approach has provided a way to optimally solve decision and control problems involving complex dynamic systems, its practical value was limited by the complexity to build a true model to capture all of the subtlety of the complex real system dynamics. Strong assumptions need to be made in the model.

Reinforcement Learning (RL) has the potential to continuously and adaptively learn from interaction with the environment without building a transition model: an autonomous agent takes an action in a state, receives a reward, moves to some next state, and repeats this procedure (Sutton and Barto, 1998b). At each step, the agent can revise its control policy with the objective of converging as quickly as possible to an optimal control policy. Model-free RL has been successfully applied to a range of challenging problems (Kober et al., 2013; Deisenroth and Rasmussen, 2011). In transportation area, Verma et al. (2017) formulate the taxi routing problem as a Markov Decision Processes (MDP) to take into account long term revenue and develop a reinforcement learning (RL) based system to

learn pick up locations from real trajectory logs of drivers. Han et al. (2016) also present an RL algorithm for an autonomous taxi to learn the existence probability of passengers from its gathered experience. However, the above mentioned studies provide decisions on highly aggregated or abstracted locations in grid-based networks instead of the physical roadway network.

Similar to Chapter 4, the optimal taxi routing problem is formulated as a Markov Decision Process (MDP) so that long-term objectives can be taken into account instead of the immediate one of meeting the next customer. The application of RL algorithm is examined to solve the problem. This method is model-free, in that no transition models are needed and the system dynamics are embedded in the observed trajectories. The algorithm is implemented and tested in a real road network of Shanghai, China, and a systematic comparison of the model-based and model-free algorithms are also provided.

## 5.1 Problem Formulation

A taxi driver's working plan and routing decisions over a time horizon on a given day is modeled as an MDP. A taxi travels in a traffic network  $G = (N, A)$ .  $N$  is the set of nodes and  $A$  the set of links. There is at most one directional link,  $a$ , from the source node  $i$  to sink node  $j$ .  $A(i)$  is the set of downstream links of  $i$ . The taxi is actively searching for, or carrying passengers during a planning horizon  $[0, T]$ . The taxi can also decide to stop and start working over the course of the horizon, more typical among ride-sourcing drivers who are usually part-time. The length of the horizon ranges from several hours to maximum working hours (usually 10-12 hours).

When a taxi is hired, the routing problem is reduced to a shortest path problem from the passenger's origin to destination, and is not studied explicitly in this paper. The state of a taxi,  $s \in S$ , is described by node  $i \in N$ , time  $t \in [0, T]$ , where  $t$  is a continuous variable, and working status  $w$ , a binary variable that is equal to 1 when the taxi is for hire, and 0 otherwise (not on the market). A taxi is associated with a depot location  $d$  that the taxi must go to if stopping working.

The action set for state  $s$  when  $w = 1$  is  $\{\text{Taking one of outgoing links}\} \cup \{\text{Waiting at the current node}\} \cup \{\text{Stopping working}\}$ . For an action in the first two categories, the taxi is actively searching for passengers, and two types of transition to a new state  $s'$  could happen. 1) The taxi is not matched with any passenger when traversing link  $a = (i, j)$ , and  $s'$  is associated with node  $j$ . 2) The taxi is matched with a passenger when traversing link  $a$  or waiting at node  $i$ , and  $s'$  is associated with the destination node of the passenger,  $i'$ . For the action of stopping working, the taxi will take a shortest path to the depot  $d$  with the working status  $w$  switched to 0. The action set for state  $s = (d, t)$  when  $w = 0$  is to either start working or continue the non-working status. By definition, any state associated with  $t \geq T$  is a terminal state.

Assume the time horizon is discretized into time intervals. The value function is constant within a given time interval, that is,  $V(s_1) = V(s_2)$ , if  $s_1$  and  $s_2$  are in the same time interval. Let  $S(s)$  be the aggregate state corresponding to state  $s$  defined on continuous time.

### 5.1.1 Model-free Reinforcement Learning without Transition Model

Research in RL aims at designing algorithms by which autonomous agents can learn to behave in some appropriate fashion in some environment, from their interaction with this

environment. The standard RL protocol considers a performance agent observing state  $s$ , taking an action  $a$ , achieving next state and receiving the instantaneous reward  $r$ . This process can be represented by a four-tuple, i.e.,  $(s, a, s', r)$ .

The  $Q$ -function represents the expected total (discounted) reward that can be obtained after taking action  $a_t$  in state  $s_t$ . The optimal  $Q$ -function  $Q^*(S(s), a)$  means the expected total reward received by an agent starting in  $s$  and picks action  $a$ , then will behave optimally afterwards. Therefore,  $Q^*(S(s), a)$  is an indication for how good it is for an agent to pick action  $a$  while being in state  $s$ . The optimal value function and  $Q$ -function have a straightforward relationship:

$$V^*(S(s)) = \max_{a \in A(s)} Q^*(S(s), a), \forall S(s) \in \mathcal{S} \quad (5.1)$$

and the optimal policy can be extracted by choosing the action  $a$  that gives maximum  $Q^*(S(s), a)$  for state  $s$ .

$$\mu^*(S(s)) = \operatorname{argmax}_{a \in A(s)} Q^*(S(s), a), \forall S(s) \in \mathcal{S} \quad (5.2)$$

The classical approach towards model-free reinforcement learning is  $Q$ -learning in which an optimal value function of state-action pairs is learned iteratively and online. This means that the  $Q$ -function is updated after each transition of the system. Typically, this approach requires thousands of iterations until successful policies are found.

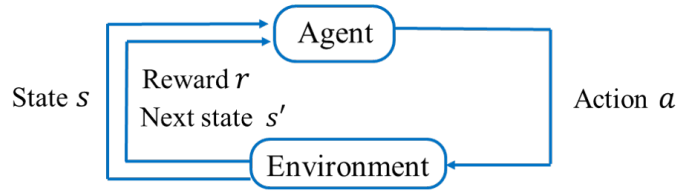


Figure 5.1: The idea of Q-learning: learn while interacting with environment

In contrast to classical on-line RL, batch reinforcement learning performs the update of the  $Q$ -function based on sets of past transitions instead of singular state transitions. Batch RL methods store and reuse information about system behavior by a set of transition tuples (state, action, successor state, reward). The  $Q$ -value function is then updated on all states (state-action pairs) simultaneously. This reuse of transition data makes batch learning methods particularly efficient.

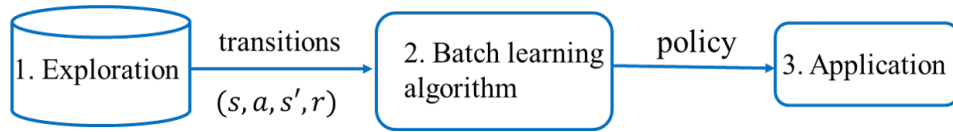


Figure 5.2: The three distinct phases of the batch reinforcement learning process: 1: Collecting transitions with an arbitrary sampling strategy. 2: Application of batch reinforcement learning algorithms in order to learn the best possible policy from the set of transitions. 3: Application of the learned policy

The fitted  $Q$ -iteration (FQI) is a batch mode RL algorithm, which yields an approximation of the  $Q$ -function on the basis of a set of transition tuples iteratively (Ernst et al., 2005). As the name suggests, FQI allows to fit (using a set of four-tuples) any (parametric or non-parametric) approximation architecture to the  $Q$ -function. It has been shown to converge for all approximators belonging to the averager class (Ernst et al., 2005). In

this chapter, the simple average is used to approximate  $Q$ -value, which is a special case of kernel-based methods.

When a large number of observations of the system transition are available, one can update the  $Q$  function directly using the observations without deriving a transition model. Observed taxi trajectories can be organized into a set of four-tuples  $(s, a, s', r)$ . At iteration  $n$ , FQI uses the greedy operator  $\max$  on the action space for improving the policy, where  $\hat{Q}^{n-1}$  are available from the previous iteration.

$$Q^n(S(s), a) = r + \gamma \max_{a' \in A(s')} \hat{Q}^{n-1}(S(s'), a'), \quad (5.3)$$

where  $\gamma$  is a discount factor ( $0 < \gamma < 1$ ) that weights short-term rewards more than long-term ones.

The trajectories can be processed in any order, however, for any trajectory, it might make sense to update starting from the end, taking advantage of the acyclic nature of the state transition in the time dimension.

The FQI algorithm for solving the vacant taxi routing problem is presented in Algorithm 4. At each iteration step, a new training set was built based on the full set of four-tuples and the estimated  $Q$ -values from previous step, then  $Q$ -values are updated accordingly. Line 8 – 10 computes the return for each  $(S(s), a)$  pair over all samples, and obtain the new  $Q$ -value. Line 11 estimated  $Q$ -value for each  $(S(s), a)$  pair based on the average  $Q$ -value over a sub-sample of the training set which corresponds to the same link.

---

**Algorithm 4** FQI for taxi routing

---

*Input:*  $D = \{(S^l(s), a^l, r^l, S^l(s')), l = 1, \dots, \#D\}$ , terminate time  $T$

*Output:*  $Q$ -values for routing to terminal time

**Initialization:**

Set  $n \leftarrow 0$

Let  $\hat{Q}^n$  be a function equal to zero everywhere on  $(\mathcal{S}, \mathcal{A})$  space

**Iterations:** Repeat until stopping conditions are met

-  $n \leftarrow n + 1$

- Build the training set  $TD = \{(S^l(s), a^l, o^l), l = 1, \dots, \#D\}$  based on  $\hat{Q}^{n-1}(\cdot)$  and the set of four-tuples  $D$

$$o^l = \begin{cases} r^l + \gamma \max_{a'} \hat{Q}^{n-1}(S^l(s'), a'), & \text{if } t < T \\ r^l, & \text{if } t \geq T \end{cases}$$

- Calculate  $\hat{Q}^n$  by taking the average of  $o^l$  over each  $(s, a)$  pair

---

The iterations stops if the difference between  $\hat{Q}^n$  and  $\hat{Q}^{n-1}$  drops below a predefined threshold.

FQI is inherently an offline method - given historical transition-tuple data, the algorithm derives the approximated  $Q$ -value and the inferred policy. When there are new samples collected with the currently best inferred policy, the FQI algorithm can be restarted based on the new data.

### 5.1.2 Discussion on Solving Time-dependent MDP

For both model-based and model-free methods, a true time-dependent problem would be solved if the time interval is small enough. However, solving an MDP in a time-dependent network renders the problem computationally challenging, since the exact dynamic programming algorithm enumerates all states and does not scale with the significant

increase in the size of the state space. In practice, it is also difficult to obtain enough reliable data if the time-interval is too small. In this thesis, approximations are applied by having large time intervals in the numerical experiment (e.g., 1h). For the model-based method, multiple static problems are solved essentially. The state transition model is static during a given interval and most of the transitions happen within the same interval. While the model-free method still keeps some system dynamics across time intervals.

The length of the time intervals should be chosen carefully based on the required level of modeling accuracy and the data available to the modeler. For both methods, if very small time intervals is used, the spatial-temporal coverage of the available data is very sparse. Thus it is far from enough to obtain the mobility patterns for the entire road network. On the other hand, if large time intervals is used, the system dynamics could be distorted.

### **5.1.3 Discussion on Data Efficiency**

The model-based and model-free approaches have distinct strengths and weaknesses.

Model-based method can help to improve data efficiency when there are small number of observations when the built model can perfectly represent the true dynamics. A model-based algorithm can immediately incorporate and expand newly gained information from the exploratory strategies into the state space and transitions, making more efficient use of information. Therefore, model-based method can quickly achieve near-optimal policy with accurately built models. However, the ground truth models are never known in reality and it is complex to build a model to capture all of the subtlety of the real system dynamics.

In contrast, model-free RL algorithms provided researchers with a way to learn complex behaviors, especially when a mathematical model of the system is unavailable. However, it often takes large amount of experiences to explore different parts of the environ-

ment in order to learn an effective solution. This can severely limit their application to real-world problems where this experience might need to be gathered directly in a real physical system.

## 5.2 Case Study

### 5.2.1 Data and Network

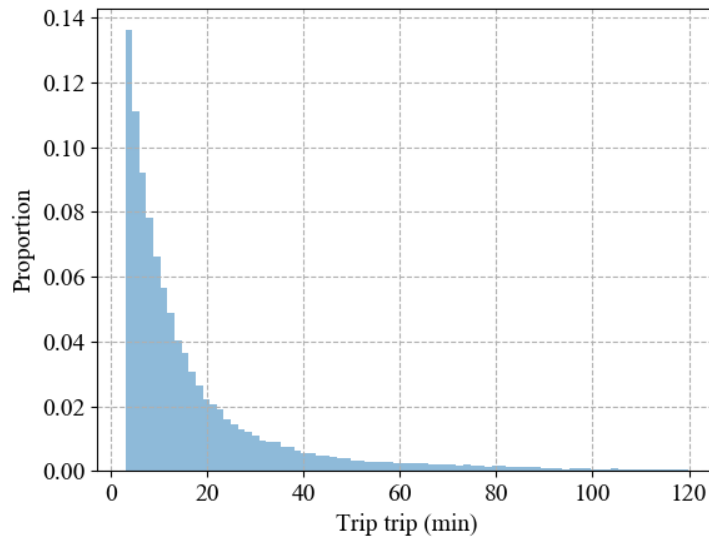
The same data set and Shanghai roadway network with Chapter 4 is used in this case study. That is, trajectories from 12,017 taxis during 5:30–11:30 am on a representative weekday in April, 2015 are extracted for the case study after initial data cleaning.

Fig 5.3 shows that trip duration and searching duration are highly skewed. About 50% of the trips are shorter than 10 min and 50% of the searchings are shorter than 15 min.

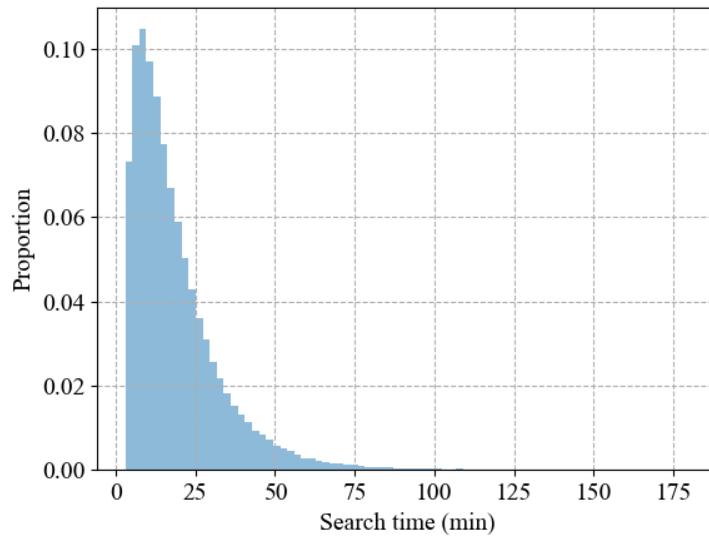
### 5.2.2 Data Pre-processing

Define a trajectory as a time-ordered sequence of location points, denoted as  $T = \{VehID; p_1, p_2, p_3 \dots p_n\}$ , where  $p_l = \{x, y, t, w, v\} (1 \leq l \leq n)$ ,  $x, y$  are longitude and latitude,  $t$  is the GPS time,  $w$  is the taxi status (empty or occupied),  $v$  is GPS speed reading.

