Route search problem considering travel time reliability and CO_2 emission in road network

ZENG, Weiliang

Route search problem considering travel time reliability and CO_2 emission in road network

Doctoral Dissertation

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Engineering

by:
ZENG, Weiliang

Academic Adviser:
Associate Professor MIWA, Tomio

Department of Civil Engineering
Nagoya University
JAPAN
February, 2016

Acknowledgments

I would like to say thank you for the kind people supporting my doctor course study and the dissertation. Without their encouragement and help, it is impossible to finish my doctor defense.

First of all, I appreciate my supervisor MIWA, Tomio sensei so much. He is a very nice professor and always encourages his students. He takes care of me so much. Though I meet some difficulties at the beginning, he provide good research topic and help me develop my doctoral research as soon as possible. He also provides many opportunities to make presentation in the international conference and aid to write good papers. I learned how to conduct research from him.

I am grateful to Professor MORIKAWA, Takayuki. He is so nice and every student likes him so much. His wide range of knowledge and great foresight inspires me so much. I always get much valuable information from his comments. Specially, his support for my new start gives me great encouragement

I am appreciated for the kind suggestions from Professor YAMAMOTO, Toshiyuki. He is very serious to research and gives me many comments to my presentation and paper writing. He is also a member of my dissertation defense committee. His valuable comments help to improve my dissertation greatly.

I would like to acknowledge associate professor KATO, Hirokazu, Professor MATSUMOTO, Yukimasa and Professor NAKANO, Masaki for serving as my dissertation defense committee members. I am grateful for the valuable comments of the committee members.

Thank you for the help of assistant professor, SATO, Hitomi. She is so nice and helps me to prepare the data and software. Without her timely help, I cannot restart my research so fast. Moreover, I would like to say thank you to all members in MORIKWA & YAMAMOTO & MIWA Lab. Gong Lei, CHU Tien Dung, Ma Danpeng, Sun Xiaohui, MOTHAFER Ghasak, Xu Gang, Mu Rui, Li Yanyan and TOSA Cristia gave me many suggestions in my study and life. Moreover, I would like to say thank you to Mrs. KAWAHARA, Hiroko. Her support and encouragement make me go out the frustration. I also want to say thank you to Professor Mizutani, Norimi. His concern and support make me continue my research.

Finally, deep thanks go to my parents. They always give me support without any complains. Their patience and understanding make me moving forward my goal. I will make effort to return for their love.

Abstract

In an effort to provide better route guidance to travelers, this study investigates the route search problem considering travel time reliability and CO_2 emission. This study adds to the emerging route guidance technology regarding reliable path finding and eco-routing in the following aspects. First, the α-reliable path problem in a stochastic network with correlated and truncated lognormal link travel times is addressed. The Lagrangian relaxation approach is applied to solve the nonlinear and non-additive problem. The Lagrangian relaxation based framework enables to handle the mean-variance α-reliable path problem, by which an intractable problem with a nonlinear and non-additive structure can be decomposed into several easy-to-solve problems. A subgradient algorithm is used to iteratively update the Lagrangian multipliers and find the approximate optimal solution. The availability of such reliable paths in a navigation system application would help travelers plan their travel time budgets with a specified on-time arrival probability efficiently. Then, travelers' risk preferences to travel time reliability are explored. The degree of risk-averse preference is formulated by comparing the on-time arrival probabilities of the pre-defined α-reliable path and the observed path under the theory of stochastic dominance. As a parallel study, the experientially reliable routing considering travel time uncertainty and driving experience of local probe vehicle drivers is proposed. Second, considering the environmental benefit, this study investigates the eco-friendly path that results in minimum CO_2 emissions while satisfying a specified budget of travel time. This eco-routing problem with travel time constraint is transformed into a bi-objective like optimization problem. Specifically, the benefit tradeoff between CO_2 emissions reduction and the travel time buffer is discussed by carrying out sensitivity analysis on a network-wide scale.

Chapter 2 offers a comprehensive literature review related to reliable routing problem and eco-routing problem.

Chapter 3 introduces the data collection method. The travel time distribution and variability at the link and path levels are analyzed based on GPS probe vehicle data. Travel time distributions at link level and path level are characterized. Several classical distributions (normal, lognormal, truncated normal, and truncated lognormal) are subjected to the K-S test, A-D test and χ^2 test. It is found that the truncated lognormal distribution enables to represents the link travel time distribution for about 90% of links. Because there is no closed-form expression for solving the joint probability distribution (sum of link

distribution) function for each path, a normal distribution is selected as a surrogate for path distribution. This is computationally tractable and shown to be acceptable in accuracy.

Chapter 4 introduces the α-reliable path problem in a stochastic network with correlated travel time. The difficulties faced in finding a solution are twofold. First, the problem is non link separable due to the cross correlation of link pairs. The explicit unknown includes two variables, which makes the problem intractable. Second, sub-path optimality does not hold due to the nonlinear term. Such a problem cannot be computed by a standard shortest-path algorithm because additivity is violated. To overcome these problems, a Cholesky decomposition method for link separation in the variance-covariance matrix and a Lagrangian relaxation approach for problem decomposition are introduced. The Lagrangian relaxation approach is applied to approximate the α-reliable path solution by closing the duality gap. Specifically, the spatial correlation of link travel times is explicitly considered by introducing a correlation coefficient matrix. The Cholesky decomposition is proposed to separate the correlation coefficient matrix and make it tractable to the α-reliable path problem. The nonlinear and non-additive problem structure is decomposed into sub-problems that can be regarded as standard shortest-path problems and series of tractable convex or concave problems. In solving the problem, the relative gap between the upper bound and lower bound of the solution is shown to decrease at each iteration and 30 iterations of the algorithm yield a small relative gap of within 2%-7%.

Chapter 5 investigated the traveler's risk-averse preference for α-reliable path problem in a transportation network. A novel data collection method for travelers' risk-averse preferences is introduced. The observed risk-averse preference is defined by using the theory of stochastic dominance. Ordered Probit model is applied to estimate the parameters of the travelers' risk preferences by considering variously individual properties (gender, age) and pre-trip information (OD distance, departure time, day of week).

Chapter 6 investigated an experientially reliable path considering travel time uncertainty and driving experience of local probe vehicle drivers. Accordingly, a two-stage route-finding procedure is proposed. First, a candidate path set is built by using the hyperpath algorithm, where the choice probability is assigned to each link with uncertain travel time. Second, the shortest path algorithm is applied to search the experientially reliable path on the graph of hyperpath where the modified link cost is penalized based on the link choice probability derived from hyperpath algorithm and the driving experience of local drivers. Four kinds of optimal path in a real-world network are compared with the observed one. It is found that the proposed path has the most similarity with the observed path and it has a higher degree of familiarity and reasonable time and distance.

Chapter 7 introduces how to determine an eco-friendly path that results in minimum CO_2 emissions while satisfying a specified budget for travel time. This eco-routing problem with travel time constraint is formulated to a bi-objective like optimization problem. The theory of Pareto-optimal optimization is then applied to solve this NP-complete routing problem. Specifically, a heuristic approach combining the weighting method and k-shortest path algorithm is developed to search the optimal path along the Pareto frontier. The performance of the proposed eco-routing strategy is verified by comparing it against other routing strategies in a real world network using real travel time and CO_2 emissions data collected by GPS-equipped probe vehicles. Specifically, the benefit tradeoff between CO_2 emissions reduction and the travel time buffer is discussed by carrying out sensitivity analysis on a network-wide scale. We compare its computation efficiency against the classic Lagrangian relaxation approach over a set of real-world networks and demonstrate its advantage in solution quality and computation time.

Finally, Chapter 8 summarizes this study and gives direction for future research. The route search problem considering travel time reliability and CO_2 emission is promising to be applied to the current navigation system easily. This work might be a particular help in the design of a more effective navigation system for individuals with various preferences to travel time and environmental benefit. Potential directions for future study in this area include improvement of the path finding algorithm and consideration of the stochastic characteristics of travel time and CO_2 emissions: (1) an efficient path-finding algorithm suitable for a real-time eco-routing navigation system needs to be developed; (2) considering that travel time and emissions are non-deterministic, the reliability of the eco-algorithm routing should be considered further; (3) in addition to spatial link travel time correlation, temporal correlation should be considered in the finding algorithm and (4) a faster path finding algorithm suitable for real-time navigation systems needs to be developed.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

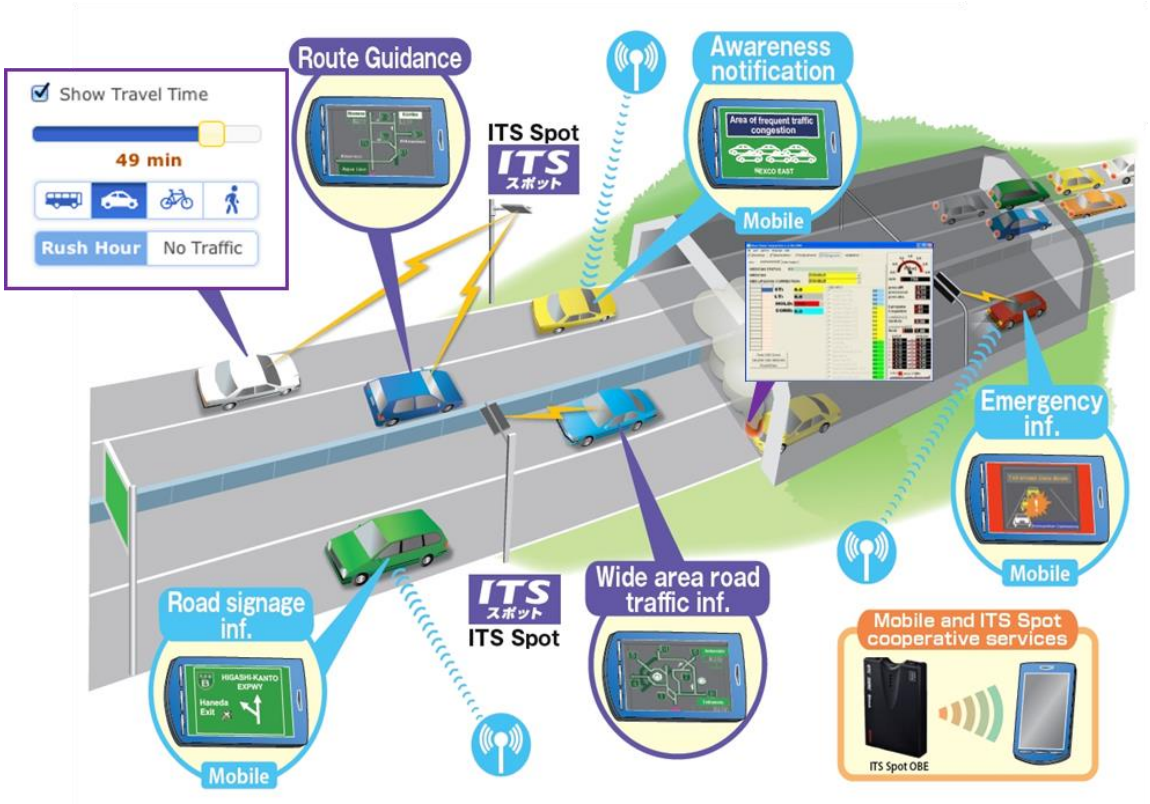
1.1 Background

With the development of the communication technology and transportation service, many vehicles have been installed the GPS-based navigation system, which enables to provide route guidance services. Travelers are becoming relying on the routing service especially for planning unfamiliar trips. Today, the routing service not only provides a shortest path, but also estimates the travel time, costs, and fuel consumption for a specified trip. It helps travelers to avoid the congestion area and select a preferable path. Due to the increasing demand of traveling, congestion is a serious problem in many countries. Traffic jam becomes worse with the increasing number of cars and uncertain travel demands. How to provide a better routing service in an uncertain transportation network is a great challenge.

1.1.1 Development of ITS

As shown in Figure 1-1, with the development of modern transportation, Intelligent Transportation Systems (ITS) aims to improve the efficiency, safety and air quality via applying a wide range of advanced communication and information technologies. Peoples are not only concerned with the real time traffic information, but also the reliability and environmental protection. For example, we are concerned about whether the shortest path provides a reliable travel time and how many fuel consumption or emission by choosing

this path. ITS covers various areas related to people's daily travel, commercial transport, and traffic management. There are some important areas that ITS focuses on: advanced navigation system, V2V and V2I communications, advanced traffic management system, application of big data, etc. In Japan, the milestone of ITS is shown in Figure 1-2 (Amano, 2015). In the first stage, it mainly focuses on the establishment of different sub-systems such as car navigation system, ETC, driving safety system, traffic management system, road management system, commercial vehicle operation system, public transportation system, pedestrian assistance system, emergency vehicle operation system. In the second stage, these sub-systems are further improved and it pays more attention to traffic safety, convenience, and standardization. Nowadays, the new challenge is how to develop a sustainable transportation system. In Japan, it is necessary to consider the aging society, sustainable fuel supply and natural disasters.



Modified from image source: www.its-jp.org

Figure 1-1 Application of Intelligent Transportation System

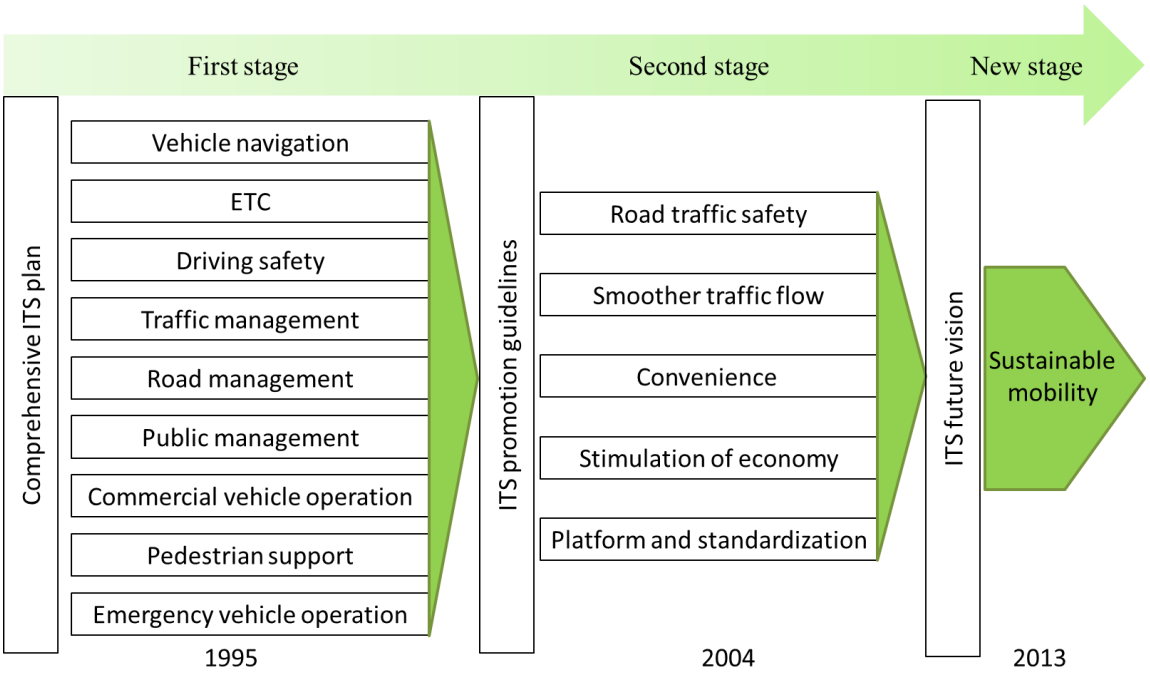


Figure 1-2 Development of ITS in Japan (Amano, 2015)

1.1.2 System-wide traffic data collection

To provide a better transportation service, for example, navigation service, an effective data collection measure is important. As shown in Figure 1-3, various sensors are applied to transportation information collection, including inductive loop detector, microwave radar, image processing vehicle detector, GPS probes, cellular phone, etc. Generally, they can be classified into three categories, i.e., point sensor, point-to-point sensor, and probe sensor. For example, the point sensor such as microwave sensor is usually used to collect the traffic flow, but it is hard for travel time collection. The point-to-point sensor such as video detector can provide the travel time estimation, but the accuracy is dependent on the identification of license plate. Recently, the probe sensor such as the probe vehicle is widely used to collect the system-wide traffic data because it enables to collect large amount of GPS data and provide reasonable travel time estimation. Probe vehicles are becoming more and more popular in transportation management and transportation research, because it has many advantages superior to other data collection methodologies. For example, the cost of installation and daily maintenance fee are relatively low. Without adding special infrastructure, any vehicles such as taxi, commercial fleet, and private cars can be used as probe vehicles. Considering the privacy, the private information such as the owner information and license plate will not be collected. The amounts of data are abundant and precise. The location, speed, time, and direction can be collected second-by-second. Then, the trajectory can be reproduced by using the GIS

(Geographic Information System) tool. Various analyses can be conducted by using the large-scale trajectory data. Some importation applications include the estimation of road speed, congestion prediction, OD estimation, and route guidance. Probe vehicle not only provides large-scale data in a wide area, but also has strong anti-interference ability. With the development of GPS, data collection by probe vehicle is robust and seldom disturbed by surrounding influences such as time of day and weather condition. Since there are many advantages by using probe vehicle, this study uses probe vehicle for data collection.

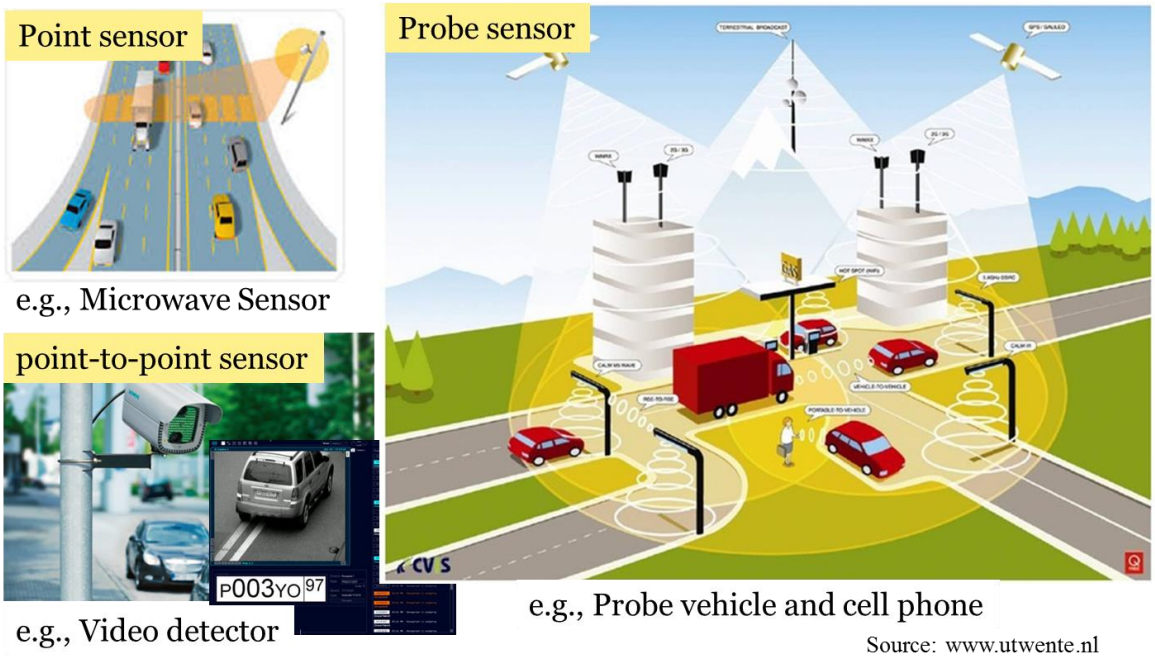


Figure 1-3 Sensors for data collection in ITS

1.1.3 Navigation system

En-route vehicle navigation system has become a standard configuration in many car models. A typical vehicle navigation system usually includes an on-board GPS and displaces the vehicle's position on a digital map (Lin and Lo, 2014). Nowadays, not only the statistical information such as lengths of links and historical travel time, but also the real-time traffic information such as traffic flow and congestion level can be obtained from various channels. Users can vary the criteria to find preferable routes. That is, users may choose least travel time, minimum cost or to avoid certain links, route sections or areas.

The navigation system can be classified into two types. One is called the centralized route guidance, and another one is called decentralized route guidance. For centralized route guidance, the large-scale data processing and computation is handled by servers in

transportation information center. Then, the optimal path is sent to the terminal device such as the cell phone and personal computer by 3G/4G wireless network. Most of the web-based map services are centralized, such as Google map and Yahoo map. Decentralized routing system usually calculates the optimal path by the build-in computation unit. The vehicle navigation system with build-in GPS device is generally the decentralized system, which enables to collect the link costs such as real-time travel time and provide the optimal path by local calculation. Many field tests demonstrate that a navigation system enables to efficiently reduce the individual travel time and avoid the congestion. In March, 1998, the center area of Tokyo was used to test the route guidance performance (JTMTA, 1999). It is found that the tested cars with dynamic route guidance system (real-time) traveled 77s (2.3%) faster than taxies, 422s (12.4%) faster than the cars with static route guidance system (based on shortest path) on average.

1.2 Problem statements

The unexpected delay is usually caused by the traffic congestion. However, the traffic congestion is difficult to mitigate due to the increasing traffic demand. The traffic congestion may cause the uncertain delay for a planned trip. On the other hand, a lot of research found that traffic congestion also increases the vehicle emission significantly. For example, the CO_2 emission for a truck will increase three times in congestion situation comparing to smooth situation.

ITS is regarded as one of the efficient techniques mitigating the traffic congestion and provide an eco-friendly society. As an important application of ITS, vehicle navigation system becomes more and more popular in our daily life. For example, many travelers usually search a shortest path by using the google navigation or yahoo navigation before departure. As shown in Figure 1-4, web-based map services such as Yahoo map and Google map may provide static information of average travel time and shortest paths. However, users do not know the travel time reliability and it is difficult to decide a better route. Reliability has become an important factor when people make decision on choosing a path especially for commercial purpose. Unexpected delay may lead to serious penalties. As we know, peoples have different risk preference on travel time reliability. A risk-averse traveler prefer to choose a path with lower travel time variance, while a risk-taking traveler prefer an unreliable path that potentially provides shortest distance or minimum travel time. For example, the shortest path may have shortest length but low on-time arrival probability. This may cause late arrivals and high penalties from travelers. On the other hand, environment problems such as global warming have become a matter of worldwide concern. It has been noted that the transportation sector accounts for approximate 23% of

total global CO_2 emissions, of which 73% are generated by road transport (JAMA, 2008; Birol, 2010). Even though alternative fuel vehicles such as all-electric vehicle will be the best solution in the future, mitigating emissions by existing gasoline vehicles is an alternative countermeasure in the near term. However, current navigation system seldom considers the travel time reliability and vehicle emission. There are questions when we use the google map or Yahoo map service. Is it really on-time arrival and is it eco-friendly?

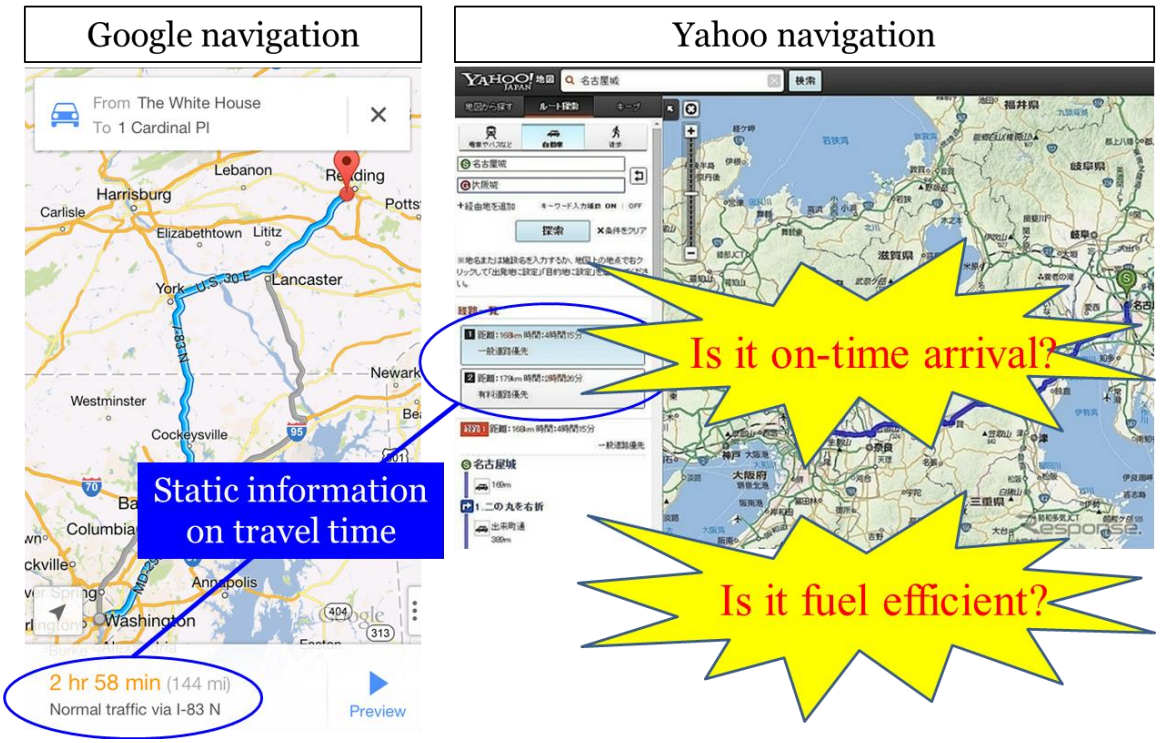


Figure 1-4 Google map and Yahoo map

1.3 Research objective

Following the problem statement, the research objective is to explore the routing methodologies considering on-time arrival probability and vehicle emission. To complete the objective, the following research is conducted.

(1) Develop a reliable routing model enabling to incorporate traveler's risk-averse preference.

(2) Explore travelers' risk-averse preference to travel time from large-scale trip records.

(3) Develop an eco-routing model considering travel time and CO_2 emission.

1.4 Research outline

The organization of this thesis is structured as follows. Chapter 1 introduces the development of ITS and navigation system. The problem statement and research objective are also given. Following this introduction, Chapter 2 offers a brief literature review related to the reliable routing problem and eco-routing problem. Chapter 3 introduces the data collection method and analyzes travel time distribution at link and path levels based on GPS probe vehicle data. Chapter 4 solves the α-reliable path problem with correlated link travel times as a nonlinear and non-additive programming problem. Chapter 5 introduces a new method for estimating the risk-averse preference. Chapter 6 investigated an experientially reliable path considering travel time uncertainty and driving experience of local probe vehicle drivers. Chapter 7 introduces how to determine an eco-friendly path that results in minimum CO_2 emissions while satisfying a specified budget for travel time. Finally, the achievements of this study and the recommendation for future research are outlined in Chapter 8.

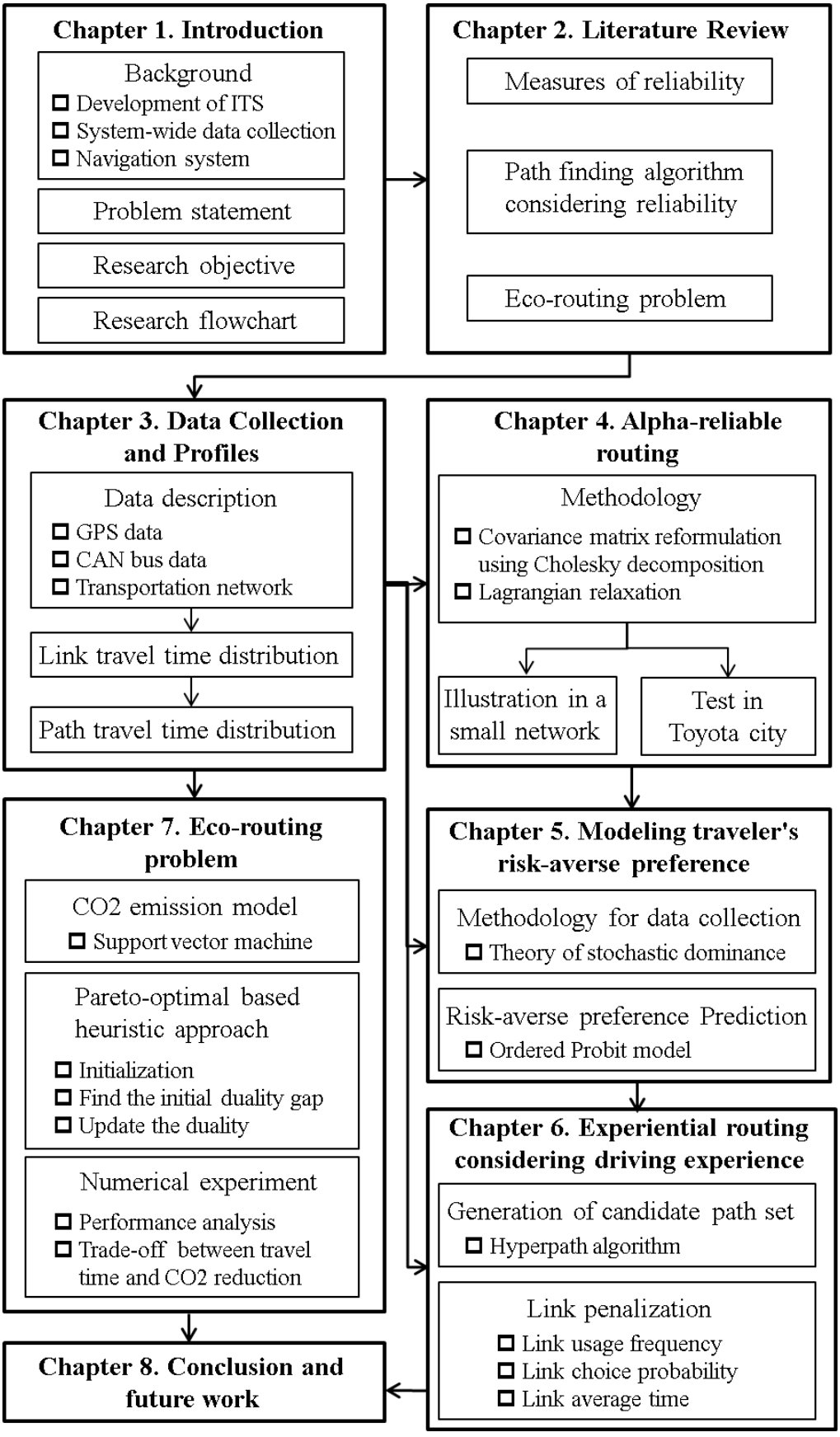


Figure 1-5 Research flowchart

1.5 References

Amano, H. (2015). Phased ITS Development in Japan, New Breeze.

Figueiredo, Lino, Isabel Jesus, JA Tenreiro Machado, J. Ferreira, and JL Martins De Carvalho. (2001). Towards the development of intelligent transportation systems. Intelligent transportation systems, vol. 88, 1206-1211.

Xiao, L., & Lo, H. K. (2014). Adaptive Vehicle Navigation With En Route Stochastic Traffic Information. IEEE Transactions on Intelligent Transportation Systems, 15(5), 1900-1912.

JTMTA. (1999). ITS developed by Japanese Police, Institute of urban traffic research.

Chapter 2

LITERATURE REVIEWS

2.1 Measure of travel time reliability

The studies for travel time reliability can be classified into two groups, i.e., network travel time reliability and path travel time reliability.

The studies for network travel time reliability mainly account for congestion effects. Chen et al. (2002) combined reliability and network equilibrium models to analyze the performance of a degradable road network. Ng and Waller (2010) characterized the travel time reliability in road networks using the fast Fourier transform when the uncertainty was given by stochastic road capacities. A bi-objective network equilibrium models for travel time reliability by combining the travel time budget (Lo et al., 2006) and the late arrival penalty (Watling, 2006) is proposed by Wang et al. (2014). Bell (2000) defined network reliability in terms of two measures, i.e., connectivity and performance reliability. A game theoretical approach was used to assess the network reliability.

The second group is the path travel time reliability. It has been recognized as a critical factor in enabling travelers to guarantee on-time journeys. There are various definitions of path travel time reliability and formulations for measuring it. It is usually represented by the probability that a trip can be made successfully within a certain travel time budget (Frank, 1969; Mirchandani, 1976). Hall (1983) measured travel reliability in terms of effective travel time, which was defined as the sum of the average travel time and a safety

margin that was the product of the standard deviation of travel time (Sivakumar and Batta, 1994; Sen et al., 2001; Shao et al., 2006). Taking account into the fact that travelers often concern about how late and how early they can arrive at their destination, Kaparias et al (2008) used the earliness and lateness probabilities as the reliability measure. Rakha et al. (2010) pointed out those key variables in measuring path travel time reliability included not only the travel time mean but also the travel time variance. And five measurements for estimating path travel time variance from the component segment travel time variances is proposed. Other reliability measures include statistical range methods (Van Lint et al., 2008; Bates et al., 2001), buffer time methods and tardy-trip measures (Lomax et al., 2003), and so on. The choice of measures usually depends on the context of study and ease of computation. In this study, the on-time arrival probability is adopted to measure the reliability because it reasonably captures empirical features of user behavior such as risk-averse preference to late arrival and can be extracted from trip records of GPS probe vehicle data.

When it goes to calculate the reliability of a path consisting of a number of links, it should be noted that the correlation among links can not be neglected, because it is likely that congestion on the downstream links will result in upstream links becoming congested. Gajewski and Rilett (2004) estimated the correlation of link travel time by using a nonparametric regression technique based on Bayesian neural cubic splines. It was found that heavier congestion reduces the correlation of travel times between links. Rachtan et al. (2013) developed three regression models to describe correlation variation by considering various combinations of variables such as spatial distance, temporal distance, traffic state and the number of lanes. They found that the primary factor in correlation is spatial distance. Jenelius and Koutsopoulos (2013) incorporated the spatial-temporal correlation of link travel times into traffic network analysis using GPS probe vehicle data with low sampling frequency. They found that attributes such as one-way streets, speed limits and signalized or non-signalized right turns and left turns had significant effects on travel time correlation.

2.2 Path finding algorithm considering reliability

Methods of finding the reliable path have attracted increasing attention recently among those investigating risk-averse navigation (Xiao and Lo, 2013) and transportation network analysis. Three categories of stochastic routing models have been proposed to present inherent travel time uncertainties and provide travelers with either pre-trip routing guidance (Samaranayake et al., 2012) or adaptive en-route guidance (Gao and Huang, 2012; Taş et al., 2014):

(1) The hyperpath model. Emanating from the study of transportation assignment (Nguyen and Pallotino, 1988; Spiess and Florian, 1989; Schmocker et al., 2013), the hyperpath represents a sequence of path selection strategies rather than a simple shortest path. Given uncertainty about link travel time, link usage probabilities are sought that minimize the driver's maximum probability to unexpected delay on each node, leading to the determination of a pessimistic travel time of arrival at the destination (Bell, 2009; Schmocker et al., 2009).

(2) The least expected travel time (LET) path model. The LET problem on stochastic and time dependent networks was first discussed by Hall (1986). Miller-Hooks and Mahmassani (1998) defined the optimal path in a stochastic and time-varying time dependent road network as the one that has the least possible travel time. Non-dominated paths under stochastic dominance rules are determined by label-correcting algorithms (Miller-Hooks and Mahmasssani, 2000; Miller-Hooks, 2001; Miller-Hooks and Mahmasssani, 2003). A heuristic algorithm based on the k-shortest path algorithm was proposed to find the LET path (Fu and Rilett, 1998). Based on the stochastic first-in-first-out property, the multiple criteria A* algorithm can be used to determine the LET path in time-dependent networks efficiently (Chen et al., 2013a; Chen et al., 2014). The LET path problem describes the decision-making under uncertainty, but it fails to take into account the risk-averse behavior of travelers.

(3) The stochastic on-time arrival or reliable a priori shortest-path model. The optimal path is defined as the one that either maximizes the reliability of arrival within a given time budget (Frank, 1969; Fan et al., 2005; Nie and Fan, 2006) or minimizes the travel time budget for a specified on-time arrival probability (Chen and Ji, 2005). However, these models usually neglect the correlation of link travel times.