In order to obtain transition-tuples, we first need to project the raw GPS points to the physical road network that consists of road segments. We matched GPS points to the nearest links with a distance less than 200 m in the network in ArcGIS 10.1. The corresponding longitudes and latitudes on the nearest links were also given. This method only considers the geometric information and that every new point in  $p_l$  is matched without considering the matched location of point  $p_{l-1}$ . Therefore, we perform subsequent topological analysis by removing the mapped link observations that are not connected to its mapped predeces-



(a) Distribution of trip time



(b) Distribution of searching time

Figure 5.3: Histograms for trip time and searching time during 5:30-11:30 am

sor and successor links based on the initial match. Finally, about 4.6% of link observations were discarded from all mapped records.

Time-dependent link travel time is calculated from GPS speed readings for each link by first removing the almost zero speed readings( below 1 km/h), and then taking the average over each 1-hour time period. Link speeds are then calculated based on the obtained link travel time and link length; if there's no GPS observation on a link during a given time interval, we fill the speed with the average speed from the previous time interval. Given the mapped trajectories and link speeds, the following steps calculates the start/end time of each transition:

- step 1: split the mapped trajectories into empty and occupied points;
- step 2: extract empty link traverses where all points are empty during a link traverse;
- step 3: calculate entry/exit time of each empty link traverse based on the position of the first /last points and the link speeds, e.g., let  $p_l^m/p_{l+n}^m$  denote the first/last mapped GPS points of the  $m^{th}$  empty traverse on link  $a = (i, j)$ , then the entry/exit time of the  $m^{th}$  empty traverse of link  $a$  is

$$t_{entry} = \frac{distance(i, p_l^m)}{\bar{v}^{a,k}} / t_{exit} = \frac{distance(p_{l+n}^m, j)}{\bar{v}^{a,k}}$$

where  $\bar{v}^{a,k}$  is the average speed on link  $a$  during time interval  $k$ ;

- step 4: map the pick up and drop off points to the nearest node; similar to step 3, calculate the node-based pick/drop off time of each occupied trips.

Taxi fares are charged based on distance traveled for both approaches. For model free approach, trip distance is calculated as the observed actual travel distance/time. The parameters for calculating taxi fare  $F(d_{h \rightarrow i'})$  in Eq. (4.6) are set as  $f_0 = 14$ ,  $d_0 = 3$ ,  $d_1 = 15$ ,  $\beta = 2.5$  and  $\gamma = 3.6$ .<sup>1</sup> The unit operating cost,  $\alpha$ , is assumed to be 0.5 CNY/min.

### 5.2.3 Experiment setting

In model-based approach, the whole study period is divided into equal time intervals. Value iteration algorithm is used to solve the Bellman equation. To improve computation efficiency, the value-iteration algorithm is translated into sparse and fixed matrix operations.

In model-free FQI approach, all action value functions are aggregated by a relatively large time interval (e.g., 15-min). After the data pre-processing, 6,453,263 records of transitions are obtained.

As discussed in Section 5.1.2, multiple static problems are solved essentially with large time-intervals. The discount factor is set as 0.95 for both methods.

### 5.2.4 Computational Performance

All optimization algorithms are coded in Python 3.5. All computations are carried out on a workstation with an eight-core 3.0GHz Xeon E5-1660 processor and 64GB RAM.

Fig. 5.4 gives an example of computational performance for model-free approach with a 15 min time interval. The running time per iteration is about 0.89 min (53.6 sec), and it takes about 85 iterations (4,556 sec) for all action values to reach convergence.

---

<sup>1</sup>source: <https://www.travelchinaguide.com/cityguides/shanghai/transportation/taxi.html>

Recall that in the model-based approach in Chapter 4, for each 2-hour time interval, the running time per iteration is about 1.98 min (118 sec) with matrix operations and in total it takes about 35 iterations (4,130 sec) to converge. The model-based algorithm can be implemented simultaneously for three time-intervals, and thus the total running time for the 6-hour study period is less than the model-free algorithm in total.

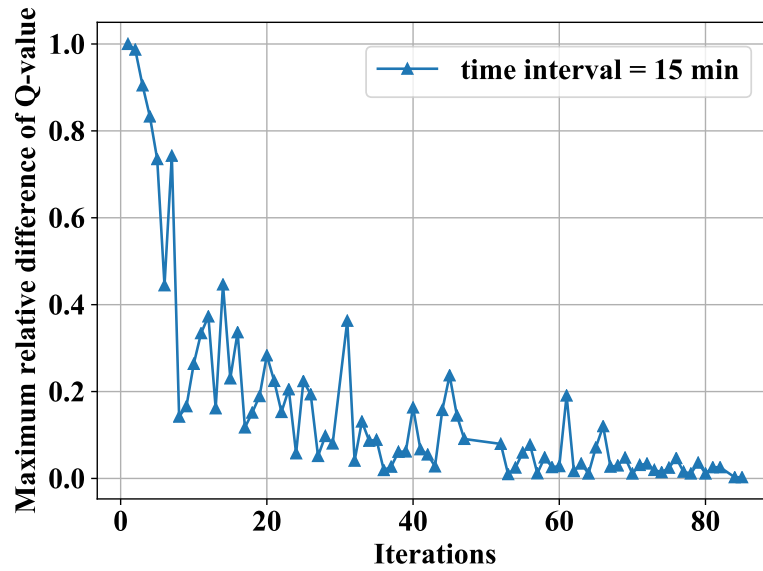


Figure 5.4: Value Function Difference vs. Iterations

The above running time does not include the data-processing time. In model-free approach, pre-processing trajectory data to get large enough transition-tuples to arrive at an effective solution usually takes time for real-world applications. In contrast, model-based approach does not need to deal with the transition data. Instead, system dynamics are represented by the model. Once the dynamics are modeled, near-optimal behavior can in principle be obtained by planning through these dynamics, and iterations are needed.

The model-based and model-free approaches have distinct strengths and weaknesses in terms of space complexity and computational complexity. In the model-based approach, the transition and payoff matrices are given as inputs to the agent. While in model-free algorithm, an agent interacts with the environments and gets reward, whose space complexity is asymptotically less than the space required to store an MDP in model-based method.

### **5.2.5 Effectiveness Evaluation**

In this subsection, average unit profit and occupancy rate obtained from the policies given by the proposed algorithms are compared during a given working period to demonstrate the benefit of the proposed algorithms to drivers.

#### **5.2.5.1 Trajectory Simulation**

As discussed in Chapter 4, the comparison of the optimal policies with the observed taxi driver routing choices cannot be made until they are implemented in the real world. Another way is to build a high-fidelity traffic simulation testbed where taxi driver behavioral models are calibrated using observed trajectory data, and then evaluate both policies against the calibrated, simulated drivers in the simulation testbed.

The metrics are generated by executing the obtained optimal policies and random walk in the network for the study period of 5:30 – 11:30 am, and the developed taxi/passenger matching model and passenger destination model are postulated as the true model based on which taxi and passenger matching is simulated, as shown in Algorithm 5.

For any given strategy, 30 trajectories are simulated from each initial state (node) where there is a positive number of observed vacant taxis at that node at 5:30 am.

---

**Algorithm 5** Simulate the next state and compute the corresponding reward

---

*Input:* - road network  $G = (N, A)$

- state  $s = (i, t)$  and action  $a$

- matching probability matrix  $M^k$ ; destination probability matrix  $D^k$ .  $k$  is used to indicate a discrete time interval index that time  $t$  falls in.

*Output:* - the simulated next state  $s' = (i', t')$  and the corresponding reward  $r'$

### 1. Compute matching probability

- Compute matching probability within the matching area, which is the sum of matching probability between link  $a$  and each of the node in the matching area,  $p = \sum_{h \in N(j)} M_{a,h}^k$

### 2. Sample next state and compute reward

2.0 Generate a random variable,  $\varpi$ , uniformly between  $[0, 1]$ .

2.1 If  $\varpi \leq p$ , the taxi is matched with a passenger

- sample a matching location  $h$  from the discrete matching probability distribution within the matching area, and the elapsed time until picking up at node  $h$  is  $\tau_{ak} + \mathcal{T}_{j \rightarrow h,k}$

- sample the passenger destination  $i'$  from the distribution of the destination nodes according to  $D_{h,:}^k$ , and the elapsed time until passenger drop off is  $\tau_{ak} + \mathcal{T}_{j \rightarrow h,k} + \mathcal{T}_{h \rightarrow i',k}$

- compute reward  $r' = g_{ss'|a}$  according to Eq.(4.14)

2.2 Elif  $\varpi > p$ , i.e. no matching happens (with probability  $1 - \sum_{h \in N(j)} p_{ah,k}$ )

- the taxi moves to node  $j$  and travel time is updated by adding  $\tau_{ak}$

- compute reward  $r' = g_{ss'|a}$  according to Eq.(4.14)

### 3. Return $s'$ and $r'$

---

#### 5.2.5.2 Batch RL Performance against Sampling Data Size

The performance of RL largely depends on the training sample. In general, the larger the sample size, the more information contained in the training sample. When the sample size is small and the spatialtemporal coverage is sparse, the agent might not be able to learn a good policy. When the size of the training dataset goes to infinity, the algorithm could learn asymptotic optimal policies. Next, the proposed RL framework is evaluated by looking at how the performance of the learned policies changes with the training sample size.

4 training sessions are performed with sample size of 160K, 320K, 480K and 640K. For each of these training sessions, 5 sub-sample are sampled from the raw transition dataset randomly, and FQI are then applied on each of the sub-sample. Trajectories are then simulated based on the obtained policies in the same test world where the developed matching/destination models are assumed as the true model. Parameters for the true model are obtained from the raw data set.

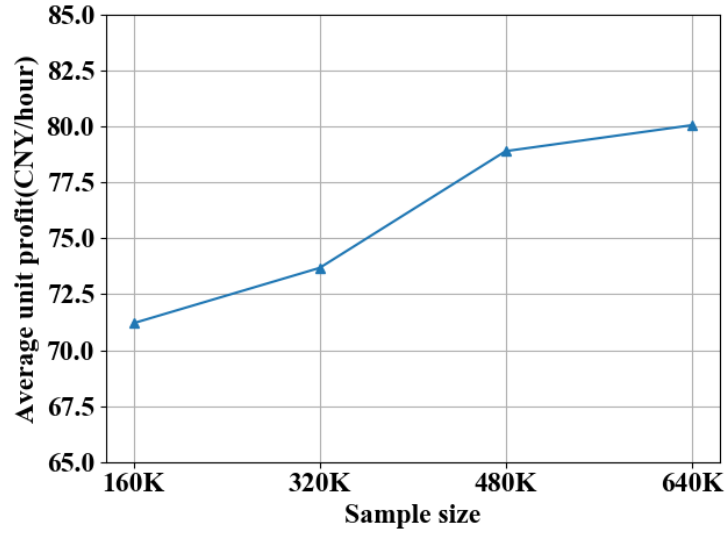


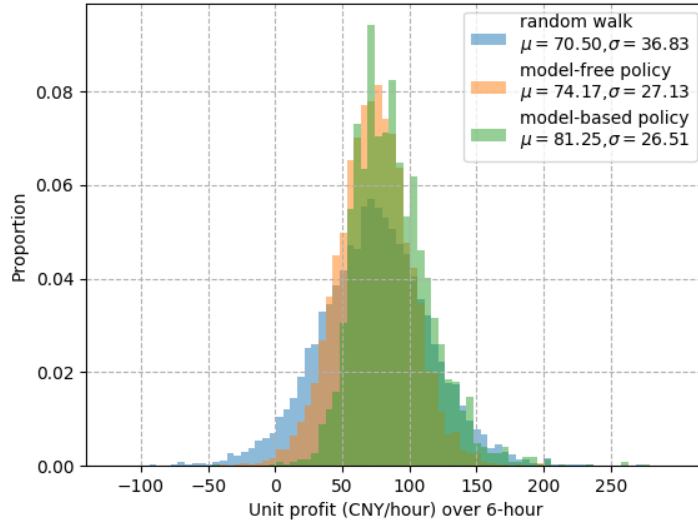
Figure 5.5: Average unit profit against the sample size for model-free approach

Fig. 5.5 plots the average unit profit for each of the given sample size. Results presented in the previous section are very promising. As expected, the performance of the learned policy increases with sample size, since more information are contained in larger training sample. The fact that the performance of batch RL with sample size 160K is slightly better than random walk indicates that the method could still learn better performance policies even from a relatively small sample size. FQI is sample efficient algorithms

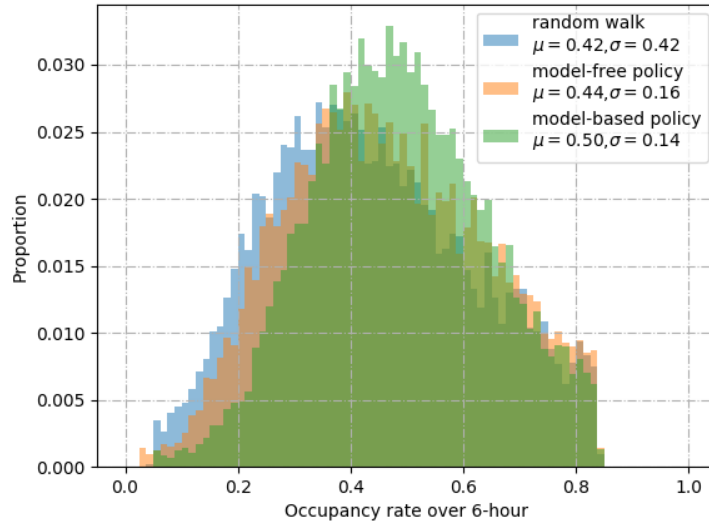
for vacant taxi routing optimization. The method is stable and do not collapse or diverge when they are learning.

### **5.2.5.3 Performance Comparison with Model-based Approach**

Fig. 5.6 shows the distribution of unit profit and occupancy rate for the 6-hour study period for both methods with 1-hour time interval. These results show that the two proposed methods achieve better performance than random walk. It can be seen that the distributions for the optimal policy from model-based method shift to the right, suggesting an across-the-board increase instead of isolated extremely large increases from certain locations. This shows that the purely data-driven model-free method are less efficient than the model-based method when the developed model perfectly reflects the true ground truth. As explained in 5.2.5.1, the developed taxi/passenger matching model and passenger destination model are postulated as the true model in trajectory simulation, and therefore the model-free best policy can perform at best as well as the model-based method. In addition, model-free method learns "nothing" if there's no data observation to update  $Q$ -value. However, if the transition-tuple dataset is large enough, the model-free learning are expected to perform comparably as well as the model-based approach.



(a) Unit profit (CNY/hour)



(b) Occupancy rate

Figure 5.6: Distribution of unit profit and occupancy rate for three strategies 5:30-11:30 am

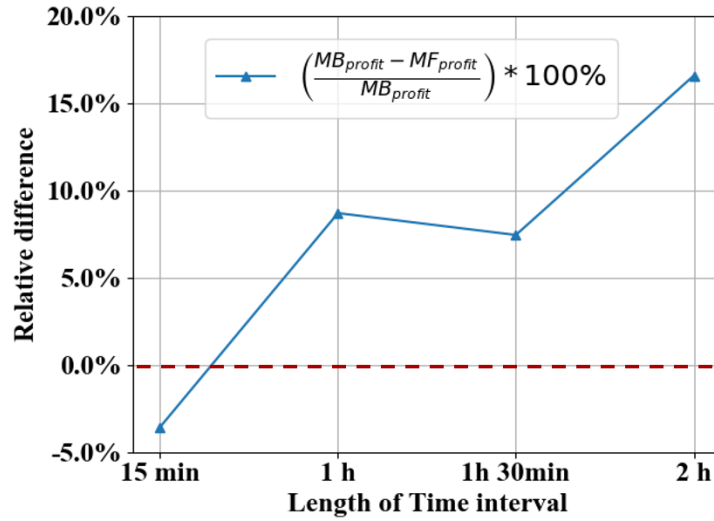


Figure 5.7: MB vs MF: average unit profit against length of time interval

Fig. 5.7 shows the relative difference of the average profit between two methods with different length of time interval. As can be seen from the figure, model-based method performs more better than the model-free method with larger time intervals. However, it is also noted that with a time interval as smaller as 15 min, model-free method performs better than model-based method. This is probably due to the loss of dynamics across time intervals in model-based method. It is more effective when the dynamics model perfectly matches the true one, but often at the cost of larger bias when the dynamics are not modeled accurately. Model-free method is less efficient but could achieve the best asymptotic performance especially for most complex problems where the true dynamics cannot be modeled accurately.

### 5.3 Summary

In this chapter, a model-free RL approach is examined to solve for the optimal vacant taxi routing problem that accounts for multiple cycles of pick-up and drop-off into the future. This method is model-free, in that no transition models are needed and the system dynamics are embedded in the observed trajectories. The approach is implemented and tested in a large-scale network of Shanghai, China, and a systematic comparison of the model-based and model-free algorithms are also provided. Results show that batch RL is a sample efficient algorithm for vacant taxi routing so as to avoid extra modeling assumptions. It could still learn better performance policies even from a relatively small training sample size. Overall, the performance of the learned policy increases with sample size.

Both policies obtained from model-based and model-free algorithms perform better than random walk despite not having any priori knowledge. Model-based method is more effective when the model perfectly matches the true dynamics but often at the cost of larger bias when the dynamics are not modeled accurately; while model-free method is less efficient but could achieve good asymptotic performance especially where the true dynamics cannot be modeled accurately.

Future work could consider to solve the problem in a true time-dependent case by improving the sampling data efficiency. Currently, the state (action) values are updated based on one-step exploration, they can be extended by bootstrapping over a longer length of time in which a significant and recognizable state change has occurred (Mahmood et al., 2017; Yang et al., 2018). For real-time application, the GPS points needs to be projected to the road network to create a path that consists of road segments, i.e., the map matching process. The reader is referred to (Zheng, 2015) for a recent review of map matching

methods. Future work can consider using a Hidden Markov Model (HMM) (Newson and Krumm, 2009) to get the mapped path.

## **CHAPTER 6**

### **CONCLUSIONS AND FUTURE DIRECTIONS**

#### **6.1 Research Summary**

This thesis investigates the dynamic routing decisions of both travelers and on-demand service providers (e.g., regular taxis, Uber, Lyft, etc).

For individual travelers, this thesis focuses on route choice behavior analysis at two time scales: day-to-day and within-day. For day-to-day route choice behavior, methodological development and empirical evidences are presented to understand the roles of learning, inertia and real-time travel information on route choices in a highly disrupted network based on data from a laboratory competitive route choice game. A learning model based on the power law of forgetting and reinforcement is applied. The learning of routing policies instead of simple paths is modeled when real-time travel information is available, where a routing policy is defined as a contingency plan that maps realized traffic conditions to path choices. Model parameter estimates are obtained from maximizing the likelihood of making the observed choices on the current day based on choices from all previous days. Prediction performance is then measured in terms of both one-step and full trajectory predictions.

The routing policy model within each day in the above day-to-day learning framework is estimated with non-recursive model which requires prior choice set generation.

In practice, sampling choice sets of routing policies is computationally costly and it does not scale well with the size of the network. In this thesis, a recursive logit model for route choice is formulated in a stochastic and time-dependent (STD) network where the choice of path is formulated as a sequence of link choices, without sampling any choice set. A decomposition algorithm is proposed for solving the value functions that relies on matrix operations. Estimation and prediction results of the proposed model are presented using a data set collected from a subnetwork of Stockholm, Sweden. Results show that the model can be estimated efficiently, and gives reasonable results for prediction.

Taxis and ride-sourcing vehicles play an important role in providing on-demand mobility in an urban transportation system. In this thesis, the single vacant taxi routing problem is investigated to maximize long-term expected profit over the complete working period. The problem is formulated as a Markov Decision Process (MDP) problem at the highest level of spatial resolution. Two approaches are proposed to solve the problem. One is the model-based approach where a model of the state transitions of the environment is obtained from queuing-theory based passenger arrival and competing taxi distribution processes. To improve computation efficiency, an enhanced value iteration algorithm for solving the MDP is proposed via matrix operations. The other is the model-free learning approach, which learns action values (and from that, the best policy) directly from observed trajectory data. This method is model-free, in that no transition models are needed and the system dynamics are embedded in the observed trajectories. Batch RL algorithm is applied to make more efficient use of the collected data by separation of the data collection and learning steps.

Both approaches is implemented and tested in a large-scale network of Shanghai, China with reasonable running time. For model-based approach, numerical experiments in a mega city suggest that matrix operation helps to achieve 8x speed-up in computation. Simulation experiments are conducted to compare the performance of the proposed strategy with a number of baseline heuristics. The MDP formulation improves unit profit by 23.0% and 8.4% over the random walk and local hotspot heuristic respectively; and improve occupancy rate by 23.8% and 8.3% respectively. Empirical spatial distributions of taxi location from a few starting locations and following the various strategies are obtained from the simulation, and the heuristics are shown to have in general more spread-out spatial patterns than the optimal policy. For model-free approach, results show that batch RL is a sample efficient algorithm for vacant taxi routing so as to avoid extra modeling assumptions. It could still learn better performance policies even from a small sample size. Overall, the performance of the learned policy increases with sample size.

A systematic comparison of the model-based and model-free algorithms are also provided. Results show that both policies perform better than random walk despite not having any priori knowledge. Model-based method is more effective when the model perfectly matches the true dynamics but struggle on building accurate models for complex tasks; while model-free method are less efficient but could achieve good asymptotic performance especially where the true dynamics cannot be modeled accurately.

## 6.2 Future Directions

We have witnessed a rapid development of on-demand ride-hailing services such as Uber, Lyft and Didi Chuxing in recent years. With the emergence of wireless communica-

tion tools, the Global Position System (GPS), and powerful mobile apps, these ride-hailing services provide significant improvements over traditional taxi systems in terms of reducing taxi cruising time and passengers' waiting time and opens up new opportunities for the exploitation of unused capacity. Meanwhile, they also provide rich information on passenger demand and taxi mobility patterns, which can benefit various research areas including demand prediction, route planning, and traffic light control.

To make fully use of such information, future research should tap on the potential of machine learning techniques. Those data-mining approaches are designed to handle the uncertainty of sparse and noisy data as it is the case for spatial data. Meanwhile, it is also acknowledged that sophisticated modelling knowledge has developed in the domain of transport planning and therefore domain expert knowledge should build the fundament when applying data-driven approaches in transportation research. These new challenges call for a multidisciplinary collaboration between transport modelers and data scientists.

Specifically, the following directions can be addressed in the future.

- **Multi-taxi routing** In Chapter 4 and 5, optimal routing decisions is provided from the viewpoint of a single taxi driver at a road network level, assuming that the movements of the taxi do not affect traffic condition, other taxis' movements or passenger demand. Future research can consider optimizing routing decisions for a fleet of vehicles. Taxis are competing with each other; at the same time, they can cooperate with each other to provide ride-sharing services. The decision of assigning requesting orders to taxis is determined by a centralized algorithm in a coordinated way from a global view.

The matching between multiple drivers and orders can be formulated as a decision-making problem in multi-agent systems and solved using a optimization algorithm that finds the global optimum in a centralized and coordinated way. The design of a centralized incentive mechanisms solution also plays an important role for achieving the centralized optimization solution.

- **Real-time optimization** In Chapter 4 and 5, simplified assumptions are made and the optimal taxi routing problem is treated as a static problem to make the method attractable and computationally efficient even for a network as large as the Shanghai network. Future work can consider to solve a true time-dependent problem needs. However, solving an MDP in a time-dependent network is computationally challenging, since the exact dynamic programming algorithm enumerates all states and does not scale with the significant increase in the size of the state space. In addition, it is also difficult to obtain enough reliable data in reality if the time-interval is too small. Approximate Dynamic Programming (ADP) is a potential tool (see, e.g., Larsen et al., 2004; George and Powell, 2006), which approximates the value function and avoids the evaluation of all possible future states, the so-called complete sweep.
- **Sample efficient batch RL** The standard procedure of batch RL is to use the complete set of samples for each training iteration. As a result, training time and memory requirement increase with the amount of collected transition tuples. A promising question is therefore whether it is possible to cleverly sample from the set of transition triples, such that the learning process is successful, even if the number of samples for learning is reduced or restricted. Future research can investigate how

to select promising subsets of the sampled data for learning so as to reduce both computation time and the memory requirements.

- **Routing in dynamic ride-sharing** Ride-sharing opens up an important opportunity to increase occupancy rates, and could substantially increase the efficiency of urban transportation systems, potentially reducing traffic congestion and fuel consumption. In any practical ride-share implementation, new riders and drivers continuously enter and leave the system. The ride-share can be established on short-notice, which can range from a few minutes to a few hours before departure time. The growing use of Internet-enabled mobile phones allows people to offer and request trips whenever they want, wherever they are. Thus, the design of efficient dispatching and routing of vehicles is the key to dynamic, on-demand ride-sharing.
- **Idle vehicle rebalancing** The empty taxi density is exogenous in the model-based vacant taxi routing optimization approach, which should be relaxed to consider the feedback loop between optimal taxi routing and the matching probability change due to multiple taxis routed to the same location, much like the congestion effect in traditional traffic assignment. Even though rich historical demand and supply data are available, using the data to seek an optimal relocation is not an easy task. One major issue is that changes in an allocation policy will impact future demand-supply. Future research can consider dynamically re balancing the differences between demand and supply, by reallocating available vehicles ahead of time, to achieve high efficiency in serving future demand.

- **Spatial matching between riders and vacant taxis** In the model-based vacant taxi routing optimization approach in Chapter 4, a vacant taxi is matched with passengers based on locality following a greedy method, i.e., finding the closest taxi to serve a passenger's request. A queuing strategy is applied to serve the passengers with the principle of first-come-first-served. In real world, such kind of greedy matching could lead to a spatiotemporal mismatch between taxis and passengers in the long run. Future study may look into a more flexible modeling of spatiotemporal matching between passengers and vacant taxis in a coordinated way.

## BIBLIOGRAPHY

- Abbeel, P. and Ng, A. Y. (2010). Inverse reinforcement learning, *Encyclopedia of machine learning*, Springer, pp. 554–558.
- Adler, M. W., van Ommeren, J. and Rietveld, P. (2013). Road congestion and incident duration, *Economics of transportation* **2**(4): 109–118.
- Anderson, J. R. and Schooler, L. J. (1991). Reflections of the environment in memory, *Psychological science* **2**(6): 396–408.
- Arcidiacono, P. (2004). Ability sorting and the returns to college major, *Journal of Econometrics* **121**(1): 343–375.
- Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings?, *Econometrica* **73**(5): 1477–1524.
- Arcidiacono, P. and Miller, R. A. (2008). Ccp estimation of dynamic discrete choice models with unobserved heterogeneity, *Tepper School of Business* p. 319.
- Ardeshiri, A., Jeihani, M. and Peeta, S. (2015). Driving simulator-based study of compliance behavior with dynamic message sign route guidance, *IET Intelligent Transportation Systems* **9**: 765–772.

- Arentze, T. and Timmermans, H. (2003). Modeling learning and adaptation processes in activity-travel choice: A framework and numerical experiment, *Transportation* **30**(1): 37–62.
- Asakura, Y., Kusakabe, T., Nguyen, L. X. and Ushiki, T. (2017). Incident detection methods using probe vehicles with on-board gps equipment, *Transportation research part C: emerging technologies* **81**: 330–341.
- Attanasio, A., Bregman, J., Ghiani, G. and Manni, E. (2007). Real-time fleet management at ecourier ltd, *Dynamic Fleet Management*, Springer, pp. 219–238.
- Attanasio, O., Goldberg, P. K. and Pavcnik, N. (2004). Trade reforms and wage inequality in colombia, *Journal of development Economics* **74**(2): 331–366.
- Avineri, E. (2006). The effect of reference point on stochastic network equilibrium, *Transportation Science* **40**(4): 409–420.
- Avineri, E. and Bovy, P. (2008). Identification of parameters for a prospect theory model for travel choice analysis, *Transportation Research Record: Journal of the Transportation Research Board* (2082): 141–147.
- Avineri, E. and Prashker, J. N. (2005). Sensitivity to travel time variability: travelers’ learning perspective, *Transportation Research Part C: Emerging Technologies* **13**(2): 157–183.
- Avineri, E. and Prashker, J. N. (2006). The impact of travel time information on traveler’s learning under uncertainty, *Transportation* **33**(4): 393–408.

- Bacache-Beauvallet, M. and Janin, L. (2012). Taxicab licence value and market regulation, *Transport Policy* **19**(1): 57–62.
- Balakrishna, R., Ben-Akiva, M., Bottom, J. and Gao, S. (2013). Information impacts on traveler behavior and network performance: State of knowledge and future directions, in S. V. Ukkusuri and K. Ozbay (eds), *Advances in Dynamic Network Modeling in Complex Transportation Systems*, Vol. 2 of *Complex Networks and Dynamic Systems*, Springer New York, pp. 193–224.
- Bamberg, S. and Schmidt, P. (2003). Incentives, morality, or habit? predicting students' car use for university routes with the models of ajzen, schwartz, and triandis, *Environment and behavior* **35**(2): 264–285.
- Ban, X. J., Ferris, M. C., Tang, L. and Lu, S. (2013). Risk-neutral second best toll pricing, *Transportation Research Part B: Methodological* **48**: 67–87.
- Barrett, R., Berry, M. W., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C. and Van der Vorst, H. (1994). *Templates for the solution of linear systems: building blocks for iterative methods*, Vol. 43, Siam.
- Barron, G. and Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions, *Journal of Behavioral Decision Making* **16**(3): 215–233.
- Bastin, F., Cirillo, C. and Hess, S. (2005). Evaluation of advanced optimisation methods for estimating mixed logit models, *Transportation Research Record* **1921**: 35–43.

- Bates, J., Polak, J., Jones, P. and Cook, A. (2001). The valuation of reliability for personal travel, *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3): 191–229.
- Battarra, M., Erdogan, G. and Vigo, D. (2014). Exact algorithms for the clustered vehicle routing problem, *Operations Research* **62**(1): 58–71.
- Beaud, M., Blayac, T. and Stéphan, M. (2016). The impact of travel time variability and travelers's risk attitudes on the values of time and reliability, *Transportation Research Part B: Methodological* **93**: 207–224.
- Bell, M. G. and Cassir, C. (2002). Risk-averse user equilibrium traffic assignment: an application of game theory, *Transportation Research Part B: Methodological* **36**(8): 671–681.
- Bellman, R. (1957). A markovian decision process, *Journal of Mathematics and Mechanics* pp. 679–684.
- Ben-Akiva, M. and Bierlaire, M. (1999a). Discrete choice methods and their applications to short-term travel decisions, in R. Hall (ed.), *Handbook of Transportation Science*, Kluwer, pp. 5–34.
- Ben-Akiva, M. and Bierlaire, M. (1999b). Discrete choice methods and their applications to short-term travel decisions, in R. Hall (ed.), *Handbook of Transportation Science*, Kluwer, pp. 5–34.

- Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H. N. and Mishalani, R. (2002). Real-time simulation of traffic demand-supply interactions within dynamit, in M. Gendreau and P. Marcotte (eds), *Transportation and Network Analysis: Current Trends*, Springer, pp. 19–36.
- Ben-Akiva, M. and Bolduc, D. (1996). *Multinomial probit with a logit kernel and a general parametric specification of the covariance structure*, D  partement d'  conomie, Universit   laval with Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.
- Ben-Akiva, M., De Palma, A. and Isam, K. (1991). Dynamic network models and driver information systems, *Transportation Research Part A: General* **25**(5): 251–266.
- Ben-Akiva, M., Gao, S., Wei, Z. and Wen, Y. (2012). A dynamic traffic assignment model for highly congested urban networks, *Transportation Research Part C* **24**: 62–68.
- Ben-Elia, E. and Avineri, E. (2015a). Response to travel information: A behavioural review, *Transport Review* **35**: 352–377.
- Ben-Elia, E. and Avineri, E. (2015b). Response to travel information: a behavioural review, *Transport reviews* **35**(3): 352–377.
- Ben-Elia, E., Di Pace, R., Bifulco, G. N. and Shiftan, Y. (2010). Modelling the impacts of atis accuracy on travellers' route-choice behaviour and risk perception.
- Ben-Elia, E., Di Pace, R., Bifulco, G. N. and Shiftan, Y. (2013). The impact of travel information's accuracy on route-choice, *Transportation Research Part C: Emerging Technologies* **26**: 146–159.

- Ben-Elia, E., Erev, I. and Shiftan, Y. (2008). The combined effect of information and experience on drivers's route-choice behavior, *Transportation* **35**(2): 165–177.
- Ben-Elia, E. and Shiftan, Y. (2010). Which road do i take? a learning-based model of route-choice behavior with real-time information, *Transportation Research Part A: Policy and Practice* **44**(4): 249–264.
- Bent, R. and Van Hentenryck, P. (2004). A two-stage hybrid local search for the vehicle routing problem with time windows, *Transportation Science* **38**(4): 515–530.
- Berbeglia, G., Cordeau, J.-F. and Laporte, G. (2012). A hybrid tabu search and constraint programming algorithm for the dynamic dial-a-ride problem, *INFORMS Journal on Computing* **24**(3): 343–355.
- Berndt, E. K., Hall, B. H., Hall, R. E. and Hausman, J. A. (1974). Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement* **3/4**: 653–665.
- Besse, P. C., Guillouet, B., Loubes, J.-M. and Royer, F. (2017). Destination prediction by trajectory distribution-based model, *IEEE Transactions on Intelligent Transportation Systems* .
- Bettman, J. R. (1979). An information processing theory of consumer choice, *Journal of Marketing* .
- Bhat, C. (1999). An analysis of evening commute stop-making behavior using repeated choice observations from a multi-day survey, *Transportation Research Part B: Methodological* **33**(7): 495–510.

- Bhat, C. R. (1998). Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling, *Transportation Research Part A: Policy and Practice* **32**(7): 495–507.
- Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model, *Transportation Research Part B: Methodological* **35**(7): 677–693.
- Bhat, C. R. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences, *Transportation Research Part B: Methodological* **37**(9): 837–855.
- Biele, G., Erev, I. and Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems, *Journal of Mathematical Psychology* **53**(3): 155–167.
- Bierlaire, M. (2003). Biogeme: a free package for the estimation of discrete choice models, *Swiss Transport Research Conference*, number TRANSP-OR-CONF-2006-048.
- Bierlaire, M. and Fethiarison, M. (2009). Estimation of discrete choice models: extending biogeme, *Swiss Transport Research Conference (STRC)*.
- Bierlaire, M. and Frejinger, E. (2008). Route choice modeling with network-free data, *Transportation Research Part C* **16**: 187–198.
- Bifulco, G. N., Di Pace, R. and Viti, F. (2014). Evaluating the effects of information reliability on travellers' route choice, *European Transport Research Review* **6**(1): 61–70.

- Bogers, E. A. I., Viti, F. and Hoogendoorn, S. P. (2006). Joint modeling of ATIS, habit and learning impacts on route choice by laboratory simulator experiments, *Transportation Research Record* (1926): 189–197.
- Bogers, E., Bierlaire, M. and Hoogendoorn, S. (2007). Modeling learning in route choice, *Transportation Research Record: Journal of the Transportation Research Board* (2014): 1–8.
- Bogers, E., Viti, F. and Hoogendoorn, S. (2005). Joint modeling of advanced travel information service, habit, and learning impacts on route choice by laboratory simulator experiments, *Transportation Research Record: Journal of the Transportation Research Board* (1926): 189–197.
- Bolduc, D. and Ben-Akiva, M. (1991). A multinomial probit formulation for large choice sets, *Proceedings of the 6th International Conference on Travel Behaviour*, Quebec, Canada.
- Boyer, S., Blandin, S. and Wynter, L. (2015). Stability of transportation networks under adaptive routing policies, *Transportation Research Part B* **81**: 886–903.
- Brownstone, D., Ghosh, A., Golob, T. F., Kazimi, C. and Van Amelsfort, D. (2003). Drivers' willingness-to-pay to reduce travel time: evidence from the san diego i-15 congestion pricing project, *Transportation Research Part A: Policy and Practice* **37**(4): 373–387.

- Brownstone, D. and Small, K. A. (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations, *Transportation Research Part A: Policy and Practice* **39**(4): 279–293.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. General considerations, *IMA Journal of Applied Mathematics* **6**(1): 76–90.
- Bussemeyer, J. R. and Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment., *Psychological review* **100**(3): 432.
- Cantarella, G. E. and Cascetta, E. (1995). Dynamic processes and equilibrium in transportation networks: Towards a unifying theory, *Transportation Science* **29**(4): 305–329.
- Cascetta, E., Nuzzolo, A., Russo, F. and Vitetta, A. (1996). A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks, *J. B. Lesort (ed.)*, Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon, France.
- Ch, J. D. C., Briones, J. et al. (2006). A diagrammatic analysis of the market for cruising taxis, *Transportation Research Part E: Logistics and Transportation Review* **42**(6): 498–526.
- Chancelier, J.-P., De Lara, M. and De Palma, A. (2007). Risk aversion, road choice, and the one-armed bandit problem, *transportation science* **41**(1): 1–14.

- Chang, H.-w., Tai, Y.-c. and Hsu, J. Y.-j. (2009). Context-aware taxi demand hotspots prediction, *International Journal of Business Intelligence and Data Mining* **5**(1): 3–18.
- Chen, A., Yang, C., Kongsomsaksakul, S. and Lee, M. (2007). Network-based accessibility measures for vulnerability analysis of degradable transportation networks, *Networks and Spatial Economics* **7**(3): 241–256.
- Chen, A. and Zhou, Z. (2010). The  $\alpha$ -reliable mean-excess traffic equilibrium model with stochastic travel times, *Transportation Research Part B: Methodological* **44**(4): 493–513.
- Chen, A., Zhou, Z., Chootinan, P., Ryu, S., Yang, C. and Wong, S. (2011). Transport network design problem under uncertainty: a review and new developments, *Transport Reviews* **31**(6): 743–768.
- Chen, A., Zhou, Z. and Lam, W. H. (2011). Modeling stochastic perception error in the mean-excess traffic equilibrium model, *Transportation Research Part B: Methodological* **45**(10): 1619–1640.
- Chen, B. Y., Lam, W. H., Sumalee, A., Li, Q., Shao, H. and Fang, Z. (2013). Finding reliable shortest paths in road networks under uncertainty, *Networks and spatial economics* **13**(2): 123–148.
- Chen, B. Y., Lam, W. H., Sumalee, A., Li, Q. and Tam, M. L. (2014). Reliable shortest path problems in stochastic time-dependent networks, *Journal of Intelligent Transportation Systems* **18**(2): 177–189.

- Chen, B. Y., Lam, W. H., Sumalee, A. and Li, Z.-l. (2012). Reliable shortest path finding in stochastic networks with spatial correlated link travel times, *International Journal of Geographical Information Science* **26**(2): 365–386.
- Chen, M., Yu, G., Chen, P. and Wang, Y. (2017). A copula-based approach for estimating the travel time reliability of urban arterial, *Transportation Research Part C: Emerging Technologies* **82**: 1–23.
- Chen, P. W. and Nie, Y. M. (2017). Connecting e-hailing to mass transit platform: Analysis of relative spatial position, *Transportation Research Part C: Emerging Technologies* **77**: 444–461.
- Chen, X. M., Zahiri, M. and Zhang, S. (2017). Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach, *Transportation Research Part C: Emerging Technologies* **76**: 51–70.
- Chorus, C. G. (2010). A new model of random regret minimization, *European Journal of Transport and Infrastructure Research* **2**(10).
- Chorus, C. G. (2012). Regret theory-based route choices and traffic equilibria, *Transportmetrica* **8**(4): 291–305.
- Chorus, C. G., Arentze, T. A. and Timmermans, H. J. (2009). Traveler compliance with advice: A bayesian utilitarian perspective, *Transportation Research Part E: Logistics and Transportation Review* **45**(3): 486–500.

- Chorus, C. G. and Dellaert, B. G. (2012). Travel choice inertia: the joint role of risk aversion and learning, *Journal of Transport Economics and Policy (JTEP)* **46**(1): 139–155.
- Chorus, C. G., Molin, E. J. E. and van Wee, B. (2006). Use and effects of advanced traveller information services (atis): A review of the literature, *Transport Reviews* **26**(2): 127–149.
- Chorus, C. G. and van Cranenburgh, S. (2018). Specification of regret-based models of choice behaviour: formal analyses and experimental design based evidence—commentary, *Transportation* **45**(1): 247–256.
- Chow, J. Y. and Regan, A. C. (2011). Network-based real option models, *Transportation Research Part B: Methodological* **45**(4): 682–695.
- Chow, J. Y. and Sayarshad, H. R. (2015). Reference policies for non-myopic sequential network design and timing problems, *Networks and Spatial Economics* pp. 1–27.
- Clark, S. and Watling, D. (2005). Modelling network travel time reliability under stochastic demand, *Transportation Research Part B: Methodological* **39**(2): 119–140.
- Connors, R. D. and Sumalee, A. (2009). A network equilibrium model with travellers’ perception of stochastic travel times, *Transportation Research Part B: Methodological* **43**(6): 614–624.
- Crainic, T. G., Nguyen, P. K. and Toulouse, M. (2016). Synchronized multi-trip multi-traffic pickup & delivery in city logistics, *Transportation Research Procedia* **12**: 26–39.

- Daly, A., Hess, S. and Dekker, T. (2014). Practical solutions for sampling alternatives in large-scale models, *Transportation Research Record* (2429): 148–156.
- De Asis, K., Hernandez-Garcia, J. F., Holland, G. Z. and Sutton, R. S. (2017). Multi-step reinforcement learning: A unifying algorithm, *arXiv preprint arXiv:1703.01327* .
- de Bekker-Grob, E. W. and Chorus, C. G. (2013). Random regret-based discrete-choice modelling: an application to healthcare, *PharmacoEconomics* **31**(7): 623–634.
- De Brébisson, A., Simon, É., Auvolet, A., Vincent, P. and Bengio, Y. (2015). Artificial neural networks applied to taxi destination prediction, *arXiv preprint arXiv:1508.00021* .
- de Moraes Ramos, G. (2015). *Dynamic Route Choice Modelling of the Effects of Travel Information using RP Data*, Ph.D. thesis, TU Delft, Delft University of Technology.
- Deisenroth, M. P., Neumann, G., Peters, J. et al. (2013). A survey on policy search for robotics, *Foundations and Trends® in Robotics* **2**(1–2): 1–142.
- Deisenroth, M. and Rasmussen, C. E. (2011). Pilco: A model-based and data-efficient approach to policy search, *Proceedings of the 28th International Conference on machine learning (ICML-11)*, pp. 465–472.
- Delle Site, P. (2018). A mixed-behaviour equilibrium model under predictive and static advanced traveller information systems (atis) and state-dependent route choice, *Transportation Research Part C: Emerging Technologies* **86**: 549–562.

- Dell'Orco, M. and Marinelli, M. (2017). Modeling the dynamic effect of information on drivers's choice behavior in the context of an advanced traveler information system, *Transportation Research Part C: Emerging Technologies* **85**: 168–183.
- Deng, Z. and Ji, M. (2011). Spatiotemporal structure of taxi services in shanghai: Using exploratory spatial data analysis, *Geoinformatics, 2011 19th International Conference on*, IEEE, pp. 1–5.
- Denrell, J. and March, J. G. (2001). Adaptation as information restriction: The hot stove effect, *Organization Science* **12**(5): 523–538.
- Ding, J. and Gao, S. (2012). An optimal adaptive routing algorithm for large-scale stochastic time-dependent networks, *Presented at TRB 91st Annual Meeting, Washington, D.C.* .
- Ding, J., Gao, S., Jenelius, E., Rahmani, M., Huang, H., Ma, L., Pereira, F. and Ben-Akiva, M. (2014). Routing policy choice set generation in stochastic time-dependent networks: Case studies for stockholm and singapore, *Transportation Research Record* **2466**: 76–86.
- Ding-Mastera, J. (2016). *Adaptive Route Choice in Stochastic Time-Dependent Networks: Routing Algorithms and Choice Modeling*, PhD thesis, Department of Civil and Environmental Engineering, University of Massachusetts Amherst.

- Ding-Mastera, J., Gao, S., Jenelius, E., Rahmani, M., Huang, H., Ma, L., Pereira, F. and Ben-Akiva, M. (2014). Routing policy choice set generation in stochastic time-dependent networks: Case studies for stockholm, sweden, and singapore, *Transportation Research Record: Journal of the Transportation Research Board* (2466): 76–86.
- Ding-Mastera, J., Gao, S., Jenelius, E., Rahmani, M., Pereira, F. and Ben-Akiva, M. (2015). Latent-class routing policy choice model with revealed-preference data, *Transportation Research Board 94th Annual Meeting*, number 15-1963.
- Dong, H., Zhang, X., Dong, Y., Chen, C. and Rao, F. (2014). Recommend a profitable cruising route for taxi drivers, *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, IEEE, pp. 2003–2008.
- Dong, Y., Wang, S., Li, L. and Zhang, Z. (2018). An empirical study on travel patterns of internet based ride-sharing, *Transportation research part C: emerging technologies* **86**: 1–22.
- Douglas, G. W. (1972). Price regulation and optimal service standards: The taxicab industry, *Journal of Transport Economics and Policy* pp. 116–127.
- Eberwein, C. and Ham, J. C. (2008). Obtaining analytic derivatives for a popular discrete-choice dynamic programming model, *Economics Letters* **101**(3): 168–171.
- Engelson, L. and Fosgerau, M. (2016). The cost of travel time variability: Three measures with properties, *Transportation Research Part B: Methodological* **91**: 555–564.