In this study, we aim to solve the α-reliable path problem in a stochastic road network with correlated link travel times. Due to the nonlinear and non-additive properties of this approach, there are few efficient routing algorithms that can contribute to solving such a problem. To date, the efficient methods for nonlinear and non-additive routing problems include: (1) the non-dominance based method (Chen et al., 2012; Chen et al., 2013b), (2) the simulation-based method (Ji et al., 2011; Zockaie, 2013; Prakash and Srinivasan, 2014), and (3) the Lagrangian relaxation approach (Xing and Zhou, 2011; Yang and Zhou, 2014; Xing and Zhou, 2013). The most reliable paths can be determined by enumerating all non-dominated paths under the first order stochastic dominance (Nie and Wu, 2009). However, it might not be efficient enough for navigation operations on large-scale networks. Chen et al. (2013b) proposed the multiple criteria A∗ algorithms for finding the α-reliable path considering various dominance conditions, which is potential to determine the most

reliable path in a large-scale road network efficiently. The simulation based method is also computationally expensive and the accuracy of the result is relying on the maximum number of simulations. As for the Lagrangian relaxation approach, Xing and Zhou (2011) showed that 10-20 calculations of standard shortest-path algorithms for the reformulated models can achieve a preferable relative gap of about 2-6% on a large-scale transportation network with 53,124 nodes and 93,900 links. However, they only approximated the correlation of link travel time by using the Monte Carlo sampling-based method without taking account into the variance-covariance matrix explicitly, which may lead to a bias estimation due to limited samples.

2.3 Eco-routing problem

Traditional navigation systems primarily find the shortest or fastest path between origin and destination based on roadway length or travel time. Intuitively, one may think that the shortest path or fastest path would also be the most eco-friendly path. However, a shortest path may take a driver through a heavy congested area, resulting in high vehicle emissions. There may be cases where a fastest path results in longer travel distance, albeit on less congested roadways. Traveling on a path at a higher speed over a longer distance will also lead to more fuel consumption compared with a shorter path (Masikos et al., 2015). Recent years have seen emergence of the eco-routing concept, in which fuel consumption or vehicle emissions are set as the objective of the routing problem, as a way to fill this gap. For example, Yao and Song (2013) developed a dynamic eco-routing model utilizing a Dijkstra shortest-path algorithm (Dijkstra, 1959) and incorporating a vehicle specific power (VSP) based model and a dynamic traffic information database. Similarly, a navigation system considering eco-route was developed and validated in the Los Angeles Metro area (Boriboonsomsin et al., 2012), in which multisource historical and real-time traffic information are integrated. Nie and Li (2013) introduced a multiple objective function to the constrained eco-routing problem where the price of travel time and fuel was considered, but no concrete method of solution was mentioned. Many studies point out that eco-routing could result in great reductions in emissions, but it naturally comes at the expense of increased travel time (Ahn and Rakha, 2008; Boriboonsomsin et al., 2012; Alam and McNabola, 2014). However, few studies on eco-routing have discussed the trade-off between emissions reduction and increased travel time (Guo et al., 2013; Boriboonsomsin et al., 2014; Aziz and Ukkusuri, 2014). Longer travel times may result in delayed travel plans, which may cause serious troubles such as missing a flight. In addition to shortest path problem, the eco-routing concept has been applied to other transportation problems such as pollution routing problem, traffic assignment problem, road pricing problem, and rail freight transports. For example, Tzeng and Chen (1993) introduced a

traffic assignment model taken into account travel time, travel distance and CO emission simultaneously. Bektas and Laporte (2011) addressed a pollution routing problem in which emissions were modelled as a function of vehicle speed and load. Koc et al. (2014) introduced a mix pollution routing problem taken into account heterogeneous vehicle fleet, which aimed to optimize the sum of vehicle fixed costs and routing cost. Franceschetti et al. (2013) proposed an integer linear programming formulation of the time-dependent pollution routing problem which taken into account traffic congestion at peak periods. Chen and Yang (2012) studied a Pareto-optimal pricing scheme that aims to take into account both vehicular congestion and CO emission. Kirschstein and Meisel (2015) developed mesoscopic GHG emission models for evaluating the ecological performance of rail freight transports.

With the travel time restriction, the eco-routing problem falls into the class of NP-hard problems (Garey and Johnson, 1979). Such NP-hard routing problems cannot be easily solved by shortest-path algorithms such as the Dijkstra algorithm (Dijkstra, 1959) and the A-star algorithm (Hart et al., 1968). Various versions of exact and approximate methods have been developed to solve this problem. The label-setting algorithms (Desrochers et al., 1988; Dumitrescu et al., 2003) or Lagrangian relaxation algorithms (Carlyle et al., 2008; Zeng et al., 2015) can be regarded as a heuristic method. Path ranking approach such as the k-shortest path algorithms (Yen, 1971; Azevedo et al., 1993; Eppstein., 1998), which enumerates all paths, can be regarded as an exact solution method (Handler and Zang, 1980). This method is applicable when the optimal path can be found for relative small k in a small-scale network. However, they are intractable for large-scale networks due to the exponential increase in computational effort required if k is a large number. Recently, Santos et al. (2007) introduced an improved approach based on the k-shortest path algorithm by identifying a more effective search direction. Lagrangian relaxation approach can efficiently search the near optimal solution. However, the optimal solution may not be found if it falls into the duality gap. To fill this gap, some approaches to close the duality gap have been developed. For example, a Lagrangian relaxation approach and closed the duality gap by using the k-shortest path algorithm is developed by Handler and Zang (1980). Beasley and Christofides (1989) solved the constrained shortest path problem by using sub-gradient optimization and developed a branch-and-bound approach to close the duality gap. Similarly, Carlyle et al. (2008) proposed a depth first branch-and-bound approach to close the duality gap, and a bisection searching technique was applied to determine the Lagrangian multipliers. Mehlhorn and Ziegelmann (2000) developed a hull based approach to solve a linear relaxation of the constrained shortest path problem and applied three approaches to close the duality gap: Hassin' algorithm

(Hassin, 1992), the k-shortest paths algorithm developed by Jimenez and Marzal (1999), and a dynamic programming algorithm.

On the other hand, the constrained eco-routing problem is similar to the bi-objective routing problem, where the travel time can be regarded as the second objective. Various versions of bi-objective routing problem have been discussed in recent literatures (Coutinho-Rodrigues et al., 1999; Xie, and Waller, 2012a; Chen and Nie, 2013; Wang et al., 2014). It is noteworthy that the optimal solution of the constrained eco-routing problem should be one of the Pareto-optimal solutions to the bi-objective routing problem. Therefore, a number of efficient approaches for finding the Pareto-optimal solution set offer promise to help find the optimal solution to the constrained eco-routing problem. For example, the non-inferior set estimation (NISE) method (Cohon, 1978) in which the weighting parameters are updated iteratively, is usually used for solving multi-objective (including bi-objective) optimization problems, and has been successfully applied to routing problems with multiple objectives. Coutinho-Rodrigues et al. (1999) proposed a combination approach applying NISE-like weighting approach and k-shortest path algorithm to find the Pareto-optimal solution set. Pugliese and Guerriero (2013) addressed the resource constrained shortest path problem based on reference point methodology (Granat and Guerriero, 2003). Raith and Ehrgott (2009) proposed a hybrid solution strategy and two-phase algorithm in which a weighting approach is applied in combination with a label setting or path ranking approach.

2.4 References

Ahn, K., & Rakha, H. (2008). The effects of route choice decisions on vehicle energy consumption and emissions. Transportation Research Part D: Transport and Environment, 13(3), 151-167.

Alam, M. S., & McNabola, A. (2014). A critical review and assessment of Eco-Driving policy & technology: Benefits & limitations. Transport Policy, 35, 42-49.

Aziz, H. A., & Ukkusuri, S. V. (2014). Exploring the trade-off between greenhouse gas emissions and travel time in daily travel decisions: Route and departure time choices. Transportation Research Part D: Transport and Environment, 32, 334-353.

Azevedo, J., Costa, M. E. O. S., Madeira, J. J. E. S., & Martins, E. Q. V. (1993). An algorithm for the ranking of shortest paths. European Journal of Operational Research, 69(1), 97-106.

Bates, J., Polak, J., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. Transportation Research Part E: Logistics and Transportation Review, 37(2), 191-229.

Bektaş, T., & Laporte, G. (2011). The pollution-routing problem. Transportation Research Part B: Methodological, 45(8), 1232-1250.

Beasley, J. E., & Christofides, N. (1989). An algorithm for the resource constrained shortest path problem. Networks, 19(4), 379-394.

Bell, M. G., 2000. A game theory approach to measuring the performance reliability of transport networks. Transportation Research Part B: Methodological, 34(6), 533-545.

Bell, M. G., 2009. Hyperstar: A multi-path Astar algorithm for risk averse vehicle navigation. Transportation Research Part B: Methodological, 43(1), 97-107.

Boriboonsomsin, K., Barth, M. J., Zhu, W., & Vu, A. (2012). Eco-routing navigation system based on multisource historical and real-time traffic information. IEEE Transactions on Intelligent Transportation Systems, 13(4), 1694-1704.

Boriboonsomsin, K., Dean, J., & Barth, M. (2014). Examination of Attributes and Value of Ecologically Friendly Route Choices. Transportation Research Record: Journal of the Transportation Research Board, 2427(1), 13-25.

Carlyle, W. M., Royset, J. O., & Kevin Wood, R. (2008). Lagrangian relaxation and enumeration for solving constrained shortest-path problems. Networks, 52(4), 256-270.

Chen, A., Ji, Z., 2005. Path finding under uncertainty. Journal of advanced transportation, 39(1), 19-37.

Chen, A., Yang, H., Lo, H. K., Tang, W. H., 2002. Capacity reliability of a road network: an assessment methodology and numerical results. Transportation Research Part B: Methodological, 36(3), 225-252.

Chen, L., & Yang, H. (2012). Managing congestion and emissions in road networks with tolls and rebates. Transportation Research Part B: Methodological, 46(8), 933-948.

Chen, B. Y., Lam, W. H., Sumalee, A., 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, B.Y., Lam, W.H.K., Li, Q.Q., Sumalee, A. Yan, K., 2013a. Shortest path finding

problem in stochastic time-dependent road networks with stochastic first-in-first-out property. IEEE Transactions on Intelligent Transportation Systems, 14, 1907-1917.

Chen, B.Y., Lam, W.H.K., Sumalee, A., Li, Q., Shao, H., Fang, Z., 2013b. 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., Tam, M. L., 2014. Reliable shortest path problems in stochastic time-dependent networks. Journal of Intelligent Transportation Systems, 18(2), 177-189.

Chen, P., & Nie, Y. M. (2013). Bicriterion shortest path problem with a general nonadditive cost. Transportation Research Part B: Methodological, 57, 419-435.

Coutinho-Rodrigues, J. M., Clímaco, J. C. N., & Current, J. R. (1999). An interactive bi-objective shortest path approach: searching for unsupported nondominated solutions. Computers and Operations Research, 26(8), 789-798.

Cohon, J. L. (1978). Multiobjective programming and planning. Mathematics in science and engineering.

Dijkstra, E. W., 1959. A note on two problems in connexion with graphs. Numerische mathematik, 1(1), 269-271.

Desrochers, M., & Soumis, F. (1988). A generalized permanent labeling algorithm for the shortest path problem with time windows. INFOR Information Systems and Operational Research.

Dumitrescu, I., & Boland, N. (2003). Improved preprocessing, labeling and scaling algorithms for the Weight‐Constrained Shortest Path Problem. Networks, 42(3), 135-153.

Frank, H., 1969. Shortest paths in probabilistic graphs. Operations Research, 17(4), 583-599.

Fu, L.P., Rilett, L.R., 1998, Expected shortest paths in dynamic and stochastic traffic networks. Transportation Research Part B-Methodological, 32, 499-516.

Fan, Y. Y., Kalaba, R. E., Moore II, J. E., 2005. Arriving on time. Journal of Optimization Theory and Applications, 127(3), 497-513.

Franceschetti, A., Honhon, D., Van Woensel, T., Bektaş, T., & Laporte, G. (2013). The

time-dependent pollution-routing problem. Transportation Research Part B: Methodological, 56, 265-293.

Garey, M., Johnson, D., (1979). Computers and Intractability: A Guide to the Theory of NP-Completeness. Freeman, San Francisco, CA.

Granat, J., & Guerriero, F. (2003). The interactive analysis of the multicriteria shortest path problem by the reference point method. European Journal of Operational Research, 151(1), 103-118.

Gajewski, B. J., Rilett, L. R., 2004. Estimating link travel time correlation: an application of Bayesian smoothing splines. Journal of Transportation and Statistics, 7(2/3), 53-70.

Gao, S., Huang, H., 2012. Real-time traveler information for optimal adaptive routing in stochastic time-dependent networks. Transportation Research Part C: Emerging Technologies, 21(1), 196-213.

Guo, L., Huang, S., & Sadek, A. W. (2013). An evaluation of environmental benefits of time-dependent green routing in the greater Buffalo-Niagara region. Journal of Intelligent Transportation Systems, 17(1), 18-30.

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. IEEE Transactions on Systems Science and Cybernetics, 4(2), 100-107.

Hall, R. W., 1983. Travel outcome and performance: the effect of uncertainty on accessibility. Transportation Research Part B: Methodological, 17(4), 275-290.

Hall, R. W., 1986. The fastest path through a network with random time-dependent travel times. Transportation science, 20(3), 182-188.

Handler, G. Y., & Zang, I. (1980). A dual algorithm for the constrained shortest path problem. Networks, 10(4), 293-309.

Hassin, R. (1992). Approximation schemes for the restricted shortest path problem. Mathematics of Operations research, 17(1), 36-42.

Ji, Z., Kim, Y. S., Chen, A., 2011. Multi-objective α-reliable path finding in stochastic networks with correlated link costs: A simulation-based multi-objective genetic algorithm approach (SMOGA). Expert Systems with Applications, 38(3), 1515-1528.

Jimenez, V. M., & Marzal, A. (1999). Computing the k shortest paths: A new algorithm

and an experimental comparison. In Algorithm engineering, Springer Berlin Heidelberg, 15-29.

Jenelius, E., Koutsopoulos, H. N., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. Transportation Research Part B: Methodological, 53, 64-81.

Kaparias, I., Bell, M. G., Belzner, H., 2008. A new measure of travel time reliability for in-vehicle navigation systems. Journal of Intelligent Transportation Systems, 12(4), 202-211.

Kirschstein, T., & Meisel, F. (2015). GHG-emission models for assessing the eco-friendliness of road and rail freight transports. Transportation Research Part B: Methodological, 73, 13-33.

Koc, C., Bektaş, T., Jabali, O., & Laporte, G. (2014). The fleet size and mix pollution-routing problem. Transportation Research Part B: Methodological, 70, 239-254.

Lo, H. K., Luo, X. W., 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.

Lomax, T., Schrank, D., Turner, S., Margiotta, R., 2003. Selecting travel reliability measures. Texas Transportation Institute, Cambridge Systematics Inc.

Masikos, M., Demestichas, K., Adamopoulou, E., & Theologou, M. (2015). Energy-efficient routing based on vehicular consumption predictions of a mesoscopic learning model. Applied Soft Computing, 28, 114-124.

Mehlhorn, K., & Ziegelmann, M. (2000). Resource constrained shortest paths. In Algorithms-ESA 2000, Springer Berlin Heidelberg, 326-337.

Mirchandani, P. B., 1976. Shortest distance and reliability of probabilistic networks. Computers & Operations Research, 3(4), 347-355.

Miller-Hooks, E. D., Mahmassani, H. S., 1998. Least possible time paths in stochastic, time-varying networks. Computers & operations research, 25(12), 1107-1125.

Miller-Hooks, E. D., Mahmassani, H. S., 2000. Least expected time paths in stochastic, time-varying transportation networks. Transportation Science, 34(2), 198-215.

Miller-Hooks, E. D., 2001. Adaptive least-expected time paths in stochastic, time-varying

transportation and data networks. Networks, 37(1), 35-52.

Miller-Hooks, E. D., Mahmassani, H., 2003. Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. European Journal of Operational Research, 146(1), 67-82.

Nie, Y., Fan, Y., 2006. Arriving-on-time problem: discrete algorithm that ensures convergence. Transportation Research Record: Journal of the Transportation Research Board, 1964(1), 193-200.

Nie, Y. M., Wu, X., 2009. Shortest path problem considering on-time arrival probability. Transportation Research Part B: Methodological, 43(6), 597-613.

Nie, Y. M., & Li, Q. (2013). An eco-routing model considering microscopic vehicle operating conditions. Transportation Research Part B: Methodological, 55, 154-170.

Nguyen, S., Pallottino, S., 1988. Equilibrium traffic assignment for large scale transit networks. European journal of operational research, 37(2), 176-186.

Ng, M., 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.

Prakash, A. A., Srinivasan, K., 2014. A Sample-Based Algorithm to Determine Minimum Robust Cost Path with Correlated Link Travel Times. Transportation Research Record: Journal of the Transportation Research Board, 2467(1), 110-119.

Pugliese, L. D. P., & Guerriero, F. (2013). A reference point approach for the resource constrained shortest path problems. Transportation Science, 47(2), 247-265.

Raith, A., & Ehrgott, M. (2009). A comparison of solution strategies for biobjective shortest path problems. Computers & Operations Research, 36(4), 1299-1331.

Rakha, H., El-Shawarby, I., Arafeh, M., 2010. Trip travel-time reliability: issues and proposed solutions. Journal of Intelligent Transportation Systems, 14(4), 232-250.

Rachtan, P., Huang, H., Gao, S., 2013. Spatiotemporal Link Speed Correlations. Transportation Research Record: Journal of the Transportation Research Board, 2390(1), 34-43.

Samaranayake, S., Blandin, S., Bayen, A., 2012. A tractable class of algorithms for reliable routing in stochastic networks. Transportation Research Part C: Emerging

Technologies, 20(1), 199-217.

Schmocker, J. D., Bell, M. G., Kurauchi, F., Shimamoto, H., 2009. A game theoretic approach to the determination of hyperpaths in transportation networks. In Transportation and Traffic Theory 2009: Golden Jubilee, 1-18.

Schmocker, J. D., Shimamoto, H., & Kurauchi, F., 2013. Generation and calibration of transit hyperpaths. Transportation Research Part C: Emerging Technologies, 36, 406-418.

Sen, S., Pillai, R., Joshi, S., Rathi, A. K., 2001. A mean-variance model for route guidance in advanced traveler information systems. Transportation Science, 35(1), 37-49.

Shao, H., Lam, W. H., Meng, Q., Tam, M. L., 2006. Demand-driven traffic assignment problem based on travel time reliability. Transportation Research Record: Journal of the Transportation Research Board, 1985(1), 220-230.

Sivakumar, R. A., Batta, R., 1994. The variance-constrained shortest path problem. Transportation Science, 28(4), 309-316.

Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. Transportation Research Part B: Methodological, 23(2), 83-102.

Taş, D., Dellaert, N., van Woensel, T., de Kok, T., 2014. The time-dependent vehicle routing problem with soft time windows and stochastic travel times. Transportation Research Part C: Emerging Technologies, 48, 66-83.

Tzeng, G. H., & Chen, C. H. (1993). Multiobjective decision making for traffic assignment. IEEE Transactions on Engineering Management, 40(2), 180-187.

Van Lint, J. W. C., van Zuylen, H. J., Tu, H., 2008. Travel time unreliability on freeways: why measures based on variance tell only half the story. Transportation Research Part A: Policy and Practice, 42(1), 258-277.

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.

Wang, J. Y., Ehrgott, M., 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.

Xiao, L., Lo, H. K., 2013. Adaptive vehicle routing for risk-averse travelers. Transportation Research Part C: Emerging Technologies, 36, 460-479.

Xie, C., & Waller, S. T. (2012a). Optimal routing with multiple objectives: efficient algorithm and application to the hazardous materials transportation problem. Computer‐Aided Civil and Infrastructure Engineering, 27(2), 77-94.

Xing, T., Zhou, X., 2011. Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach. Transportation Research Part B: Methodological, 45(10), 1660-1679.

Xing, T., Zhou, X., 2013. Reformulation and solution algorithms for absolute and percentile robust shortest path problems. IEEE Transactions on Intelligent Transportation Systems, 14(2), 943-954.

Yao, E., & Song, Y. (2013). Study on eco-route planning algorithm and environmental impact assessment. Journal of Intelligent Transportation Systems, 17(1), 42-53.

Yang, L., Zhou, X., 2014. Constraint reformulation and a Lagrangian relaxation-based solution algorithm for a least expected time path problem. Transportation Research Part B: Methodological, 59, 22-44.

Yen, J. Y. (1971). Finding the k shortest loopless paths in a network. Management Science, 17(11), 712-716.

Zeng, W., Miwa, T., Wakita, Y., & Morikawa, T. (2015). Application of Lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times. Transportation Research Part C: Emerging Technologies, 56, 309-334.

Zockaie, A., Nie, Y. M., Wu, X., Mahmassani, H. S., 2013. Impacts of Correlations on Reliable Shortest Path Finding. Transportation Research Record: Journal of the Transportation Research Board, 2334(1), 1-9.

Chapter 3

DATA COLLECTION METHOD

3.1 Data description

3.1.1 GPS data

Probe vehicles with GPS equipment have been proven to be one of the most efficient tools for traffic information collection, since they provide spatial-temporal information with high accuracy and reliability at a low cost. GPS data is therefore often used when analyzing travelers' behavior and traffic patterns because much data exists. GPS data contain a vehicle's latitude and longitude position, speed and direction with certain frequency. The coordinates can be matched to road segments on a digital map by using the map-matching technology (Miwa et al., 2012). With such information it is possible to analyze which route the driver chooses and the travel information (e.g., travel distance, average travel speed) along the trip.

3.1.2 CAN Bus data

A Controller Area Network (CAN) bus is a control network for the vehicle electronic equipment. It was designed as a communication medium of control units in vehicles (Voss, 2005). A CAN bus connects actuators with sensors enabling to detect the health of vehicle from various vehicle-related information (Othman et al., 2006), e.g., engines RPM (rounds

per minute) and fuel consumption. Fuel consumption is available in different formats, e.g. the fuel level in the tank, the instantaneous fuel consumption and the accumulated fuel consumption. Instantaneous fuel consumption is estimated based on RPM and fuel flow. As shown in Figure 3-1, through the OBD device, CAN bus data are logged second-by-second. To calculate the CO_2 emission from fuel consumption, the carbon emission can be derived by multiplying the ratio of the molecular weight of CO_2 by the molecular weight of carbon (Coe, 2005). According to the previous study by the U.S. Environmental Protection Agency, CO_2 emission from a liter of gasoline is 2.32kg. The SD card can be used as the data logger which records the GPS data and OBD data. And then all the data can be uploaded to a central server. Considering the applicability and robustness of the routing model for a navigation system, we aggregate the emission data and travel time in a link-based level after map-matching.

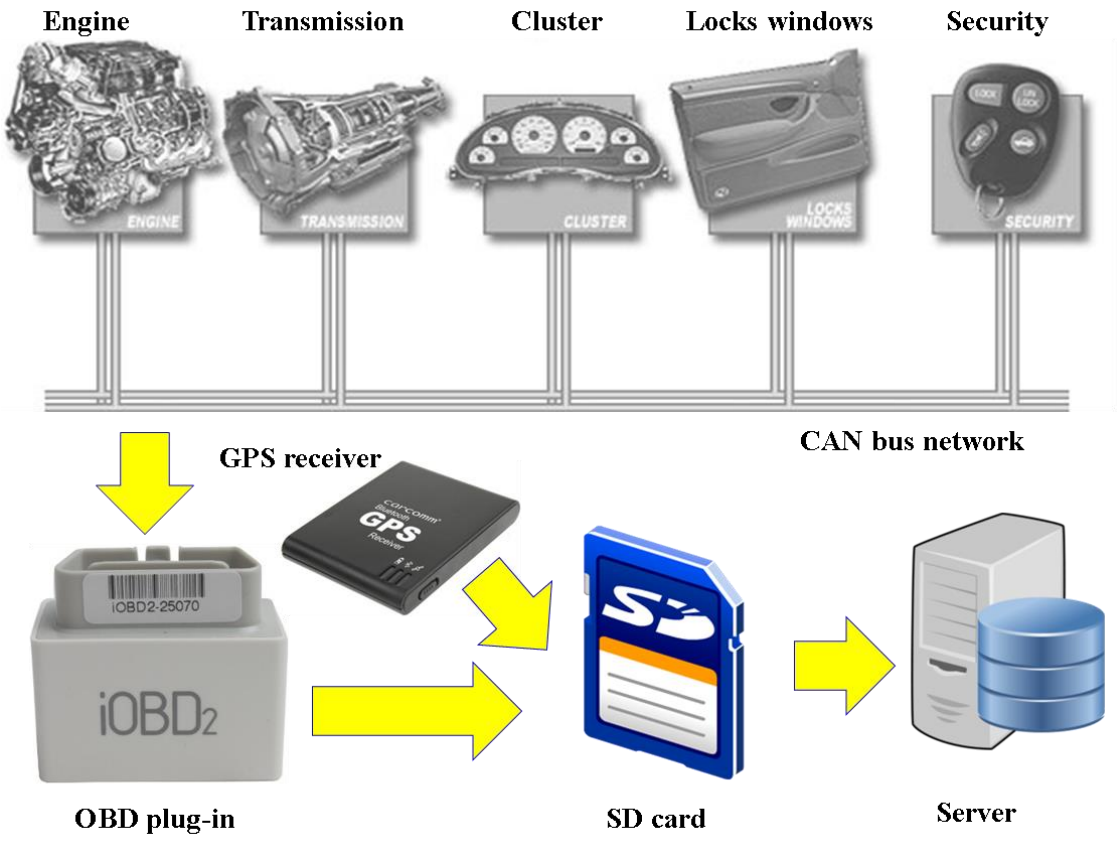


Figure 3-1 Data collection for vehicle CO_2 emission

3.1.3 Available data in transportation network

The real-world network with 4072 nodes and 12,877 links in Toyota city, Japan is used as the test area. The study area that the drivers operate in and the link usage frequency can be seen in Figure 3-2. The probe data cover about 80% of all links in the tested network. Link usage frequency is higher than five per week in 43.5% of the links, while it is zero in 19.9% of the links. This could be explained by the fact that some of the links are not attractive due to small sample size of probe vehicles in the experiment. Another potential explanation for the links with zero usage frequency is that local drivers are not willing to choose unfamiliar paths for their commutes.

Figure 3-3 gives an example of the dataset for trip records. A trip is composed by a group of links that the traveler selected from the origin to destination. The origin and destination can be usually defined from when the engine is turned on and turned off, respectively. To correct some unreasonable trips with very short travel time due to their temporal stopping at red signal, we connect adjacent trips with short stop less than 2 minutes. Combining the GPS data, CAN bus data and vehicle type, the trip fuel consumption, trip distance, number of intersection, average travel speed, coefficient of variance (COV) of link speed, and engine displacement for each vehicle, can be obtained.

Figure 3-4 gives a basic description of the collected dataset. About 64% of the trip OD distance is less than 5km, indicating that travelers are familiar with the road network in a relatively small activity range. The average travel speed of all the trips is 6.8m/s (24.5km/h) and normally distributed from 0 to 30m/s, indicating a medium delay level in the city. There are average 7.5 intersections per kilometer along one trip, indicating that travelers may need to stop due to the traffic light or conflicting traffic flow frequently. The average of COV of link speed is about 0.5 and 90% of the trips have a relatively low value of COV of link speed (less than 0.8), indicating a relatively stable traffic condition. The average engine displacement of 153 vehicles with 22 models is 1945cc. And the trip fuel consumption is distributed in a wide range from 0 to 2900ml with an average value of 330ml.

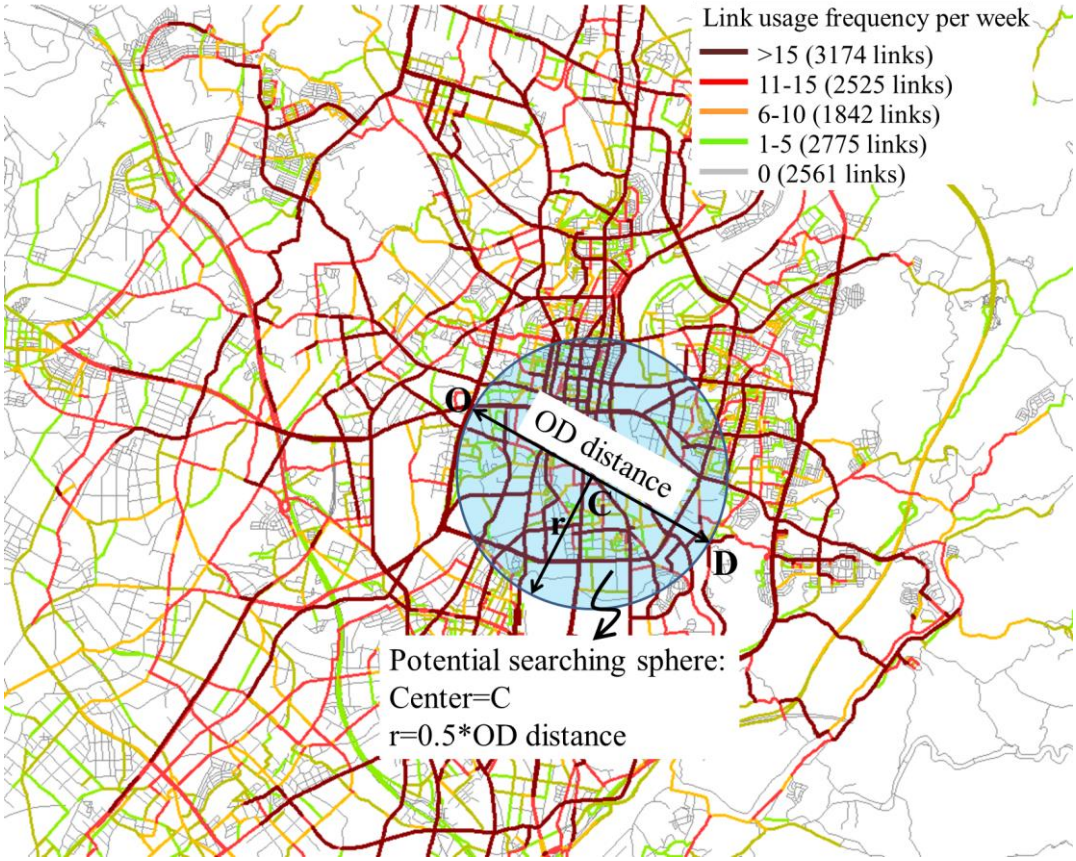
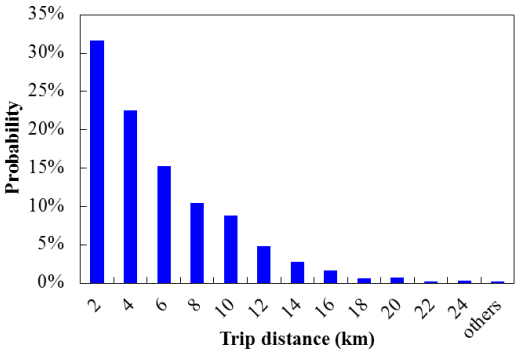


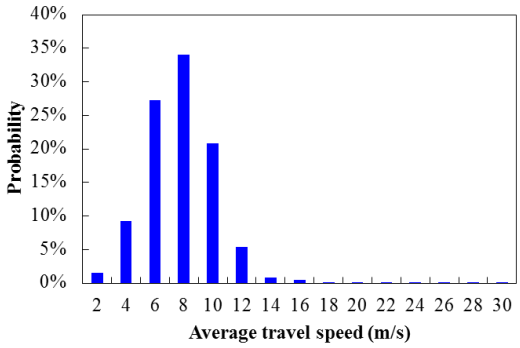
Figure 3-2 Tested network with link usage frequency in Toyota city, Japan

Trip ID	Link ID	Vehicle Model	Displacement (cc)	Link Sequence	Time	Distance (m)	Fuel Consumption (ml)
001	1122	Toyota Vitz	1496	1	2011-03-27 14:53:34	186	7.0905
001	1125	Toyota Vitz	1496	2	2011-03-27 14:53:45	230	6.0765
001	1021	Toyota Vitz	1496	3	2011-03-27 14:54:30	131	10.2515
001	…	Toyota Vitz	…	…	…	…	…
001	1094	Toyota Vitz	1496	45	2011-03-27 15:20:10	250	7.594

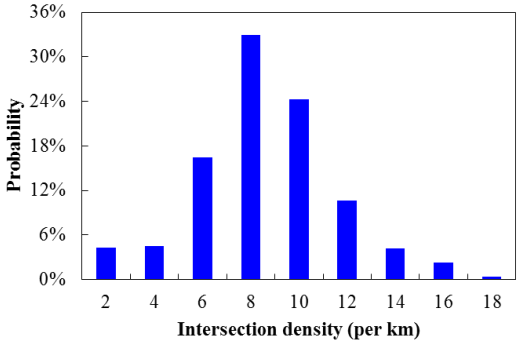
Figure 3-3 A trip record



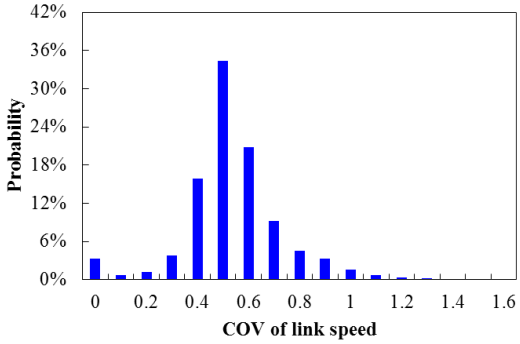
(a) Distribution of trip distance



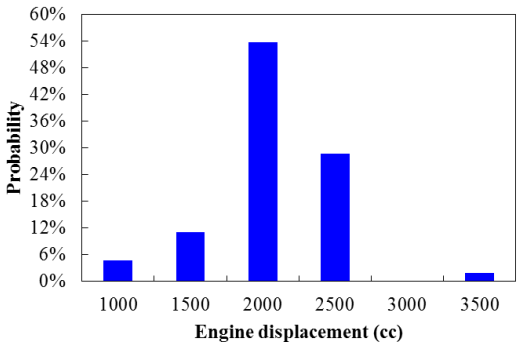
(b) Distribution of average travel speed



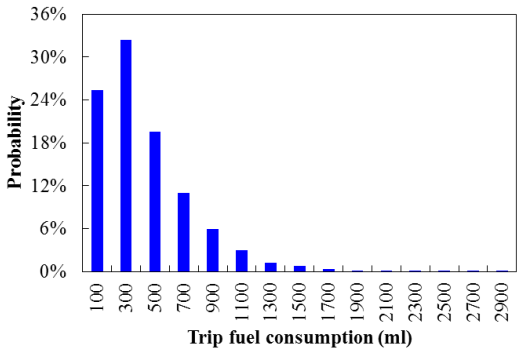
(c) Distribution of intersection density



(d) Distribution of COV of link speed



(e) Distribution of engine displacement



(f) Distribution of trip fuel consumption

Figure 3-4 Description of data set

3.2 Link travel time distribution

Links with sufficient travel time records (link usage frequency > 15 per week) are extracted for link travel time distribution analysis. The period of analysis is divided into four groups, i.e., peak hours (8:00-10:00 and 17:00-20:00) on weekday, off-peak hours (0:00-8:00, 10:00-17:00 and 20:00-24:00) on weekday, peak hours on holiday, and off-peak hours on holiday. Since all influencing factors such as weather and traffic accident cannot be considered completely, we assume that link travel times are I.I.D. in each period.

among data in each period. Based on the prior literature, several classical distributions are considered: normal, lognormal, truncated normal, and truncated lognormal. The goodness of fit of each distribution is evaluated using standard statistical tests such as K-S test, A-D test, and χ^2 test at 5% significance level. The χ^2 test is sensitive to the choice of the number of intervals. However, there is no optimal choice for the width of the interval. In this study, the initial number of interval is set as ten (the default value in Matlab). For the χ^2 to be valid, the expected count in each interval should be at least five (Croarkin and Tobias, 2006). The test may be not valid for small samples, and if some of the counts are less than five, some intervals will be combined to guarantee sufficient samples. As shown in Table 3-1, the truncated lognormal distribution is acceptable for modeling the travel time distribution for about 90% of the tested links, which outperforms the other three distributions. Therefore, it is reasonable to choose a truncated lognormal distribution to approximate the observed link travel time distribution.