- Engle, R. F. (1984). Wald, likelihood ratio, and lagrange multiplier tests in econometrics, *Handbook of econometrics* **2**: 775–826.
- Erev, I. and Barron, G. (2005a). On adaptation, maximization, and reinforcement learning among cognitive strategies, *Psychological Review* **112**(4): 912–931.
- Erev, I. and Barron, G. (2005b). On adaptation, maximization, and reinforcement learning among cognitive strategies., *Psychological review* **112**(4): 912.
- Erev, I., Ert, E. and Roth, A. E. (2010). A choice prediction competition for market entry games: An introduction, *Games* **1**(2): 117–136.
- Ernst, D., Geurts, P. and Wehenkel, L. (2005). Tree-based batch mode reinforcement learning, *Journal of Machine Learning Research* **6**(Apr): 503–556.
- Estes, W. K. (2014). *Handbook of Learning and Cognitive Processes (Volume 4): Attention and Memory*, Psychology Press.
- Fan, Y., Kalaba, R. and Moore, J. (2005). Arriving on time, *Journal of Optimization Theory and Applications* **127**(3): 497–513.
- Fan, Y. and Nie, Y. (2006). Optimal routing for maximizing the travel time reliability, *Networks and Spatial Economics* **6**(3-4): 333–344.
- Ferrucci, F. and Bock, S. (2015). A general approach for controlling vehicle en-route diversions in dynamic vehicle routing problems, *Transportation Research Part B: Methodological* **77**: 76–87.

- Ferrucci, F. and Bock, S. (2016). Pro-active real-time routing in applications with multiple request patterns, *European Journal of Operational Research* **253**(2): 356–371.
- Ferrucci, F., Bock, S. and Gendreau, M. (2013). A pro-active real-time control approach for dynamic vehicle routing problems dealing with the delivery of urgent goods, *European Journal of Operational Research* **225**(1): 130–141.
- Figliozzi, M. A., Mahmassani, H. S. and Jaillet, P. (2007). Pricing in dynamic vehicle routing problems, *Transportation Science* **41**(3): 302–318.
- Fletcher, R. (1970). A new approach to variable metric algorithms, *The Computer Journal* **13**(3): 317–322.
- Flötteröd, G. and Bierlaire, M. (2013). Metropolis–hastings sampling of paths, *Transportation Research Part B* **48**: 53–66.
- Fosgerau, M. and Bierlaire, M. (2009). Discrete choice models with multiplicative error terms, *Transportation Research Part B* **43**(5): 494–505.
- Fosgerau, M., Frejinger, E. and Karlstrom, A. (2011). A dynamic discrete choice approach for consistent route choice model estimation, *Presented at the Swiss Transport Research Conference (STRC)* .
- Fosgerau, M., Frejinger, E. and Karlstrom, A. (2013a). A link based network route choice model with unrestricted choice set, *Transportation Research Part B* **56**: 70–80.
- Fosgerau, M., Frejinger, E. and Karlstrom, A. (2013b). A link based network route choice model with unrestricted choice set, *Transportation Research Part B: Methodological* **56**: 70–80.

- Fosgerau, M. and Fukuda, D. (2012). Valuing travel time variability: Characteristics of the travel time distribution on an urban road, *Transportation Research Part C: Emerging Technologies* **24**: 83–101.
- Fosgerau, M. and Karlström, A. (2010). The value of reliability, *Transportation Research Part B: Methodological* **44**(1): 38–49.
- Frejinger, E. (2008). *Route choice analysis: data, models, algorithms and applications*, PhD thesis, École Polytechnique Federale de Lausanne.
- Frejinger, E. and Bierlaire, M. (2007a). Capturing correlation with subnetworks in route choice models, *Transportation Research Part B* **41**: 363–378.
- Frejinger, E. and Bierlaire, M. (2007b). Capturing correlation with subnetworks in route choice models, *Transportation Research Part B* **41**(3): 363–378.
- Frejinger, E., Bierlaire, M. and Ben-Akiva, M. (2009a). Sampling of alternatives for route choice modeling, *Transportation Research Part B* **43**(10).
- Frejinger, E., Bierlaire, M. and Ben-Akiva, M. (2009b). Sampling of alternatives for route choice modeling, *Transportation Research Part B: Methodological* **43**(10): 984–994.
- Gao, S. (2005a). *Optimal Adaptive Routing and Traffic Assignment in Stochastic Time-Dependent Networks*, PhD thesis, Massachusetts Institute of Technology.
- Gao, S. (2005b). *Optimal adaptive routing and traffic assignment in stochastic time-dependent networks*, PhD thesis, Massachusetts Institute of Technology.

- Gao, S. (2012a). Modeling strategic route choice and real-time information impacts in stochastic and time-dependent networks, *IEEE Transactions on Intelligent Transportation Systems* **13**(3): 1298–1311.
- Gao, S. (2012b). Modeling strategic route choice and real-time information impacts in stochastic and time-dependent networks, *IEEE Transactions on Intelligent Transportation Systems* **13**(3): 1298–1311.
- Gao, S. and Chabini, I. (2002). Best routing policy problem in stochastic time-dependent networks, *Transportation Research Record: Journal of the Transportation Research Board* **1783**(1): 188–196.
- Gao, S. and Chabini, I. (2006a). Optimal routing policy problems in stochastic time-dependent networks, *Transportation Research Part B* **40**(2): 93–122.
- Gao, S. and Chabini, I. (2006b). Optimal routing policy problems in stochastic time-dependent networks, *Transportation Research Part B: Methodological* **40**(2): 93–122.
- Gao, S., Frejinger, E. and Ben-Akiva, M. (2008a). Adaptive route choice models in stochastic time-dependent networks, *Transportation Research Record: Journal of the Transportation Research Board* **2085**(1): 136–143.
- Gao, S., Frejinger, E. and Ben-Akiva, M. (2008b). Adaptive route choice models in stochastic time dependent networks, **2085**: 136–143.

- Gao, S., Frejinger, E. and Ben-Akiva, M. (2010). Adaptive route choices in risky traffic networks: A prospect theory approach, *Transportation research part C: emerging technologies* **18**(5): 727–740.
- Gao, S., Frejinger, E. and Ben-Akiva, M. (2011). Cognitive cost in route choice with real-time information: An exploratory analysis, *Transportation Research Part A* **45**: 916–926.
- Gao, S. and Huang, H. (2009). Is more information better for routing in an uncertain network?, *Transportation Research Board 88th Annual Meeting*, number 09-1315.
- Gao, S. and Huang, H. (2012a). Real-time traveler information for optimal adaptive routing in stochastic time-dependent networks, *Transportation Research Part C* **21**(1): 196–213.
- Gao, S. and Huang, H. (2012b). Real-time traveler information for optimal adaptive routing in stochastic time-dependent networks, *Transportation Research Part C: Emerging Technologies* **21**(1): 196–213.
- Gardner, B. (2009). Modelling motivation and habit in stable travel mode contexts, *Transportation Research Part F: Traffic Psychology and Behaviour* **12**(1): 68–76.
- Ge, Y., Xiong, H., Tuzhilin, A., Xiao, K., Gruteser, M. and Pazzani, M. (2010). An energy-efficient mobile recommender system, *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 899–908.

- George, A. P. and Powell, W. B. (2006). Adaptive stepsizes for recursive estimation with applications in approximate dynamic programming, *Machine learning* **65**(1): 167–198.
- Ghiani, G., Manni, E., Quaranta, A. and Triki, C. (2009). Anticipatory algorithms for same-day courier dispatching, *Transportation Research Part E: Logistics and Transportation Review* **45**(1): 96–106.
- Ghiani, G., Manni, E. and Thomas, B. W. (2012). A comparison of anticipatory algorithms for the dynamic and stochastic traveling salesman problem, *Transportation Science* **46**(3): 374–387.
- Ghosh, A. (2001). Valuing time and reliability: commuters’ mode choice from a real time congestion pricing experiment.
- Gilbert, J. R., Moler, C. and Schreiber, R. (1992). Sparse matrices in matlab: design and implementation, *SIAM Journal on Matrix Analysis and Applications* **13**(1): 333–356.
- Goldfarb, D. (1970). A family of variable metric updates derived by variational means, *Mathematics of Computation* **24**(109): 23–26.
- Gonzalez, C. and Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience., *Psychological review* **118**(4): 523.
- Gonzalez, C., Lerch, J. F. and Lebiere, C. (2003). Instance-based learning in dynamic decision making, *Cognitive Science* **27**(4): 591–635.

- Gromicho, J., van Hoorn, J. J., Kok, A. L. and Schutten, J. (2012). Restricted dynamic programming: a flexible framework for solving realistic vrps, *Computers & operations research* **39**(5): 902–909.
- Guevara, C. A., Chorus, C. G. and Ben-Akiva, M. E. (2014). Sampling of alternatives in random regret minimization models, *Transportation Science* pp. 1–16.
- Guevara, C. A., Tang, Y. and Gao, S. (Forthcoming). The initial condition problem with complete history dependency in learning models for travel choices, *Transportation Research Part B: Methodological* .
- Hall, R. W. (1986). The fastest path through a network with random time-dependent travel times, *Transportation science* **20**(3): 182–188.
- Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numerische Mathematik* **2**(1): 84–90.
- Han, M., Senellart, P., Bressan, S. and Wu, H. (2016). Routing an autonomous taxi with reinforcement learning, *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, ACM, pp. 2421–2424.
- Han, Q., Arentze, T., Timmermans, H., Janssens, D. and Wets, G. (2011). The effects of social networks on choice set dynamics: Results of numerical simulations using an agent-based approach, *Transportation Research Part A* **45**: 310–322.
- Hasnat, M. M. and Hasan, S. (2018). Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data, *Transportation Research Part C: Emerging Technologies* **96**: 38–54.

- Hauser, J. R. and Gaskin, S. P. (1984). Application of the “defender” consumer model, *Marketing Science* **3**(4): 327–351.
- Hauser, J. R. and Wernerfelt, B. (1990). An evaluation cost model of consideration sets, *Journal of consumer research* pp. 393–408.
- Hausman, J. A., Leonard, G. K. and McFadden, D. (1995). A utility-consistent, combined discrete choice and count data model assessing recreational use losses due to natural resource damage, *Journal of Public Economics* **56**(1): 1–30.
- He, F. and Shen, Z.-J. M. (2015). Modeling taxi services with smartphone-based e-hailing applications, *Transportation Research Part C: Emerging Technologies* **58**: 93–106.
- He, X. and Liu, H. X. (2012). Modeling the day-to-day traffic evolution process after an unexpected network disruption, *Transportation Research Part B: Methodological* **46**(1): 50–71.
- Helbing, D., Schönhof, M., Stark, H.-U. and Holyst, J. A. (2005). How individuals learn to take turns: Emergence of alternating cooperation in a congestion game and the prisoner’s dilemma, *Advances in Complex Systems* **8**(01): 87–116.
- Hensher, D. A. and Greene, W. H. (2003). The mixed logit model: the state of practice, *Transportation* **30**(2): 133–176.
- Hensher, D. A., Greene, W. H. and Chorus, C. G. (2013). Random regret minimization or random utility maximization: an exploratory analysis in the context of automobile fuel choice, *Journal of Advanced Transportation* **47**(7): 667–678.

- Hensher, D. A., Greene, W. H. and Li, Z. (2011). Embedding risk attitude and decision weights in non-linear logit to accommodate time variability in the value of expected travel time savings, *Transportation research part B: methodological* **45**(7): 954–972.
- Hertwig, R., Barron, G., Weber, E. U. and Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice, *Psychological science* **15**(8): 534–539.
- Hong, Z., Chen, Y., Mahmassani, H. S. and Xu, S. (2017). Commuter ride-sharing using topology-based vehicle trajectory clustering: Methodology, application and impact evaluation, *Transportation Research Part C: Emerging Technologies* **85**: 573–590.
- Horowitz, J. L. (1984). The stability of stochastic equilibrium in a two-link transportation network, *Transportation Research Part B: Methodological* **18**(1): 13–28.
- Horowitz, J. L. and Louviere, J. J. (1995). What is the role of consideration sets in choice modeling?, *International Journal of Research in Marketing* **12**(1): 39–54.
- Hosni, H., Naoum-Sawaya, J. and Artail, H. (2014). The shared-taxi problem: Formulation and solution methods, *Transportation Research Part B: Methodological* **70**: 303–318.
- Hou, Y., Li, X., Zhao, Y., Jia, X., Sadek, A. W., Hulme, K. and Qiao, C. (2013). Towards efficient vacant taxis cruising guidance, *Global Communications Conference (GLOBECOM), 2013 IEEE*, IEEE, pp. 54–59.
- Hu et al., H. (2012a). Pick-up tree based route recommendation from taxi trajectories, *International Conference on Web-Age Information Management*, Springer, pp. 471–483.

- Hu et al., X. (2012b). Modeling routing behavior for vacant taxicabs in urban traffic networks, *Transportation Research Record: Journal of the Transportation Research Board* pp. 81–88.
- Hu, H., Wu, Z., Mao, B., Zhuang, Y., Cao, J. and Pan, J. (2012). Pick-up tree based route recommendation from taxi trajectories, *International Conference on Web-Age Information Management*, Springer, pp. 471–483.
- Hu, L., Dong, J., Lin, Z. and Yang, J. (2018). Analyzing battery electric vehicle feasibility from taxi travel patterns: The case study of new york city, usa, *Transportation Research Part C: Emerging Technologies* **87**: 91–104.
- Hu, T.-Y. and Mahmassani, H. S. (1995). Evolution of network flows under real-time information: Day-to-day dynamic simulation assignment framework, *Transportation Research Record* pp. 46–56.
- Hu, X., Gao, S., Chiu, Y.-C. and Lin, D.-Y. (2012). Modeling routing behavior for vacant taxicabs in urban traffic networks, number 2284, Transportation Research Board of the National Academies, pp. 81–88.
- Huang, J., Huangfu, X., Sun, H., Li, H., Zhao, P., Cheng, H. and Song, Q. (2015). Backward path growth for efficient mobile sequential recommendation, *IEEE Transactions on Knowledge and Data Engineering* **27**(1): 46–60.
- Huang, Z., Ling, X., Wang, P., Zhang, F., Mao, Y., Lin, T. and Wang, F.-Y. (2018). Modeling real-time human mobility based on mobile phone and transportation data fusion, *Transportation Research Part C: Emerging Technologies* **96**: 251–269.

- Huang, Z., Zhao, Z., Shijia, E., Yu, C., Shan, G., Li, T., Cheng, J., Sun, J. and Xiang, Y. (2017). Prace: A taxi recommender for finding passengers with deep learning approaches, *International Conference on Intelligent Computing*, Springer, pp. 759–770.
- Hvattum, L. M., Løkketangen, A. and Laporte, G. (2006). Solving a dynamic and stochastic vehicle routing problem with a sample scenario hedging heuristic, *Transportation Science* **40**(4): 421–438.
- Hwang, R.-H., Hsueh, Y.-L. and Chen, Y.-T. (2015). An effective taxi recommender system based on a spatio-temporal factor analysis model, *Information Sciences* **314**: 28–40.
- Ichoua, S., Gendreau, M. and Potvin, J.-Y. (2006). Exploiting knowledge about future demands for real-time vehicle dispatching, *Transportation Science* **40**(2): 211–225.
- Iskhakov, F., Rust, J., Schjerning, B. and Lee, J. (2015). Constrained optimization approaches to estimation of structural models: Comment, *Available at SSRN* .
- Jenelius, E. and Mattsson, L.-G. (2012). Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study, *Transportation research part A: policy and practice* **46**(5): 746–760.
- Jha, M., Madanat, S. and Peeta, S. (1998). Perception updating and day-to-day travel choice dynamics in traffic networks with information provision, *Transportation Research Part C: Emerging Technologies* **6**(3): 189–212.

- Jiang, B., Yin, J. and Zhao, S. (2009). Characterizing the human mobility pattern in a large street network, *Physical Review E* **80**(2): 021136.
- Jin, S. T., Kong, H., Wu, R. and Sui, D. Z. (2018). Ridesourcing, the sharing economy, and the future of cities, *Cities* .
- Jin, X., Hossan, M. S. and Asgari, H. (2015). Investigating the value of time and value of reliability for investigating the value of time and value of reliability for managed lanes, *Technical Report FDOT Project Number BDV29-977-12*, Florida International University.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica* **47**(2): 263–292.
- Kaysi, I. (1991). *Framework and models for provision of driver information system*, PhD thesis, PhD-thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Ke, J., Zheng, H., Yang, H. and Chen, X. M. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach, *Transportation Research Part C: Emerging Technologies* **85**: 591–608.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* ., Prentice-Hall, Inc.
- Kim, H., Oh, J.-S. and Jayakrishnan, R. (2005). Effect of taxi information system on efficiency and quality of taxi services, *Transportation Research Record: Journal of the Transportation Research Board* (1903): 96–104.

- Kim, J. and Mahmassani, H. S. (2015). Compound gamma representation for modeling travel time variability in a traffic network, *Transportation Research Part B: Methodological* **80**: 40–63.
- Kirchler, D. and Calvo, R. W. (2013). A granular tabu search algorithm for the dial-a-ride problem, *Transportation Research Part B: Methodological* **56**: 120–135.
- Klein, I. and Ben-Elia, E. (2017). System optimal atis as a congestion management instrument—game-based experiment and agent based model, *Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2017 5th IEEE International Conference on*, IEEE, pp. 816–820.
- Klein, I., Levy, N. and Ben-Elia, E. (2018). An agent-based model of the emergence of cooperation and a fair and stable system optimum using atis on a simple road network, *Transportation research part C: emerging technologies* **86**: 183–201.
- Kober, J., Bagnell, J. A. and Peters, J. (2013). Reinforcement learning in robotics: A survey, *The International Journal of Robotics Research* **32**(11): 1238–1274.
- Kok, A. L., Hans, E. and Schutten, J. (2012). Vehicle routing under time-dependent travel times: the impact of congestion avoidance, *Computers & Operations Research* **39**(5): 910–918.
- Kou, W., Chen, X., Yu, L., Qi, Y. and Wang, Y. (2017). Urban commuters’s valuation of travel time reliability based on stated preference survey: A case study of beijing, *Transportation Research Part A: Policy and Practice* **95**: 372–380.

- Kraan, M., Mahmassani, H. and Huynh, N. (2000). Traveler responses to advanced traveler information systems for shopping trips: Interactive survey approach, *Transportation Research Record: Journal of the Transportation Research Board* (1725): 116–123.
- Laha, A. K. and Putatunda, S. (2018). Real time location prediction with taxi-gps data streams, *Transportation Research Part C: Emerging Technologies* **92**: 298–322.
- Lai, X. and Bierlaire, M. (2015a). Specification of the cross-nested logit model with sampling of alternatives for route choice models, *Transportation Research Part B* **80**: 220–234.
- Lai, X. and Bierlaire, M. (2015b). Specification of the cross-nested logit model with sampling of alternatives for route choice models, *Transportation Research Part B: Methodological* **80**: 220–234.
- Lam, T. C. (2000). The effect of variability of travel time on route and time-of-day choice, *Technical report*.
- Lam, T. C. and Small, K. A. (2001). The value of time and reliability: measurement from a value pricing experiment, *Transportation Research Part E: Logistics and Transportation Review* **37**(2-3): 231–251.
- Lam, W. H., Shao, H. and Sumalee, A. (2008). Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply, *Transportation research part B: methodological* **42**(10): 890–910.
- Larsen, A., Madsen, O. B. and Solomon, M. M. (2004). The a priori dynamic traveling salesman problem with time windows, *Transportation Science* **38**(4): 459–472.

- Larsen, A., Madsen, O. and Solomon, M. (2002). Partially dynamic vehicle routing—models and algorithms, *Journal of the operational research society* **53**(6): 637–646.
- Larson, R. C. and Odoni, A. R. (1981). *Urban operations research*, Prentice-Hall.
- Le Cam, L. and Yang, G. L. (2012). *Asymptotics in statistics: some basic concepts*, Springer Science & Business Media.
- Lee, A. and Savelsbergh, M. (2015). Dynamic ridesharing: Is there a role for dedicated drivers?, *Transportation Research Part B: Methodological* **81**: 483–497.
- Lejarraga, T., Dutt, V. and Gonzalez, C. (2012). Instance-based learning: A general model of repeated binary choice, *Journal of Behavioral Decision Making* **25**(2): 143–153.
- Li, D., Miwa, T. and Morikawa, T. (2013). Dynamic route choice behavior analysis considering en route learning and choices, *Transportation Research Record: Journal of the Transportation Research Board* (2383): 1–9.
- Li, M. and Huang, H.-J. (2017). A regret theory-based route choice model, *Transportmetrica A: Transport Science* **13**(3): 250–272.
- Li, M., Rouphail, N. M., Mahmoudi, M., Liu, J. and Zhou, X. (2017). Multi-scenario optimization approach for assessing the impacts of advanced traffic information under realistic stochastic capacity distributions, *Transportation Research Part C: Emerging Technologies* **77**: 113–133.
- Li, Q., Zeng, Z., Yang, B. and Zhang, T. (2009). Hierarchical route planning based on taxi gps-trajectories, *2009 17th International Conference on Geoinformatics*, IEEE, pp. 1–5.

- Li, R., Liu, Z. and Zhang, R. (2018). Studying the benefits of carpooling in an urban area using automatic vehicle identification data, *Transportation Research Part C: Emerging Technologies* **93**: 367–380.
- Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., Zhang, W. and Wang, Z. (2012). Prediction of urban human mobility using large-scale taxi traces and its applications, *Frontiers of Computer Science* **6**(1): 111–121.
- Li, Z., Hensher, D. A. and Rose, J. M. (2010). Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence, *Transportation research part E: logistics and transportation review* **46**(3): 384–403.
- Liu, S., Ni, L. M. and Krishnan, R. (2014). Fraud detection from taxis' driving behaviors, *IEEE Transactions on Vehicular Technology* **63**(1): 464–472.
- Lo, H. K., Luo, X. and Siu, B. W. (2006). Degradable transport network: travel time budget of travelers with heterogeneous risk aversion, *Transportation Research Part B: Methodological* **40**(9): 792–806.
- Lokhandwala, M. and Cai, H. (2018). Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of nyc, *Transportation Research Part C: Emerging Technologies* **97**: 45–60.
- Lomax, T., Shrank, D. and Eiselee, B. (2011). Urban mobility report powered by inrix traffic data, *College Station, Texas* .

- Long, J., Szeto, W., Du, J. and Wong, R. (2017). A dynamic taxi traffic assignment model: A two-level continuum transportation system approach, *Transportation Research Part B: Methodological* **100**: 222–254.
- Lowalekar, M., Varakantham, P. and Jaillet, P. (2016). Online spatio-temporal matching in stochastic and dynamic domains., *AAAI*, pp. 3271–3277.
- Lu, X., Gao, S. and Ben-Elia, E. (2011a). Information impacts on route choice and learning behavior in a congested network: An experimental approach, *Transportation Research Record* (Forthcoming).
- Lu, X., Gao, S. and Ben-Elia, E. (2011b). Information impacts on route choice and learning behavior in a congested network: experimental approach, *Transportation Research Record: Journal of the Transportation Research Board* (2243): 89–98.
- Lu, X., Gao, S., Ben-Elia, E. and Pothering, R. (2014). Travelers’ day-to-day route choice behavior with real-time information in a congested risky network, *Mathematical Population Studies* **21**(4): 205–219.
- Lv, Z., Wu, J., Yao, S. and Zhu, L. (2017). Fcd-based analysis of taxi operation characteristics: A case of shanghai, *Journal of East China Normal University* **3**: 133–144.
- Ma, S., Zheng, Y. and Wolfson, O. (2013). T-share: A large-scale dynamic taxi ridesharing service, *Data Engineering (ICDE), 2013 IEEE 29th International Conference on*, IEEE, pp. 410–421.

- Ma, T.-Y. and Di Pace, R. (2017). Comparing paradigms for strategy learning of route choice with traffic information under uncertainty, *Expert Systems with Applications* **88**: 352–367.
- Ma, Z., Koutsopoulos, H. N., Ferreira, L. and Mesbah, M. (2017). Estimation of trip travel time distribution using a generalized markov chain approach, *Transportation Research Part C: Emerging Technologies* **74**: 1–21.
- Maciejewski, M. and Nagel, K. (2013). Simulation and dynamic optimization of taxi services in matsim, *Technical report*, VSP Working Paper 13-05, TU Berlin, Transport Systems Planning and Transport Telematics. [www.vsp.tu-berlin.de/publications](http://www.vsp.tu-berlin.de/publications).
- Mahmassani, H. S. (2001). Dynamic network traffic assignment and simulation methodology for advanced system management applications, *Network and Spatial Economics* **1**: 267–292.
- Mahmassani, H. S., Huynh, N. N., Srinivasan, K. and Kraan, M. (2003). Tripmaker choice behavior for shopping trips under real-time information: model formulation and results of stated-preference internet-based interactive experiments, *Journal of Retailing and Consumer Services* **10**(6): 311–321.
- Mahmood, A. R., Yu, H. and Sutton, R. S. (2017). Multi-step off-policy learning without importance sampling ratios, *arXiv preprint arXiv:1702.03006* .
- Mai, T. (2015). Recursive network mev model for route choice analysis, *CIRRELT Technical Report, N° 37* .

- Mai, T. (2016). A method of integrating correlation structures for a generalized recursive route choice model, *Transportation Research Part B* **93**(1): 146–161.
- Mai, T., Bastin, F. and Toulouse, M. (2014). On optimization algorithms for the maximum likelihood estimation, *CIRRELT Technical Report, N° 64* .
- Mai, T., Fosgerau, M. and Frejinger, E. (2015a). A nested recursive logit model for route choice analysis, *Transportation Research Part B* **75**: 100–112.
- Mai, T., Fosgerau, M. and Frejinger, E. (2015b). A nested recursive logit model for route choice analysis, *Transportation Research Part B: Methodological* **75**: 100–112.
- Mai, T., Fosgerau, M. and Frejinger, E. (2015c). A nested recursive logit model for route choice analysis, *Transportation Research Part B* **75**(0): 100 – 112.
- Mak, V., Gisches, E. J. and Rapoport, A. (2015). Route vs. segment: An experiment on real-time travel information in congestible networks, *Production and Operations Management* **24**(6): 947–960.
- Maling, D. H. (2013). *Coordinate systems and map projections*, Elsevier.
- Masoud, N. and Jayakrishnan, R. (2017a). A decomposition algorithm to solve the multi-hop peer-to-peer ride-matching problem, *Transportation Research Part B: Methodological* **99**: 1–29.
- Masoud, N. and Jayakrishnan, R. (2017b). A real-time algorithm to solve the peer-to-peer ride-matching problem in a flexible ridesharing system, *Transportation Research Part B: Methodological* **106**: 218–236.

- Mayeres, I., Ochelen, S. and Proost, S. (1996). The marginal external costs of urban transport, *Transportation Research Part D: Transport and Environment* **1**(2): 111–130.
- McFadden, D. (2001). Economic choices, *American Economic Review* pp. 351–378.
- McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response, *Journal of applied Econometrics* pp. 447–470.
- McFadden, D. et al. (1978). *Modelling the choice of residential location*, Institute of Transportation Studies, University of California.
- Miller-Hooks, E. D. and Mahmassani, H. S. (2000). Least expected time paths in stochastic, time-varying transportation networks, *Transportation Science* **34**(2): 198–215.
- Miller, R. A. (1984). Job matching and occupational choice, *The Journal of Political Economy* pp. 1086–1120.
- Mirchandani, P. and Soroush, H. (1987). Generalized traffic equilibrium with probabilistic travel times and perceptions, *Transportation Science* **21**(3): 133–152.
- Mitrović-Minić, S., Krishnamurti, R. and Laporte, G. (2004). Double-horizon based heuristics for the dynamic pickup and delivery problem with time windows, *Transportation Research Part B: Methodological* **38**(8): 669–685.
- Moreira-Matias, L., Gama, J., Ferreira, M. and Damas, L. (2012). A predictive model for the passenger demand on a taxi network, *2012 15th International IEEE Conference on Intelligent Transportation Systems*, IEEE, pp. 1014–1019.

- Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J. and Damas, L. (2013). Predicting taxi-passenger demand using streaming data, *IEEE Transactions on Intelligent Transportation Systems* **14**(3): 1393–1402.
- Morikawa, T. and Miwa, T. (2006). Preliminary analysis on dynamic route choice behavior: Using probe-vehicle data, *Journal of advanced transportation* **40**(2): 140–163.
- Nagurney, A. and Dong, J. (2002). A multiclass, multicriteria traffic network equilibrium model with elastic demand, *Transportation Research Part B: Methodological* **36**(5): 445–469.
- Nakayama, S., Kitamura, R. and Fujii, S. (2001). Drivers' route choice rules and network behavior: Do drivers become rational and homogeneous through learning?, *Transportation Research Record: Journal of the Transportation Research Board* (1752): 62–68.
- Nassir, N., Ziebarth, J., Sall, E. and Zorn, L. (2014). Choice set generation algorithm suitable for measuring route choice accessibility, *Transportation Research Record: Journal of the Transportation Research Board* (2430): 170–181.
- Newell, A. and Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice, *Cognitive skills and their acquisition* **1**: 1–55.
- Newson, P. and Krumm, J. (2009). Hidden markov map matching through noise and sparseness, *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, ACM, pp. 336–343.

- Ng, M., Szeto, W. and Waller, S. T. (2011). Distribution-free travel time reliability assessment with probability inequalities, *Transportation Research Part B: Methodological* **45**(6): 852–866.
- Ng, M. and Waller, S. T. (2010). A computationally efficient methodology to characterize travel time reliability using the fast fourier transform, *Transportation Research Part B: Methodological* **44**(10): 1202–1219.
- Nie, Y. M. (2017). How can the taxi industry survive the tide of ridesourcing? evidence from shenzhen, china, *Transportation Research Part C: Emerging Technologies* **79**: 242–256.
- Nie, Y. M. and Wu, X. (2009). Shortest path problem considering on-time arrival probability, *Transportation Research Part B: Methodological* **43**(6): 597–613.
- Nie, Y. M., Wu, X. and Homem-de Mello, T. (2012). Optimal path problems with second-order stochastic dominance constraints, *Networks and spatial Economics* **12**(4): 561–587.
- Nikolova, E. and Stier-Moses, N. E. (2014). A mean-risk model for the traffic assignment problem with stochastic travel times, *Operations Research* **62**(2): 366–382.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*, 2nd edn, Springer, New York, NY, USA.
- Noland, R. B. and Polak, J. W. (2002). Travel time variability: a review of theoretical and empirical issues, *Transport reviews* **22**(1): 39–54.