Table 3-1 Comparison of goodness-of-fit of distributions for link travel times

Distribution	Accepted by K-S	Accepted by A-D	Accepted by χ^2
Normal	65%	62%	70%
Lognormal	72%	70%	75%
Truncated normal	78%	75%	86%
Truncated lognormal	91%	88%	92%

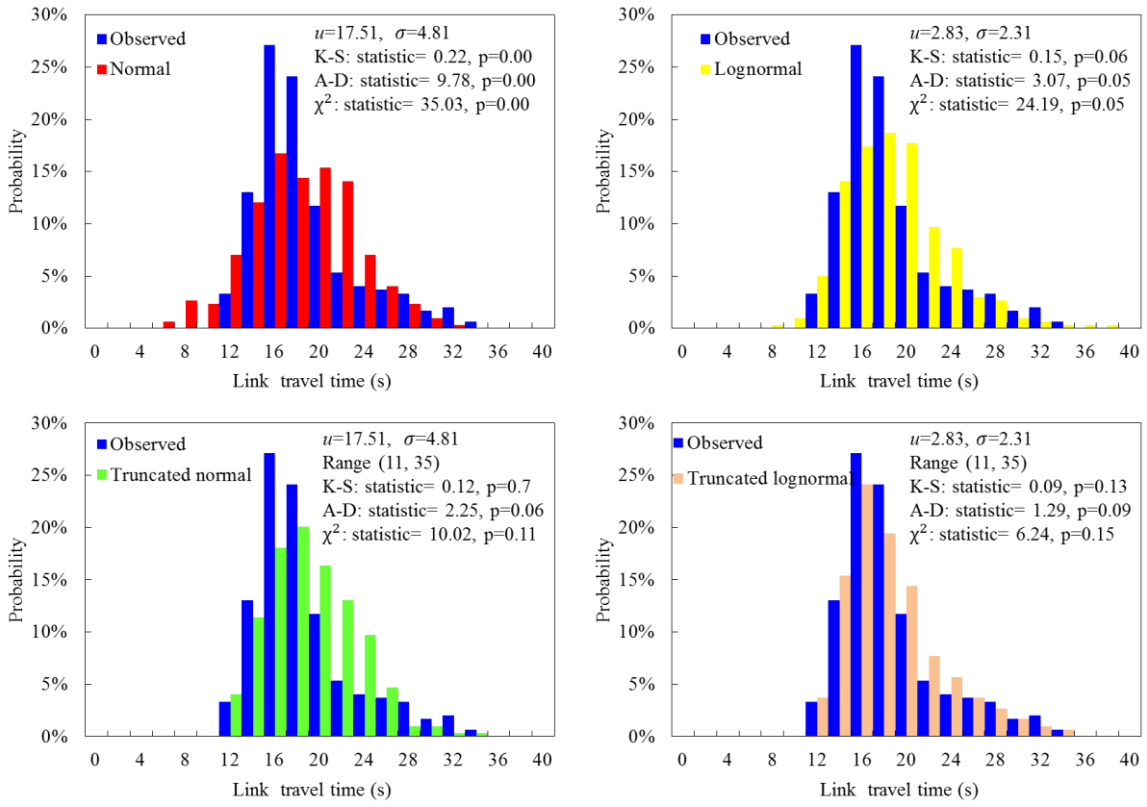


Figure 3-5 Observed and estimated distributions of link travel time

Table 3-2 Parameter estimation result of truncated lognormal link travel time

Period of Analysis		Weekday (Off-peak hour)		Weekday (Peak hour)		Holiday (Off-peak hour)		Holiday (Peak hour)	
Parameter	Variable	Coef.	p-value	Coef.	p-value	Coef.	p-value	Coef.	p-value
u_{ij}	Length (l_{ij})	0.0052	0.00	0.0063	0.00	0.0049	0.00	0.0054	0.00
	LaneCount (n_{ij})	0.012	0.01	0.017	0.01	0.011	0.02	0.015	0.03
	Constant (c_1)	1.45	0.02	1.55	0.02	1.41	0.01	1.52	0.02
σ_{ij}	Length (l_{ij})	-0.00063	0.02	-0.00086	0.00	-0.00065	0.02	-0.00051	0.02
	Constant (c_2)	0.69	0.00	0.85	0.00	0.61	0.00	0.67	0.00
Sample size		38976		16704		19448		8439	
Adjusted R-square		0.62		0.46		0.52		0.47	

To further illustrate the advantage of the truncated lognormal distribution, one of the tested links is extracted for analysis. As shown in Figure 3-5, the results of the K-S, A-D, and χ^2 tests indicate that the hypotheses that the normal distribution for modeling travel time variability along the tested link should be rejected ($p < 0.05$). According to the p value, the lognormal, truncated normal and truncated lognormal distributions are acceptable for representing the observed link travel time distribution ($p \geq 0.05$). As expected, the truncated lognormal distribution has the best result of goodness-of-fit, because it not only models the skewness characteristic of link travel time, but also limits travel time to a reasonable range.

When no observations are available for certain links in the road network, we roughly estimate travel time distributions. Assuming that travel time in each link can be approximately fitted using the truncated lognormal distribution, we take the travel time mean and travel time variance on a link to be related to link length (l_{ij}) and lane count (n_{ij}). The simple linear regression model can be written as:

$$u_{ij} = a_1 l_{ij} + b_1 n_{ij} + c_1 \tag{3-1}$$

$$\sigma_{ij} = a_2 l_{ij} + b_2 n_{ij} + c_2 \tag{3-2}$$

where a_1, b_1, c_1, a_2, b_2 and c_2 are coefficients to be estimated.

The estimation results are shown in Table 3-2. The positive signs of "Length" and "LaneCount" for u_{ij} suggest that link length and the number of lane have a positive effect on link travel time. A negative sign of "Length" for σ_{ij} means that link travel time variance decreases with an increase in link length. By using this model, link travel times can be estimated for links with no observed data.

3.3 Spatial correlation

With abundant and sufficient probe data, the correlation coefficient matrix of link travel time can be obtained directly. However, the correlation coefficients of some link pairs cannot be calculated due to missing data. Here, we propose a regression model for correlation coefficient estimation. It is assumed that correlation coefficients follow the normal distribution. Link attributes such as connection dummy, movement direction dummy, link length, and lane count are considered in the regression model. Because the variance of the correlation coefficient depends on the sample size (so a smaller sample size may result in larger variance), the inverse of sample size is considered for variance estimation. The linear regression model can be written as:

$$\rho_{ij,kl} = a_1 y_1 + a_2 y_2 + \cdots + a_{n-1} y_{n-1} + a_n + e \tag{3-3}$$

$$e \sim N\left(0, \left(\sigma_{ij,kl}^{\rho}\right)^2\right) \tag{3-4}$$

$$\sigma_{ij,kl}^{\rho} = b_1 \frac{1}{N_{ij,kl}} + b_2 \tag{3-5}$$

where $y_1, y_2, \ldots, y_{n-1}$ are independent variables of influence factors and $a_1, a_2, \ldots, a_n$ and b_1, b_2 are the model coefficients, while $N_{ij,kl}$ is the sample size of the link pair. It should be noted that the correlation coefficients are bounded between -1 and 1 and are not exactly normally distributed. The tails of normal distribution may exceed the bound of the correlation coefficients. To make the normal distribution applicable to the estimation of correlation coefficient, we bound the estimated value with the range between -1 and 1. For example, if the estimated correlation coefficient exceeds 1, the estimated value is set to 1. According to our estimation, less than 1% of the estimated results exceed the range. Therefore, from a practical point of view, normal distribution is still reasonable to be applied to approximate the correlation coefficients.

Table 3-3 Parameter estimation result of correlation coefficient

Parameters	Variables	Coef.	p-value	Note
$u_{ij,kl}^{\rho}$	Connected	0.0015	0.00	Dummy, 1 if the link pair is connected, 0 otherwise
	Straight	0.039	0.04	Dummy, 1 if the traffic movement on the link pair is straight, 0 otherwise
	Left-turn	0.013	0.04	Dummy, 1 if the traffic movement on the link pair is left-turn, 0 otherwise
	LengthSum	0.000076	0.01	The combined length of the two links
	LengthDiff	-0.000060	0.00	The length difference of the two links
	LaneSum	0.0019	0.01	The total lane count of the two links
	LaneDiff	-0.0039	0.00	The lane count difference of the two links
	Constant	-0.017	0.02	
$\sigma_{ij,kl}^{\rho}$	Sample scale	0.66	0.00	Reciprocal of sample size of the corresponding link pair
	Constant	0.13	0.00	
Number of link pairs		417548		
Adjusted R-square value		0.44		

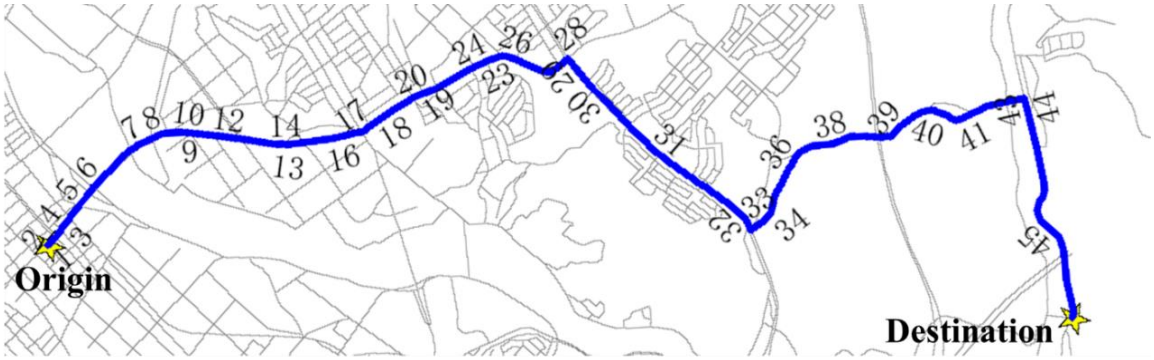


Figure 3-6 A path example for link travel time correlation analysis

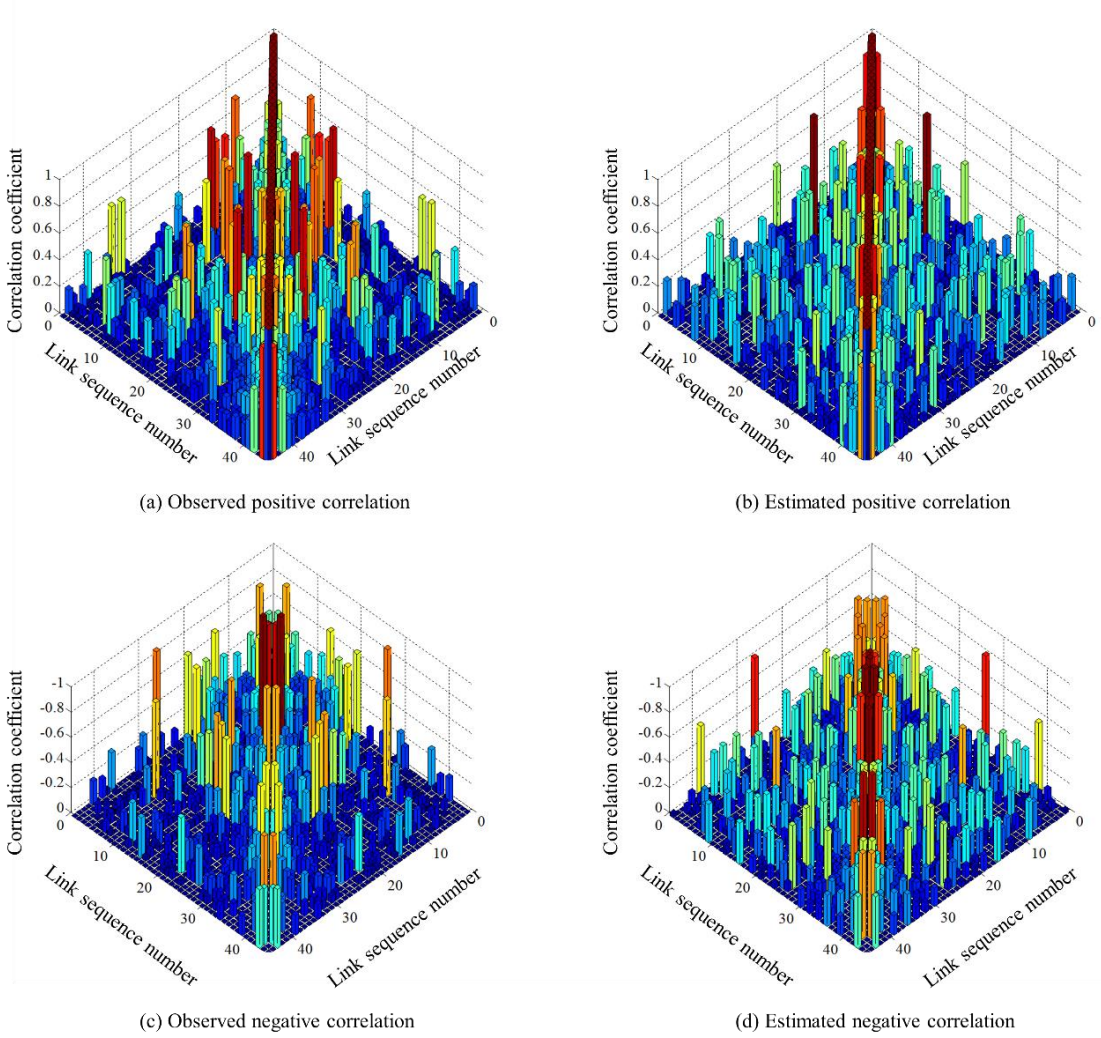


Figure 3-7 Illustration of link travel time correlation for one path

Various combinations of variables are used to estimate the regression model for link travel time correlation coefficients. As shown in Table 3-3, all the variables in the regression model are significant according to p-values at the 95% confidence level. A positive sign suggests that the correlation increases with an increase of the corresponding variable. As expected, the correlation increases if the link pair is connected. Link pairs with straight or left-turn traffic movement may have higher correlation because they share a common signal phase in Japan (where driving is on the left). The estimation results also show that more lanes and longer links increase the correlation of the link pair. This suggests that link pairs consisting of high quality roads may have stronger correlation. However, the correlation decreases if the length or lane difference of the two links in the pair increases. This suggests that link pair heterogeneity may decrease their correlation.

As shown in Figure 3-6, one of the paths extracted from the probe vehicle trips is used for spatial correlation analysis. The tested path comprises 45 links and covers both urban

and suburban areas. The observed and estimated link correlation coefficients are shown in Figure 3-7, where the axis x and y coordinates represent the number of link sequence along the path and the axis z represents the value of link correlation coefficient. This demonstrates that correlation coefficients are indeed associated with spatial distance. A larger spatial distance between two links tends to decrease the correlation effect. On the other hand, the estimated results do roughly represent the link correlation when no observations are available. However, more extensive is still needed to reveal the true correlation.

3.4 Path travel time

Given the distribution of link travel time, the distribution of path travel time could be derived theoretically by computing the sum of link travel time distribution. However, a tractable solution for the PDF or CDF of the truncated lognormal sum is not available. To make the computation of the α-reliable paths tractable, we assume that the normal distribution can be used to approximate the distribution of path travel times as composed of truncated lognormal link travel times.

Though versions of the central limit theorem (CLT) exist for independent stochastic variables, there is no CLT for the sum of dependent truncated or non-truncated lognormal distribution. Actually, the shape of path travel time distribution is relying on the variance-covariance matrix of link travel times. However, there is no tractable solution for the sum of dependent variables with arbitrary distributions except normal distribution. In such case, Monte-Carlo simulation can be regarded as an alternative tool for conducting sampling experiments on stochastic analysis. In this study, we use Monte-Carlo simulation to justify our assumption. The essential role of the simulation procedure is to generate a number of sets of scenarios (travel time realization) so that in each network scenario, the link travel time can be sampled and the path travel time can be obtained by summating the link travel times on the selected path. The travel time on each link follows a predefined truncated lognormal distribution ($LN(\mu,\sigma)$, $E(x)=e^{\mu+\frac{\sigma^2}{2}}$, $Var(x)=e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$). Without loss of generality, the lognormal distribution is truncated by 5% and 95% confidence level in this simulation study. We use the following Matlab function to generate link travel time samples from multivariate lognormal distribution:

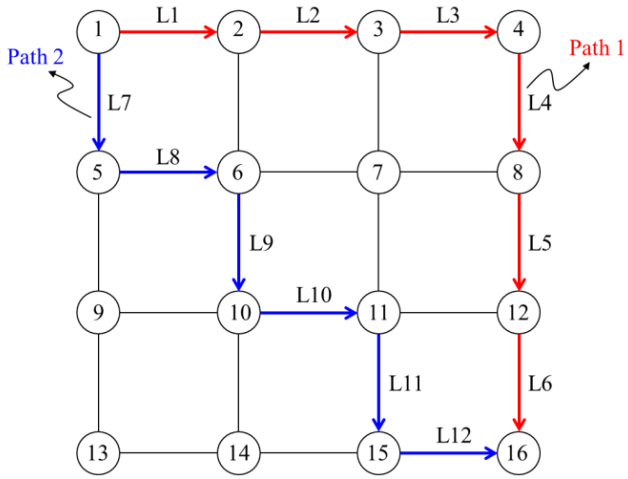
$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_{12}^2 & \rho_{12,23}\sigma_{12}\sigma_{23} & \rho_{12,34}\sigma_{12}\sigma_{34}\\ \rho_{12,23}\sigma_{12}\sigma_{23} & \sigma_{23}^2 & \rho_{23,34}\sigma_{23}\sigma_{34}\\ \rho_{12,34}\sigma_{12}\sigma_{34} & \rho_{23,34}\sigma_{23}\sigma_{34} & \sigma_{34}^2\end{bmatrix} \tag{3-6}$$

$$\mathbf{X}_N = \text{mvnrnd}([\mu_{12} \quad \mu_{23} \quad \mu_{34}], \mathbf{\Sigma}, n) \tag{3-7}$$

$$\mathbf{X}_{LN} = \exp(\mathbf{X}_N) \tag{3-8}$$

where $\mathbf{\Sigma}$ is the variance-covariance matrix, ρ is the correlation coefficient for each link pair, σ is the standard deviation, u is the mean, the subscript denotes the link index, n is the sample size, $\mathbf{X}_N$ is the matrix of dependent random variables with normal distribution, $\mathbf{X}_{LN}$ is the matrix of dependent random variables with lognormal distribution.

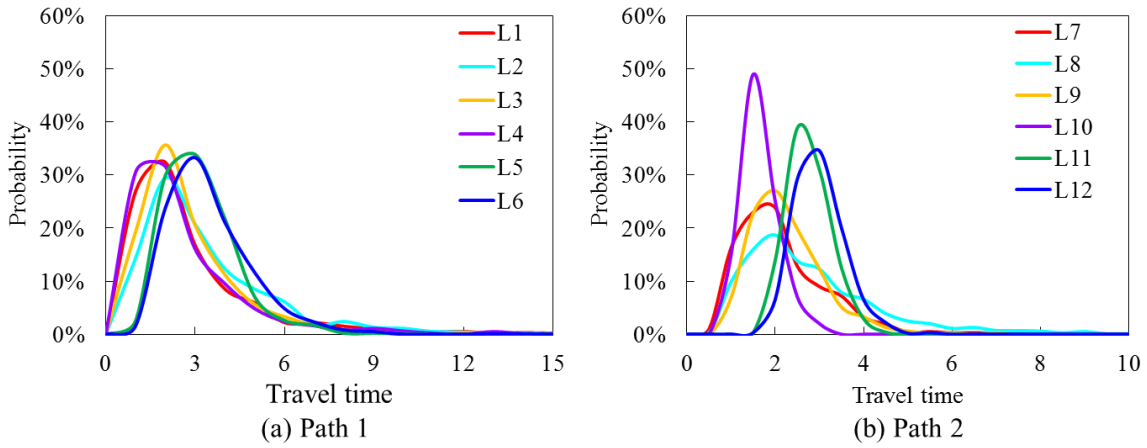
To verify our assumption that the normal distribution can be applied to approximate the sum of dependent lognormal distribution, we compare two paths in which link travel times have smaller correlation and larger correlation, respectively. As shown in Figure 3-8, Path 1 and Path 2 have the same link count and average link travel time, while their travel time variances and link travel time correlations are different. Table 3-4 and Table 3-5 show the link travel time correlation coefficient matrix for Path 1 and Path 2, respectively. Path 1 is set to have smaller link travel time correlation coefficients which are less than 0.5, while Path 2 is set to have larger link travel time correlation where certain links can have larger correlation coefficients than others. The larger link travel time correlation coefficients are highlighted in Table 3-5. To estimate the distribution of path travel time, 1000 random draws dependent on the setting of average link travel time and the correlation coefficient matrix are extracted for each link. The travel time distribution for each link is shown in Figure 3-9. Then, 1000 samples of path travel time are obtained by summating the link travel times. Figure 3-10 shows the comparison of the curves of the probability density function between Monte-Carlo simulation and normal approximation. The K-S, A-D and χ^2 tests are used to determine whether the normal distribution can approximate the simulated distribution. The statistical results show that the normal distribution can be acceptable for approximating the path travel time distribution with correlated link travel time in two cases. Comparing Path 1 and Path 2, it is found that the goodness-of-fit for Path 1 is better than that for Path 2. It indicates that larger correlation among links may cause certain bias of normal approximation. However, we found that most of the links are weak dependent ($\rho<0.2$) except the neighboring links in the real-world transportation network as shown in Figure 3-7. Therefore, it is reasonable to apply the normal distribution to approximate the sum of link travel times.



Parameter setting for each link

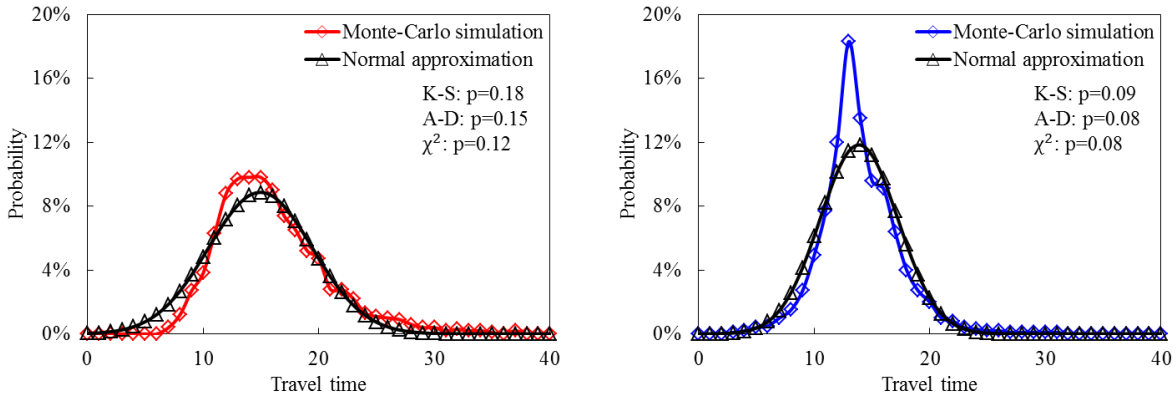
Path 1	L1	L2	L3	L4	L5	L6
μ_{Ln}	0.5	0.8	0.6	0.4	0.9	1
σ_{Ln}	0.59	0.70	0.65	0.64	0.51	0.58
Path 2	L7	L8	L9	L10	L11	L12
μ_{Ln}	0.5	0.8	0.6	0.4	0.9	1
σ_{Ln}	0.51	0.71	0.51	0.31	0.23	0.42

Figure 3-8 Tested network for simulation study



(a) Path 1 (b) Path 2

Figure 3-9 Link travel time distributions by Monte-Carlo simulation



(a) Path 1 (b) Path 2

Figure 3-10 Path travel time distribution by Monte-Carlo simulation and normal approximation

Table 3-4 Correlation coefficient matrix for path 1

Coefficient correlation	L1	L2	L3	L4	L5	L6
L1	1.00	0.05	0.10	0.24	0.26	0.20
L2	0.05	1.00	0.17	0.07	-0.21	-0.04
L3	0.11	0.17	1.00	0.12	0.07	-0.23
L4	0.24	0.07	0.12	1.00	0.10	0.10
L5	0.26	-0.21	0.07	0.10	1.00	-0.42
L6	0.20	-0.04	-0.23	0.10	-0.42	1.00

Table 3-5 Correlation coefficient matrix for path 2

Coefficient correlation	L7	L8	L9	L10	L11	L12
L7	1.00	0.90	0.59	-0.12	0.20	0.30
L8	0.90	1.00	0.43	0.23	0.43	0.01
L9	0.59	0.43	1.00	-0.30	-0.33	0.01
L10	-0.12	0.23	-0.30	1.00	0.54	-0.78
L11	0.20	0.43	-0.33	0.54	1.00	-0.22
L12	0.30	0.01	0.01	-0.78	-0.22	1.00

Table 3-6 Goodness-of-fit of normal distribution for path travel time

Length (km)	Sample	Mean γ	Accepted percentage		
			K-S test	A-D test	χ^2 test
<5	15	0.52	62%	54%	62%
5-10	25	0.35	76%	72%	72%
>10	21	0.21	81%	81%	81%

To further test the accuracy of this approximation, 61 paths with more than 50 trip records each are extracted for travel time analysis. First, we use a skewness index to evaluate the symmetry of the observed distribution. Then, the K-S, A-D and χ^2 tests are used to determine whether the normal distribution can approximate the observed distribution. Skewness measures the asymmetry about its mean for a given distribution with random variables. A positive statistic indicates that the tail of the distribution on the right side is longer than that on the left side. The skewness index can be written as:

$$\gamma = \mathrm{E}\left[\left(\frac{\mathbf{T}_p - u_p}{\sigma_p}\right)^3\right] = \frac{E[\mathbf{T}_p^3] - 3u_p\sigma_p^2 - u_p^3}{\sigma_p^3} \tag{3-9}$$

As indicated by the mean value of γ, as given in Table 3-6, the observed path travel time distributions usually have skewness on the right side, while travel times along shorter paths have more significant skewness than those along longer paths. A normal distribution

would be acceptable for modeling the path travel time distribution for about 62% of tested paths that are less than 5km length, about 76% of tested paths with a length of 5-10km, and about 81% of tested paths longer than 10km. Despite some studies that assume a lognormal distribution for path travel time modeling for reasons of the skewness of the distribution, obtaining the sum of dependent lognormal distribution functions for each path is an intractable problem. According to the literature reviews, there are no adequate methods for solving such an intractable problem in the transportation field or in fact in any other fields. Therefore, we select a normal distribution as a surrogate for path travel time distribution. According to the empirical test results above, the normal distribution is acceptable for most of the path travel time distributions, which indicates that the assumption is reasonable and practical for the routing problem.

3.5 Summary

In this chapter, the data collection for travel time and fuel consumption is introduced. Link travel time and path travel time distributions are characterized using empirical probe vehicle data. Several classical distributions (normal, lognormal, truncated normal, and truncated lognormal) are subjected to K-S test, A-D test and χ^2 test. It is found that the truncated lognormal distribution enables to approximate the link travel time distribution for about 90% of the links. Because there is no tractable method for solving the sum of link distribution function, the normal distribution is chosen as a surrogate for path travel time distribution. This is computationally tractable and shown to be acceptable in accuracy. Regression models are established for link correlation and link travel time estimation for cases where insufficient probe vehicle data are available. A number of variables, including link length, lane count, movement direction, and connection relationship, are considered in the regression models. The results demonstrate that correlation coefficients are indeed associated with spatial distance. A larger spatial distance between two links tends to decrease the correlation effect.

3.6 References

Coe, E., (2005). Average carbon dioxide emissions resulting from gasoline and diesel fuel. Tech. rep., United States Environmental Protection Agency.

Croarkin, C., Tobias, P., 2006. NIST/SEMATECH e-handbook of statistical methods. National Institute of Standards and Technology, URL http://www. itl. nist.

gov/div898/handbook.

Miwa, T., Kiuchi, D., Yamamoto, T., Morikawa, T., (2012). Development of map matching algorithm for low frequency probe data. Transportation Research Part C: Emerging Technologies, 22, 132-145.

Voss, W. (2005). A comprehensible guide to controller area network. Copperhill Media.

Othman HF, Aji YR, Fakhreddin FT, Al-Ali AR (2006) Controller area networks: evolution and applications. In: 2nd Information and Communication Technologies, ICTTA'06, vol 2, 3088–3093.

Chapter 4

ALPHA-RELIABLE PATH FINDING PROBELM

4.1 Introduction

In-vehicle navigation systems, as important applications in the field of intelligent transportation systems, have made great advances in recent years. Current systems allow users to vary the criteria used to build routes (that is, users may choose least travel time, minimum cost or to avoid certain links, route sections or areas). In developing a better trip plan, a reliable path within a given travel time budget seems more attractive than the shortest path in a stochastic network. It is widely recognized that transportation users place a high value on reliability when making their travel decisions (Lam and Small, 2001; Bates et al, 2001). For example, large travel time variations may lead to late arrivals and unexpected penalties from travelers (such as a missed a job interview). Therefore, there is an increasing demand for measuring, modeling and optimizing the reliability of travel time in developing a reliable navigation system.

For traditional routing criteria, such as average travel time and distance, the generalized objective functions are linear and additive for all the links. Such deterministic optimization problems can be efficiently solved by shortest-path algorithms such as Dijkstra algorithm and A-star algorithm. However, road networks usually have significant degree of uncertainty attributed to various factors such as varied traffic demand, signal control, and physical bottlenecks. These uncertainties lead to a non-deterministic cost on a

path. Though the shortest-path algorithms for deterministic road networks can be applied to find the least expected travel time path in an uncertain network efficiently, it may not be optimal for users who pay more attention to the travel time reliability. To find a reliable path in a stochastic network, the uncertainty of link travel times and their associated distribution functions need to be taken into account explicitly. In an early study, Frank (1969) proposed finding a reliable path by maximizing the on-time arrival probability for a given travel time budget. Mirchandani (1976) introduced a recursive algorithm that enumerated all paths and all travel time possibilities to solve Frank's problem by discretizing the travel time. By considering the variance, the reliable path problem can be formulated as a multi-objective path finding problem with a combination of travel time mean and variance (Sivakumar and Batta, 1994; Sen et al, 2001; Chang et al., 2005). Chen and Ji (2005) proposed the concept of α-reliable path. It provides a reliable path with the minimum travel time budget allowing travelers specifying a confidence level α of travel time. However, path travel time reliability measures usually lead to nonlinear and non-additive objective functions because variance is involved and there is a spatio-temporal correlation. These problems greatly increase the computational difficulty and impose challenges for the routing procedure.

The stochastic routing problem addressed in this study aims to search the α-reliable path in a stochastic transportation network with link travel time correlation. The spatial correlation of link travel time is explicitly considered by using a variance-covariance matrix, which is incorporated into the α-reliable path problem by Cholesky decomposition. The criterion for the α-reliable path problem with a given on-time probability is established. The Lagrangian relaxation is applied to calculate the lower bound of the α-reliable path solution. Specifically, the nonlinear and non-additive problem structure is decomposed into sub-problems that can be efficiently solved as a standard shortest-path problem, a series of univariate convex minimization problems and a univariate concave minimization problem, respectively. A sub-gradient algorithm is used to iteratively update the Lagrangian multipliers and find approximate optimal solutions by reducing the relative gap between upper bound and lower bound of each solution.

4.2 Problem statement

We wish to find a path with minimal travel time within a specified on-time arrival probability between two nodes in a stochastic network. In this section, we first formulate the problem and then, in the remainder of the section, discuss our motivation for solving it.

4.2.1 Variable definition

Throughout this chapter, the following notation will be used:

N: Set of nodes in the whole network;

A: Set of links in the whole network;

r: Origin node;

s: Destination node;

m: Number of links in the whole network;

a_{ij}: Directed link from node i to j;

x_{ij}: Binary variable that indicates the selection of link a_{ij};

α: On-time arrival probability;

w_{kl}: Sum of the column value of link a_{kl} in the Cholesky matrix;

y: Travel time variance of the candidate path;

T_p: Stochastic travel time of path p;

u_p: Average travel time of path p;

σ_p: Standard deviation of the travel time of path p;

u_{ij}: Mean of the link travel time with truncated lognormal distribution;

σ_{ij}: Standard deviation of the link travel time with truncated lognormal distribution;

$\rho_{ij,kl}$: Correlation coefficient of the link pair (a_{ij}, a_{kl}).

Q: Variance-covariance matrix of link travel times;

L: Lower triangular matrix of **Q**;

$Z(\alpha)$: Inverse cumulative density function of standard normal distribution at α confidence level;

ε': Relative gap between upper bound and lower bound of the candidate solution.

4.2.2 Problem formulation

Given the link travel time distribution, the path travel time distribution could be derived theoretically by computing the sum of link travel time distribution. However, a tractable solution for the PDF or CDF of the truncated lognormal sum is not available. To make the computation of the α-reliable paths tractable, we assume that the normal distribution can be used to approximate the distribution of path travel times as composed of truncated lognormal link travel times. The assumption is justified by simulation study shown in Chapter 3. Accordingly, the reliability of path travel time can be represented by the probability that the trip travel time is less than a given travel time (T_p) as follows.

$$\alpha(T_p) = \Phi\left(\frac{T_p - u_p}{\sigma_p}\right) \tag{4-1}$$

$$u_p = \sum_{ij \in p} u_{ij} \tag{4-2}$$

$$\sigma_p = \sqrt{\sum_{ij \in p} \sigma_{ij}^2 + \sum_{ij \in p, kl \in p, ij \neq kl} \rho_{ij,kl} \sigma_{ij} \sigma_{kl}} \tag{4-3}$$

where,

$\alpha(T_p)$: Reliability of path travel time;

T_p: Stochastic travel time of path p;

u_p: Mean travel time of path p;

σ_p: Standard deviation of travel time of path p;

u_{ij}: Modified mean of the link travel time with truncated lognormal distribution;

σ_{ij}, σ_{kl}: Modified standard deviation of the link travel time with truncated lognormal distribution;

$\rho_{ij,kl}$: Correlation coefficient of the link pair (a_{ij}, a_{kl}).

The transportation network is modeled as a directed graph $G(N, A)$, where $N = \{1,2,\dots,n\}$ denotes the node set and $A=\{a_{12}, a_{23}, \dots, a_{mn}\}$ represents the link set. This study addresses a time-invariant context, and the α-reliable path is considered to be a pre-trip path.

Given an on-time arrival probability α, a path $p_{rs}^{*} \in P_{rs}$ is defined as the α-reliable path if $\Phi_{T_{rs}^{*}}^{-1}(\alpha) < \Phi_{T_{rs}}^{-1}(\alpha)$ for any other path $p_{rs} \in P_{rs}$ (Chen and Ji, 2005). P_{rs} is the path set from origin *r* to destination *s*. $\Phi_{T_{rs}^{*}}^{-1}(\alpha)$ and $\Phi_{T_{rs}}^{-1}(\alpha)$ are the inverse functions of Eq.(4-1) for paths p_{rs} and p_{rs}^{*}, respectively. Therefore, the α-reliable path problem for a specified OD pair can be represented as:

$$\mathrm{P}_1\colon \mathrm{z}^{*} = \mathrm{Min}\left\{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sigma_p\right\}$$

$$= \mathrm{Min}\left\{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{\sum_{ij\in A}\sigma_{ij}^{2}x_{ij} + \sum_{ij\in A, kl\in A, ij\neq kl}\rho_{ij,kl}\sigma_{ij}\sigma_{kl}\,x_{ij}x_{kl}}\right\} \quad (4\text{-}4)$$

Subject to

$$\sum_{(i,j)\in A} x_{ij} - \sum_{(j,i)\in A} x_{ji} = g \quad (4\text{-}5)$$

$$g = \begin{cases} 1 & i = r \\ 0 & i \in N - \{r, s\} \\ -1 & i = s \end{cases} \quad (4\text{-}6)$$

where $x_{ij} \in \{0,1\}$ indicates a link on the selected path, g denotes the direction of flow for each node, and $Z(\alpha)$ denotes the inverse CDF of the standard normal distribution at α confidence level, which can be obtained from the standard normal table.