- Ordóñez, F. and Stier-Moses, N. E. (2010). Wardrop equilibria with risk-averse users, *Transportation Science* **44**(1): 63–86.
- Pakes, A. (1986). Patents as options: Some estimates of the value of holding european patent stocks, *Technical report*, National Bureau of Economic Research.
- Papinski, D. and Scott, D. (2011). A gis-based toolkit for route choice analysis, *Journal of Transport Geography* **19**(3): 434–442.
- Paquette, J., Cordeau, J.-F., Laporte, G. and Pascoal, M. M. (2013). Combining multicriteria analysis and tabu search for dial-a-ride problems, *Transportation Research Part B: Methodological* **52**: 1–16.
- Parragh, S. N., Cordeau, J.-F., Doerner, K. F. and Hartl, R. F. (2012). Models and algorithms for the heterogeneous dial-a-ride problem with driver-related constraints, *OR spectrum* **34**(3): 593–633.
- Parragh, S. N., Doerner, K. F. and Hartl, R. F. (2008). A survey on pickup and delivery problems, *Journal für Betriebswirtschaft* **58**(1): 21–51.
- Peeta, S. and Yu, J. W. (2005). A hybrid model for driver route choice incorporating en-route attributes and real-time information effects, *Networks and Spatial Economics* **5**: 21–40.
- Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A. and Ratti, C. (2010). Taxi-aware map: Identifying and predicting vacant taxis in the city, *International Joint Conference on Ambient Intelligence*, Springer, pp. 86–95.

- Pillat, J. Mandir, E. and Friedrich, M. (2011). Dynamic choice set generation based on global positioning system trajectories and stated preference data, *Transportation Research Record: Journal of the Transportation Research Board* **2231**: 18–26.
- Powell, J. W., Huang, Y., Bastani, F. and Ji, M. (2011). Towards reducing taxicab cruising time using spatio-temporal profitability maps., *SSTD*, Springer, pp. 242–260.
- Powell, W. B. (2011). *Approximate dynamic programming: Solving the curses of dimensionality*, 2nd edn, Wiley.
- Prato, C. G. (2009). Route choice modeling: past, present and future research directions, *Journal of choice modelling* **2**(1): 65–100.
- Prato, C. G. (2012). Estimating random regret minimization models in the route choice context, *Proceedings of the 13th International Conference on Travel Behaviour Research*, Toronto, Canada.
- Prato, C. G. (2014). Expanding the applicability of random regret minimization for route choice analysis, *Transportation* **41**(2): 351–375.
- Prato, C., Rasmussen, T. and Nielsen, O. (2014). Estimating value of congestion and of reliability from observation of route choice behavior of car drivers, *Transportation Research Record: Journal of the Transportation Research Board* (2412): 20–27.
- Psaraftis, H. N. (1980). A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem, *Transportation Science* **14**(2): 130–154.

- Qi, H., Ma, S., Jia, N. and Wang, G. (2018). Individual response modes to pre-trip information in congestible networks: laboratory experiment, *Transportmetrica A: Transport Science* pp. 1–20.
- Qi, J., Sim, M., Sun, D. and Yuan, X. (2016). Preferences for travel time under risk and ambiguity: Implications in path selection and network equilibrium, *Transportation Research Part B: Methodological* **94**: 264–284.
- Qian, X. and Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using gps data, *Applied geography* **59**: 31–42.
- Qian, X. and Ukkusuri, S. V. (2017a). Taxi market equilibrium with third-party hailing service, *Transportation Research Part B: Methodological* **100**: 43–63.
- Qian, X. and Ukkusuri, S. V. (2017b). Time-of-day pricing in taxi markets, *IEEE Transactions on Intelligent Transportation Systems* **18**(6): 1610–1622.
- Qian, X., Zhang, W., Ukkusuri, S. V. and Yang, C. (2017). Optimal assignment and incentive design in the taxi group ride problem, *Transportation Research Part B: Methodological* **103**: 208–226.
- Qian, Z. and Zhan, H. (2013). A hybrid route choice model for dynamic traffic assignment, *Networks and Spatial Economics* **13**(2): 183–203.
- Qin, G., Huang, Z., Xiang, Y. and Sun, J. (2019). Probdetect: A choice probability-based taxi trip anomaly detection model considering traffic variability, *Transportation Research Part C: Emerging Technologies* **98**: 221–238.

- Qin, G., Li, T., Yu, B., Wang, Y., Huang, Z. and Sun, J. (2017). Mining factors affecting taxi drivers's incomes using gps trajectories, *Transportation Research Part C: Emerging Technologies* **79**: 103–118.
- Qu, M., Zhu, H., Liu, J., Liu, G. and Xiong, H. (2014). A cost-effective recommender system for taxi drivers, *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 45–54.
- Quddus, M. and Washington, S. (2015). Shortest path and vehicle trajectory aided map-matching for low frequency gps data, *Transportation Research Part C: Emerging Technologies* **55**: 328–339.
- Rahmani, M., Jenelius, E. and Koutsopoulos, H. N. (2015). Non-parametric estimation of route travel time distributions from low-frequency floating car data, *Transportation Research Part C: Emerging Technologies* **58**: 343–362.
- Rahmani, M. and Koutsopoulos, H. N. (2013). Path inference from sparse floating car data for urban networks, *Transportation Research Part C: Emerging Technologies* **30**: 41–54.
- Ramming, M. S. (2002). *Network knowledge and route choice*, PhD thesis, Massachusetts Institute of Technology.
- Ramos, G. M., Daamen, W. and Hoogendoorn, S. (2014). A state-of-the-art review: developments in utility theory, prospect theory and regret theory to investigate travellers' behaviour in situations involving travel time uncertainty, *Transport Reviews* **34**(1): 46–67.

- Rapoport, A., Gisches, E. J., Daniel, T. and Lindsey, R. (2014). Pre-trip information and route-choice decisions with stochastic travel conditions: Experiment, *Transportation Research Part B: Methodological* **68**: 154–172.
- Rasouli, S. and Timmermans, H. (2017). Specification of regret-based models of choice behaviour: formal analyses and experimental design based evidence, *Transportation* **44**(6): 1555–1576.
- Ratchford, B. T. (1980). The value of information for selected appliances, *Journal of Marketing Research* pp. 14–25.
- Ratliff, R. M., Rao, B. V., Narayan, C. P. and Yellepeddi, K. (2008). A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings, *Journal of Revenue and Pricing Management* **7**(2): 153–171.
- Rayle, L., Dai, D., Chan, N., Cervero, R. and Shaheen, S. (2016). Just a better taxi? a survey-based comparison of taxis, transit, and ridesourcing services in san francisco, *Transport Policy* **45**: 168–178.
- Razo, M. and Gao, S. (2010). Strategic thinking and risk attitudes in route choice: A stated preference approach, *Transportation Research Record* **2085**: 136–143.
- Razo, M. and Gao, S. (2013a). A rank-dependent expected utility model for strategic route choice with stated preference data, *Transportation Research Part C* **27**: 117–130.
- Razo, M. and Gao, S. (2013b). A rank-dependent expected utility model for strategic route choice with stated preference data, *Transportation Research Part C: Emerging Technologies* **27**: 117–130.

- Rilett, L. and Park, D. (2001). Incorporating uncertainty and multiple objectives in real-time route selection, *Journal of transportation engineering* **127**(6): 531–539.
- Rubin, D. C. and Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention., *Psychological review* **103**(4): 734.
- Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher, *Econometrica* **55**(5): 999–1033.
- Rust, J. (1988). Maximum likelihood estimation of discrete control processes, *SIAM Journal on Control and Optimization* **26**(5): 1006–1024.
- Saad, Y. (2003). *Iterative methods for sparse linear systems*, Siam.
- Saadi, I., Wong, M., Farooq, B., Teller, J. and Cools, M. (2017). An investigation into machine learning approaches for forecasting spatio-temporal demand in ride-hailing service, *arXiv preprint arXiv:1703.02433* .
- Saeedmanesh, M. and Geroliminis, N. (2017). Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks, *Transportation Research Part B: Methodological* **105**: 193–211.
- Sáez, D., Cortés, C. E. and Núñez, A. (2008). Hybrid adaptive predictive control for the multi-vehicle dynamic pick-up and delivery problem based on genetic algorithms and fuzzy clustering, *Computers & Operations Research* **35**(11): 3412–3438.
- Salanova, J. M., Estrada, M., Aifadopoulou, G. and Mitsakis, E. (2011). A review of the modeling of taxi services, *Procedia-Social and Behavioral Sciences* **20**: 150–161.

- Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making, *Journal of risk and uncertainty* **1**(1): 7–59.
- Sayarshad, H. R. and Chow, J. Y. (2015). A scalable non-myopic dynamic dial-a-ride and pricing problem, *Transportation Research Part B: Methodological* **81**: 539–554.
- Sayarshad, H. R. and Chow, J. Y. (2017). Non-myopic relocation of idle mobility-on-demand vehicles as a dynamic location-allocation-queueing problem, *Transportation Research Part E: Logistics and Transportation Review* **106**: 60–77.
- Schilde, M., Doerner, K. F. and Hartl, R. F. (2011). Metaheuristics for the dynamic stochastic dial-a-ride problem with expected return transports, *Computers & operations research* **38**(12): 1719–1730.
- Schrank, D., Eisele, B., Lomax, T. and Bak, J. (2015). 2015 urban mobility scorecard, *Technical report*, Texas A&M Transportation Institute.
- Secomandi, N. (2000). Comparing neuro-dynamic programming algorithms for the vehicle routing problem with stochastic demands, *Computers & Operations Research* **27**(11): 1201–1225.
- Secomandi, N. (2001). A rollout policy for the vehicle routing problem with stochastic demands, *Operations Research* **49**(5): 796–802.
- Selten, R., Chmura, T., Pitz, T., Kube, S. and Schreckenberg, M. (2007a). Commuters route choice behaviour, *Games and Economic Behavior* **58**(2): 394–406.
- Selten, R., Chmura, T., Pitz, T., Kube, S. and Schreckenberg, M. (2007b). Commuters route choice behaviour, *Games and Economic Behavior* **58**(2): 394–406.

- Seo, T., Kusakabe, T., Gotoh, H. and Asakura, Y. (2017). Interactive online machine learning approach for activity-travel survey, *Transportation Research Part B: Methodological* .
- Seow, K. T., Dang, N. H. and Lee, D.-H. (2010). A collaborative multiagent taxi-dispatch system, *IEEE Transactions on Automation Science and Engineering* **7**(3): 607–616.
- Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization, *Mathematics of Computation* **24**(111): 647–656.
- Shiftan, Y., Bekhor, S. and Albert, G. (2011). Route choice behaviour with pre-trip travel time information, *IET intelligent transport systems* **5**(3): 183–189.
- Shou, Z. and Di, X. (2018). Similarity analysis of frequent sequential activity pattern mining, *Transportation Research Part C: Emerging Technologies* **96**: 122–143.
- Simon, H. A. (1955). A behavioral model of rational choice, *The quarterly journal of economics* pp. 99–118.
- Sirisoma, R., Wong, S., Lam, W. H., Wang, D., Yang, H. and Zhang, P. (2010). Empirical evidence for taxi customer-search model, *Proceedings of the Institution of Civil Engineers: Transport*, Thomas Telford (ICE Publishing). The Journal's web site is located at <http://www.transport-ice.com>.
- Small, E. E., Giorgi, F. and Sloan, L. C. (1999). Regional climate model simulation of precipitation in central asia: Mean and interannual variability, *Journal of Geophysical Research: Atmospheres* **104**(D6): 6563–6582.

- Sohn, J. (2006). Evaluating the significance of highway network links under the flood damage: An accessibility approach, *Transportation research part A: policy and practice* **40**(6): 491–506.
- Spivey, M. Z. and Powell, W. B. (2004). The dynamic assignment problem, *Transportation Science* **38**(4): 399–419.
- Srinivasan, K. and Mahmassani, H. (2000). Modeling inertia and compliance mechanisms in route choice behavior under real-time information, *Transportation Research Record: Journal of the Transportation Research Board* (1725): 45–53.
- Su, C.-L. and Judd, K. L. (2012). Constrained optimization approaches to estimation of structural models, *Econometrica* **80**(5): 2213–2230.
- Susilawati, S., Taylor, M. A. and Somenahalli, S. V. (2013). Distributions of travel time variability on urban roads, *Journal of Advanced Transportation* **47**(8): 720–736.
- Sutton, R. S. and Barto, A. G. (1998a). *Introduction to reinforcement learning*, Vol. 135, MIT press Cambridge.
- Sutton, R. S. and Barto, A. G. (1998b). *Reinforcement learning: An introduction*, Vol. 1, MIT press Cambridge.
- Sutton, R. S. and Barto, A. G. (1998c). *Reinforcement Learning: An Introduction*, The MIT Press, Cambridge, MA.
- Swihart, M. R. and Papastavrou, J. D. (1999). A stochastic and dynamic model for the single-vehicle pick-up and delivery problem, *European Journal of Operational Research* **114**(3): 447–464.

- Szeto, W., O'Brien, L. and O'Mahony, M. (2006). Risk-averse traffic assignment with elastic demands: Ncp formulation and solution method for assessing performance reliability, *Networks and Spatial Economics* **6**(3-4): 313–332.
- Szeto, W. Y., Wong, R. C. P., Wong, S. C. and Yang, H. (2013). A time-dependent logit-based taxi customer-search model, *International Journal of Urban Sciences* **17**(2): 184–198.
- Tan, Z. and Yang, H. (2012). Flexible build-operate-transfer contracts for road franchising under demand uncertainty, *Transportation Research Part B: Methodological* **46**(10): 1419–1439.
- Tan, Z., Yang, H. and Guo, R. (2014). Pareto efficiency of reliability-based traffic equilibria and risk-taking behavior of travelers, *Transportation research part B: methodological* **66**: 16–31.
- Tang, Y. and Gao, S. (2015). An exploratory study of instance-based learning for network route choice, *Transportation Research Board 94th Annual Meeting*, number 15-5503.
- Tang, Y., Gao, S. and Ben-Elia, E. (2017). An exploratory study of instance-based learning for route choice with random travel times, *Journal of choice modelling* **24**: 22–35.
- Taylor, M. A. (2013). Travel through time: the story of research on travel time reliability, *Transportmetrica B: transport dynamics* **1**(3): 174–194.
- Taylor, M. A. (2017). Fosgerau's travel time reliability ratio and the burr distribution, *Transportation Research Part B: Methodological* **97**: 50–63.

- Thomas, B. W. (2007). Waiting strategies for anticipating service requests from known customer locations, *Transportation Science* **41**(3): 319–331.
- Thomas, B. W. and White, C. C. (2004). Anticipatory route selection, *Transportation Science* **38**(4): 473–487.
- Tong, Y., Chen, Y., Zhou, Z., Chen, L., Wang, J., Yang, Q., Ye, J. and Lv, W. (2017). The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 1653–1662.
- Toth, P. and Vigo, D. (2001). Branch-and-bound algorithms for the capacitated vrp, *The vehicle routing problem*, Society for Industrial and Applied Mathematics, pp. 29–51.
- Train, K. (2000). Halton sequences for mixed logit, *Department of Economics, UCB* .
- Train, K. (2001). A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit, *University of California, Berkeley* pp. 1–13.
- Train, K. E. (2002). *Discrete choice methods with simulation*, Vol. 8, Cambridge University Press Cambridge.
- Train, K. E. (2009). *Discrete choice methods with simulation*, Cambridge university press.
- Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and uncertainty* **5**(4): 297–323.

- van der Walt, S. v. d., Colbert, S. C. and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation, *Computing in Science & Engineering* **13**(2): 22–30.
- Van Lint, J. and Van Zuylen, H. (2005). Monitoring and predicting freeway travel time reliability: Using width and skew of day-to-day travel time distribution, *Transportation Research Record: Journal of the Transportation Research Board* (1917): 54–62.
- Verma, T., Varakantham, P., Kraus, S. and Lau, H. C. (2017). Augmenting decisions of taxi drivers through reinforcement learning for improving revenues, *Proceedings of the International Conference on Automated Planning and Scheduling, Pittsburgh, PA, USA*, pp. 18–23.
- Wang, J. and Lampert, B. (2014). Improving taxi revenue with reinforcement learning.
- Wang, J. Y., Ehrgott, M. and Chen, A. (2014). A bi-objective user equilibrium model of travel time reliability in a road network, *Transportation Research Part B: Methodological* **66**: 4–15.
- Wang, X., He, F., Yang, H. and Gao, H. O. (2016). Pricing strategies for a taxi-hailing platform, *Transportation Research Part E: Logistics and Transportation Review* **93**: 212–231.
- Watling, D. (2006). User equilibrium traffic network assignment with stochastic travel times and late arrival penalty, *European journal of operational research* **175**(3): 1539–1556.

- Wei, D., , C., Liu, H., Wu, D. and Kumfer, W. (2016). The impact of service refusal to the supply–demand equilibrium in the taxicab market, *Networks and Spatial Economics* pp. 1–29.
- Whelan, G., Batley, R., Fowkes, T. and Daly, A. (2002). Flexible models for analyzing route and departure time choice, *Publication of: Association for European Transport* .
- Wolpin, K. I. (1984). An estimable dynamic stochastic model of fertility and child mortality, *The Journal of Political Economy* pp. 852–874.
- Wong, K., Wong, S. C., Bell, M. G. and Yang, H. (2005). Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing markov chain approach, *Journal of advanced transportation* **39**(1): 81–104.
- Wong, K., Wong, S. C. and Yang, H. (2001). Modeling urban taxi services in congested road networks with elastic demand, *Transportation Research Part B: Methodological* **35**(9): 819–842.
- Wong, K., Wong, S. C., Yang, H. and Wu, J. (2008). Modeling urban taxi services with multiple user classes and vehicle modes, *Transportation Research Part B: Methodological* **42**(10): 985–1007.
- Wong, R. C. P., Szeto, W. Y., Wong, S. and Yang, H. (2014). Modelling multi-period customer-searching behaviour of taxi drivers, *Transportmetrica B: Transport Dynamics* **2**(1): 40–59.

- Wong, R. C., Szeto, W. and Wong, S. (2015a). Sequential logit approach to modeling the customer-search decisions of taxi drivers, *Asian Transport Studies* **3**(4).
- Wong, R., Szeto, W. and Wong, S. (2014). A cell-based logit-opportunity taxi customer-search model, *Transportation Research Part C: Emerging Technologies* **48**: 84–96.
- Wong, R., Szeto, W. and Wong, S. (2015b). A two-stage approach to modeling vacant taxi movements, *Transportation Research Part C: Emerging Technologies* **59**: 147–163.
- Wong, S. and Yang, H. (1998). Network model of urban taxi services: improved algorithm, *Transportation Research Record: Journal of the Transportation Research Board* (1623): 27–30.
- Wong, T. J., Cokely, E. T. and Schooler, L. J. (2010). An online database of act-r parameters: Towards a transparent community-based approach to model development, *Proceedings of the Tenth International Conference on Cognitive Modeling, Philadelphia, PA, USA*, Citeseer, pp. 282–286.
- Woodard, D., Nogin, G., Koch, P., Racz, D., Goldszmidt, M. and Horvitz, E. (2017). Predicting travel time reliability using mobile phone gps data, *Transportation Research Part C: Emerging Technologies* **75**: 30–44.
- Wright, P. and Barbour, F. (1977). *Phased decision strategies: Sequels to an initial screening*, Graduate School of Business, Stanford University.
- Wu, X. (2015). Study on mean-standard deviation shortest path problem in stochastic and time-dependent networks: A stochastic dominance based approach, *Transportation Research Part B: Methodological* **80**: 275–290.

- Wu, X. and Nie, Y. M. (2011). Modeling heterogeneous risk-taking behavior in route choice: A stochastic dominance approach, *Transportation research part A: policy and practice* **45**(9): 896–915.
- Xiao, G., Juan, Z. and Zhang, C. (2015). Travel mode detection based on gps track data and bayesian networks, *Computers, Environment and Urban Systems* **54**: 14–22.
- Xiao, Y., Coulombel, N. and De Palma, A. (2017). The valuation of travel time reliability: does congestion matter?, *Transportation Research Part B: Methodological* **97**: 113–141.
- Xie, C. and Liu, Z. (2014). On the stochastic network equilibrium with heterogeneous choice inertia, *Transportation Research Part B: Methodological* **66**: 90–109.
- Xie, J., Nie, Y. M. and Liu, X. (2017). Testing the proportionality condition with taxi trajectory data, *Transportation Research Part B: Methodological* **104**: 583–601.
- Xu, H., Zhou, J. and Xu, W. (2011). A decision-making rule for modeling travelers's route choice behavior based on cumulative prospect theory, *Transportation Research Part C: Emerging Technologies* **19**(2): 218–228.
- Xu, J., Rahmatizadeh, R., Bölöni, L. and Turgut, D. (2017). Real-time prediction of taxi demand using recurrent neural networks, *IEEE Transactions on Intelligent Transportation Systems* .
- Xu, X., Chen, A., Cheng, L. and Lo, H. K. (2014). Modeling distribution tail in network performance assessment: A mean-excess total travel time risk measure and analytical estimation method, *Transportation Research Part B: Methodological* **66**: 32–49.

- Xu, X., Chen, A., Cheng, L. and Yang, C. (2017). A link-based mean-excess traffic equilibrium model under uncertainty, *Transportation Research Part B: Methodological* **95**: 53–75.
- Xu, Z., Yin, Y. and Zha, L. (2017). Optimal parking provision for ride-sourcing services, *Transportation Research Part B: Methodological* **105**: 559–578.
- Yai, T., Iwakura, S. and Morichi, S. (2002). Real-time simulation of traffic demand-supply interactions within dynamit, *Transportation Wid Network Analysis: Current Trends. Miscellenea in Honor ofMichael Florian (M. Gendreau and P. Marcotte, eds.), Kluwer, Netherlands* pp. 19–36.
- Yamamoto, K., Uesugi, K. and Watanabe, T. (2008). Adaptive routing of cruising taxis by mutual exchange of pathways, *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, Springer, pp. 559–566.
- Yang, F. and Zhang, D. (2009). Day-to-day stationary link flow pattern, *Transportation Research Part B* **43**: 119–126.
- Yang, H., Leung, C. W., Wong, S. C. and Bell, M. G. (2010). Equilibria of bilateral taxi–customer searching and meeting on networks, *Transportation Research Part B: Methodological* **44**(8-9): 1067–1083.
- Yang, H., Wang, Z., Xie, K. and Dai, D. (2017). Use of ubiquitous probe vehicle data for identifying secondary crashes, *Transportation research part C: emerging technologies* **82**: 138–160.

- Yang, H. and Wong, S. C. (1998). A network model of urban taxi services, *Transportation Research Part B: Methodological* **32**(4): 235–246.
- Yang, H., Wong, S. C. and Wong, K. (2002). Demand–supply equilibrium of taxi services in a network under competition and regulation, *Transportation Research Part B: Methodological* **36**(9): 799–819.
- Yang, H. and Yang, T. (2011). Equilibrium properties of taxi markets with search frictions, *Transportation Research Part B: Methodological* **45**(4): 696–713.
- Yang, H., Ye, M., Tang, W. H.-C. and Wong, S. C. (2005a). A multiperiod dynamic model of taxi services with endogenous service intensity, *Operations research* **53**(3): 501–515.
- Yang, H., Ye, M., Tang, W. H. and Wong, S. C. (2005b). Regulating taxi services in the presence of congestion externality, *Transportation Research Part A: Policy and Practice* **39**(1): 17–40.
- Yang, J., Dong, J. and Hu, L. (2017). A data-driven optimization-based approach for siting and sizing of electric taxi charging stations, *Transportation Research Part C: Emerging Technologies* **77**: 462–477.
- Yang, J., Jaillet, P. and Mahmassani, H. (2004). Real-time multivehicle truckload pickup and delivery problems, *Transportation Science* **38**(2): 135–148.
- Yang, J. and Jiang, G. (2014). Development of an enhanced route choice model based on cumulative prospect theory, *Transportation Research Part C: Emerging Technologies* **47**: 168–178.

- Yang, L., Kwan, M.-P., Pan, X., Wan, B. and Zhou, S. (2017). Scalable space-time trajectory cube for path-finding: A study using big taxi trajectory data, *Transportation Research Part B: Methodological* **101**: 1–27.
- Yang, L., Shi, M., Zheng, Q., Meng, W. and Pan, G. (2018). A unified approach for multi-step temporal-difference learning with eligibility traces in reinforcement learning, *arXiv preprint arXiv:1802.03171* .
- Yang, T., Yang, H. and Wong, S. C. (2014). Taxi services with search frictions and congestion externalities, *Journal of Advanced Transportation* **48**(6): 575–587.
- Yang, W., Wang, X., Rahimi, S. M. and Luo, J. (2015). Recommending profitable taxi travel routes based on big taxi trajectories data, *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, pp. 370–382.
- Yin, Y. and Ieda, H. (2001). Assessing performance reliability of road networks under nonrecurrent congestion, *Transportation Research Record: Journal of the Transportation Research Board* (1771): 148–155.
- Yin, Y., Lam, W. H. and Ieda, H. (2004). New technology and the modeling of risk-taking behavior in congested road networks, *Transportation Research Part C: Emerging Technologies* **12**(3-4): 171–192.
- Ying, J. J.-C., Lu, E. H.-C., Lee, W.-C., Weng, T.-C. and Tseng, V. S. (2010). Mining user similarity from semantic trajectories, *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, ACM, pp. 19–26.

- Yu, X., Gao, S., Hu, X. and Park, H. (2018). Multi-cycle optimal taxi routing with e-hailing, *The 97th Annual Meeting of Transportation Research Board Compendium of Papers, Washington, DC, USA*.
- Yu, X., Mai, T., Gao, S., Ding, J. and Frejinger, E. (2015). A recursive logit model for routing policy choice, *Proceedings of the 4th International Choice Modeling Conference (ICMC)*, Austin, Texas, U.S.A.
- Yuan, J., Sun, G., Tian, Y., Chen, G. and Liu, Z. (2009). Selective-nra algorithms for top-k queries., *APWeb/WAIM*, Springer, pp. 15–26.
- Yuan, J., Zheng, Y., Xie, X. and Sun, G. (2013). T-drive: Enhancing driving directions with taxi drivers' intelligence, *IEEE Transactions on Knowledge and Data Engineering* **25**(1): 220–232.
- Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X., Sun, G. and Huang, Y. (2010a). T-drive: driving directions based on taxi trajectories, *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, ACM, pp. 99–108.
- Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X., Sun, G. and Huang, Y. (2010b). T-drive: driving directions based on taxi trajectories, *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, ACM, pp. 99–108.

- Yuan, J., Zheng, Y., Zhang, L., Xie, X. and Sun, G. (2011). Where to find my next passenger, *Proceedings of the 13th international conference on Ubiquitous computing*, ACM, pp. 109–118.
- Yuan, N. J., Zheng, Y., Zhang, L. and Xie, X. (2013). T-finder: A recommender system for finding passengers and vacant taxis, *IEEE Transactions on Knowledge and Data Engineering* **25**(10): 2390–2403.
- Zang, Z., Xu, X., Yang, C. and Chen, A. (2018). A distribution-fitting-free approach to calculating travel time reliability ratio, *Transportation Research Part C: Emerging Technologies* **89**: 83–95.
- Zha, L., Yin, Y. and Du, Y. (2017). Surge pricing and labor supply in the ride-sourcing market, *Transportation Research Procedia* **23**: 2–21.
- Zha, L., Yin, Y. and Yang, H. (2016). Economic analysis of ride-sourcing markets, *Transportation Research Part C: Emerging Technologies* **71**: 249–266.
- Zhang, C., Chen, X. and Sumalee, A. (2011). Robust wardrop’s user equilibrium assignment under stochastic demand and supply: expected residual minimization approach, *Transportation Research Part B: Methodological* **45**(3): 534–552.
- Zhang, C., Liu, T.-L., Huang, H.-J. and Chen, J. (2018). A cumulative prospect theory approach to commuters’ day-to-day route-choice modeling with friends’ travel information, *Transportation Research Part C: Emerging Technologies* **86**: 527–548.
- Zhang, D. and He, T. (2012). pCruise: Reducing cruising miles for taxicab networks, *IEEE 33rd Real-Time Systems Symposium (RTSS)*, IEEE, pp. 85–94.

- Zhang, D., He, T., Lin, S., Munir, S. and Stankovic, J. A. (2015). Online cruising mile reduction in large-scale taxicab networks, *IEEE Transactions on parallel and distributed systems* **26**(11): 3122–3135.
- Zhang, D. and Wang, X. C. (2018). Understanding many-to-many matching relationship and its correlation with joint response, *Transportation Research Part B: Methodological* **108**: 249–260.
- Zhang, J. and Yang, H. (2015). Modeling route choice inertia in network equilibrium with heterogeneous prevailing choice sets, *Transportation Research Part C: Emerging Technologies* **57**: 42–54.
- Zhang, K., Feng, Z., Chen, S., Huang, K. and Wang, G. (2016). A framework for passengers demand prediction and recommendation, *IEEE International Conference on Services Computing (SCC)*, pp. 340–347.
- Zhang, W., Ukkusuri, S. and Yang, C. (2018). Modeling the taxi drivers’ customer-searching behaviors outside downtown areas, *Sustainability* **10**(9): 3003.
- Zhao, K., Khryashchev, D., Freire, J., Silva, C. and Vo, H. (2016). Predicting taxi demand at high spatial resolution: Approaching the limit of predictability, *Big Data (Big Data), 2016 IEEE International Conference on*, IEEE, pp. 833–842.
- Zhao, Z., Koutsopoulos, H. N. and Zhao, J. (2018a). Detecting pattern changes in individual travel behavior: A bayesian approach, *Transportation Research Part B: Methodological* **112**: 73–88.

- Zhao, Z., Koutsopoulos, H. N. and Zhao, J. (2018b). Individual mobility prediction using transit smart card data, *Transportation research part C: emerging technologies* **89**: 19–34.
- Zheng, Y. (2015). Trajectory data mining: an overview, *ACM Transactions on Intelligent Systems and Technology (TIST)* **6**(3): 29.
- Zhu, J. and Ye, X. (2018). Development of destination choice model with pairwise district-level constants using taxi gps data, *Transportation Research Part C: Emerging Technologies* **93**: 410–424.
- Ziebart, B. D., Maas, A. L., Dey, A. K. and Bagnell, J. A. (2008). Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior, *Proceedings of the 10th international conference on Ubiquitous computing*, ACM, pp. 322–331.
- Ziegelmeyer, A., Koessler, F. and Denant-Boemont, L. (2008). Road traffic congestion and public information: an experimental investigation, *Journal of Transport Economics and Policy (JTEP)* **42**(1): 43–82.
- Zong, F., Tian, Y., He, Y., Tang, J. and Lv, J. (2019). Trip destination prediction based on multi-day gps data, *Physica A: Statistical Mechanics and its Applications* **515**: 258–269.
- Zong, F., Wu, T. and Jia, H. (2018). Taxi drivers’ cruising patterns-insights from taxi gps traces, *IEEE Transactions on Intelligent Transportation Systems* (99): 1–12.