4.2.3 Sources of difficulty

As shown in Eq.(4-4), the α-reliable path problem is a nonlinear problem because of the standard deviation term. The difficulties faced in finding a solution are twofold. First, the problem is non link separable due to the cross correlation of link pairs. The explicit unknown variables include x_{ij} and x_{kl}, which makes the problem intractable. Second, sub-path optimality does not hold due to the nonlinear term. Such a problem cannot be computed by a standard shortest-path algorithm because additivity is violated. To overcome these problems, a Cholesky decomposition method for link separation in the variance-covariance matrix and a Lagrangian relaxation approach for problem decomposition are introduced in the following section.

4.3 Methodology for finding the α-reliable path

4.3.1 Covariance matrix reformulation using Cholesky decomposition

The variance-covariance matrix of link travel time can be written as follows:

$$\mathbf{Q} = \begin{bmatrix} \mathrm{Q}_{a_{i1},a_{i1}} & \mathrm{Q}_{a_{j1},a_{j2}} & \cdots & \mathrm{Q}_{a_{kn},a_{km}} \\ \mathrm{Q}_{a_{j1},a_{j2}} & \ddots & & \vdots \\ \vdots & & & \\ \mathrm{Q}_{a_{kn},a_{km}} & \cdots & & \mathrm{Q}_{a_{mn},a_{mn}} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2_{a_{i1},a_{i1}} & \rho_{a_{j1},a_{j2}}\sigma_{a_{j1}}\sigma_{a_{j2}} & \cdots & \rho_{a_{kn},a_{km}}\sigma_{a_{kn}}\sigma_{a_{km}} \\ \rho_{a_{j1},a_{j2}}\sigma_{a_{j1}}\sigma_{a_{j2}} & \ddots & & \vdots \\ \vdots & & & \\ \rho_{a_{kn},a_{km}}\sigma_{a_{kn}}\sigma_{a_{km}} & \cdots & & \sigma^2_{a_{mn},a_{mn}} \end{bmatrix} \quad (4\text{-}7)$$

where,

$\sigma_{a_{kn}}$: Standard deviation of travel time of link a_{kn};

$\rho_{a_{kn},a_{km}}$: Correlation coefficient of travel time for link a_{kn} and link a_{km}.

Since the link travel time variance-covariance matrix (**Q**) is a positive semi-definite matrix (Ng et al., 2010), it can be decomposed to a lower triangle matrix (**L**) and an upper triangle matrix by Cholesky decomposition (Golub and Loan, 1989). The Cholesky decomposition can be applied to decompose the correlations between variables and to efficiently solve the linear problems that contain a covariance matrix.

$$\mathbf{Q} = \mathbf{L}\mathbf{L}^{\mathrm{T}} = \begin{bmatrix} l_{1,1} & 0 & \ldots & 0 \\ l_{2,1} & \ddots & & \vdots \\ \vdots & & & \\ l_{m,1} & \cdots & & l_{m,m} \end{bmatrix} \begin{bmatrix} l_{1,1} & l_{2,1} & \ldots & l_{m,1} \\ 0 & \ddots & & \vdots \\ \vdots & & & \\ 0 & \cdots & & l_{m,m} \end{bmatrix} \quad (4\text{-}8)$$

where **Q** represents the variance-covariance matrix of link travel time. It is symmetric so that $\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}$. The diagonal elements of **Q** satisfy $\mathrm{Q}_{i,i} \geq 0$. It is positive semi-definite so that $\mathbf{b}^{\mathrm{T}}\mathbf{Q}\mathbf{b} \geq 0$ for all $\mathbf{b} \in \mathbf{R}^n$. **L** is the lower triangular matrix with positive diagonal entries. **L** is also called the square root of the variance-covariance matrix. The elements of **L** can be obtained by the Cholesky-Banachiewicz algorithm (Golub and Loan, 1989) recursively starting from the upper left corner of the lower triangular matrix **L**.

$$l_{i,i} = \sqrt{Q_{i,i} - \sum_{k=1}^{i-1} l_{k,i}^2} \quad (4\text{-}9)$$

$$l_{i,j} = \frac{Q_{i,j} - \sum_{k=1}^{i-1} l_{k,i} l_{k,j}}{l_{i,i}} \quad (4\text{-}10)$$

This algorithm is sometimes called the inner product formulation because the sums in step 3 of the following solution procedure are inner product.

Step 1: Let $l_{1,1} = Q_{1,1}$ (the first coefficient in the upper left corner);

Step 2: For *j*=2,…, *m*, let $l_{1,j} = \frac{Q_{1,j}}{l_{1,1}}$

Step 3: For *i*=2,…, *m*,

{

Let $l_{i,i} = \sqrt{Q_{i,i} - \sum_{k=1}^{i-1} l_{k,i}^2}$

For *j*=*i*+1,…, *m*

{

Let $l_{i,j} = \frac{Q_{i,j} - \sum_{k=1}^{i-1} l_{k,i} l_{k,j}}{l_{i,i}}$

}

}

The variance of path travel time σ_p^2 can be expressed as:

$$\sigma_p^2 = \mathbf{X}^{\mathrm{T}}\mathbf{Q}\mathbf{X} = \mathbf{X}^{\mathrm{T}}\mathbf{L}\mathbf{L}^{\mathrm{T}}\mathbf{X} = \sum_{kl \in A}\left(\sum_{ij \in A} l_{ij,kl} x_{ij}\right)^2 \quad (4\text{-}11)$$

$$\mathbf{X} = \left[x_{12}, x_{23}, \ldots, x_{ij}\right]^{\mathrm{T}} \quad (4\text{-}12)$$

For example, considering a path (p_0) consisting of three links a_{12}, a_{23}, and a_{34}. The path travel time variance is given by:

$$\sigma_{p_0}^2 = \sigma_{12}^2 + \sigma_{23}^2 + \sigma_{34}^2 + 2\rho_{12,23}\sigma_{12}\sigma_{23} + 2\rho_{12,34}\sigma_{12}\sigma_{34} + 2\rho_{23,34}\sigma_{23}\sigma_{34} \quad (4\text{-}13)$$

And the variance-covariance matrix for the link travel time is given by:

$$\mathbf{Q}_0 = \mathbf{L}_0\mathbf{L}_0^{\mathrm{T}}$$

$$= \begin{bmatrix} l_{a_{12},a_{12}} & 0 & 0 \\ l_{a_{12},a_{23}} & l_{a_{23},a_{23}} & 0 \\ l_{a_{12},a_{34}} & l_{a_{23},a_{34}} & l_{a_{34},a_{34}} \end{bmatrix} \begin{bmatrix} l_{a_{12},a_{12}} & l_{a_{23},a_{12}} & l_{a_{34},a_{12}} \\ 0 & l_{a_{23},a_{23}} & l_{a_{34},a_{23}} \\ 0 & 0 & l_{a_{34},a_{34}} \end{bmatrix} \quad (4\text{-}14)$$

According to the relation between the variance-covariance matrix and the associated Cholesky coefficients, the variance of path travel time can be expressed as:

$$\sigma_{p_0}^2 = \left(l_{a_{12},a_{12}} + l_{a_{12},a_{23}} + l_{a_{12},a_{34}}\right)^2 + \left(l_{a_{23},a_{23}} + l_{a_{23},a_{34}}\right)^2 + \left(l_{a_{34},a_{34}}\right)^2 \quad (4\text{-}15)$$

The above expression indicates that the path travel time variance can be presented by the sum of squares of the Cholesky coefficients in each column of the lower triangle matrix. Following this simplification, the α-reliable path problem with correlation term can be expressed as:

$$\mathrm{P}_1\text{: } z^* = \mathrm{Min}\left\{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{\sum_{kl\in A}\left(\sum_{ij\in A} l_{ij,kl}x_{ij}\right)^2}\right\} \quad (4\text{-}16)$$

With the Cholesky composition of the variance-covariance matrix, P_1 is simplified in terms of link separable components.

4.3.2 Lagrangian relaxation

In this section, the Lagrangian relaxation method is applied to solve the nonlinear and non-additive objective function in Eq.(4-16). The Lagrangian relaxation method has been applied to various integer programming problems (Fisher, 1981). It is a well-known solution to address the computational problems caused by nonlinear function in constrained shortest-path problems (Sivakumar and Batta, 1994; Handler and Zang, 1980; Xing and Zhou, 2011). One of the important application using Lagrangian relaxation decomposition is the variable splitting (Guignard and Kim, 1987; Joernsten and Naesberg, 1986), which enables to split the original variable *x* into a pair of variables (*x*, *y*), and then link the auxiliary variable *y* with *x* via a constraint $bx = y$. To solve the α-reliable path problem P_1 with a linear function, a three-step Lagrangian relaxation method is implemented.

Step 1: Substitute nonlinear and non-additive term by auxiliary variables

Let $y = \sum_{kl\in A}\left(\sum_{ij\in A} l_{ij,kl}x_{ij}\right)^2$ and $w_{kl} = \sum_{ij\in A} l_{ij,kl}x_{ij}, \forall kl \in A$. $1 + m$ auxiliary variables are introduced. The variable *y* denotes the variance of path travel time. The variable w_{kl} represents the sum of the coefficients for link a_{kl} (the sum of the column value of link a_{kl} in the Cholesky matrix). P_1 can be expressed as:

$$P_2: z^* = \text{Min}\left\{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y}\right\} \quad (4\text{-}17)$$

Subject to

$$\sum_{(i,j)\in A} x_{ij} - \sum_{(j,i)\in A} x_{ji} = g \quad (4\text{-}18)$$

$$\sum_{kl\in A}\left(\sum_{ij\in A} l_{ij,kl}x_{ij}\right)^2 = y \quad (4\text{-}19)$$

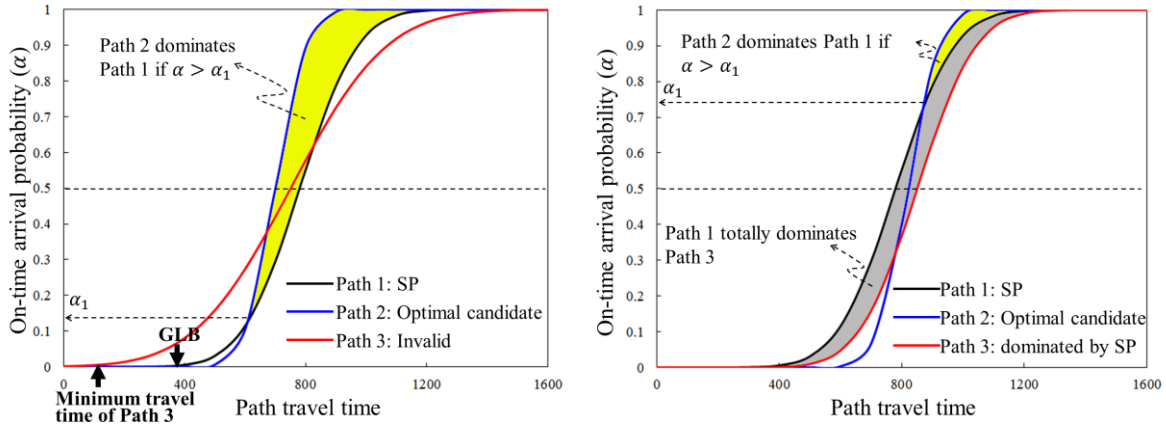
$$\sum_{ij\in A} l_{ij,kl}x_{ij} = w_{kl}, \ \ \forall kl \in A \quad (4\text{-}20)$$

$$0 \le y \le y' \quad (4\text{-}21)$$

where y' is the variance of travel time of the shortest distance path.

Lemma 1. The variance of the optimal path travel time is between 0 and the variance of the shortest distance path (SP).

Proof. We assume that the maximum travel speed in the network is a constant (e.g., 100km/h), SP has the potential of becoming the least travel time path if the driving speed can achieve to the maximum travel speed in the network. Therefore, the global lower bound (GLB) of the path travel time for a specified OD may occur in SP rather than other paths. The next step is to justify whether the travel time variance of SP is the boundary value of the variance for the α-reliable path. For any candidate solution to the α-reliable path problem, the path travel time variances for two cases need to be justified. The first case is shown in Figure 4-1 (a). If the mean travel time of the candidate α-reliable path (e.g., Path 2) is less than that of SP (Path 1), the travel time variance of the candidate (e.g., variance of Path 2) must be less than that of SP. Otherwise, the candidate path is invalid or nonexistent (e.g., Path 3 in Figure 4-1 (a)), because the minimum travel time of this path will be less than the GLB. The second case is shown in Figure 4-1 (b). If the mean travel time of the α-reliable path is no less than that of SP, the variance of path travel time also must be less than that of SP (Path 1 in Figure 4-1 (b)). Otherwise, this path (e.g., Path 3 in Figure 4-1 (b)) has larger average travel time and larger travel time variance compared to SP, which means that it is totally dominated by SP and cannot be the solution for the α-reliable path problem ($\text{Min}\{u_p + Z(\alpha)\sigma_p\}$). In summary, the above cases indicate that the variance of the travel time of the α-reliable path should be less than that of SP. Since the variance of the path travel time is always larger than or equal to zero, the feasible solution of *y* is between zero and the variance (y') of the SP.



(a) Case 1 : mean travel time less than that of SP (b) Case 2 : mean travel time no less than that of SP

Figure 4-1 Illustration of stochastic dominance

Lemma 2. If α is more than 0.5, the travel time variance of the optimal path is between zero and the variance of the least expected time path.

Proof. See Lemma 1 in Xing and Zhou (2011) and Proposition 2 in Chen et al (2013).

Step 2: Remove constraints by Lagrangian relaxation

To further relax the constraints in Eqs.(4-19) and (4-20), a set of Lagrangian multipliers, represented by $\boldsymbol{\lambda}$ and υ, is incorporated into the objective function (Eq.(4-17)).

$$L(\boldsymbol{\lambda},\upsilon) = \text{Min}\{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y} + \sum_{kl\in A}\lambda_{kl}\left(\sum_{ij\in A} l_{ij,kl}x_{ij} - w_{kl}\right) + \upsilon(\sum_{kl\in A} w_{kl}^2 - y)\} \quad (4\text{-}22)$$

Subject to Eqs.(4-5) and (4-21).

Lemma 3 (Principle of Lagrangian bounding). Given a set of Lagrangian multipliers $(\boldsymbol{\lambda},\upsilon)$, the value of the Lagrangian function $L(\boldsymbol{\lambda},\upsilon)$ is a lower bound on the value of the objective function (z^*) for the primal problem (P_2).

Proof. For every feasible solution to (P_2), we have

$$z^* = \text{Min}\left\{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y}: \sum_{kl\in A}\left(\sum_{ij\in A} l_{ij,kl}x_{ij}\right)^2 = y; \sum_{ij\in A} l_{ij,kl}x_{ij} = w_{kl}, \forall kl \in A;\ \sum_{(i,j)\in A} x_{ij} - \sum_{(j,i)\in A} x_{ji} = g;\ 0 \le y \le y'\right\}$$

$$= \text{Min}\left\{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y} + \sum_{kl\in A}\lambda_{kl}\left(\sum_{ij\in A} l_{ij,kl}x_{ij} - w_{kl}\right) + \upsilon\left(\sum_{kl\in A} w_{kl}^2 - y\right) : \sum_{kl\in A}\left(\sum_{ij\in A} l_{ij,kl}x_{ij}\right)^2 = \text{y}; \sum_{ij\in A} l_{ij,kl}x_{ij} = w_{kl}, \forall kl \in A; \sum_{(i,j)\in A} x_{ij} - \sum_{(j,i)\in A} x_{ji} = g; 0 \le y \le y'\right\}$$

Since removing the constraints $\sum_{kl\in A}\left(\sum_{ij\in A} l_{ij,kl}x_{ij}\right)^2 = y$ and $\sum_{ij\in A} l_{ij,kl}x_{ij} = w_{kl}, \forall kl \in A$ from the second equation cannot result in an increasing value of the objective function, i.e.,

$$\text{z}^* \ge \text{Min}\left\{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y} + \sum_{kl\in A}\lambda_{kl}\left(\sum_{ij\in A} l_{ij,kl}x_{ij} - w_{kl}\right) + \upsilon\left(\sum_{kl\in A} w_{kl}^2 - y\right) : \sum_{(i,j)\in A} x_{ij} - \sum_{(j,i)\in A} x_{ji} = g;\ 0 \le \text{y} \le \text{y}'\right\} = L(\boldsymbol{\lambda}, \upsilon)$$

Therefore, the value of the Lagrangian function can be seen as a lower bound for solution of the primal problem.

Step 3: Problem decomposition

By regrouping the variables in Eq.(4-22), a clear view of the components of the dual problem can be given as:

$$L(\boldsymbol{\lambda}, \upsilon) = \text{Min}\left\{\sum_{ij\in A}\left(u_{ij} + \sum_{kl\in A}\lambda_{kl}l_{ij,kl}\right)x_{ij} + \sum_{kl\in A}(\upsilon w_{kl}^2 - \lambda_{kl}w_{kl}) + Z(\alpha)\sqrt{y} - \nu y\right\} \quad (4\text{-}23)$$

Noting that x_{ij}, w_{kl} and y in Eq.(4-23) are subject to independent constraints. Therefore, the dual function $L(\boldsymbol{\lambda}, \upsilon)$ can be decomposed and separately solved by the three independent sub-functions.

$$L_x(\boldsymbol{\lambda}, \upsilon) = \text{Min}\left\{\sum_{ij\in A}\left(u_{ij} + \sum_{kl\in A}\lambda_{kl}l_{ij,kl}\right)x_{ij} : \sum_{(i,j)\in A} x_{ij} - \sum_{(j,i)\in A} x_{ji} = g\right\} \quad (4\text{-}24)$$

$$L_w(\boldsymbol{\lambda}, \upsilon) = \text{Min}\left\{\sum_{kl\in A}(\upsilon w_{kl}^2 - \lambda_{kl}w_{kl})\right\}\ \forall kl \in A \quad (4\text{-}25)$$

$$L_y(\boldsymbol{\lambda}, \upsilon) = \text{Min}\left\{Z(\alpha)\sqrt{y} - \nu y : 0 \le y \le y'\right\} \quad (4\text{-}26)$$

Solution for Eq.(4-24**)**: this equation can be solved using the shortest-path algorithm. The cost function for each link is a combination of average link travel time and the weighted sum of the components of the Cholesky matrix.

Solution for Eq.(4-25**)**: the auxiliary variable w_{kl} can be easily obtained by implementing the first order gradient.

$$\frac{\partial L_w(\lambda_{kl},\upsilon)}{\partial w_{kl}} = 2\upsilon w_{kl} - \lambda_{kl} = 0 \tag{4-27}$$

$$w_{kl} = \frac{\lambda_{kl}}{2\upsilon} \tag{4-28}$$

Solution for Eq.(4-26**)**: the minimum of this concave function is Min $\{0, Z(\alpha)\sqrt{y'} - \nu y'\}$ since the minimum of a concave function can be obtained at one of the extreme points in the feasible region.

4.3.3 Determining the Lagrangian multiplier

To optimize the primal problem (P_2), we need to obtain the maximum lower bound of the Lagrangian multiplier problem:

$$L^* = max\, L(\boldsymbol{\lambda}, \upsilon) \tag{4-29}$$

According to the Lagrangian bounding principle, L^* is always a lower bound of the objective function value for the primal problem (P_2), i.e., $L^* \leq z^*$. Based on Eqs.(4-17) and (4-29) and the Lagrangian bounding principle, an inequality can be obtained as follows:

$$L(\boldsymbol{\lambda}, \upsilon) \leq L^* \leq z^* \leq \sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y} \tag{4-30}$$

This inequality provides us with valid bounds for comparing objective function values of the Lagrange problem and optimization for any given Lagrange multipliers $(\boldsymbol{\lambda}, \upsilon)$ and any feasible solutions to the primal problem. Even if $L(\boldsymbol{\lambda}, \upsilon) \leq \sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y}$, the lower bound allows to estimate a bound on how far a candidate solution is from the actual optimality. For example, if $\frac{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y} - L(\boldsymbol{\lambda},\upsilon)}{\sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y}} < 0.1$, we know that the value of the objective function for the candidate solution is less than 10% from the optimality. This property permits us to assess the degree of sub-optimality of given solutions and it permits us to terminate the search for an optimal solution when the gap from optimality is small enough.

ε is defined as the gap between the lower bound (LB) and the upper bound (UB) of the actual optimal solution, where UB is calculated by the feasible solution. Because the value of the actual optimal solution must be between the value of LB and UB, the error of the candidate solution (corresponding to UB) will be no larger than the ε associated to the primal optimal value (z^*). For each feasible solution, LB, UB and the relative value of the approximation error (ε') can be written as:

$$LB = L(\boldsymbol{\lambda}, \upsilon) \tag{4-31}$$

$$UB = \sum_{ij\in A} u_{ij}x_{ij} + Z(\alpha)\sqrt{y} \tag{4-32}$$

$$\varepsilon' = \frac{UB-LB}{UB} \tag{4-33}$$

To minimize the estimation error (ε), the Lagrangian multipliers ($\boldsymbol{\lambda}, \upsilon$) and the auxiliary variable (y) are computed iteratively. The search direction for each Lagrangian multiplier is found by the sub-gradient optimization technique (Held et al; 1974; Ahuja et al, 1993). For solving the optimization problem $L(\boldsymbol{\lambda}, \upsilon)$ with $m+1$ dimensional vector ($\boldsymbol{\lambda}, \upsilon$), the following classical idea is applied: forming the gradient $\nabla L(\boldsymbol{\lambda}, \upsilon)$ of L defined as a row vector with components ($\frac{\partial L(\boldsymbol{\lambda},\upsilon)}{d\lambda_1}$, $\frac{\partial L(\boldsymbol{\lambda},\upsilon)}{d\lambda_2}$,…, $\frac{\partial L(\boldsymbol{\lambda},\upsilon)}{d\lambda_m}$, $\frac{\partial L(\boldsymbol{\lambda},\upsilon)}{d\upsilon}$).

$$\nabla L(\boldsymbol{\lambda}, \upsilon) = \left\{\left(\sum_{ij\in A} l_{ij,kl}x_{ij} - w_{kl}\,, \forall kl \in A\right), \sum_{kl\in A} w_{kl}^2 - y\right\} \tag{4-34}$$

Initializing from any feasible Lagrangian multipliers, the sub-function (4-24), (4-25) and (4-26) must be solved orderly to update the Lagrangian multipliers ($\boldsymbol{\lambda}, \upsilon$) at iteration k. After obtaining solutions, w_{kl} and y at iteration k, and the Lagrangian multipliers, are updated as follows.

$$\lambda_{kl}^{k+1} = \lambda_{kl}^{k} + \theta_{\lambda_{kl}}^{k}\left(\sum_{ij\in A} l_{ij,kl}x_{ij}^{k} - w_{kl}^{k}\right), \forall kl \in A \tag{4-35}$$

$$\upsilon^{k+1} = \upsilon^{k} + \theta_{\upsilon}^{k}\left[\sum_{kl\in A}\left(w_{kl}^{k}\right)^2 - y^{k}\right] \tag{4-36}$$

We update the step of each iteration k by using the following heuristic algorithm:

$$\theta_{\lambda_{kl}}^{k} = \frac{\delta_{kl}^{k}(UB-LB)}{\left\|\sum_{ij\in A} l_{ij,kl}x_{ij}^{k} - w_{kl}^{k}\right\|^2} \tag{4-37}$$

$$\theta_{\upsilon}^{k} = \frac{\delta_{\upsilon}^{k}(UB-LB)}{\left\|\sum_{kl\in A}\left(w_{kl}^{k}\right)^2 - y^{k}\right\|^2} \tag{4-38}$$

where δ_{kl}^{k} is a scalar chosen between 0 and 2 (Ahuja et al, 1993). At the beginning, UB is the objective function value of any known feasible solution to P_2, e.g., the solution of the shortest path. As the algorithm proceeds, when a better feasible solution is generated, we use the objective function value of this solution to replace UB.

4.3.4 Solution algorithm

Based on the discussion presented from section 4.3.1 to section 4.3.3, pseudo code for the solution algorithm is presented as follows.

Step 1: Initialization

Randomly generate the initial values for Lagrangian multipliers ($\boldsymbol{\lambda}, \upsilon$);

Initialize iteration number k=1;

Set the maximum speed as the regulation speed for each link;

Solve the static shortest-path problem with minimum link travel time;

Set y' as the travel time variance of the shortest path;

Calculate UB using Eq.(4-32).

Step 2: Solve the decomposed dual problems

Obtain the solution of the first sub-function (4-24) using A-star algorithm;

Obtain the solution of the second sub-function (4-25) using Eq.(4-28);

Obtain the solution of the third sub-function (4-26) with Min $\{0, Z(\alpha)\sqrt{y'} - \nu y'\}$;

Calculate LB, UB and ε' using Eqs.(4-31), (4-32) and (4-33), respectively.

Step 3: Update the Lagrangian multipliers

Update Lagrangian multipliers with Eqs.(4-35)-(4-38)

Step 4: Termination condition test

If $k > k_{max}$ or the relative gap is no larger than the specified tolerable gap, stop the algorithm, otherwise go back to the second step.

4.4 Numerical experiment

4.4.1 Illustrative example

To demonstrate the solution algorithm presented in section 5.4, a simple network with only four links is used as an example. As shown in Figure 4-2, there are only two paths from node 1 to node 4 in this network: Path 1 (1-2-4) and Path 2 (1-2-3-4). The calculation result of the algorithm is shown in Table 4-1. The intermediate computational procedures in the first few interactions are illustrated as follows.

Initialization.

Randomly generate multipliers (λ_{12}, λ_{24}, λ_{23}, λ_{34}, υ), set the on-time arrival probability α, and set the scalar of step δ_{kl}^{k}.

$\lambda_{12} = 0.718$; $\lambda_{24} = 0.958$; $\lambda_{23} = 0.957$; $\lambda_{34} = 0.852$; $\upsilon = 0.055$;

α=0.841 or $Z(\alpha) = 1$; $\delta_{kl}^{k} = 0.1$

Since α>0.5, set y' as the variance of the travel time of the least expected time path (Path 2)

$y' = 0.37$

Calculate UB using Eq.(4-32)

$UB = 3.7 + \sqrt{0.37} = 4.308$

Iteration 1.

(1) Calculate $L_x = Min\{L_x(P1), L_x(P2)\}$ using Eq.(4-24).

$L_x(P1) = 3.7 + (0.718 \times 0.4 + 0.958 \times 0.15 + 0.957 \times 0.2 + 0.852 \times 0.06) + (0.718 \times 0 + 0.958 \times 0.26 + 0.957 \times 0.115 + 0.852 \times 0.012) = 4.742$

$L_x(P2) = 3.6 + (0.718 \times 0.4 + 0.958 \times 0.15 + 0.957 \times 0.2 + 0.852 \times 0.06) + (0.718 \times 0 + 0.958 \times 0 + 0.957 \times 0.443 + 0.852 \times 0.083) + (0.718 \times 0 + 0.958 \times 0 + 0.957 \times 0 + 0.852 \times 0.172) = 4.914$

Because $L_x(P1) < L_x(P2)$, Path 1 is the candidate optimal, $L_x = 4.742$

(2) Calculate w_{12}, w_{24}, w_{23}, and w_{34} using Eq.(4-28).

$w_{12} = \frac{\lambda_{12}}{2\upsilon} = 6.539$

$w_{24} = \frac{\lambda_{24}}{2\upsilon} = 8.732$

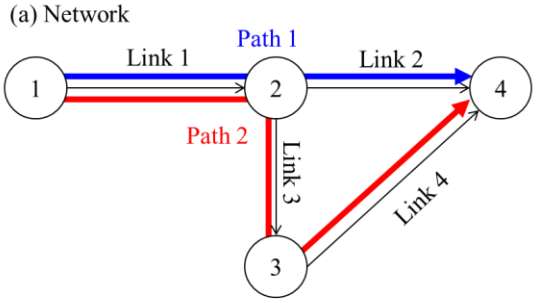
$w_{23} = \frac{\lambda_{23}}{2\upsilon} = 8.722$

$w_{34} = \frac{\lambda_{34}}{2\upsilon} = 7.765$

(3) Calculate L_w using Eq.(4-25)

$L_w = \upsilon {w_{12}}^2 - \lambda_{12} w_{12} + \upsilon {w_{24}}^2 - \lambda_{24} w_{24} + \upsilon {w_{23}}^2 - \lambda_{23} w_{23} + \upsilon {w_{34}}^2 - \lambda_{34} w_{34} = -14.010$

(a) Network



(b) Mean and standard deviation of link travel time

Link	Mean	Standard deviation
1	1.2	0.4
2	2.5	0.3
3	1.2	0.5
4	1.2	0.2

(c) Mean and variance of path travel time

Path	Mean	Variance
Path 1 (1-2-4)	3.7	0.37
Path 2 (1-2-3-4)	3.6	0.758

(d) Correlation coefficient matrix

Link	1	2	3	4
1	1	0.5	0.4	0.3
2	0.5	1	0.4	0.2
3	0.4	0.4	1	0.5
4	0.3	0.2	0.5	1

(e) Variance-covariance matrix

Link	1	2	3	4
1	0.16	0.06	0.08	0.024
2	0.06	0.09	0.06	0.012
3	0.08	0.06	0.25	0.05
4	0.024	0.012	0.05	0.04

(f) Lower triangular matrix of Cholesky decomposition

Link	1	2	3	4
1	0.400	0.000	0.00	0.000
2	0.150	0.260	0.00	0.000
3	0.200	0.115	0.443	0.000
4	0.060	0.012	0.083	0.172
Column sum for Path 1	0.550	0.260	0.000	0.000
Column sum for Path 2	0.660	0.127	0.526	0.172

Figure 4-2 A simple network for algorithm illustration

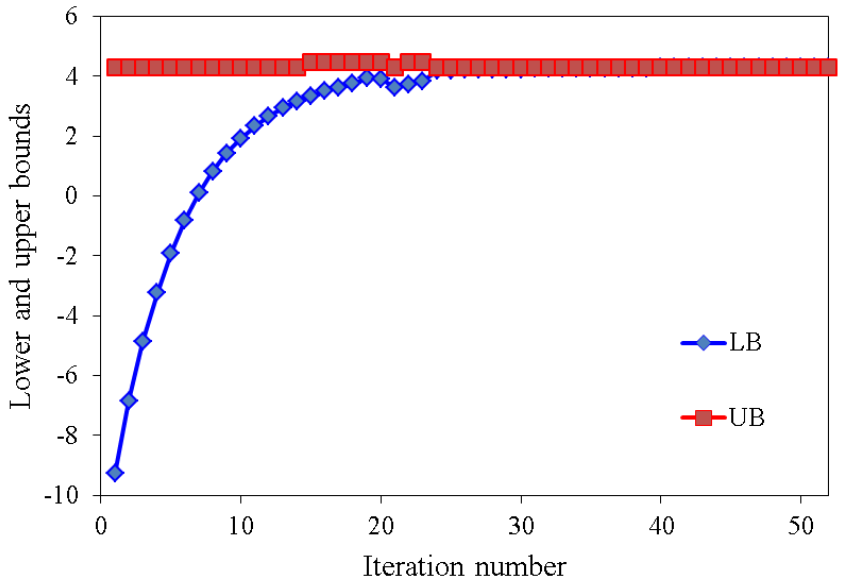


Figure 4-3 Evolution of the gap value

Table 4-1 Calculation result of each iteration

Iteration	1	2	3	4	5	6	7	8	9	10	11	12	13
λ_{12}	0.718	0.667	0.618	0.571	0.526	0.481	0.439	0.397	0.357	0.319	0.283	0.250	0.225
λ_{24}	0.958	0.919	0.881	0.846	0.813	0.782	0.753	0.725	0.699	0.674	0.651	0.629	0.608
λ_{23}	0.957	0.916	0.878	0.841	0.806	0.773	0.742	0.711	0.683	0.655	0.628	0.602	0.577
λ_{34}	0.852	0.806	0.762	0.719	0.678	0.638	0.599	0.561	0.523	0.486	0.448	0.408	0.367
υ	0.055	0.060	0.066	0.072	0.079	0.087	0.096	0.105	0.115	0.126	0.139	0.153	0.170
$L_x(P1)$	4.742	4.690	4.640	4.592	4.546	4.503	4.461	4.421	4.382	4.345	4.310	4.277	4.247
$L_x(P2)$	4.914	4.847	4.784	4.722	4.664	4.608	4.554	4.502	4.451	4.403	4.356	4.310	4.268
L_x	4.742	4.690	4.640	4.592	4.546	4.503	4.461	4.421	4.382	4.345	4.310	4.277	4.247
w_{12}	6.539	5.541	4.682	3.944	3.309	2.764	2.296	1.894	1.551	1.260	1.015	0.816	0.663
w_{24}	8.732	7.631	6.676	5.844	5.121	4.491	3.941	3.459	3.037	2.667	2.339	2.049	1.791
w_{23}	8.722	7.612	6.647	5.807	5.076	4.438	3.881	3.394	2.966	2.589	2.257	1.962	1.699
w_{34}	7.765	6.696	5.769	4.966	4.269	3.663	3.136	2.676	2.273	1.920	1.608	1.331	1.082
y	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370
θ_{12}	0.008	0.010	0.011	0.013	0.016	0.019	0.024	0.030	0.038	0.051	0.069	0.096	0.130
θ_{24}	0.005	0.005	0.005	0.006	0.006	0.007	0.008	0.008	0.009	0.010	0.011	0.012	0.013
θ_{23}	0.005	0.005	0.006	0.006	0.007	0.007	0.008	0.008	0.009	0.010	0.012	0.013	0.015
θ_{34}	0.006	0.007	0.007	0.008	0.009	0.011	0.012	0.014	0.017	0.020	0.024	0.031	0.041
θ_{υ}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003
L_w	-14.010	-11.538	-9.503	-7.827	-6.445	-5.306	-4.365	-3.588	-2.946	-2.414	-1.974	-1.609	-1.307
L_y	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LB	-9.268	-6.848	-4.863	-3.235	-1.899	-0.803	0.096	0.833	1.436	1.931	2.336	2.668	2.940
UB	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308
gap	13.576	11.156	9.171	7.543	6.207	5.111	4.213	3.476	2.872	2.377	1.972	1.640	1.368
ε'	3.151	2.590	2.129	1.751	1.441	1.186	0.978	0.807	0.667	0.552	0.458	0.381	0.318
Iteration	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
λ_{12}	0.210	0.209	0.242	0.271	0.306	0.354	0.448	0.483	0.472	0.467	0.477	0.491	0.504
λ_{24}	0.587	0.568	0.543	0.515	0.484	0.447	0.407	0.364	0.338	0.291	0.244	0.252	0.259
λ_{23}	0.551	0.526	0.508	0.491	0.477	0.469	0.500	0.507	0.479	0.467	0.465	0.467	0.468
λ_{34}	0.322	0.271	0.217	0.157	0.121	0.125	0.141	0.151	0.068	0.088	0.105	0.106	0.106
υ	0.188	0.210	0.244	0.293	0.376	0.661	0.539	0.275	0.314	0.423	-2.870	-2.903	-2.934
$L_x(P1)$	4.222	4.201	4.195	4.186	4.180	4.182	4.214	4.213	4.184	4.160	4.145	4.154	4.163
$L_x(P2)$	4.228	4.192	4.173	4.151	4.140	4.150	4.206	4.222	4.170	4.159	4.159	4.168	4.175
L_x	4.222	4.192	4.173	4.151	4.140	4.150	4.206	4.213	4.170	4.159	4.145	4.154	4.163
w_{12}	0.558	0.497	0.495	0.463	0.406	0.268	0.415	0.878	0.752	0.552	-0.083	-0.085	-0.086
w_{24}	1.560	1.351	1.110	0.879	0.644	0.338	0.377	0.662	0.539	0.344	-0.042	-0.043	-0.044
w_{23}	1.463	1.251	1.039	0.838	0.634	0.355	0.463	0.921	0.764	0.552	-0.081	-0.080	-0.080
w_{34}	0.856	0.646	0.444	0.268	0.161	0.094	0.131	0.274	0.109	0.104	-0.018	-0.018	-0.018
y	0.370	0.758	0.758	0.758	0.758	0.758	0.758	0.370	0.758	0.758	0.370	0.370	0.370
θ_{12}	0.161	0.204	0.177	0.175	0.191	0.239	0.145	0.034	0.053	0.090	0.022	0.020	0.018
θ_{24}	0.015	0.021	0.028	0.041	0.072	0.191	0.170	0.065	0.115	0.218	0.027	0.024	0.022
θ_{23}	0.017	0.025	0.032	0.046	0.075	0.178	0.117	0.030	0.052	0.090	0.023	0.021	0.019
θ_{34}	0.060	0.114	0.220	0.374	0.333	0.211	0.247	0.301	0.314	0.255	0.030	0.027	0.025
θ_{υ}	0.004	0.010	0.024	0.083	1.144	0.279	1.248	0.022	0.157	173.636	0.095	0.086	0.078
L_w	-1.058	-0.852	-0.673	-0.516	-0.379	-0.212	-0.295	-0.587	-0.455	-0.313	0.045	0.046	0.047
L_y	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LB	3.164	3.340	3.500	3.635	3.761	3.938	3.911	3.626	3.714	3.846	4.189	4.200	4.210
UB	4.308	4.471	4.471	4.471	4.471	4.471	4.471	4.308	4.471	4.471	4.308	4.308	4.308
gap	1.145	1.130	0.970	0.836	0.710	0.533	0.559	0.682	0.757	0.624	0.119	0.108	0.098
ε'	0.266	0.253	0.217	0.187	0.159	0.119	0.125	0.158	0.169	0.140	0.028	0.025	0.023

Iteration	27	28	29	30	31	32	33	34	35	36	37	38	39
λ_{12}	0.515	0.526	0.536	0.544	0.552	0.559	0.566	0.572	0.577	0.582	0.586	0.590	0.594
λ_{24}	0.266	0.272	0.277	0.282	0.287	0.291	0.295	0.298	0.301	0.304	0.307	0.309	0.311
λ_{23}	0.470	0.471	0.472	0.473	0.474	0.475	0.476	0.477	0.478	0.478	0.479	0.479	0.480
λ_{34}	0.107	0.107	0.108	0.108	0.108	0.108	0.109	0.109	0.109	0.109	0.109	0.110	0.110
v	-2.962	-2.987	-3.010	-3.031	-3.050	-3.067	-3.083	-3.098	-3.111	-3.123	-3.134	-3.144	-3.153
$L_x(P1)$	4.171	4.178	4.184	4.190	4.196	4.201	4.205	4.209	4.213	4.216	4.219	4.222	4.224
$L_x(P2)$	4.182	4.188	4.193	4.199	4.203	4.207	4.211	4.214	4.218	4.220	4.223	4.225	4.228
L_x	4.171	4.178	4.184	4.190	4.196	4.201	4.205	4.209	4.213	4.216	4.219	4.222	4.224
w_{12}	-0.087	-0.088	-0.089	-0.090	-0.091	-0.091	-0.092	-0.092	-0.093	-0.093	-0.094	-0.094	-0.094
w_{24}	-0.045	-0.045	-0.046	-0.047	-0.047	-0.047	-0.048	-0.048	-0.048	-0.049	-0.049	-0.049	-0.049
w_{23}	-0.079	-0.079	-0.078	-0.078	-0.078	-0.077	-0.077	-0.077	-0.077	-0.077	-0.076	-0.076	-0.076
w_{34}	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	-0.017	-0.017	-0.017
y	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370
θ_{12}	0.017	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.006	0.005
θ_{24}	0.020	0.018	0.016	0.015	0.013	0.012	0.011	0.010	0.009	0.008	0.008	0.007	0.006
θ_{23}	0.017	0.016	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.007	0.006	0.006
θ_{34}	0.022	0.020	0.019	0.017	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007
θ_v	0.071	0.065	0.059	0.054	0.049	0.045	0.041	0.037	0.034	0.031	0.028	0.026	0.023
L_w	0.048	0.049	0.050	0.050	0.051	0.052	0.052	0.053	0.053	0.054	0.054	0.055	0.055
L_y	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LB	4.219	4.227	4.234	4.241	4.247	4.252	4.257	4.262	4.266	4.270	4.273	4.276	4.279
UB	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308
gap	0.089	0.081	0.074	0.067	0.061	0.056	0.051	0.046	0.042	0.038	0.035	0.032	0.029
ε'	0.021	0.019	0.017	0.016	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.007
Iteration	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**
λ_{12}	0.598	0.601	0.603	0.606	0.608	0.610	0.612	0.614	0.615	0.617	0.618	0.619	0.621
λ_{24}	0.313	0.315	0.316	0.318	0.319	0.320	0.322	0.323	0.324	0.324	0.325	0.326	0.326
λ_{23}	0.480	0.481	0.481	0.481	0.482	0.482	0.482	0.482	0.483	0.483	0.483	0.483	0.483
λ_{34}	0.110	0.110	0.110	0.110	0.110	0.110	0.110	0.111	0.111	0.111	0.111	0.111	0.111
v	-3.161	-3.168	-3.175	-3.181	-3.187	-3.192	-3.197	-3.201	-3.205	-3.209	-3.212	-3.215	-3.218
$L_x(P1)$	4.227	4.229	4.231	4.232	4.234	4.235	4.237	4.238	4.239	4.240	4.241	4.242	4.242
$L_x(P2)$	4.229	4.231	4.233	4.234	4.236	4.237	4.238	4.239	4.240	4.241	4.242	4.242	4.243
L_x	4.227	4.229	4.231	4.232	4.234	4.235	4.237	4.238	4.239	4.240	4.241	4.242	4.242
w_{12}	-0.095	-0.095	-0.095	-0.095	-0.095	-0.096	-0.096	-0.096	-0.096	-0.096	-0.096	-0.096	-0.096
w_{24}	-0.050	-0.050	-0.050	-0.050	-0.050	-0.050	-0.050	-0.050	-0.050	-0.051	-0.051	-0.051	-0.051
w_{23}	-0.076	-0.076	-0.076	-0.076	-0.076	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075	-0.075
w_{34}	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017
y	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370	0.370
θ_{12}	0.005	0.004	0.004	0.004	0.003	0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.002
θ_{24}	0.006	0.005	0.005	0.004	0.004	0.004	0.003	0.003	0.003	0.002	0.002	0.002	0.002
θ_{23}	0.005	0.005	0.004	0.004	0.004	0.003	0.003	0.003	0.002	0.002	0.002	0.002	0.002
θ_{34}	0.007	0.006	0.005	0.005	0.005	0.004	0.004	0.003	0.003	0.003	0.003	0.002	0.002
θ_v	0.021	0.019	0.018	0.016	0.015	0.013	0.012	0.011	0.010	0.009	0.008	0.008	0.007
L_w	0.055	0.055	0.056	0.056	0.056	0.056	0.057	0.057	0.057	0.057	0.057	0.057	0.057
L_y	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LB	4.282	4.284	4.286	4.288	4.290	4.292	4.293	4.295	4.296	4.297	4.298	4.299	4.300
UB	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308	4.308
gap	0.026	0.024	0.022	0.020	0.018	0.017	0.015	0.014	0.013	0.011	0.010	0.009	0.009
ε'	0.006	0.006	0.005	0.005	0.004	0.004	0.004	0.003	0.003	0.003	0.002	0.002	0.002

(4) Calculate y using Eq.(4-19).

$y = 0.550^2 + 0.260^2 = 0.370$

(5) Calculate L_y using Eq.(4-26).

$L_y = \text{Min}\left\{0, Z(\alpha)\sqrt{y'} - \nu y'\right\} = 0$

(6) Update LB, UB, and ε' using Eqs.(4-31)-(4-33)

$\text{LB} = L_x + L_w + L_y = -9.268$

$UB = 3.7 + \sqrt{0.37} = 4.308$

$\varepsilon' = \frac{UB - LB}{UB} = 3.151$

(7) Calculate iteration steps θ_{12}, θ_{24}, θ_{23}, θ_{34}, and θ_υ using Eqs.(4-37) and (4-38)

$\theta_{12} = \frac{\delta_{kl}^k (UB - LB)}{(0.550 - w_{12})^2 + (0.260 - w_{12})^2 + (0.000 - w_{12})^2 + (0.000 - w_{12})^2} = 0.008$

$\theta_{24} = \frac{\delta_{kl}^k (UB - LB)}{(0.550 - w_{24})^2 + (0.260 - w_{24})^2 + (0.000 - w_{24})^2 + (0.000 - w_{24})^2} = 0.005$

$\theta_{23} = \frac{\delta_{kl}^k (UB - LB)}{(0.550 - w_{23})^2 + (0.260 - w_{23})^2 + (0.000 - w_{23})^2 + (0.000 - w_{23})^2} = 0.005$

$\theta_{34} = \frac{\delta_{kl}^k (UB - LB)}{(0.550 - w_{34})^2 + (0.260 - w_{34})^2 + (0.000 - w_{34})^2 + (0.000 - w_{34})^2} = 0.006$

$\theta_\upsilon = \frac{\delta_{kl}^k (UB - LB)}{\left(w_{12}^2 + w_{24}^2 + w_{23}^2 + w_{34}^2 - y\right)^2} = 0.000$

Iteration 2.

(1) Update multipliers λ_{12}, λ_{24}, λ_{23}, λ_{34}, and υ by using Eqs.(4-35) and (4-36)

$\lambda_{12} = \lambda_{12} + \theta_{12}(0.550 - w_{12}) = 0.667$

$\lambda_{24} = \lambda_{24} + \theta_{24} \times (0.260 - w_{24}) = 0.919$

$\lambda_{23} = \lambda_{23} + \theta_{23} \times (0.000 - w_{23}) = 0.916$

$\lambda_{34} = \lambda_{34} + \theta_{34} \times (0.000 - w_{34}) = 0.806$

$\upsilon = \upsilon + \theta_\upsilon (w_{12}^2 + w_{24}^2 + w_{23}^2 + w_{34}^2) = 0.060$

(2) Calculate $L_x = Min\{L_x(P1), L_x(P2)\}$ using Eq.(4-24).

$L_x(P1) = 4.690$

$L_x(P2) = 4.847$

Hence, Path 1 is the candidate optimal, $L_x = 4.690$

(3) Calculate w_{12}, w_{24}, w_{23}, and w_{34} using Eq.(4-28).

$w_{12} = \frac{\lambda_{12}}{2\upsilon} = 5.541$

$w_{24} = \frac{\lambda_{24}}{2\upsilon} = 7.631$

$w_{23} = \frac{\lambda_{23}}{2\upsilon} = 7.612$

$w_{34} = \frac{\lambda_{34}}{2\upsilon} = 6.696$

(4) Calculate L_w using Eq.(4-25)

$$L_w = \upsilon {w_{12}}^2 - \lambda_{12} w_{12} + \upsilon {w_{24}}^2 - \lambda_{24} w_{24} + \upsilon {w_{23}}^2 - \lambda_{23} w_{23} + \upsilon {w_{34}}^2 - \lambda_{34} w_{34}$$
$$= -11.538$$

(5) Calculate y using Eq.(4-19).

$y = 0.550^2 + 0.260^2 = 0.37$

(6) Calculate L_y using Eq.(4-26).

$L_y = \text{Min}\left\{0, Z(\alpha)\sqrt{y'} - \nu y'\right\} = 0$

(7) Update LB, UB, and ε' using Eqs.(4-31)-(4-33)

$\text{LB} = L_x + L_w + L_y = -6.848$

$UB = 3.7 + \sqrt{0.37} = 4.308$

$\varepsilon' = \frac{UB - LB}{UB} = 2.590$

(8) Calculate iteration steps θ_{12}, θ_{24}, θ_{23}, θ_{34}, and θ_υ using Eqs.(4-37) and (4-38)

$$\theta_{12} = \frac{\delta_{kl}^k (UB - LB)}{(0.550 - w_{12})^2 + (0.260 - w_{12})^2 + (0.000 - w_{12})^2 + (0.000 - w_{12})^2} = 0.010$$

$$\theta_{24} = \frac{\delta_{kl}^k (UB - LB)}{(0.550 - w_{24})^2 + (0.260 - w_{24})^2 + (0.000 - w_{24})^2 + (0.000 - w_{24})^2} = 0.005$$

$$\theta_{23} = \frac{\delta_{kl}^k (UB - LB)}{(0.550 - w_{23})^2 + (0.260 - w_{23})^2 + (0.000 - w_{23})^2 + (0.000 - w_{23})^2} = 0.005$$

$$\theta_{34} = \frac{\delta_{kl}^k (UB - LB)}{(0.550 - w_{34})^2 + (0.260 - w_{34})^2 + (0.000 - w_{34})^2 + (0.000 - w_{34})^2} = 0.007$$

$$\theta_\upsilon = \frac{\delta_{kl}^k (UB - LB)}{\left(w_{12}^2 + w_{24}^2 + w_{23}^2 + w_{34}^2 - y\right)^2} = 0.000$$

Thereafter, the computational procedure for iterations 3-52 is similar to that of iteration 2. In this example, the proposed algorithm successfully finds the optimal solution (Path 1). Starting with random positive values, the Lagrangian multipliers are iteratively updated to increase the lower bound of the primal problem. As shown in Table 4-1, Path 1 is regarded as the optimal solution with the lowest upper bound. Figure 4-3 shows the convergent trend of the upper and lower bounds at the various iterations of the sub-gradient algorithm. This clearly shows that the relative gap gradually reduces to a very small value as the iteration proceeds, demonstrating that the optimal path is determined.

4.4.2 Implementation in a real-world network

(1) Convergence analysis

To verify the convergence of the α-reliable path-finding algorithm, 1000 random OD pairs are extracted from the observed trips for testing. These extracted OD pairs satisfy two rules: (1) the linear distance separating the OD pair is 5km or more; (2) the percentage of links with a usage frequency of 10 times per week or more is higher than 30% in the

potential path searching sphere, which is shown in Figure 3-2. These extraction rules can decrease the path-finding bias due to the insufficiency of link travel time records. The performance of the proposed algorithm is evaluated by the convergence of average relative gap ($\bar{\varepsilon}' = \frac{1}{100}\sum_{n=1}^{100}\varepsilon_n'$), which is the average value for all 100 OD pairs under a pre-specified maximum iteration number and on-time arrival probability (α). As shown in Figure 4-4, the average relative gap between the upper and lower bounds decreases as the number of iterations increases. It is found that 30 iterations are sufficient for the algorithm to converge to a relatively small gap (2%-7%) under different reliability settings. However, a significant pattern in the relationship between reliability setting and convergence could not be determined.

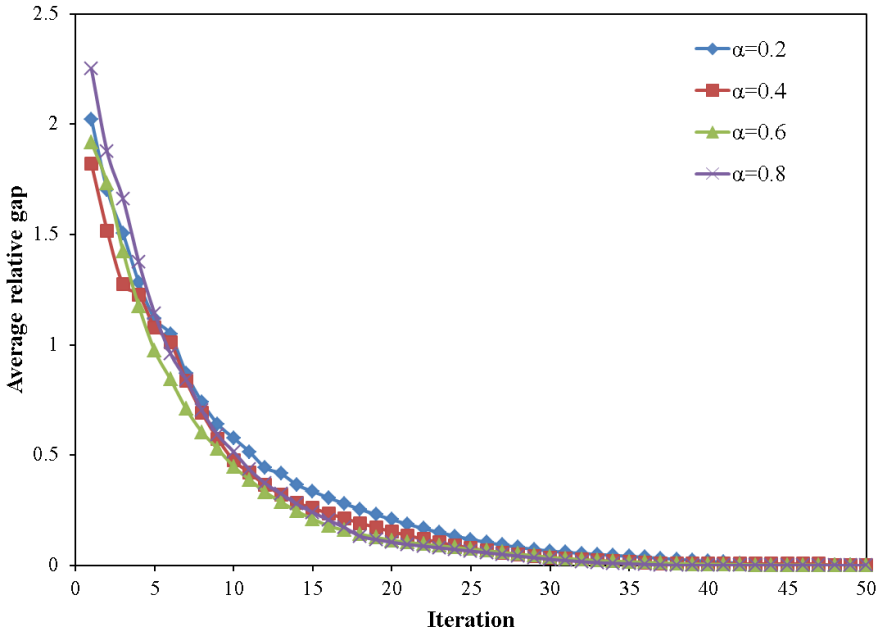


Figure 4-4 Average relative gap for different on-time arrival probability settings

Table 4-2 Computation time

OD Distance	<5 km	5-10 km	10-15km	>15km
Sample size	25	25	25	25
Average link number	30	51	74	152
Average running time	2.5s	6.3s	9.8s	19.2s

Table 4-3 Properties of the examined paths

Path	Path 1	Path 2	Path 3	Path 4
Expected travel time (s)	850	832	780	800
Variance of travel time (s^2)	9025	13556	40023	62501
Path length (m)	9280	9366	8705	8239
Link usage frequency per week	9.2	9.1	3.5	5.2

(2) Running time analysis

The algorithm is programmed by C# program on the Windows 7 platform and run on a PC with Intel Dual-Core 2.1GHz CPU and 4GB memory. The large-scale network with 4072 nodes and 12,877 links is the one shown in Figure 3-2. The most computationally expensive step of the proposed algorithm is the shortest-path computation at each iteration. Therefore, the time complexity of the proposed algorithm approximates to $O[km\log(m)]$, where *k* is the maximum iteration number and *m* is the link count of the graph. As shown in Table 4-2, the average running time increases as the OD distance increases because more links will be considered in the shortest-path algorithm. For different OD settings, the average running time is about 9.5 seconds, which can be considered acceptable for path finding in a large-scale network.

(3) Case study

In Figure 4-5, one real-world OD is extracted to demonstrate the applicability of the α-reliable path finding method. Three on-time arrival probabilities (α=0.2, 0.5, 0.8) are set, representing the decision-making strategies of risk-taking, risk-neutral, and risk-averse travelers, respectively. Accordingly, three paths are found, one corresponding to each strategy. Figure 4-6 shows the probability cumulative function curves of the three estimated paths and the observed path. Each curve represents the optimal path under a certain reliability setting. For example, when the on-time arrival probability is set to 0.2, the travel time budgets for the four paths are $T_1^{0.2}$, $T_2^{0.2}$, $T_3^{0.2}$ and $T_4^{0.2}$, respectively. Path 4 is recommended because its travel time budget is less than that of the other paths ($T_4^{0.2} < T_3^{0.2} < T_2^{0.2} < T_1^{0.2}$). However, if a traveler is more concerned with the variability of the risk of path travel time (i.e., the probability that the path travel time will exceed a given travel time budget), Path 1 and Path 2 are recommended. Actually, Path 1 and Path 2 are very similar because they share 80% of the links, as shown in Figure 4-5. If the on-time arrival probability is set to 0.85 or above, Path 1 dominates Path 2. This example demonstrates that the observed traveler was risk-averse to on-time arrival. That is, as shown in Table 4-3, though Path 1 has the maximum expected travel time among the examined paths, it has the least variance of travel time. That means the traveler preferred a highly reliable path though it may be longer. Further, it is found that the component links of Path 1 have the highest average link usage frequency. This indirectly indicates that Path 1 is the most reliable path because rational drivers usually do not choose unreliable links frequently.

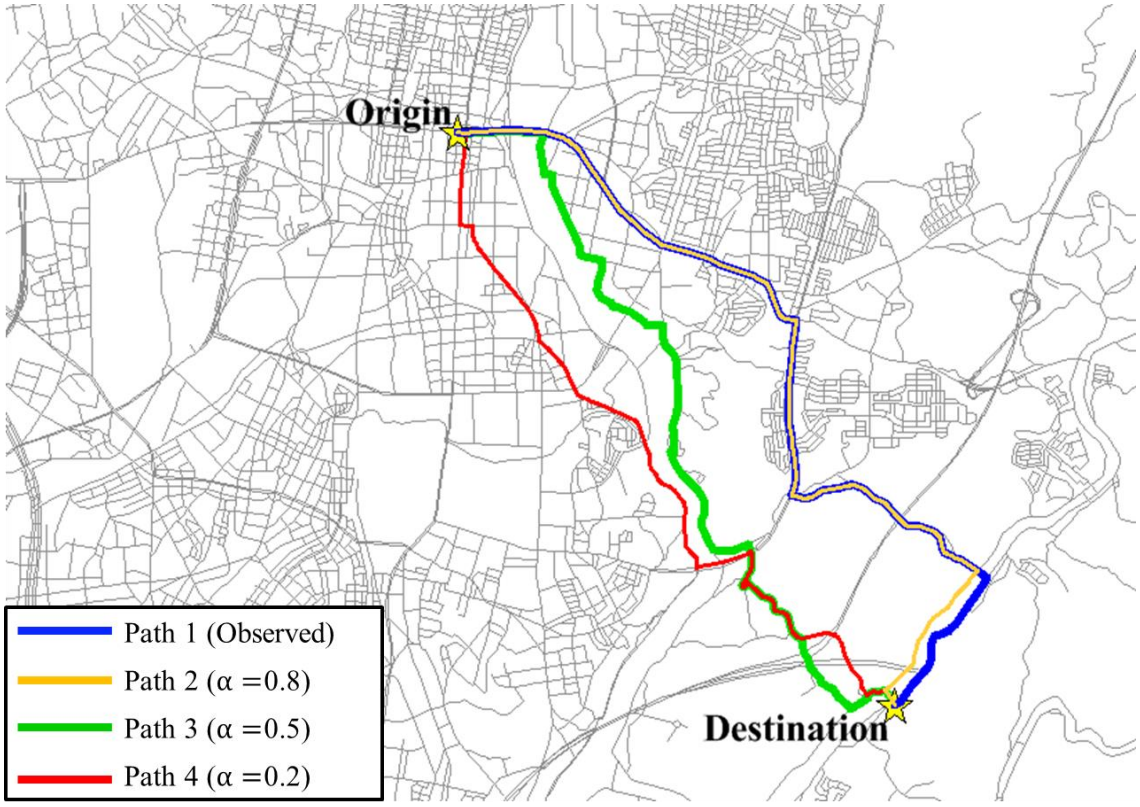


Figure 4-5 Path finding for three settings of on-time arrival probability

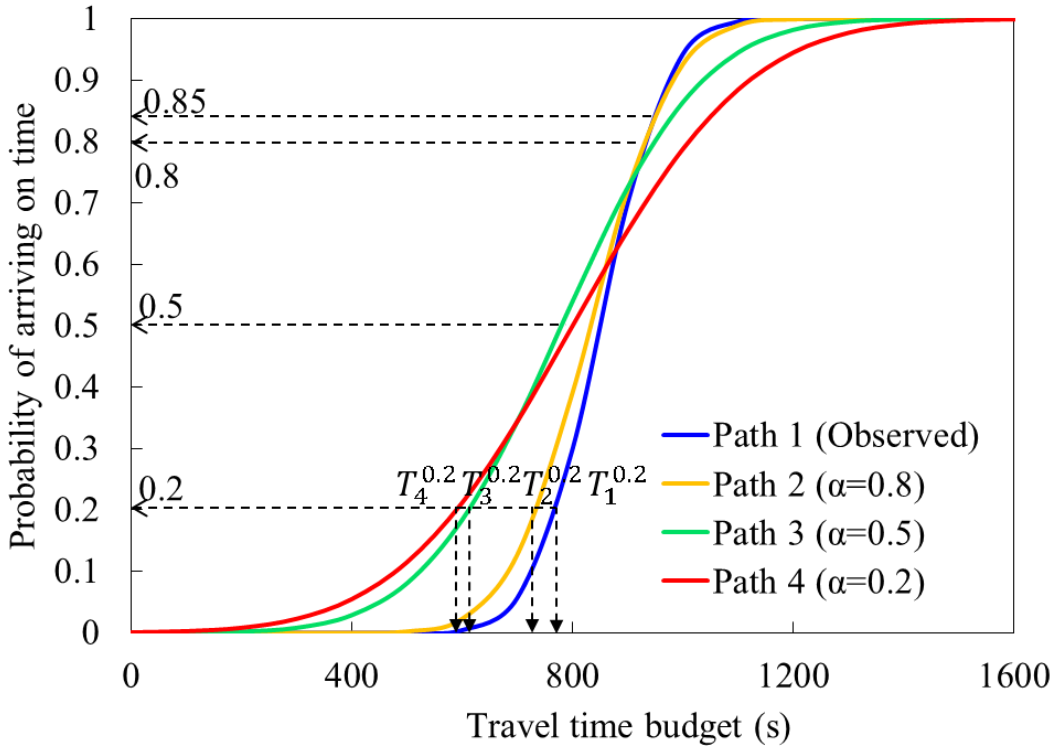


Figure 4-6 On-time arrival probabilities of four paths with different time budgets

4.5 Summary

This study discussed the α-reliable path problem in a stochastic network with correlated and truncated lognormal link travel times. The Lagrangian relaxation approach is applied to solve the nonlinear and non-additive problem. The availability of such reliable paths in a navigation system application would help travelers plan their travel time budgets with a given on-time arrival probability efficiently. The proposed α-reliable path-finding algorithm was applied to a large-scale real-world network in Toyota city, Japan. The performance of the proposed method was shown to be accurate within a reasonable computation time. The major achievements of this study are summarized as follows.

(1) The spatial correlation of link travel times is explicitly considered by introducing a correlation coefficient matrix. The Cholesky decomposition is proposed to separate the correlation coefficient matrix and make it tractable to the α-reliable path problem.

(2) The Lagrangian relaxation approach is applied to estimate the lower bound of the α-reliable path solution. Specifically, the nonlinear and non-additive problem structure is decomposed into sub-problems that can be solved as standard shortest-path problems and series of tractable convex or concave problems. In solving the problem, the relative gap between the upper and lower bound of the solution is shown to decrease at every iteration and 30 iterations of the algorithm yield a small relative gap of within 2%-7%.

(3) A case study using the probe vehicle data is carried out with the assumption of pre-trip navigation planning under different reliability settings. This demonstrates that the proposed α-reliable path finding algorithm enables the decision-making strategies of risk-taking, risk-neutral and risk-averse travelers to be modeled.

Potential directions for further study include extending the proposed algorithm, improvement of path travel time estimation, and link travel time correlation estimation: (1) α-reliable path finding could be extended to a time-varying stochastic network; (2) a more accurate approximation method for path travel time distribution estimation could be developed by considering the skewness characteristic; (3) in addition to spatial link travel time correlation, temporal correlation should be considered in the finding algorithm and (4) a faster path finding algorithm suitable for real-time navigation systems needs to be developed.

4.6 References

Ahuja, R. K., Magnanti, T. L., Orlin, J. B., 1993. Network flows: theory, algorithms, and

applications, Prentice-Hall, Upper Saddle River, NJ, 598-648.

Bates, J., Polak, J., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. Transportation Research Part E: Logistics and Transportation Review, 37(2), 191-229.

Chen, A., Ji, Z., 2005. Path finding under uncertainty. Journal of advanced transportation, 39(1), 19-37.

Chen, B. Y., Lam, W.H.K., Sumalee, A., Li, Q., Shao, H., Fang, Z., 2013. Finding reliable shortest paths in road networks under uncertainty. Networks and spatial economics, 13(2), 123-148.

Fisher, M. L., 1981. The Lagrangian Relaxation Method for Solving Integer Programming Problems. Management Science, 27(1), 1-18.

Golub, G. H., Loan, C. V., 1989. Matrix computations. Johns Hopkins series in the mathematical sciences.

Guignard, M., Kim, S., 1987. Lagrangean decomposition: a model yielding stronger Lagrangean bounds. Mathematical programming, 39(2), 215-228.

Handler, G. Y., Zang, I., 1980. A dual algorithm for the constrained shortest path problem. Networks, 10(4), 293-309.

Held, M., Wolfe, P., Crowder, H. P., 1974. Validation of subgradient optimization. Mathematical programming, 6(1), 62-88.

Jornsten, K., Nasberg, M., 1986. A new Lagrangian relaxation approach to the generalized assignment problem. European Journal of Operational Research, 27(3), 313-323.

Lam, T. C., 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), 231-251.

Ng, M., Kockelman, K., Waller, S., 2010. Relaxing the multivariate normality assumption in the simulation of transportation system dependencies: an old technique in a new domain. Transportation Letters, 2(2), 63-74.

Sivakumar, R. A., Batta, R., 1994. The variance-constrained shortest path problem. Transportation Science, 28(4), 309-316.

Xing, T., Zhou, X., 2011. Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach. Transportation Research Part B: Methodological, 45(10), 1660-1679.

Chapter 5

MODELING TRAVELER'S RISK PREFERENCE TO TRAVEL TIME RELIABILITY

5.1 Introduction

Travel time reliability has been regarded as an important factor in traveler's route choice decisions. In general, learning travelers' risk preferences requires repeated interaction with the travelers. Although revealed or stated preference surveys are possible to obtain user preferences directly, data collection and analysis for large-scale SP and RP surveys are time-consuming and expensive. This study explores travelers' risk preferences to travel time from GPS trip data.

Methods of finding the reliable path have attracted increasing attention for risk-averse navigation. In developing a better trip plan, the most reliable shortest path (Nie and Wu, 2009) within a given travel time budget seems more attractive than the shortest path in a stochastic network. However, travelers could exhibit different route choice behaviors when face with the stochastic transportation network depending on their risk preferences. Risk-averse travelers are likely to choose a more reliable path with lower travel time variance. And risk-taking travelers are likely to prefer an unreliable path that potentially provides a shortest distance or minimum travel time (Ben-Elia et al, 2013). To our knowledge, most of the current studies about the most reliable shortest path problem assumed that travelers

are rational and homogeneous. The varied nature of travelers' risk-averse preferences is seldom taken into account. Alternatively, Chen and Ji (2005) proposed the concept of the α-reliable path that allows travelers to specify a confidence level α for finding a reliable path with the minimum travel time budget. The confidence level α can be regarded as a surrogate index of travelers' risk-averse preferences. The α-reliable path definition requires the travelers to express their expected risk-averse preference toward travel time uncertainty (Chen et al, 2013). However, travelers may feel confused to define a suitable confidence level α prior to their trips without any references or default value provided by the navigation system. However, if the traveler does not have an idea of how to specify the α value, the α reliable-path based navigation system may provide a sub-optimal or dissatisfactory path to the traveler by using a random α value.

A reasonable approach in improving the automatic routing for risk-averse navigation is to incorporate travelers' risk-averse preferences into the routing processes. This study aims to acquire knowledge on travelers' risk-averse preferences by learning repeated route choice behavior from large-scale trip records of probe vehicles. The probe vehicles with GPS devices can effectively provide detailed travel information on start and end time, precise observation of route, and OD information. With the investigation of the travelers' personal information, an ordered probit model is applied to learn and predict the travelers' risk preferences by considering various individual properties (gender, age) and pre-trip information (OD distance, departure time).

Route choice data for travelers' preference analysis can be collected by using revealed or stated preference surveys or through GPS record. Using stated preference, Abdel-Aty et al. (1995) investigated the effect of travel time variability on route choice. The results indicated the significance of the degree of travel time reliability on decision-making process for route selection. Small et al. (2005) identified the varied nature of travels' preferences for both travel time and reliability by combining both stated preference (SP) and reveal preference (RP) survey on the observed behavior of commuters. The study indicated that travelers had substantial heterogeneity in their preferences of travel time reliability. However, both SP and RP survey have limitations. For example, respondents have to assume choice set instead of experiencing the route choice practically in SP survey. They may simply answer questions that they would not realistically pursue. Though RP survey could reflect the decision-making preference more realistically, few studies use RP data investigating travel time reliability because almost no examples of the investigation settings with significant difference of the variation of travel time across at least two choices (Carrion and Levinson, 2012). On the other hand, data collection and analysis for SP and RP surveys are time-consuming and expensive. Alternatively, advancements in traffic information collection technologies, including trip records from probe vehicles, can

facilitate the investigation of influences dominates route choice decisions. Li et al. (2005) showed how GPS enables to record observed route choice information effectively by collecting 182 travelers over a 10 day period. Papinski et al. (2009) explored the decision-making process of route choice by comparing the observed and planned routes obtained from personal based GPS. Papinski and Scott (2011) developed a GIS-based toolkit for route choice analysis based on GPS data. Even though traffic information such as travel time and average speed can be collected in an economical and timely way, concerns of GPS data include the device reliability, respondent accuracy, and urban canyon effects (Wolf and Thompson, 2003). To fill this gap, map-matching technology for GPS points can be regarded as a powerful tool to provide the precise route in terms of junctions and links. The map-matching algorithm of Miwa et al. (2012) resulted in an 80% correct identification of routes traveled.

A practical way to discover users' preferences is to find regularities by observing substantial route choice behavior. Previous studies showed that both discrete choice models based on random utility theory (Ben-Akiva and Lerman, 1985) and machine learning (Dougherty, 1995; Yamamoto et al., 2002) are available to reveal the route choice preferences. Machine learning requires neither estimating any parameters describing the distribution of variables nor assuming any particular form of functions (Yamamoto et al., 2002). Park et al (2007) developed a decision tree model in the C4.5 algorithm to learn user preferences of route choice behavior by considering travel distance, travel time, travel time reliability, and so on. They concluded that the decision tree learning algorithm can outperform random utility models.

In summary, there is growing interest in analyzing route choice behavior by using the emerging techniques such as probe vehicle. However, few studies discuss how to quantify travelers' risk-averse preferences. The measure of risk-averse preference to travel time reliability is quite new to the field of reliable path finding problem for risk-averse navigation. This study attempts to address this gap by proposing a new method for measuring the degree of risk-averse preference and applying ordered probit model to learn travelers' risk preferences to travel time reliability from large-scale GPS trip records.

5.2 Problem statement

This section postulates that a rational traveler chooses the α-reliable shortest path based on the desired risk-averse attitude. Given an on-time arrival probability α, a path $p_{rs}^{*} \in P_{rs}$ is defined as the α-reliable shortest path if $\Phi_{T_{rs}^{*}}^{-1}(\alpha) < \Phi_{T_{rs}}^{-1}(\alpha)$ for any other path $p_{rs} \in P_{rs}$ (Chapter 4). P_{rs} is the path set from origin r to destination s. $\Phi_{T_{rs}^{*}}^{-1}(\alpha)$ and

$\Phi_{T_{rs}}^{-1}(\alpha)$ are the inverse cumulative distribution functions (CDF) of travel time for paths p_{rs} and p_{rs}^{*}, respectively. It is assumed that path travel time follow normal distribution. Consequently, the α-reliable path problem can be described as below:

$$\Phi_{T_{rs}^{*}}^{-1}(\alpha) = \text{Min}\{u_p + Z_\alpha \sigma_p\} \tag{5-1}$$

where Z_α is the inverse CDF of the standard normal distribution at α confidence level, u_p is the path mean travel time, and σ_p is the standard deviation of path travel time. To find the α-reliable shortest path, it is necessary to specify the confidence level (α) by the traveler. Here, the variable α can be explained as the on-time arrival probability of a certain travel time budget. Eq.(5-1) is the objective function that travelers want to minimize the travel time on a given on-time arrival probability.

Definition 1: The degree of traveler's risk-averse preference (DR) toward travel time is defined as the value of on-time arrival probability of the selected path under a specified travel time budget.

As shown in Figure 5-1, assumed that the travel time budget is 800s, the DR is 0.94 if Path 2 is chosen. Noted that the DR is related to the travel time distribution and the specified travel time budget, it is necessary to understand the path travel time distribution and the individual travel time budget. The distribution of path travel time can be estimated from the observed dataset. However, it is quite difficult to get the travel time budget exactly through the observed trip travel time. Actually, the travel time budget is usually larger than the actual travel time, but it can not be derived without asking the traveler. Such problem makes the estimation of DR intractable. Alternatively, the bound value of DR can be estimated by comparing a series of α-reliable paths if travelers are rational (never choose the inferior paths dominated by other paths).

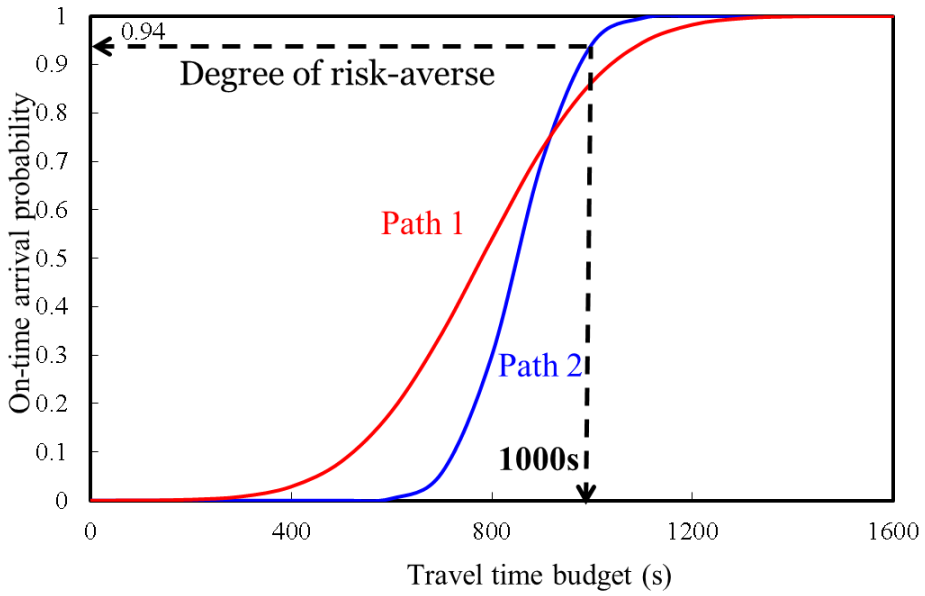


Figure 5-1 Illustration of degree of risk-averse preference

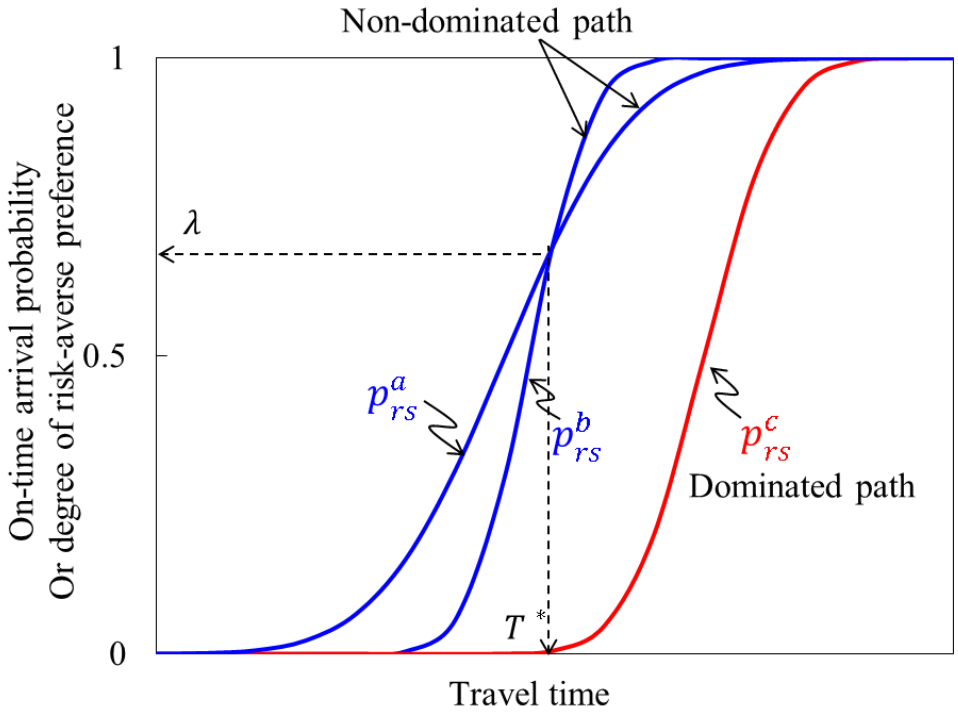


Figure 5-2 Stochastic dominance

5.3 Methodology for data collection

In this study, we propose a novel method to collect the travelers' risk-averse preferences by applying the method of stochastic dominance.

Definition 2 (First order stochastic dominance): given two paths $p_{rs}^a \neq p_{rs}^b \in \boldsymbol{P}_{rs}$ where r is origin and s is destination, p_{rs}^a dominates p_{rs}^b (denoted by $p_{rs}^a \succ p_{rs}^b$) if $\Phi_{T_{rs}^a}^{-1}(\alpha) < \Phi_{T_{rs}^b}^{-1}(\alpha)$ for any confidence level $0 < \alpha < 1$.

Definition 3 (Mean-Variance dominance): Given an on-time arrival probability α and two paths ($p_{rs}^a \neq p_{rs}^b \in \boldsymbol{P}_{rs}$), $p_{rs}^a \succ p_{rs}^b$ if p_{rs}^a and p_{rs}^b satisfy one of the following conditions:

(1) $u_a \leq u_b$ and $Z_\alpha \sigma_a < Z_\alpha \sigma_b$ or

(2) $u_a < u_b$ and $Z_\alpha \sigma_a \leq Z_\alpha \sigma_b$

where

u_a, u_b: Path travel time mean for p_{rs}^a and p_{rs}^b, respectively.

σ_a, σ_b: Standard deviation of path travel time for p_{rs}^a and p_{rs}^b, respectively.

Z_α: Inverse CDF of the standard normal distribution at α confidence level.

Figure 5-2 shows an example of stochastic dominance. Path p_{rs}^a and path p_{rs}^b dominate path p_{rs}^c, but path p_{rs}^b is not dominated by path p_{rs}^a for any confidence level $0 < \alpha < 1$. p_{rs}^a and path p_{rs}^b are called the non-dominated paths (NDP), and path p_{rs}^c is called the dominated path or inferior path. A rational traveler will never choose path p_{rs}^c as the preferable path. There is a cross point (T^*, λ) between the CDF curves of the two NDPs. If the degree of traveler's risk-averse preference (α) is less than λ ($0 < \alpha < \lambda$), then path p_{rs}^a is selected, while path p_{rs}^b is selected if the degree of traveler's risk-averse preference is more than λ ($\lambda < \alpha < 1$). Consequently, the chosen paths can reflect the travelers' risk-averse preferences.

Both first order stochastic dominance (FSD) and Mean-Variance dominance (MVD) can be used to identify the non-dominated path between two paths, but it is difficult to apply FSD to identify the non-dominated path in a path set (more than three paths). Figure 5-3 shows an example of the limitation of FSD in a path set. According to FSD, Path 2 dominates Path 1 if $\alpha < 0.09$ and Path 2 dominates Path 3 if $\alpha > 0.16$. That is, Path 2 is a non-dominated path with respect to Path 1 or Path 3. However, it is also found that Path 2 is dominated by Path 1 if $\alpha > 0.09$ and it is dominated by Path 3 if $\alpha < 0.16$. That is, Path 2 is dominated by Path 1 or Path 3 in the whole range of α ($0 \leq \alpha \leq 1$). Consequently, a rational traveler will not choose Path 2 regardless of the degree of desired risk-averse preference.

Definition 4 (Non-dominated path): A path $p_{rs}^{a} \in \boldsymbol{P}_{rs}$ is a non-dominated path, if and only if p_{rs}^{a} dominates any paths $\forall p_{rs}^{i} \in \boldsymbol{P}_{rs}$ at certain confidence level $0 < \lambda_1 < \alpha < \lambda_2 < 1$.

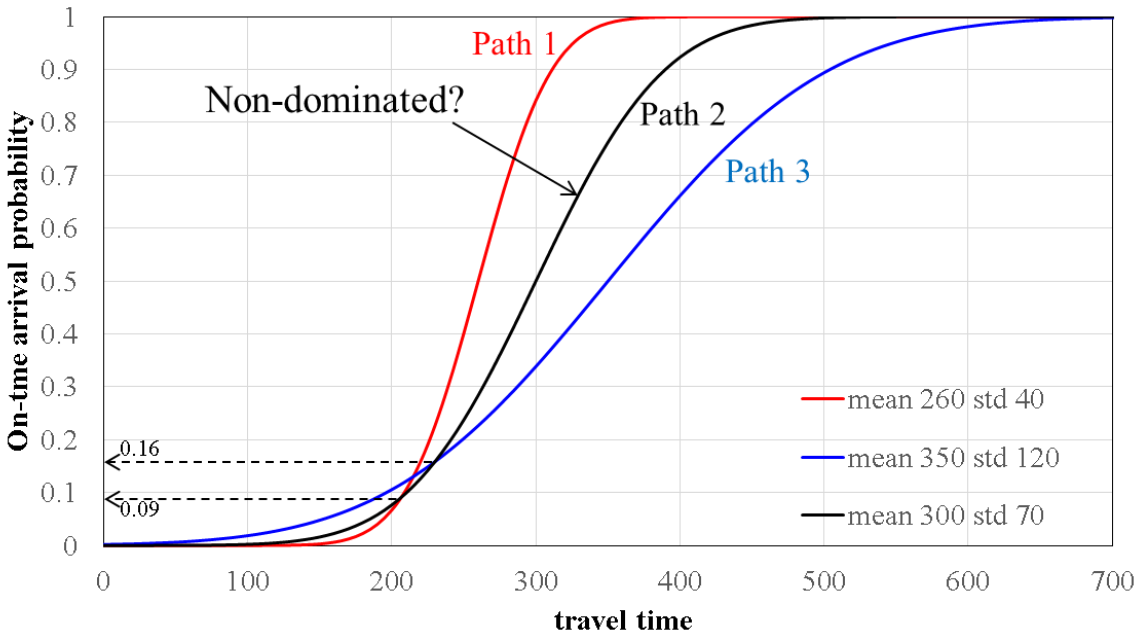


Figure 5-3 Limitation of FSD

Here, we use **Definition 4** to judge whether the observed path is non-dominated path in a path set. Since the α-reliable paths are the non-dominated paths between OD nodes based on **Definition 4**, we don't need to generate all paths between OD nodes. Instead, the α-reliable paths can be used as the path set ($\boldsymbol{p}_{rs}^{\alpha}$). And then, Mean-Variance dominance is used to judge whether the observed path is dominated by the α-reliable paths.

First, we check whether the observed path p_{rs}^{obs} (with mean travel time u_{obs} and standard deviation of travel time σ_{obs}) is dominated by any of the α-reliable paths ($\forall p_{rs}^{\alpha} \in \boldsymbol{p}_{rs}^{\alpha}$) in risk-averse condition ($\alpha > 0.5$). Since $\alpha > 0.5$, $Z_{\alpha} > 0$, we need to check whether any α-reliable paths (with mean travel time u_{α} and standard deviation of travel time σ_{α}) exist satisfy: (1) $u_{\alpha} \le u_{obs}$ and $\sigma_{\alpha} < \sigma_{obs}$ or (2) $u_{\alpha} < u_{obs}$ and $\sigma_{\alpha} \le \sigma_{obs}$. If any of the two conditions satisfy, the observed path is regarded as the dominated path in risk-averse condition.

Algorithm 5-1 Test algorithm for non-dominated path

1	**Step 1**: Generate the α-reliable path set: $\boldsymbol{p}_{rs}^{\alpha}$
2	**Step 2**: Check whether the observed path (p_{rs}^{obs}) is dominated at risk-averse condition
3	Check Mean-Variance dominance for any of the α-reliable paths ($\forall p_{rs}^{\alpha} \in \boldsymbol{p}_{rs}^{\alpha}$):

4	(1) $u_\alpha \leq u_{obs}$ and $\sigma_\alpha < \sigma_{obs}$ or (2) $u_\alpha < u_{obs}$ and $\sigma_\alpha \leq \sigma_{obs}$
5	**Step 3**: Check whether the observed path (p_{rs}^{obs}) is dominated at risk-taking condition
6	Check Mean-Variance dominance for any of α-reliable paths ($\forall p_{rs}^{\alpha} \in \boldsymbol{p}_{rs}^{\alpha}$):
7	(1) $u_\alpha \leq u_{obs}$ and $\sigma_\alpha > \sigma_{obs}$ or (2) $u_\alpha < u_{obs}$ and $\sigma_\alpha \geq \sigma_{obs}$
8	**Step 4**: Determine the non-dominated path
9	Dominated path: satisfy step 2 or step 3
10	Non-dominated path: not satisfy step 2 and step 3

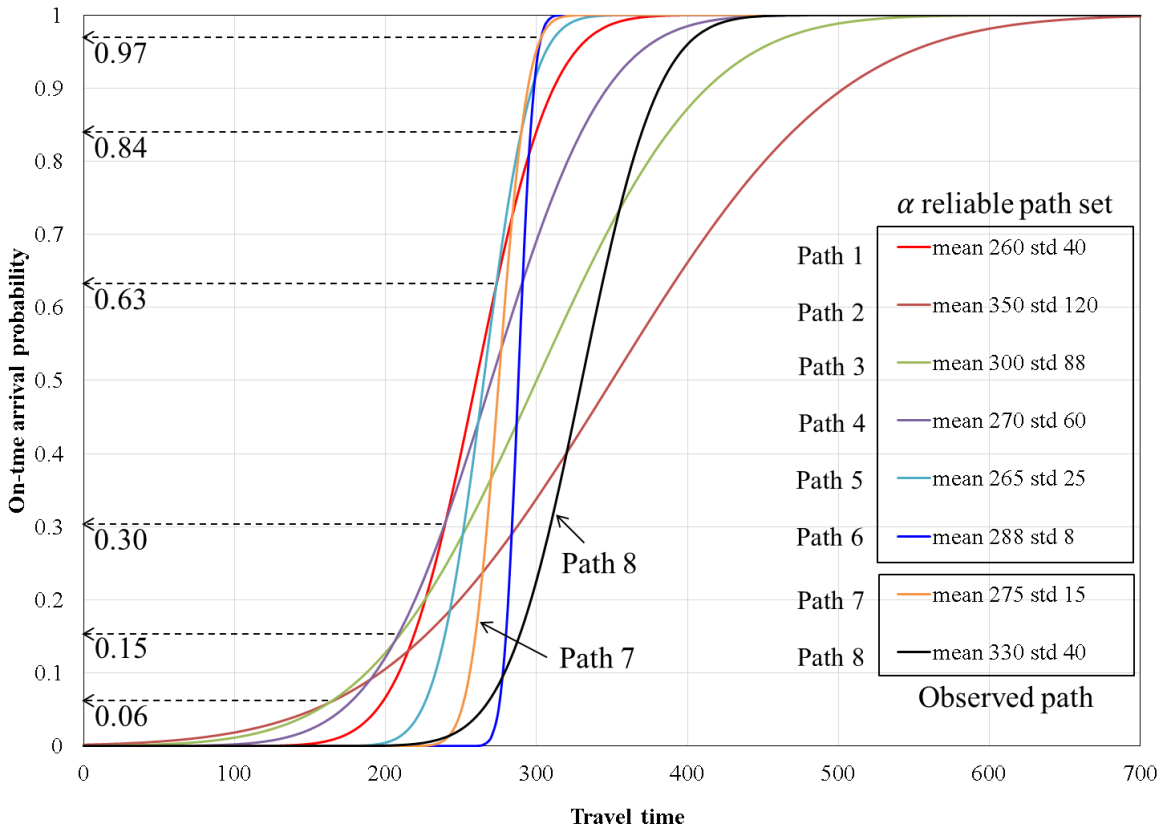


Figure 5-4 Illustration of determining the observed non-dominated path

Second, we check whether the observed path p_{rs}^{obs} is dominated by any of the α-reliable paths ($\forall p_{rs}^{\alpha} \in \boldsymbol{p}_{rs}^{\alpha}$) in risk-taking condition ($\alpha < 0.5$). Since $\alpha < 0.5$, $Z_\alpha < 0$, we need to check whether any α-reliable paths exist satisfy: (1) $u_\alpha \leq u_{obs}$ and $\sigma_\alpha > \sigma_{obs}$ or (2) $u_\alpha < u_{obs}$ and $\sigma_\alpha \geq \sigma_{obs}$. If any of the two conditions satisfy, the observed path is regarded as the dominated path in risk-taking condition.

According to Mean-Variance dominance, the observed path is a dominated path if it is dominated by any of the α-reliable paths in risk-averse condition or risk-taking condition. That is, the observed path is the non-dominated path if it is not dominated by any of the α-

reliable paths in risk-averse condition and risk-taking condition. Based on the discussion above, pseudo code for determining the non-dominated path is given in Algorithm 5-1.

An example of determining the observed non-dominated path by using Algorithm 5-1 is illustrated in Figure 5-4. The α-reliable path set includes six paths and two observed paths (i.e., Path 7 and Path 8) need to test. Let's test Path 7 first. Because the mean travel time of Path 7 is smaller than that of Path 2, Path 3 and Path 6, Path 7 is not dominated by the three paths. For risk-averse condition, we select the α-reliable path (among Path 1, Path 4 and Path 5) with minimum travel time variance as the candidate (i.e., Path 5). Because $\sigma_5 > \sigma_7$ which not satisfies the Mean-Variance dominance, Path 7 is not dominated by Path 5. Because the travel time variances of Path 1 and Path 4 are larger than that of Path 5, it can be concluded that Path 7 is not dominated by any of α-reliable paths in risk-averse condition. Since the observed path is not dominated in risk-averse condition, it is not necessary to test it in risk-taking condition. Therefore, Path 7 is the non-dominated path. Another example is Path 8. Since the mean travel time of Path 8 is smaller than that of Path 2, Path 8 is not dominated by Path 2. For risk-averse condition, we select Path 6 which has the minimum travel time variance in the candidate set for test. It is found that $\sigma_6 < \sigma_8$ which satisfies the second condition of Mean-Variance dominance. Therefore, Path 8 is dominated by Path 6 in risk-averse condition. Without considering the risk-averse condition, it can be concluded that Path 8 is not the non-dominated path.

Assumed that travelers are rational (they never choose the dominated paths), we extract the non-dominated observed paths by using Algorithm 5-1. Those dominated observed paths will be excluded for risk-averse preference estimation. Because the α-reliable paths and the observed path are non-dominated path, the cross point of the CDF curves between the non-dominated observed path and each α-reliable path must exist. And the value of cross point determines boundary of the dominance condition between the two tested paths. The bound value of the risk-averse preference can be estimated by checking the cross points. The degree of risk-averse preference (λ) for each cross point can be formulated as follows:

$$\lambda = \Phi\left(\frac{u_{obs}\sigma_\alpha - u_\alpha\sigma_{obs}}{u_{obs}(\sigma_\alpha - \sigma_{obs})} - 1\right) \tag{5-2}$$

where

u_α, u_{obs}: mean path travel times for the α-reliable path and the observed path, respectively.

σ_α, σ_{obs}: standard deviations of path travel time for the α-reliable path and observed path, respectively.

Φ: Cumulative density function (CDF) of the standard normal distribution.

The lower bound (λ_{LB}) and upper bound (λ_{UB}) of the risk-averse preference can be determined by comparing the observed path with the α-reliable paths. Obviously, the observed path dominates the α-reliable path with larger travel time variance (i.e., $\sigma_\alpha > \sigma_{obs}$) if $\alpha' > \lambda$ where α' is the desired risk-averse preference. Therefore, the maximum value of λ can be regarded as the lower bound of the observed risk-averse preference.

$$\lambda_{LB} = \text{Max}\left\{\Phi\left(\frac{u_{obs}\sigma_\alpha - u_\alpha\sigma_{obs}}{u_{obs}(\sigma_\alpha - \sigma_{obs})} - 1\right)\right\}, if\ \sigma_\alpha > \sigma_{obs} \quad (5\text{-}3)$$

Similarly, the observed path dominates the α-reliable path with smaller travel time variance (i.e., $\sigma_\alpha < \sigma_{obs}$) if $\alpha' < \lambda$. Therefore, the minimum value of λ can be regarded as the upper bound of the observed risk-averse preference.

$$\lambda_{UB} = \text{Min}\left\{\Phi\left(\frac{u_{obs}\sigma_\alpha - u_\alpha\sigma_{obs}}{u_{obs}(\sigma_\alpha - \sigma_{obs})} - 1\right)\right\}, if\ \sigma_\alpha < \sigma_{obs} \quad (5\text{-}4)$$

The observed path dominates all the α-reliable paths if the desired risk-averse preference is set as $\lambda_{LB} < \alpha' < \lambda_{UB}$. The degree of traveler's risk-averse preference (λ_{DR}) can be approximated by the average of the lower bound and upper bound.

$$\lambda_{DR} = \frac{1}{2}(\lambda_{LB} + \lambda_{UB}) \quad (5\text{-}5)$$

An example of approximating the degree of traveler's risk-averse preference is given in Figure 5-5. To estimate the traveler's risk-averse preference from the observed path, we first calculate cross points with the four α-reliable paths. Because the travel time variances for Path 1, Path 2 and Path 3 are larger than that of the observed path, the lower bound of traveler's risk-averse preference is selected from the cross points of the CDF curves between the observed path and these three α-reliable path. Because the cross points of the CDF curves between the observed path and Path 1 has the maximum value of λ, i.e., $\lambda = 0.63$, the lower bound of traveler's risk-averse preference is 0.63. Because only Path 4 has smaller travel time variance than the observed one, the cross points of the CDF curves between the observed path and Path 4 is used to estimate the upper bound of traveler's risk-averse preference, i.e., $\lambda_{UB} = 0.91$. Consequently, the estimation of traveler's risk-averse preference is $(0.63 + 0.91)/2 = 0.77$.

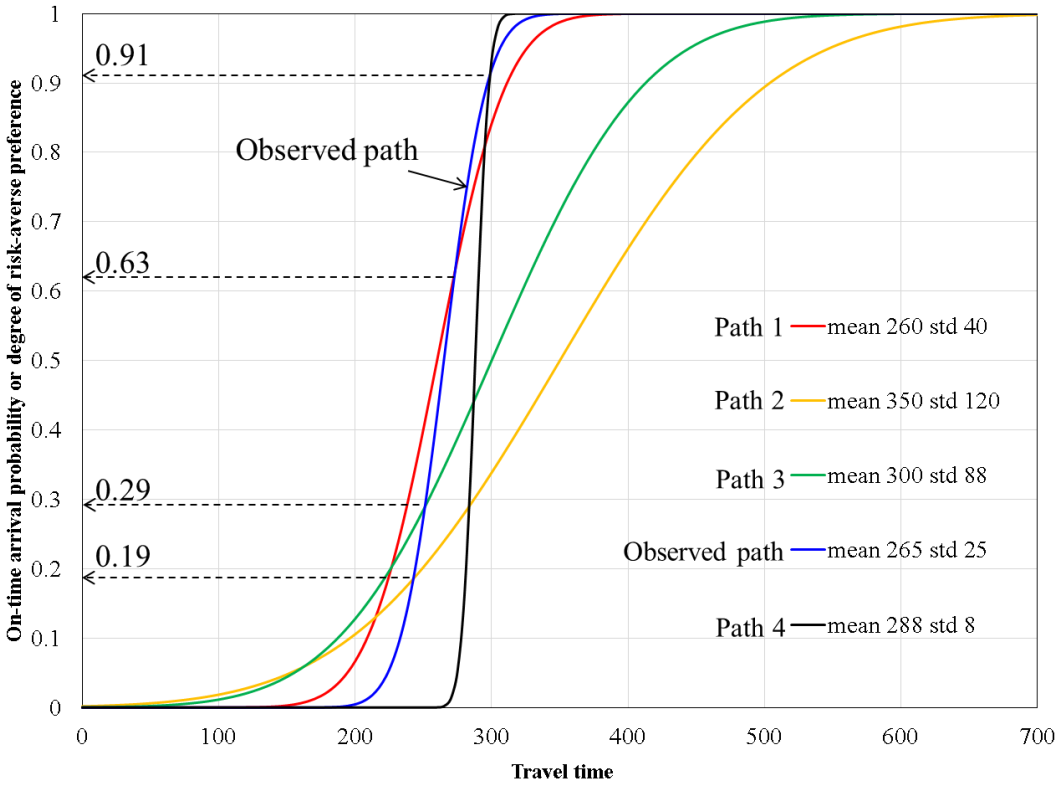


Figure 5-5 Estimation of traveler's risk-averse preference

5.4 Prediction of traveler's risk-averse preference

The risk-averse navigation system requires the input of the travelers' risk-averse attitudes toward the stochastic travel time. To make the navigation system more user-friendly, it is necessary to learn the travelers' risk-averse preferences from their trip records. To simplify the input of the risk attitudes, the estimated degrees of travelers' risk-averse preferences are aggregated to four levels, i.e., high risk-taking ($0 \leq \lambda_{DR} < 0.25$), slight risk-taking ($0.25 \leq \lambda_{DR} < 0.5$), slight risk-averse ($0.5 \leq \lambda_{DR} < 0.75$), and high risk-averse ($0.75 \leq \lambda_{DR} < 1$). In this study, the ordered probit model is applied to predict the level of risk-averse preference with various explanatory variables. The ordered probit model is suitable to estimate the parameters for the level of risk-averse preference with a categorical nature. Furthermore, it is already embed in several commercially statistic software such as STATA, which is convenient to apply to data analysis. The general model specification is shown as follows:

$$y_i^* = \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i \tag{5-6}$$

where y_i^* is a latent variable expressing the surrogate degree of the i^{th} traveler's risk-averse preference; $\boldsymbol{x}_i$ denotes the vector of observed explanatory variables; $\boldsymbol{\beta}$ is the

parameter vector; ε_i is the random error term, which is assumed to follow a normal distribution with zero mean and unit variance.

For any trips, it is reasonable to assume that a high surrogate degree of risk-averse preference, i.e., y_i^*, will represent a high level of risk-averse preference, i.e., y_i. Therefore, the coded discrete level of risk-averse preference, y_i, is determined as follows:

$$y_i = \begin{cases} 1\ if -\infty \le y_i^* < u_1\ \ (high\ risk - taking) \\ 2\ \ if\ u_1 \le y_i^* < u_2\ \ (slight\ risk - taking) \\ 3\ \ if\ u_2 \le y_i^* < u_3\ \ (slight\ risk - averse) \\ 4\ \ if\ u_3 \le y_i^* < +\infty\ \ (hight\ risk - averse) \end{cases} \tag{5-7}$$

where the u_i denote thresholds to be estimated.

The probabilities associated with y_i are shown as follows:

$$P_i(1) = Pr(y_i = 1) = Pr(y_i^* < u_1) = Pr(x_i\beta + \varepsilon_i < u_1)$$

$$= Pr(\varepsilon_i < u_1 - x_i\beta) = \phi(u_1 - x_i\beta) \tag{5-8}$$

$$P_i(2) = Pr(y_i = 2) = Pr(u_1 < y_i^* \le u_2)$$

$$= Pr(\varepsilon_i \le u_2 - x_i\beta) - Pr(\varepsilon_i \le u_1 - x_i\beta)$$

$$= \phi(u_2 - x_i\beta) - \phi(u_1 - x_i\beta) \tag{5-9}$$

$$P_i(k) = Pr(y_i = k) = Pr(u_{k-1} < y_i^* \le u_k)$$

$$= Pr(\varepsilon_i \le u_k - x_i\beta) - Pr(\varepsilon_i \le u_{k-1} - x_i\beta)$$

$$= \phi(u_k - \boldsymbol{x}_i\boldsymbol{\beta}) - \phi(u_{k-1} - \boldsymbol{x}_i\boldsymbol{\beta}) \tag{5-10}$$

$$P_i(K) = Pr(y_i = K) = Pr(u_k < y_i^*) = 1 - \phi(u_k - \boldsymbol{x}_i\boldsymbol{\beta}) \tag{5-11}$$

Where i is a traveler, k is a response level of risk-averse preference, $Pr(y_i = k)$ is the probability that the traveler i responses to k, and $\phi(*)$ is the standard normal cumulative distribution function.

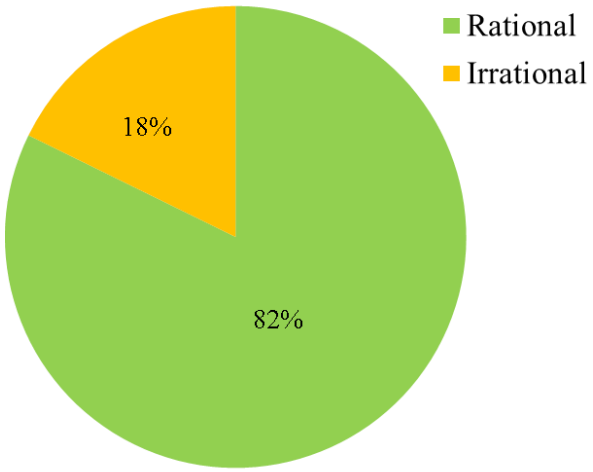


Figure 5-6 Proportion of rational traveler

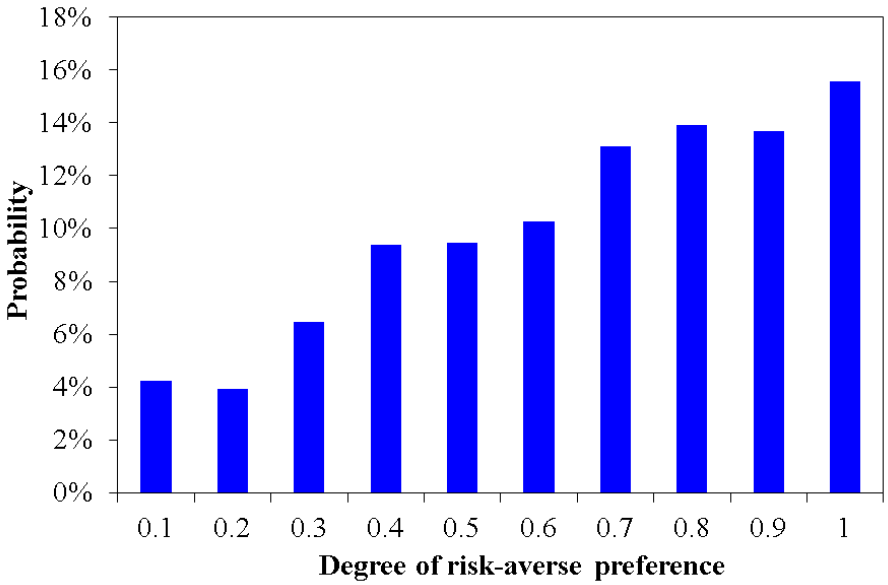


Figure 5-7 Distribution of degree of risk-averse preference

5.5 Data analysis

The GPS data in this study are collected from 153 probe vehicles in Toyota City, Japan. After map-matching (Miwa et al, 2012) and some basic data cleaning work, 4032 trip records with different OD pairs are obtained for one month (2011/3). The observed lower bound or upper bound of the risk-averse preference is collected by using the theory of stochastic dominance. The risk-taking travelers are willing to take risk of delay to pursue the minimum travel time budget while the risk-averse travelers try to minimize the risk of delay by increasing his/her travel time budget.

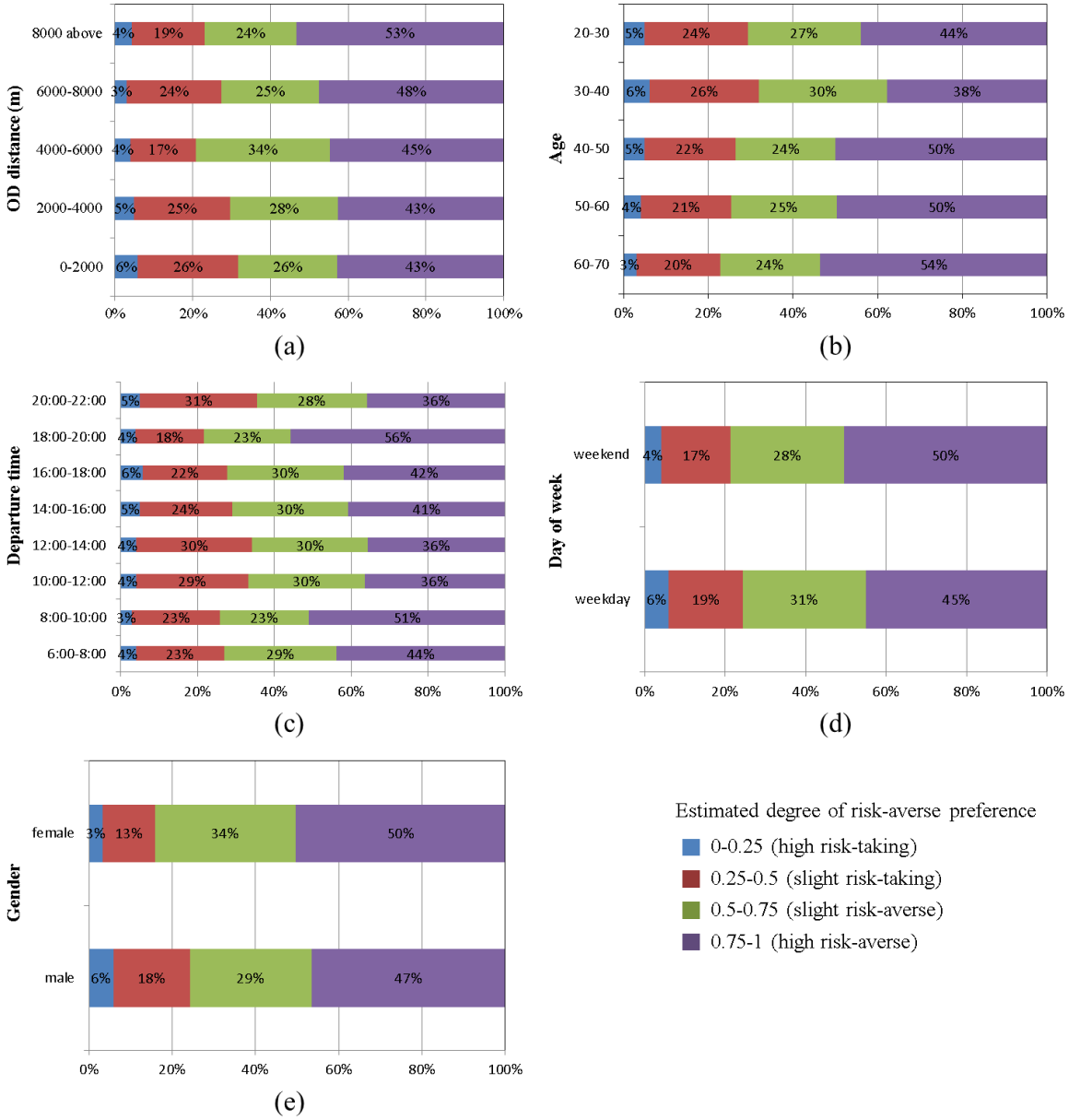


Figure 5-8 Statistic of degree of risk-averse preference in different categories

As shown in Figure 5-6, 82% of the trips are rational, who selected non-dominated paths. It indicates that most of the local drivers are familiar with the traffic condition and the road network. Figure 5-7 shows the overall distribution of degree of risk-averse preference for all the trips. It is found that the frequency increases as the degree of risk-averse preference increases. It indicates that travelers prefer the risk-averse routes in most of the trips.

Figure 5-8 shows the statistic of degree of risk-averse preference in different categories. As shown in Figure 5-8 (a), it is found that as the OD distance increases, the proportion of high risk-averse preference increases. It indicates travelers have higher risk-

averse preference when they plan a longer trip. Figure 5-8 (b) shows that 54% of the aged people (age>60) prefer high risk-averse routes, while only 44% of the young people (age<30) prefer the high risk-averse routes. About 5% of the high risk-taking routes are chosen by young people, while only 3% of the high risk-taking routes are chosen by the aged people. It indicates that aged people have higher risk-averse attitude. One possible reason is that the aged people have more driving experience than the young people. Figure 5-8 (c) shows the risk attitudes in different departure time. It is found that people are more risk-averse in peak hours (8:00-10:00 and 18:00-20:00). The risk-averse attitudes decrease and the risk-taking attitudes increase after the peak hours. Figure 5-8 (d) shows that more people prefer the risk-averse routes in weekend. Figure 5-8 (e) shows that female is more risk-averse than the male.

Table 5-1 Parameter estimation results of ordered probit model

Risk-averse level	Description	Coef.	z	$p > \|z\|$
OD distance	The straight distance between origin and destination	0.007	2.05	0.04
Age	Age of driver (year)	0.001	2.33	0.02
Departure time	Departure time (off-peak hours=0, peak hours=1)	0.174	2.51	0.01
Day of week	Day of week (weekday=0, weekend=1)	0.245	7.25	0.00
Gender	Gender (male=0, female=1)	0.067	1.98	0.05
Thresholds				
u_1	-1.08			
u_2	-0.18			
u_3	0.70			
Sample size	4032			
Log likelihood	-5522			

5.6 Model results

We estimate the parameters in the ordered probit model by using maximum likelihood estimation (MLE). The explanatory variables include the OD distance, age, departure time, day of week and gender. It should be noted that the departure time, day of week and gender are dummy variables. The estimation results are shown in Table 5-1 shows the coefficients, z-tests and their associated p-value. All of the explanatory variables are statistically significant at 95% confidence level. For OD distance, we would expect that a 0.007 increase in the logarithm odds of being in a higher level of risk-averse preference. For a one unit increase in age, we would expect a 0.001 increase in the logarithm odds of being

in a higher risk-averse level, given that all of the other variables in the model are held constant. Similarly, we would expect a 0.174, 0.245, and 0.067 increases in the logarithm odds of being in a higher risk-averse level for departure time, day of week, and gender, respectively. The threshold shown at the bottom of the Table 5-1 indicates where y_i^* is cut to make the four groups that observed in our dataset.

5.7 Summary

This section investigated the traveler's risk-averse preference prediction for α-reliable shortest path problem in stochastic network. A novel data collection methodology for travelers' risk-averse preferences is introduced. The observed lower bound or upper bound of the risk-averse preference is collected by using the theory of stochastic dominance. Ordered probit model is applied to learn and predict the travelers' risk preferences by considering variously individual properties (gender, age) and pre-trip information (OD distance, departure time, day of week). The parameter estimation results show that the ordered probit model enables to explain how the explanatory variables influence the level of risk-averse preference.

In this study, only the GPS data and basic personal information were used to estimate the traveler's risk-averse preference. In the future, we need to consider more influences such as traffic condition, trip purpose and whether condition so as to improve the accuracy of the proposed estimation models.

5.8 References

Abdel-Aty, M. A., Kitamura, R., & Jovanis, P. P. (1995). "Investigating Effect of Travel Time Variability on Route Choice Using Repeated-measurement Stated Preference data". Transportation Research Record, (1493), 39-45.

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand (Vol. 9). MIT press.

Ben-Elia, E., Di Pace, R., Bifulco, G. N., & Shiftan, Y. (2013). "The Impact of Travel Information's Accuracy on Route-choice". Transportation Research Part C: Emerging Technologies, 26, 146-159.

Carrion, C., & Levinson, D. (2012). "Value of Travel Time Reliability: A review of Current Evidence". Transportation Research Part A: Policy and Practice, 46(4), 720-741.

Chen, A. & Ji, Z., (2005). “Path Finding Under Uncertainty.” Journal of advanced transportation, 39(1), 19-37.

Chen, B. Y., Lam, W. H., Sumalee, A., Li, Q., Shao, H., & Fang, Z., (2013). “Finding Reliable Shortest Paths in Road Networks under Uncertainty”. Networks and spatial economics, 13(2), 123-148.

Dougherty, M. (1995). “A Review of Neural Networks Applied to Transport”. Transportation Research Part C: Emerging Technologies, 3(4), 247-260.

Li, H., Guensler, R., & Ogle, J., (2005). “Analysis of Morning Commute Route Choice Patterns Using Global Positioning System-based Vehicle Activity Data.” Transportation Research Record: Journal of the Transportation Research Board, 1926, 162-170.

Miwa, T., Kiuchi, D., Yamamoto, T., & Morikawa, T., (2012). “Development of Map Matching Algorithm for Low Frequency Probe Data.” Transportation Research Part C: Emerging Technologies, 22, 132-145.

Nie, Y. M., & Wu, X., (2009). “Shortest Path Problem Considering On-time Arrival Probability”. Transportation Research Part B: Methodological, 43(6), 597-613.

Park, K., Bell, M., Kaparias, I., & Bogenberger, K. (2007). “Learning User Preferences of Route Choice Behaviour for Adaptive Route Guidance.” IET Intelligent Transport Systems, 1(2), 159-166.

Papinski, D., Scott, D.M., & Doherty, S.T., (2009). “Exploring the Route Choice Decision-making Process: a Comparison of Planned and Observed Routes Obtained Using Person-based GPS”. Transportation Research Part F: Traffic Psychology and Behaviour, 12, 347–358.

Papinski, D., & Scott, D. M. (2011). “A GIS-based Toolkit for Route Choice Analysis.” Journal of Transport Geography, 19(3), 434-442

Small, K. A., Winston, C., & Yan, J. (2005). “Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability”. Econometrica, 73(4), 1367-1382.

Wolf, J., Oliveira, M., & Thompson, M. (2003). “Impact of Underreporting on Mileage and Travel Time Estimates: Results from Global Positioning System-enhanced Household Travel Survey”. Transportation Research Record: Journal of the Transportation Research Board, 1854(1), 189-198.

Yamamoto, T., Kitamura, R., & Fujii, J. (2002). “Drivers' Route Choice Behavior: Analysis by Data Mining Algorithms.” Transportation Research Record: Journal of the Transportation Research Board, 1807(1), 59-66.

Chapter 6

EXPERIENTIAL ROUTING CONSIDERING DRIVING EXPERIENCE

6.1 Introduction

Finding the fastest and most reliable path for risk-averse navigation is an important application in intelligent transportation system. The uncertainty factors include traffic demand variation, signal control, and physical bottlenecks. These uncertainties may lead to a non-deterministic delay. A good navigation service should consider these aspects properly. Unfortunately, current dynamic navigation system based on real-time traffic information are not mature both in the software and hardware fields (Tang et al., 2010). One of the most concerns is the short-term forecast accuracy on travel time. In an urban transportation network, not all of the link travel times can be collected because most traffic information collection technologies such as automatic vehicle identification and loop detectors cover only limited and specific road sections. Even though probe vehicles with GPS equipment have proven to be an efficient tool for traffic information collection covering a wider area, it is still difficult to guarantee the accuracy and instantaneity. On the other hand, several studies found that the observed paths are usually longer in terms of time than their shortest travel time path alternatives (Papinski and Scott, 2011; Papinski et al., 2009). A possible reason is that the actual travel time distributes stochastically and the estimated travel time has error, which results in the overestimate or underestimate. Another potential explanation is that experienced drivers can find a real shortest path based on their

long-term driving experience. When people choose a route, they usually consider multiple factors such as distance, time-variant traffic flows, road gradient and signals (Yuan et al., 2011; Yuan et al., 2013). These influences can be learned by experienced drivers but they are difficult to be incorporated into current navigation systems. Consequently, those roads selected by drivers with abundant driving experience can provide us with a valuable resource to better plan an optimal path for a new trip.

Therefore, it is promising to improve the performance of current navigation system by integrating the estimated travel time and the driving experience of local drivers. The route-finding problem addressed in this study aims to find an experientially reliable path that minimizes the path travel time by considering the uncertain delay and driving experience. The contribution of our work lies in four aspects.

(1) A path set based on a risk-averse view of uncertain link delays is built by the hyperpath concept, which minimizes the expected arrival time at the destination and all intermediate nodes.

(2) The experientially reliable path is generated from the hyperpath where the modified link cost is penalized by the link choice probability and driving experience defined by the degree of familiarity.

(3) A sensitivity analysis is given to demonstrate how the output can be attributed to different variations in the inputs for the penalized link cost function.

(4) The performance of the proposed path is evaluated by comparing other optimal paths and the observed path in a large-scale transportation network.

6.2 Methodology

6.2.1 Framework

This study aims to find an experientially reliable path for risk-averse navigation. To fully utilize the objective travel time estimated by the probe vehicles and the subjective knowledge of experienced drivers, a two-stage path finding procedure is proposed as shown in Figure 6-1. First, the potential optimal path set is determined by using the travel time estimated by probe vehicles. In this stage, the hyperpath algorithm is applied to find a candidate path set, reflecting an optimization strategy that minimizes the uncertain delay. A choice probability for each link in the candidate path set is assigned by the hyperpath operation. In the second stage, the degree of familiarity for each link is formulated by

using the link usage frequency learned from the historical trips of experienced drivers. Then, the shortest-path algorithm is applied to find an experientially reliable path in the candidate path set where the links are penalized by the degree of familiarity and the link choice probability. This path finding procedure is robust because it enables to work when the probe vehicle data are not sufficient to estimate the link travel time distribution, which is more practical than the most reliable path models (Nie and Wu, 2009; Chen and Ji, 2005; Zeng et al., 2015; Xing and Zhou, 2011) that strictly require precise link travel time estimation.

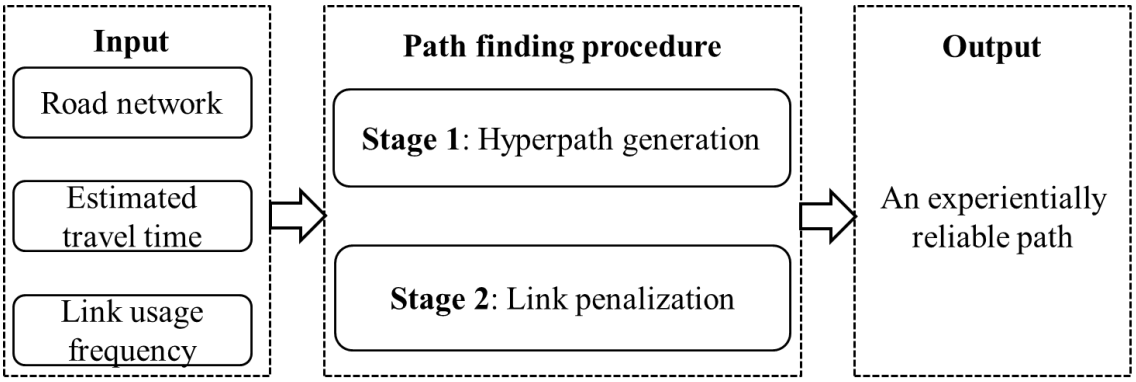


Figure 6-1 Framework of the path finding procedure

6.2.2 Definition of notation

A: Set of links in the whole network;

N: Set of nodes in the whole network;

r: Origin node;

s: Destination node;

A_i^+: Set of outgoing links from node i;

A_i^-: Set of incoming links to node i;

a_{ij}: Directed link from node i to node j;

c_{ij}: Average free-flow travel time on link a_{ij};

d_{ij}: Average delay on link a_{ij};

p_{ij}: Probability that link a_{ij} is chosen;

w_i: Expected delay at node i;

p_i: Probability that node i is chosen;

6.2.3 Application of hyperpath algorithm to generate the candidate path set

The hyperpath concept was originally proposed to generate multiple paths in the context of routing problem in Bell's study (Bell, 2009). Here, this concept is applied to generate a candidate path set for risk-averse navigation. Instead of recommending a specific shortest path, hyperpath provides a set of potential optimal paths with choice probability based on the exposure to delay, which minimizes the expected travel time in the stochastic network. More specifically, to avoid the risk of stochastic delay at node *i*, all the attractive outgoing links of node i will be considered. Since the delay is stochastic, the best way is to assign outgoing links choice probabilities so as to minimize the exposure to delay on each path.

In traffic assignment problem (Nguyen and Pallottino, 1988; Spiess and Florian, 1989), the traffic demand from origin *r* to destination *s* is assigned to the network yielding link volumes by the optimal strategy. In the navigation problem, the assigned link volume can be explained as the probability (p_{ij}) that link a_{ij} is chosen, because the traffic demand for a specific individual driver from origin to destination is one. Then the hyperpath formulation for navigation problem can be given as follows. A detail explanation for hyperpath is given in Bell's studies (Bell, 2009; Bell et al., 2012).

$$f = Min\left\{\sum_{(i,j)\in A} c_{ij}p_{ij} + \sum_{i\in N} w_i\right\} \tag{6-1}$$

$$\text{s.t. } \sum_{(i,j)\in A_i^+} p_{ij} - \sum_{(i,j)\in A_i^-} p_{ij} = g_i \tag{6-2}$$

$$g_i = \begin{cases} 1 & i = r \\ 0 & i \in N - \{r,s\} \\ -1 & i = s \end{cases} \tag{6-3}$$

$$p_{ij}d_{ij} \le w_i, (i,j) \in A_i^+, i \in N \tag{6-4}$$

$$p_{ij} \ge 0 \tag{6-5}$$

The objective function can be explained as the expected travel time from origin to destination. c_{ij} is the average free-flow travel time on link a_{ij}, w_i is the expected delay at node i, p_{ij} is the probability that link a_{ij} is chosen, and d_{ij} is the average delay on link a_{ij}. Note that the expected node delay w_i can be seen as the "combined waiting time" for alternative links at node i, which means that if there is only one alternative link a_{ij} of the outgoing link set of node i, the traveler will be fully suffered by average link delay (d_{ij}), whereas more alternative links enables to alleviate the exposure to delay at node i. In Bell's study (Bell, 2009), the expected node delay is explained as the exposure to maximum link delay to a pessimistic driver. Constraint (6-2) and (6-3) are the flow conservations. Constraint (6-4) ensures that the choice probability (p_{ij}) of link a_{ij} is inversely proportion to its expected delay (d_{ij}). Schmocker et al. (2013) interpreted the constraint (6-4) as a specific game where the traveler fears a fictive opponent who may lead to delays on outgoing links.

The analytical solution of w_i and p_{ij} are dependent on the distribution pattern of the stochastic delay to traverse link a_{ij}. Here, we assumed that the stochastic delay to traverse link a_{ij} follows the exponential distribution. Accordingly, w_i and p_{ij} can be expressed by Eq.(6-6) and Eq.(6-7), respectively. The detail of derivation can be found in (Kanturska et al., 2013).

$$w_i = \frac{1}{\sum_{(i,j)\in A_i^+} \frac{1}{d_{ij}}} \tag{6-6}$$

$$p_{ij} = \frac{\frac{1}{d_{ij}}}{\sum_{(i,j)\in A_i^+} \frac{1}{d_{ij}}} \tag{6-7}$$

The data requirement for hyperpath problem are the average free-flow travel time (c_{ij}) and the average link delay (d_{ij}). As shown in Figure 6-2, the average free-flow travel time (c_{ij}) is defined as the average of the stochastic value from 0 to 50th percentile of the link travel time distribution. Accordingly, the average link delay (d_{ij}) is defined as the mean difference between the stochastic link travel time with delay (from 50th to 100th percentile) and the 50th percentile travel time. Then, the expected delay (w_i) at node i and the link choice probability (p_{ij}) can be obtained by Eq.(6-6) and (6-7). The formulations for c_{ij} and d_{ij} are given as follows.

$$c_{ij} = E\left[t_{ij}\right], \forall\, t_{ij} \le t_{ij}^{(50)} \tag{6-8}$$

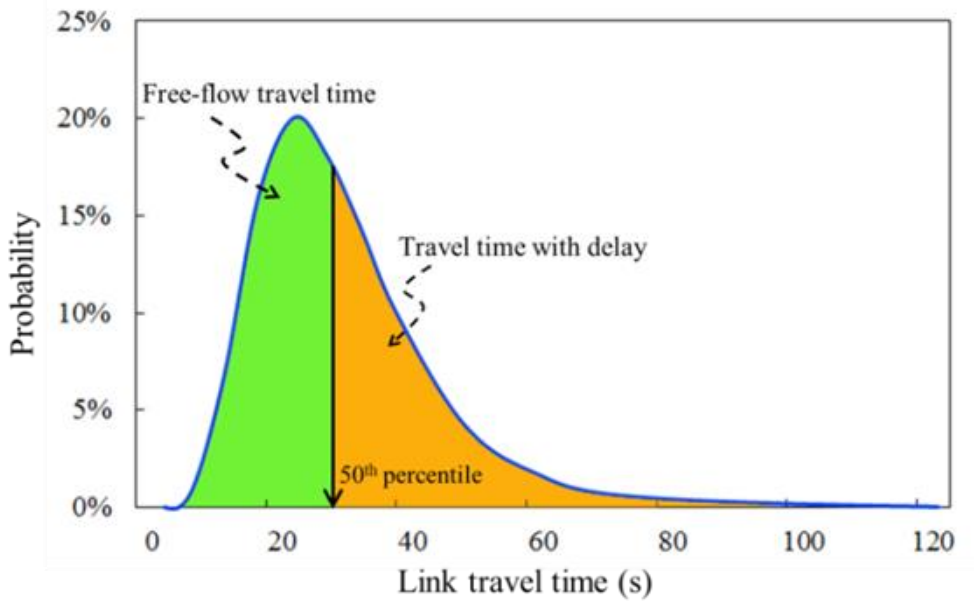


Figure 6-2 Illustration of free-flow travel time and delay

$$d_{ij} = E\left[t_{ij} - t_{ij}^{(50)}\right], \forall\, t_{ij} > t_{ij}^{(50)} \tag{6-9}$$

where t_{ij} is the stochastic link travel time, and $t_{ij}^{(50)}$ is the 50th percentile travel time.

6.2.4 Link penalization and experiential reliable path construction

The hyperpath algorithm generates a set of paths and assigns a choice probability to each link according to the expected delay. Actually, the paths in the optimal hyperpath are restricted by the optimistic (all links with no-delay) and pessimistic (all links with expected delay) travel times (Ma et al., 2013). Thus, it is reasonable to recommend the paths generated by the hyperpath algorithm to users as alternative paths with acceptable travel times. However, suggesting multiple paths leaves the drivers with the problem of selecting a single path for a trip.

Local commuters are good at choosing a reliable path due to their experiences accumulated over years. When facing multiple route choices, most of the people who are unfamiliar with the local traffic condition may simply select the shortest path while the local commuters may utilize their experience to choose the best one with less uncertain delay. Obviously, there are some advantages to utilize the driving experience of the local commuters. First, because local commuters are very familiar with the regular traffic condition, they can avoid traffic congestion during peak hour and save significant amount of travel time. Second, local commuters rarely choose unfamiliar or low grade roads with

poor accessibility. This makes the path travel time more reliable. Naturally, the link usage frequency reflects the driving experience which can be expressed by degree of familiarity. Therefore, we use the link usage frequency to formulate the degree of familiarity on each link. To normalize the degree of familiarity from zero to one, a sigmoid learning curve (Yuan et al., 2011; Leibowitz et al., 2010) is used to model the degree of familiarity of local drivers.

$$r_{ij}^{k} = \frac{1}{1+e^{-\left(an_{ij}^{k}+b\right)}} \tag{6-10}$$

where n_{ij}^{k} is the link usage frequency that driver k traverses the link in a certain period such as peak hours; $an_{ij}^{k} + b$ is the linear transformation that maps n_{ij}^{k} from $[0, maximum]$ to $[-6, 6]$, where maximum represents the maximal individual trip count.

The average degree of familiarity on link a_{ij} can be expressed as Eq.(6-11).

$$r_{ij} = \frac{1}{m}\sum_{k=1}^{m} r_{ij}^{k} \tag{6-11}$$

where *m* is the number of probe vehicle drivers on link a_{ij}.

Given that most travelers prefer a reliable path with a more accurate travel time estimate, it is necessary to find an optimal path from the hyperpath. In general, a common approach to find a reliable path is to penalize unreliable links (Kaparias et al., 2007; Chen et al., 2007). Hence, the idea of link travel time penalty is introduced. To avoid the selection of links with larger uncertain delay and lower degree of familiarity, link travel time in hyperpath is penalized by the link choice probability and the degree of familiarity. The penalized link travel time can be expressed as Eq. (6-12).

$$T'_{ij} = T_{ij} + \theta d_{ij}\left[w_1\left(1-p_{ij}\right) + w_2\left(1-r_{ij}\right)\right] \tag{6-12}$$

where $T_{ij} = E[t_{ij}]$ denotes the average travel time on link a_{ij}, w_1 and w_2 respectively denote the penalty weight for the link choice probability and the degree of familiarity, and θ denotes the penalty scale. Setting a higher value of θ will lead to a larger penalty on the link travel time, which can be specified by users. The sum of w_1 and w_2 equals to one. In case study, θ, w_1 and w_2 are respectively set to 2, 0.5, and 0.5. With this formulation, smaller link choice probability and lower degree of familiarity will lead to a larger penalization on the link travel time.

After penalizing the travel time of each link in the hyperpath link set, the well-known shortest path algorithm such as A* algorithm (Hart et al., 1968) is applied to find the experiential reliable path.

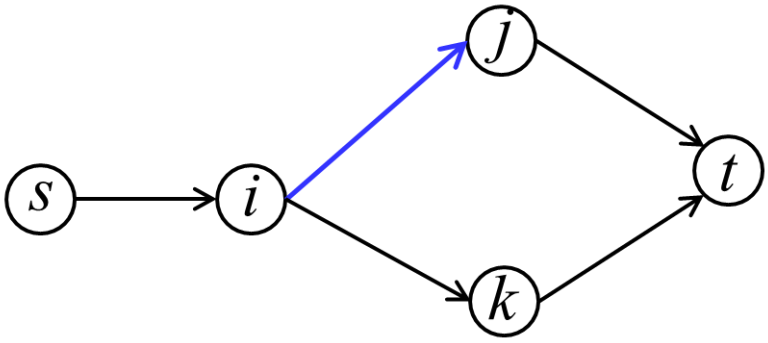


Figure 6-3 A simple network for sensitivity analysis

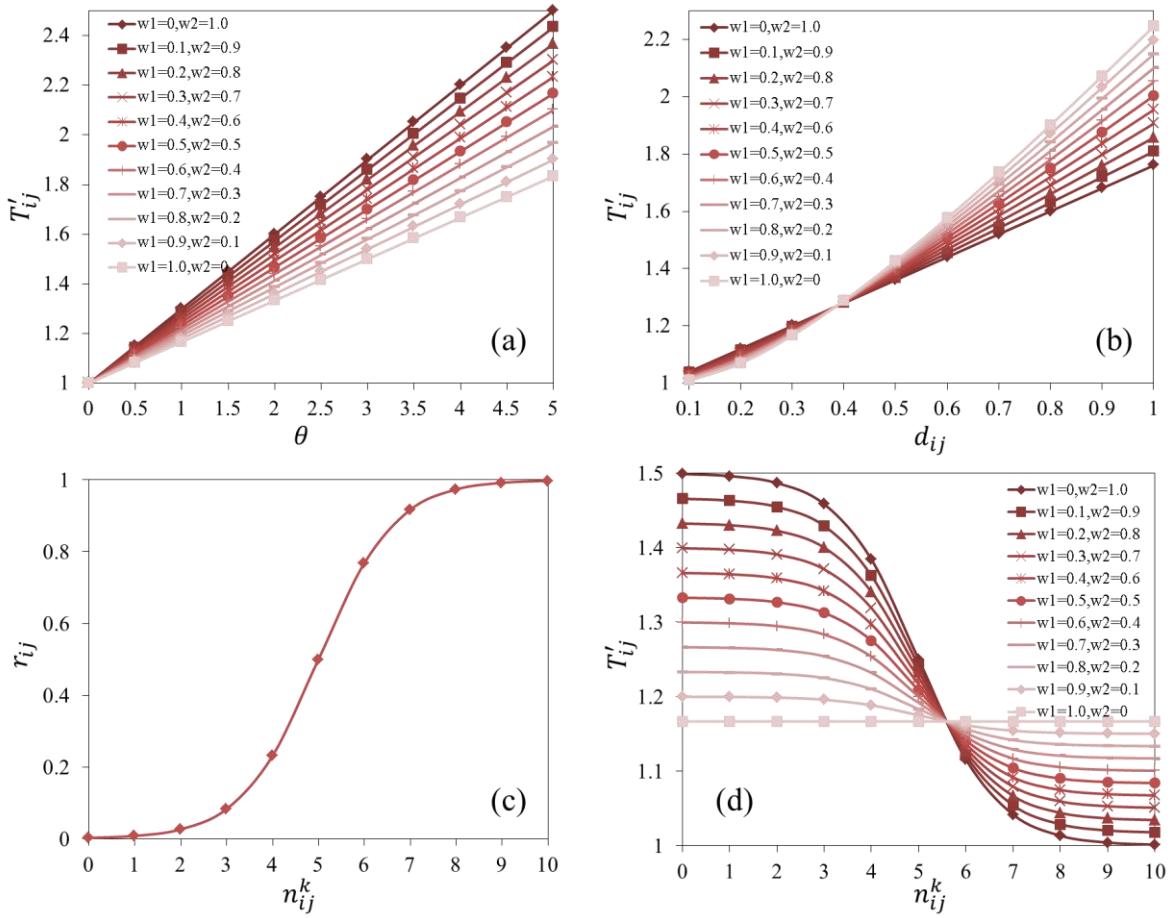


Figure 6-4 Sensitivity analysis for link penalization function

6.3 Numerical analysis

6.3.1 Sensitivity analysis for link penalization function

To better understand how the link penalization function behaves with the change of the three parameters (i.e., θ, w_1 and w_2) and how the variation in the outcome can be attributed to different variations in the input variables (i.e., d_{ij}, r_{ij}), a sensitivity analysis is carried out. To simplify the calculation for p_{ij}, we use a simple network for sensitivity analysis as shown in Figure 6-3. And link (i, j) is chosen as the target link. Given default values for $T_{ij} = 1, d_{ij} = 0.5, d_{ik} = 1$ and $r_{ij} = 0.4$ in Eq. (6-12), Figure 6-4(a) shows the variation of the penalized link travel time in different settings of the three parameters. It indicates that as the parameter θ increases, the penalized link travel time T'_{ij} increases linearly in different combinations of w_1 and w_2. On the other hand, T'_{ij} increases as w_1 increases (w_2 decreases correspondingly) when θ is fixed to a certain value. This is because p_{ij} is larger than r_{ij} in this case ($p_{ij} = 0.67, r_{ij} = 0.4$). In contrast, if p_{ij} is smaller than r_{ij}, T'_{ij} will decrease as w_1 increases. Figure 6-4(b) shows the variation of the penalized link travel time with the change of delay d_{ij} in different combinations of w_1 and w_2. The default values for T_{ij}, d_{ik}, r_{ij}, and θ are set as 1, 1, 0.6, and 1, respectively. It is found that as d_{ij} increases, T'_{ij} increases. It should be noted that the increasing curve is not exactly linear because $p_{ij} = \frac{1}{d_{ij}} / (\frac{1}{d_{ij}} + \frac{1}{d_{ik}})$ in Eq. (6-7). Figure 6-4(c) shows how the link usage frequency n_{ij}^k impacts on the degree of familiarity r_{ij}. To simplify the calculation, we set the maximum individual trip count as 10 and assume that only one traveler goes through link (i, j). It is found that as n_{ij}^k increases, the rate of increase for r_{ij} increases fast in the first half of the curve, while the rate of increase for r_{ij} decreases in the second half of the curve. Figure 6-4(d) shows the variation of the penalized link travel time with the change of the link usage frequency n_{ij}^k. The default values for T_{ij}, d_{ik}, d_{ij}, and θ are set as 1, 1, 0.5, and 1, respectively. Other settings are the same as that in Figure 6-4(c). It is found that as n_{ij}^k increases, T'_{ij} decreases. It also found that the rate of increase for T'_{ij} is sensitive to the combination of w_1 and w_2. As the weight for r_{ij} (i.e., w_2) increases, the variation of T'_{ij} becomes significant when n_{ij}^k changes.

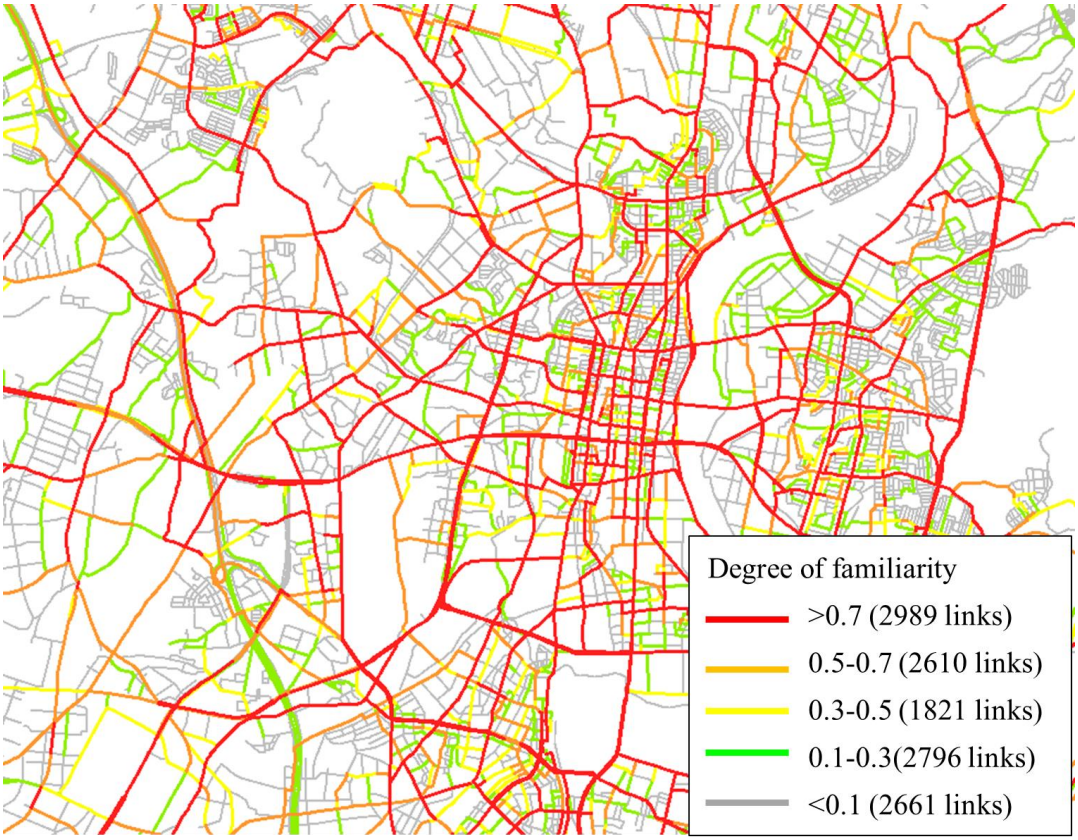


Figure 6-5 Tested network and degree of familiarity on each link

6.3.2 Path performance analysis

As shown in Figure 6-5, the tested network with 4072 nodes and 12,877 links is in Toyota city, Japan. Path records are obtained from probe vehicle data after map-matching and link travel time estimation, which is introduced in our previous studies (Miwa et al., 2012; Miwa et al., 2008). A completed path record includes the ordering link sequence, the detail of link information, and the traveler information. Since all of the drivers are local, their driving experiences are valuable to risk-averse strangers who are unfamiliar with the network. As introduced in section 3.4, we use the degree of familiarity to present the driving experience. The statistical period is one week in peak hours (7:00-10:00, 17:00-20:00). It is found that the degree of familiarity is more than 0.5 in 43.5% of the links, while it is less than 0.1 in 20.6% of the links. This could be explained by the fact that some of the links are not attractive due to insufficient probe vehicles in the network or the poor road condition. Another potential explanation is that local drivers are not willing to take risk to choose unfamiliar paths for their commutes.

To demonstrate the performance of the proposed path, four kinds of optimal paths (shortest distance, average shortest time, penalized shortest time and the proposed) are compared with the observed path. The shortest distance path, shortest time path, and

penalized shortest time path are generated by the A* algorithm with the criteria of distance, travel time, and penalized travel time, respectively. It should be noted that the link penalty of the penalized shortest time path is different from the proposed one, which is only related to the degree of familiarity which is directly applied to the whole network (θ is set as the same as the proposed method).

Here, several performance indices are used, i.e., detour index (DI) (Barthelemy, 2011), cosine similarity index (CSI) (Fonzone et al., 2012), average degree of familiarity, and total travel time.

DI is a measure of the efficiency of a path in terms of how well it overcomes distance. It can be defined as the ratio of straight distance and the actual distance. The closer the detour index gets to 1, the more the path is spatially efficient.

$$DI = \frac{D(S)}{D(A)} \tag{6-13}$$

where $\mathrm{D(S)}$ is the straight distance of the OD pair, $\mathrm{D(A)}$ is the actual distance of a specified path of the OD pair.

CSI represents how much similarity there is between the observed path and the paths with different criteria, which can be applied to evaluate the overlap degree of two paths.

$$CSI = \frac{E_1 E_2}{\|E_1\| \|E_2\|} \tag{6-14}$$

where $E_1 = [e_{11}, e_{12}, \dots, e_{1n}]$ and $E_2 = [e_{21}, e_{22}, \dots, e_{2n}]$ are the link length vectors for the optimal path and the observed path, respectively. n is the total link count of the two paths (excluding duplicate calculations for the same link in two paths). e denotes the link length. For example, $\mathrm{e_{1n}}$ denotes the length of link $\mathrm{e_{1n}}$. Noted that if link $\mathrm{e_{1n}}$ is not included in the path, e_{1n} is set to 0.

To better understand the paths driven by the probe vehicles and test the proposed path-finding method, historical paths from 7240 OD pairs are extracted for analysis. The observed OD pairs are divided into four categories based on the straight OD distance, which is defined as the Euclidean distance between the origin and destination. The sample sizes are 4246, 2088, 703 and 205 for OD distances less than 5km, between 5km and 10km, between 10km and 15km, more than 15km, respectively.

Figure 6-6 shows the curves of cumulative density function (CDF) of the average degree of familiarity on the five paths. It shows that the CDF curve of average degree of familiarity for the observed path shifts from the left to the right as the OD distance

increases, which indicates that the average degree of familiarity of the observed path increases as the OD distance increases. This can be explained in two possible ways. First, the observed paths with longer distance usually have higher percentage of corridor routes and expressways, which have a relatively higher usage frequency. Second, most drivers pay more attention to the degree of familiarity as travel distance increases. On the other hand, it is found that the proposed path dominates the observed path, which indicates that the proposed method can provide feasible path with more expert experience than the observed one. It can be explained by the fact that the proposed method enables to take account of the network-wide driving experience of all the drivers explicitly, and the sub-objective of the penalized link cost function is to maximize the weighted degree of average link familiarity. The shortest distance path and the shortest time path provide a shorter distance or less travel time, but the average degree of familiarity is relatively low. That means drivers might suffer an unpredicted delay if such paths are chosen.

Figure 6-7(a) compares the DI of four kinds of paths for historical OD pairs with different traveling distances. As expected, the shortest distance path has the highest value of detour index, while the penalized shortest time path has the lowest one. Though the penalized shortest time path has the highest degree of familiarity, it may cost an unexpected long distance and takes a detour in the network. The detour index of the proposed path is higher than that of the observed path if the OD distance is larger than 5km. It indicates that the proposed path has a better performance than the observed path on distance efficiency for a larger travel distance.

Figure 6-7(b) compares the CSI of different paths. It shows that the proposed path is most similar to the observed one, while the shortest distance path and the penalized shortest time path has relatively low similarity for all historical OD pairs. The CSI of the proposed path increases as the OD distance increases, which indicates that the proposed method enables to find an applicable path based on hyperpath strategy and driving experience for a larger OD distance. On the other hand, it further verifies that drivers become more risk-averse when the travel distance increases. The CSI of the shortest time path also increases as the OD distance increases, which indicate that drivers are also concerned with the travel time budget if the travel distance becomes greater. However, the CSI of the shortest distance path decreases as the OD distance increases. It falls below 0.5 when the OD distance increases to 5km. This can be explained because the components of the shortest paths may include some links with poor road condition or minor roads that few drivers are familiar with. And the traffic condition of the shortest path is uncertain when the OD distance increases. Therefore, the shortest distance paths should not be recommended to drivers if the travel distance is more than 5km.

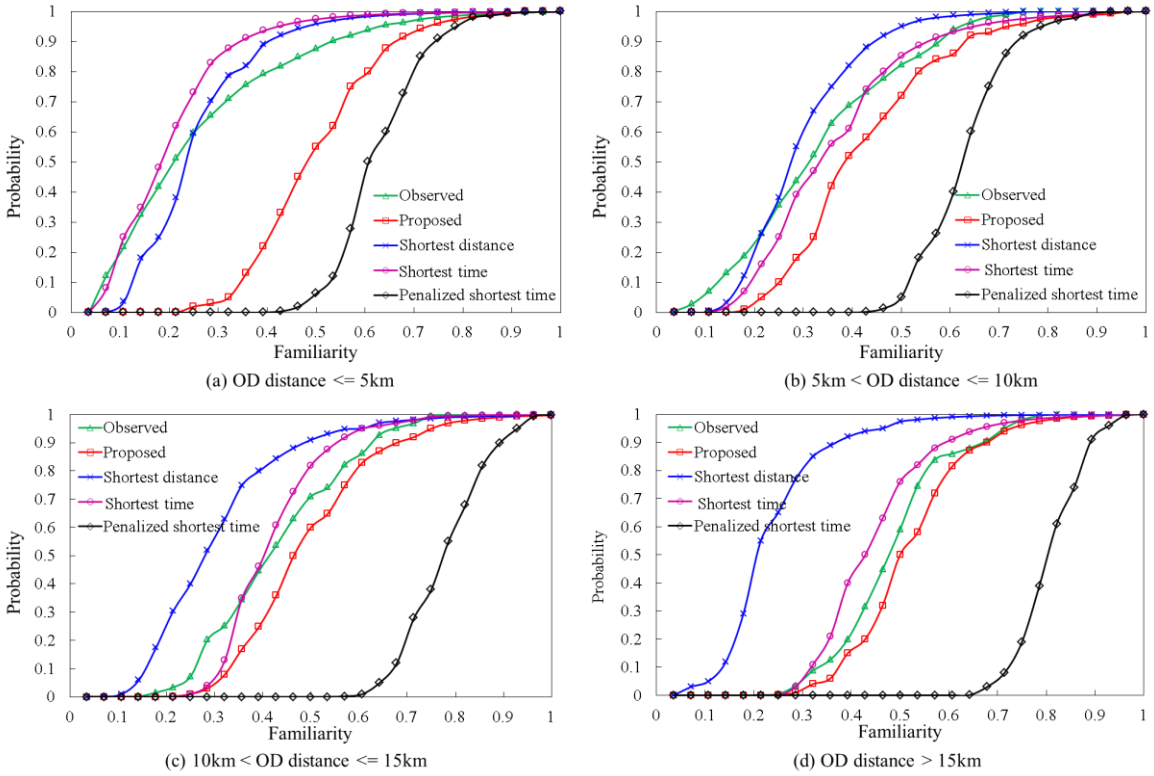


Figure 6-6 CDF of degree of familiarity

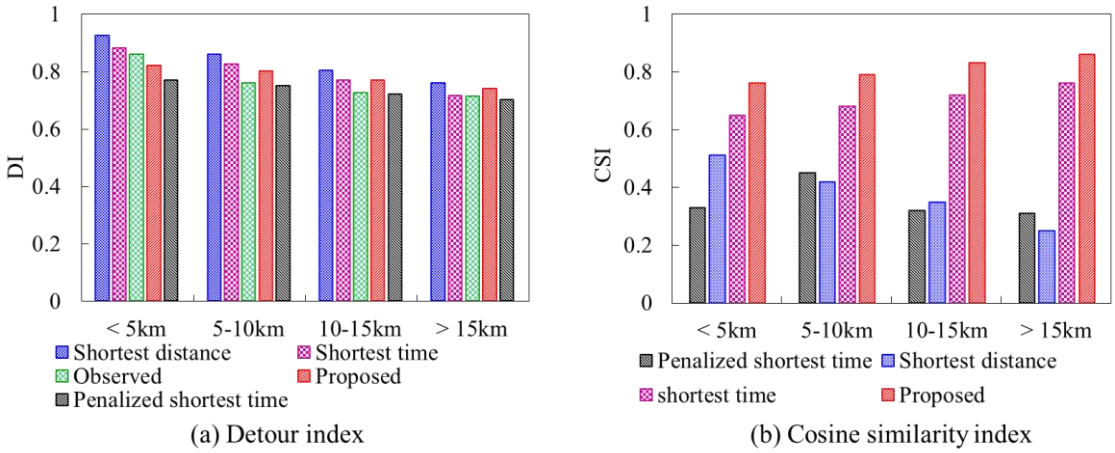
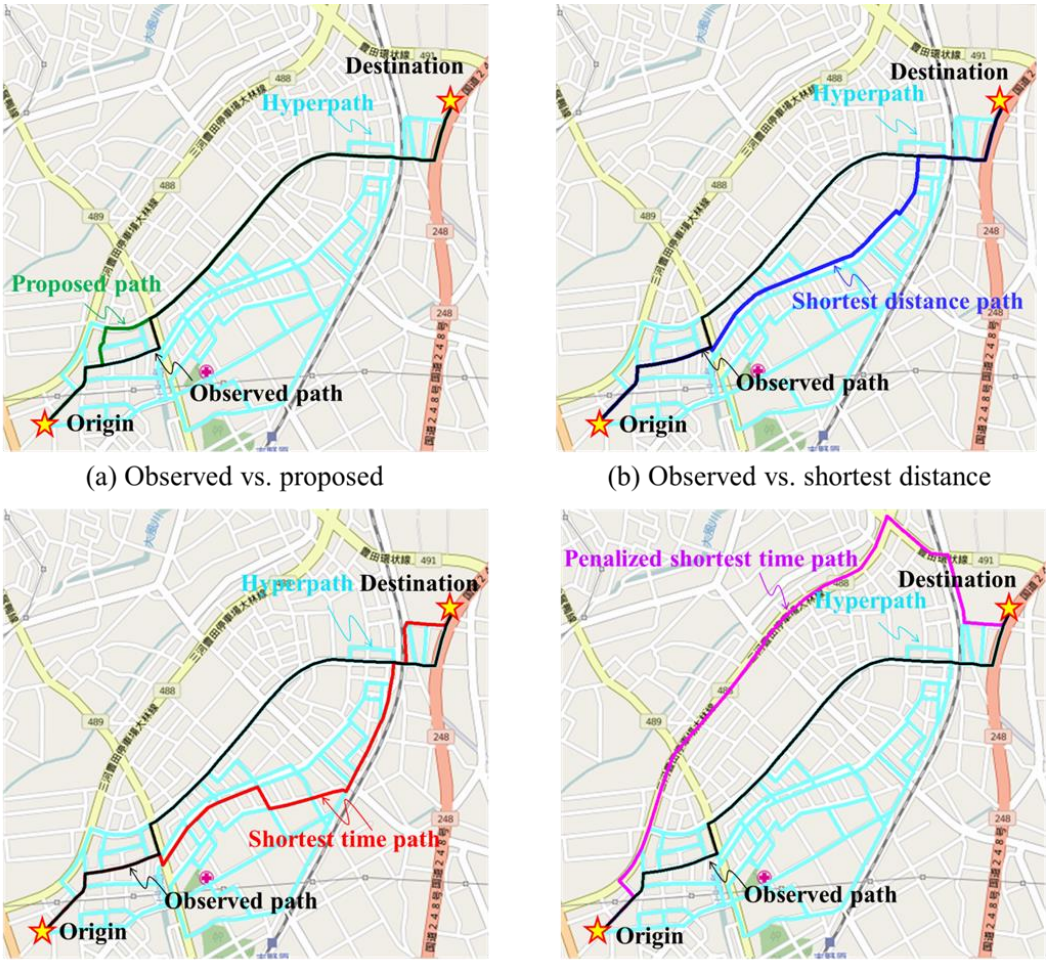


Figure 6-7 Comparison of detour index and cosine similarity index

6.3.3 Case study

In Figure 6-8, one real-world OD in the central area of Toyota city is extracted to demonstrate the performance of the proposed path finding procedure. The results, shown in Table 6-1, demonstrate the performance of different paths of the specified OD pair. As expected, the proposed path is most similar to the observed path, which has the highest CSI. Interestingly, the degree of familiarity of the proposed path is higher than that of the

observed one. That means the proposed method finds a feasible path with more expert experience through the use of other drivers' experience, and this can be of value in guiding the driver. As expected, the penalized shortest time path has the highest average degree of familiarity, but has lowest DI and CSI. Drivers may not prefer the penalized shortest time path because it will take an unexpected long distance and more travel time. Note that the proposed path has higher average degree of familiarity, but it is slower than the observed one. For a risk-averse driver, we recommend the proposed path because it has higher usage frequency by all the experienced drivers, which may have less risk on stochastic delay and other unexpected incidents such as traffic jam or accidents. For a risk-neutral driver, we recommend the observed path because it has similar degree of familiarity and less average travel time. In the practical application, hyperpath and the five optimal paths could be provided to users, allowing them to make a choice depending on individual preference and the path performance.



(a) Observed vs. proposed

(b) Observed vs. shortest distance

(c) Observed vs. shortest time

(d) Observed vs. penalized shortest time

Figure 6-8 Case study for path comparison

Table 6-1 Path performance comparison

Index	Observed	Proposed	Shortest distance	Shortest time	Penalized shortest time
CSI	1	0.79	0.38	0.32	0.11
Average link familiarity	0.63	0.67	0.33	0.39	0.72
DI	0.82	0.84	0.85	0.77	0.67
Total distance	2246m	2204m	2197m	2407m	2776m
Total travel time	321s	382s	302s	263s	425s

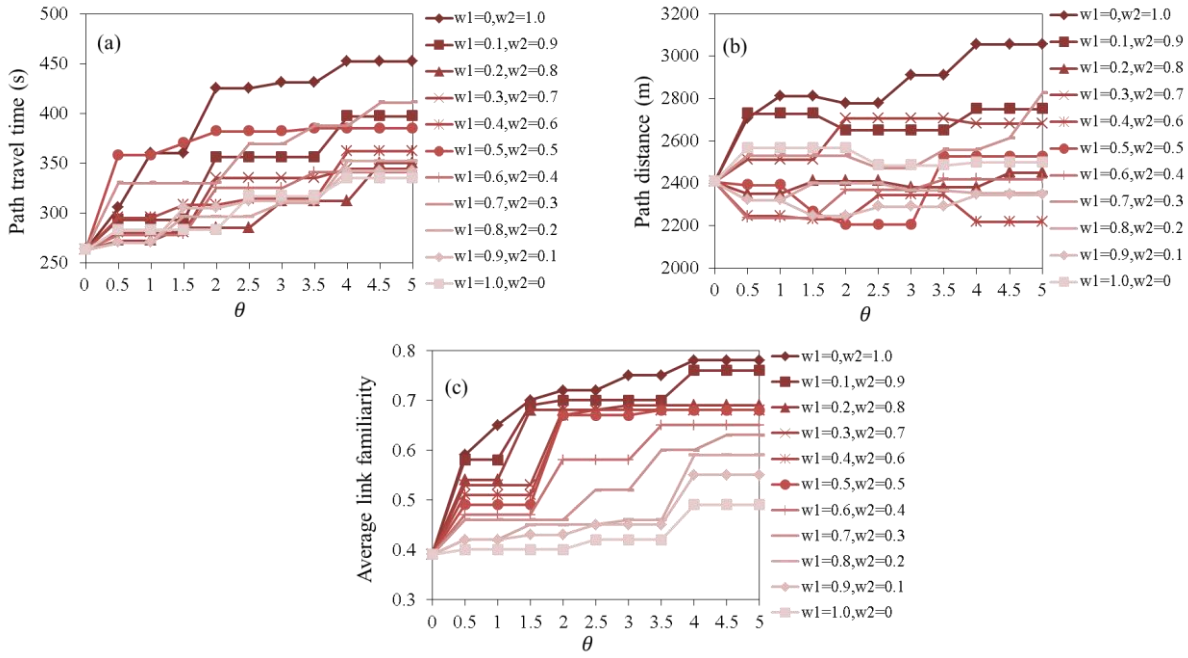


Figure 6-9 Relation between parameter settings and path performance

A sensitivity analysis for different path performance indices is carried out by changing the combination of the three parameters. Figure 6-9 gives the relation between the parameter settings and the path performance. As shown in Figure 6-9(a), it is found that as the penalty scale θ increases, the path travel time increases. It is because the objective of the routing algorithm is no longer to minimize the travel time. Instead, as the θ increases, the weight for the link choice probability p_{ij} and the degree of familiarity r_{ij} increases, this will be at the cost of increasing in travel time. However, there is no significant pattern in the relationship between the path travel time and the combination of w_1 and w_2 when θ is fixed to a certain value. Figure 6-9(b) shows how the parameter combination impacts on the path distance. It is found that as the penalty scale θ increases, the increase of the path distance is not significant. Similar to path travel time, no significant pattern in the

relationship between the path distance and the combination of w_1 and w_2 is found. It indicates that increasing the weight for the expert knowledge does not necessarily increase the path distance. Figure 6-9(c) shows how the parameter combinations impact on the average link familiarity. It is found that as the penalty scale θ increases, the average link familiarity increases. And a significant pattern in the relationship between the average link familiarity and the combination of w_1 and w_2 is found, i.e., the average link familiarity increases as w_2 increases (w_1 decreases correspondingly). It is because the routing objective function tends to minimize the average link familiarity when θ and w_2 increase.

The three parameters (i.e., θ, w_1 and w_2) for different groups of route guidance users can be calibrated by using a grid-like search which is similar to the parameter estimation in machine learning (Chang and Lin, 2002). In grid-like search, all combinations of (θ, w_1 and w_2) are tried and the one with the minimum relative error is picked up. The relative error (e) can be defined as follows:

$$e = k_1 \left|\frac{t-t_0}{t_0}\right| + k_2 \left|\frac{d-d_0}{d_0}\right| + k_3 \left|\frac{r-r_0}{r_0}\right| \quad (6\text{-}15)$$

where t, d and r are the travel time, distance and average link familiarity of the candidate path by using a combination of the three parameters, respectively; t_0, d_0 and r_0 are the travel time, distance and average link familiarity of the observed path chosen by the local driver, respectively; k_1, k_2 and k_3 are the preference values for each performance index, respectively. Here, we set k_1, k_2 and k_3 as 1/3. In the case in Figure 6-8, the calibrated parameters are $\theta = 3$, $w_1 = 0.5$ and $w_2 = 0.5$.

6.4 Summary

The ultimate aim of this research is to better help travelers plan their trips and avoid the risk of uncertain travel time. A two-stage path finding procedure is developed to find an experientially reliable path. Firstly, the hyperpath concept helps to determine a set of potential optimal paths. Secondly, the links in the graph of hyperpath are penalized based on the choice probability and the degree of familiarity, and then the A-star algorithm is applied to search the optimal path.

There are two advantages to this routing method. First, any reasonable detours are taken into account and the link choice probability for each outgoing link from an attractive node is estimated according to the expected delay, which guarantees the potential optimal paths are included. Second, the solution provides not only the hyperpath with the

recommended link choice probability, but also a shortest path with high degree of familiarity. This can help travelers plan their trips effectively.

By comparing the performance of the proposed path with other optimal paths, it is found that the proposed path is most similar to the observed path (in the probe vehicle data). It also indicates that the proposed method yields a highly reliable path to drivers.

However, there remain several problems needing further attention. For example, turn restrictions and direction constraints should be considered in a practical network. In addition, a faster path finding algorithm suitable for a time-dependent stochastic network needs to be developed. These limitations will be improved by a more comprehensive modeling approach and a stricter validation procedure in our future work.

6.5 References

Bell, M. G. (2009). Hyperstar: A multi-path Astar algorithm for risk averse vehicle navigation. Transportation Research Part B: Methodological, 43(1), 97-107.

Bell, M. G., Trozzi, V., Hosseinloo, S. H., Gentile, G., & Fonzone, A. (2012). Time-dependent Hyperstar algorithm for robust vehicle navigation. Transportation Research Part A: Policy and Practice, 46(5), 790-800.

Barthelemy, M. (2011). Spatial networks. Physics Reports, 499(1), 1-101.

Chang, C., & Lin, C. (2002). Training v-support vector regression: Theory and algorithms. Neural Computation, 14, 1959–1977.

Chen, A., & Ji, Z. (2005). Path finding under uncertainty. Journal of advanced transportation, 39(1), 19-37.

Chen, Y., Bell, M. G., & Bogenberger, K. (2007). Reliable pretrip multipath planning and dynamic adaptation for a centralized road navigation system. IEEE Transactions on Intelligent Transportation Systems, 8(1), 14-20.

Fonzone, A., Schmocker, J. D., Ma, J., & Fukuda, D. (2012). Link-Based Route Choice Considering Risk Aversion, Disappointment, and Regret. Transportation Research Record: Journal of the Transportation Research Board, 2322(1), 119-128.

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. IEEE Transactions on Systems Science and Cybernetics, 4(2), 100-107.

Kanturska, U., Trozzi, V., & Bell, M. G. (2013). Scheduled Hyperpath. Transportation Research Record: Journal of the Transportation Research Board, 2378(1), 99-109.

Kaparias, I., Bell, M. G. H., Chen, Y., & Bogenberger, K. (2007). ICNavS: a tool for reliable dynamic route guidance. IET Intelligent Transport Systems, 1(4), 225-233.

Leibowitz, N., Baum, B., Enden, G., & Karniel, A. (2010). The exponential learning equation as a function of successful trials results in sigmoid performance. Journal of Mathematical Psychology, 54(3), 338-340.

Ma, J., Fukuda, D., & Schmocker, J. D. (2013). Faster hyperpath generating algorithms for vehicle navigation. Transportmetrica A: Transport Science, 9(10), 925-948.

Miwa, T., Kiuchi, D., Yamamoto, T., & Morikawa, T. (2012). Development of map matching algorithm for low frequency probe data. Transportation Research Part C: Emerging Technologies, 22, 132-145.

Miwa, T., Sakai, T., Morikawa, T. (2008). Route identification and travel time prediction using probe-car data, International Journal of ITS Research, 2(1), 21-28.

Nie, Y. M., & Wu, X. (2009). Shortest path problem considering on-time arrival probability. Transportation Research Part B: Methodological, 43(6), 597-613.

Nguyen, S., & Pallottino, S. (1988). Equilibrium traffic assignment for large scale transit networks. European journal of operational research, 37(2), 176-186.

Papinski, D., Scott, D. M. (2011). A GIS-based toolkit for route choice analysis. Journal of Transport Geography, 19(3), 434-442.

Papinski, D., Scott, D. M., Doherty, S. T. (2009). Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. Transportation research part F: traffic psychology and behaviour, 12(4), 347-358.

Spiess, H., & Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. Transportation Research Part B: Methodological, 23(2), 83-102.

Schmocker, J. D., Bell, M. G., Kurauchi, F., & Shimamoto, H. (2009). A game theoretic approach to the determination of hyperpaths in transportation networks. In Transportation and Traffic Theory 2009: Golden Jubilee, Springer US, 1-18.

Tang, L., Li, Q., Chang, X., Shaw, S., Zhao, Z. (2010). Modeling of taxi drivers'

experience for routing applications. Science China Technological Sciences, 53(1), 44-51.

Xing, T., & Zhou, X. (2011). Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach. Transportation Research Part B: Methodological, 45(10), 1660-1679.

Yuan, J., Zheng, Y., Xie, X., Sun, G. (2011). Driving with knowledge from the physical world. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 316-324.

Yuan, J., Zheng, Y., Xie, X., Sun, G. (2013). T-Drive: Enhancing driving directions with taxi drivers' intelligence. IEEE Transactions on Knowledge and Data Engineering, 25(1), 220-232.

Zeng, W., Miwa, T., Wakita, Y., & Morikawa, T. (2015). Application of Lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times. Transportation Research Part C: Emerging Technologies, 56, 309-334.

Chapter 7

ECO-ROUTING PROBLEM CONSIDERING CO_2 EMISSION AND TRAVEL TIME

7.1 Introduction

With travel demand continuing to grow, fuel consumption and greenhouse gas (GHG) emissions are increasing unceasingly. Vehicle emissions contribute substantially to CO, CO_2, HC and NOx. It has been noted that the transportation sector accounts for approximate 23% of total global CO_2 emissions, of which 73% are generated by road transport (JAMA, 2008; Birol, 2010). Urban traffic emission modeling and control have been attracted more and more attention (Tan and Gao, 2015; Csikos et al., 2015).

Even though alternative fuel vehicles such as all-electric and fuel cell vehicles will be the best solution, mitigating emissions by existing gasoline vehicles is an alternative countermeasure in the near term. Eco-driving, a term used for emerging driving assistance techniques that support the driver in optimizing route choice and driving behavior to reduce vehicle emissions, has been showing significant benefit in fuel saving and air quality improvement (Beusen et al., 2009; Mensing et al., 2014). Eco-driving techniques can be classified into three decision-making levels (Sivak and Schoettle, 2012): strategic level (vehicle maintenance), tactical level (pre-trip eco-routing), and operational level (on-board driving assistance). In pre-trip eco-routing, a navigation system attempts to find the most eco-friendly path from origin to destination based on estimates of vehicle emission for all possible paths, while on-board driving assistance systems analyze drivers' behavior

and instantaneous fuel consumption, and provide valuable feedback that helps them adjust their driving style to a more eco-friendly style. The focus of this study is the tactical level, in which pre-trip route planning determines an eco-friendly path for a driver.

Traditional navigation systems usually provide the shortest path based on total travel time or distance, without considering vehicle emissions. Intuitively, one may think that the shortest path or fastest path would also be the most eco-friendly path. However, a shortest path may take a driver through a heavy congested area, resulting in high vehicle emissions. On the other hand, there may be cases where a fastest path results in longer travel distance, albeit on less congested roadways. Traveling on a path at a higher speed over a longer distance will also result in higher vehicle emission compared with a shorter path (Masikos et al., 2015). With the environmental problems outlined above becoming matters of urgent concern, eco-routing is an application that promises reduced emissions (Boriboonsomsin et al., 2012; Yao and Song, 2013; Guo et al., 2013; Boriboonsomsin et al., 2014). An investigation conducted in Sweden (Ericsson et al., 2006) found that 46% of trips based on the spontaneous route choice of the traveler were not the most eco-friendly. Vehicle emissions on these trips could be reduced by 8.2% if the most eco-friendly route were chosen. Similarly, Ahn and Rakha (2008) reported that a 4%-20% reduction in vehicle emission can be achieved if an eco-routing strategy is adopted. On the other hand, eco-routing could result in significant reductions in emissions, but it naturally comes at the expense of increased travel time. A field study in Japan (Kono et al., 2008) found that the vehicle emission of the eco-friendly path is 9% lower than that of the least travel time path, while travel time is 9% longer. In such cases, an eco-routing navigation system might suggest the most eco-friendly path with lower vehicle emissions, but the travel time may exceed the travel time budget.

Actually, establishing an efficient and preferable eco-routing navigation system is a great challenge because many aspects of this problem need to be considered. For example, what kind of tools can efficiently collect and estimate emissions in a large-scale transportation network? How should an eco-routing strategy be developed and how will such a strategy impact travel time and the environment? To fill these gaps, this study aims to find an eco-friendly path that produces minimum CO_2 emissions while satisfying travel time budget. In particular, the contribution of this study is as follows:

(1)Probe cars with GPS and OBD (on-board diagnostics) device are used to collect the CO_2 emission and travel time, which is an efficient tool to collect large-scale data.

(2)A vehicle CO_2 emission model is developed by using support vector machine (SVM). Specifically, the explanatory variables include average speed, average acceleration,

road gradient, and vehicle displacement, which are easily to be extracted from Google API and probe vehicles.

(3)The eco-routing problem considering CO_2 emission and travel time is transformed to a bi-objective like optimization problem. The method of Pareto-optimal optimization is applied to solve this routing problem.

(4)The performance of the proposed CO_2 emission model and eco-routing strategy is validated by comparing the actual driving records. A sensitivity analysis is conducted to indicate the relationship between the vehicle CO_2 emission and the explanatory variables.

(5)The sensitivity of the network-wide benefit of eco-routing is quantified for different travel time buffers and OD distances.

7.2 State-of-art models for of vehicle fuel consumption

Applicable vehicle emission model is critical to the development of an eco-routing navigation system. Models in literatures can be classified into macroscopic, mesoscopic or microscopic models depending on the level of detail that the models incorporate in the calculation procedure. Macroscopic models usually estimate average fuel consumption or vehicle emission rate from aggregated parameters, e.g., MOBILE6 model (EPA, 2003). Such model is intended to estimate the average emission for a large area, but it is not well suited for dynamic emission assessment (Barth et al., 2001). In contrast, microscopic models enable to take continuous retrieval of microscopic parameters such as instantaneous speed and instantaneous engine condition. For example, the Comprehensive Modal Emissions Model (CMEM) estimates vehicle emissions based on the instantaneous engine output power that determined by various vehicle-related parameters (Barth et al., 1996, Barth et al., 2000). The CMEM model is shown as follows.

$$f = \frac{\phi P}{\lambda} \tag{7-1}$$

Where f is the fuel rate, ϕ is the air-fuel equivalence ratio (usually set to 1), and λ is the lower heating value.

$$P = \sum_{i=0}^{3} \alpha_i v^i + \beta a v \tag{7-2}$$

$$\alpha_0 = \frac{P_a}{\eta} \tag{7-3}$$

$$\alpha_1 = Zg\frac{G+c_1}{\eta\epsilon} + c_4 K_0 V\theta\left(\underline{r} + c_3 v_h^2\right) \tag{7-4}$$

$$\alpha_2 = Zg\frac{c_1}{c_2\eta\epsilon} - 2c_3c_4K_0Vv_h\theta \quad (7\text{-}5)$$

$$\alpha_3 = \frac{\rho C_d A}{2\eta\epsilon} + c_3c_4K_0V\theta \quad (7\text{-}6)$$

$$\beta = \frac{Z(1+e_0)}{\eta\epsilon} \quad (7\text{-}7)$$

Where the detail of the parameter is shown in Table 7-1.

Table 7-1 Parameters for CMEM model (Nie and Li, 2013)

Parameter	Description	Unit	Default value
g	Acceleration of gravity	m/s^2	9.81
c_1	Const		0.01
c_2	Const	m/s	44.73
η	Engine efficiency		0.4
ρ	Air density	kg/m^3	1.247
A	Frontal area	m^2	2
C_d	Drag coefficient		0.3
K_0	Const	J/rev/l	200
V	Engine displacement	l	2.0
$\underline{r}$	Const		2
v_h	Const	m/s	35
λ	Lower heating value	J/g	44000
G	Grade		0
ϵ	Drivetrain efficiency		0.85
Z	Axillary power	W	1000
a	Acceleration	m/s^2	
v	Speed	m/s	

Another model, known as Virginia Tech Microscopic Energy and Emission Model (VT-Micro), was developed as a third-order regression model that estimates emission rates as a function of the instantaneous speed and acceleration (Rakha et al., 2004).

$$MOE_e = \sum_{i=0}^{3}\sum_{j=0}^{3}\left(K_{i,j}^{e}u^i a^j\right) \quad (7\text{-}8)$$

Where MOE_e is the instantaneous fuel consumption, $K_{i,j}^{e}$ is the coefficient to be estimated, u is the instantaneous speed, and a is the instantaneous acceleration.

Considering dynamic driving parameters, Jimenez-Palacios (1998) proposed a Vehicle Specific Power (VSP) based model, which had been incorporated into Motor Vehicle Emission Simulator (MOVES) (Koupal et al., 2002). The real-time vehicle

emission can be estimated by binning instantaneous VSP (Bandeira et al., 2013). The VSP model is shown as follows.

$$VSP = v \times (a(1 + \varepsilon_i) + g \times grade + g \times C_R) + \frac{1}{2}\rho_a \frac{C_D \times A}{m}(v + v_w)^2 v \quad (7\text{-}9)$$

Where the detail of the parameter is shown in Table 7-2.

Table 7-2 Parameters for VSP model (Jimenez-Palacios, 1998)

Parameter	Description	Unit	Default value
g	Acceleration of gravity	m/s^2	9.81
a	Acceleration	m/s^2	
v	Speed	m/s	
m	Vehicle mass	kg	
$grade$	Road gradient		
ε_i	Mass factor		
A	Frontal area	m^2	
C_R	Coefficient of rolling resistance		
C_D	Drag coefficient		
ρ_a	Ambient air density	kg/m^3	1.207
v_w	Headwind into the vehicle	m/s	

Because it is quite difficult to obtain instantaneous information such as second-by-second speed profiles except in an ideal laboratory, microscopic models may not be applicable for routing problems. To overcome this limitation, some studies proposed mesoscopic models. Unlike microscopic and macroscopic models, mesoscopic models usually estimate fuel consumption or vehicle emissions on a link and it is not necessary to collect substantial amount of second-by-second speed profiles. For example, Minett et al. (2011) generated synthetic speed profiles based on historical link speed data and used them for calculating fuel consumption per link. In recent years, machine learning was introduced to establish the learning based model for fuel consumption or vehicle emission (Masikos et al., 2013; Masikos et al., 2015a; Masikos et al., 2015b). It was found that the machine learning technique outperformed the multivariate regression technique (Boriboonsomsin et al., 2012) because it was capable of generating forecasts properly and of adequately identifying the nonlinearities underlying vehicle fuel consumption process.

7.3 CO_2 emission model based on SVM

To develop an eco-routing navigation system, it is necessary to predict vehicle CO_2 emission per kilometer. Support vector machine (SVM), a novel supervised learning method used for regression, has been recently proved to be a promising tool in the field of

emission prediction. It has shown very high performance in solving various forecasting problems such as ozone concentration (Ortiz-Garcia et al., 2010; Luna et al., 2014), CO concentration (Yeganeh et al., 2012), and urban air quality (Singh et al., 2013), etc. These successful applications motivate us to apply SVM to learn vehicle CO_2 emission from large-scale GPS and OBD data.

A number of factors based on the characteristics of traffic, vehicle, and street configuration are found to affect vehicular fuel consumption which is linear related to CO_2 emission (Park and Rakha, 2006; Ahn et al., 2002; Pandian et al., 2009; Guo et al., 2014). These variables can be roughly grouped into six categories, i.e., vehicle-related (e.g., acceleration, power demands, and engine displacement), roadway related (e.g., road gradient), traffic-related (e.g., congestion and average speed), and driver-related factors (e.g., risk-taking and aggressive driving). In general, the selections of these factors to estimate CO_2 emissions usually depends on the context of the study, ease of measurement and ease of computation. According to the state-of-art for modeling fuel consumption and CO_2 emission (Barth et al., 1996; Nie and Li, 2013; Rakha et al., 2004; Jimenez-Palacios, 1998), a number of explanatory variables that includes the average speed, average acceleration, road gradient, and vehicle displacement are considered as the input for the SVM model. The structure of the SVM based model is shown in Figure 7-1.

Given a set of data points $\{(x_1,y_1), (x_2,y_2), \ldots, (x_l,y_l)\}$ randomly and independently generated from an unknown function, SVM approximates the function using the following form (Cristianini and Taylor, 2000; Vapnik, 1999):

$$f(x) = \omega \cdot \phi(x) + b \tag{7-10}$$

where $\phi(x)$ denotes the high dimensional feature spaces which are non-linearly mapped from the input space x, i.e., average speed, average acceleration, road gradient, and vehicle displacement; The road gradient can be extracted by using the Google Elevation API; Average acceleration is calculated as the change between the entering speed and the outgoing speed on a link; $f(x)$ is the target value, i.e., vehicle CO_2 emission per kilometre; *l* is the number of training samples; The coefficients ω and b are estimated by minimizing the regularized risk function:

$$\text{Minimize } \frac{1}{2}\|\omega\|^2 + C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon\left(y_i, f(x_i)\right) \tag{7-11}$$

$$L_\varepsilon\left(y_i, f(x_i)\right) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{7-12}$$

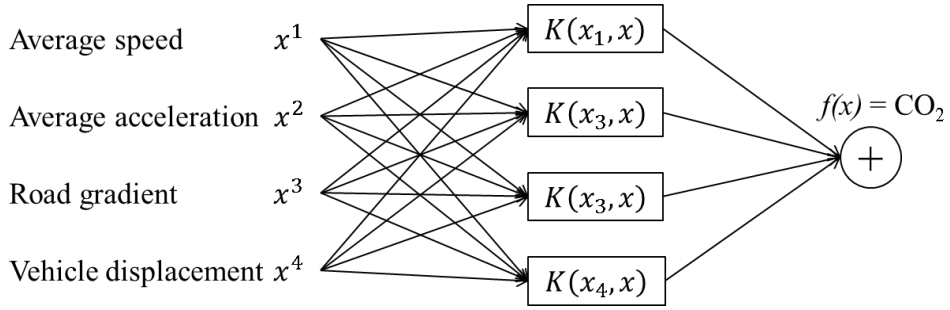


Figure 7-1 Structure of SVM model

where $\|\omega\|^2$ denotes the regularized term. Minimizing $\|\omega\|^2$ will make a function as flat as possible, which plays the role of controlling the generalization capacity. The second term $C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon\big(y_i, f(x_i)\big)$ is the empirical error measured by the ε-insensitive (ε-SV) loss function (Vapnik, 2000). The constant C > 0 determines the tradeoff between the flatness of f(x) and the amount up to which deviations larger than ε are allowed. Increasing the value of C will lead to the relative importance of the empirical risk with respect to the regularization term to increase.

To get the estimation of $\boldsymbol{\omega}$ and b, Eq.(7-11) is transformed to the primal objective function by introducing the positive slack variables ξ_i and ξ_i^*. Then the above problem can be formalized as:

$$\text{Minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{7-13}$$

Subject to

$$\begin{cases} y_i - \omega \cdot \phi(x_i) - b \le \varepsilon + \xi_i \\ \omega \cdot \phi(x_i) + b - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{7-14}$$

The key to solve the optimization problem is to construct a Lagrangian function from the objective function and the corresponding constraints. By introducing Lagrangian multipliers and exploiting the constraints, the Lagrangian formulation has the following form:

$$L = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) - \sum_{i=1}^{l} a_i(\varepsilon + \xi_i - y_i + \omega \cdot \phi(x_i) + b) - \sum_{i=1}^{l} a_i^*(\varepsilon + \xi_i + y_i - \omega \cdot \phi(x_i) - b) \tag{7-15}$$

where η_i, η_i^*, a_i and a_i^* are Lagrangian multipliers. The dual variables in Eq.(7-15) have to satisfy positive constraints:

$$\eta_i^{(*)}, a_i^{(*)} \geq 0 \tag{7-16}$$

where $(*)$ denotes variables with and without *.

To get the optimality, take the partial derivatives of L with respect to the primal variables (ω, b, ξ_i, ξ_i^*):.

$$\partial_b L = \sum_{i=1}^{l}(a_i^* - a_i) = 0 \tag{7-17}$$

$$\partial_\omega L = \omega - \sum_{i=1}^{l}(a_i - a_i^*)\phi(x_i) = 0 \tag{7-18}$$

$$\partial_{\xi_i^{(*)}} L = C - a_i^{(*)} - \eta_i^{(*)} = 0 \tag{7-19}$$

Substituting Eq.(7-17), Eq. (7-18) and Eq. (7-19) into Eq.(7-15), the dual optimization problem can be written as:

$$\text{Minimize } W(a_i, a_i^*) = -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(a_i - a_i^*)(a_j - a_j^*)\,K(x_i, x_j) - \varepsilon\sum_{i=1}^{l}(a_i^* + a_i) + \sum_{i=1}^{l} y_i(a_i - a_i^*) \tag{7-20}$$

Subject to

$$\begin{cases}\sum_{i=1}^{l}(a_i - a_i^*) = 0 \\ a_i, a_i^* \in [0, C]\end{cases} \tag{7-21}$$

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \tag{7-22}$$

where $K(x_i, x_j)$ is the kernel function. The value of the kernel is equal to the inner product of two vectors and in the feature space $\phi(x_i)$ and $\phi(x_j)$. Kernel functions enable the dot product to be performed in high-dimensional feature space using low-dimensional space data input without having to compute the map explicitly. Most of the previous researches selected radial basis function (RBF) as the kernel model for regression because it maps samples into a higher dimensional space and it has less numerical difficulties in contrast to polynomial kernels whose values may go to infinity or zero. Thus, RBF kernel is used for the SVM model in this study. There are three parameters while using RBF kernel: (C, ε, γ). However, it is not known a priori which C, ε, and γ are the best choice for the problem. To properly select the three parameters, the grid-search method is used for parameter determination. In grid-search, all pairs of (C, ε, γ) are tested and the one with the best performance is picked up (Chang and Lin, 2011).

7.4 Eco-routing problem

7.4.1 Problem statement

Our objective is to find the minimum CO_2 emissions path between two nodes in a transportation network within the constraint of a certain travel time budget. The transportation network is modeled as a directed graph $G(N, A)$, where N=$\{n_1, n_2, \ldots, n_n\}$ represents the set of nodes and A=$\{a_{12}, a_{23}, \ldots, a_{mn}\}$ represents the set of links. Unlike a traditional network that only has a single attribute assigned to each link, such as travel time or length, the network considered in this study has two attributes: link CO_2 emissions c_{ij} and link travel time t_{ij}. Here, we wish to search the most eco-friendly path within the travel time budget (T). Consequently, the eco-routing problem from origin r to destination s can be described as the following problem:

$$\text{P1: } Min\ Z(x) = \sum_{ij \in A} c_{ij} x_{ij} \tag{7-23}$$

Subject to

$$\sum_{ij \in A} t_{ij} x_{ij} \leq T \tag{7-24}$$

$$\sum_{(i,j) \in A} x_{ij} - \sum_{(j,i) \in A} x_{ji} = g \tag{7-25}$$

$$g = \begin{cases} 1 & i = r \\ 0 & i \in N - \{r, s\} \\ -1 & i = s \end{cases} \tag{7-26}$$

where $x_{ij} \in \{0,1\}$ indicates a link on the selected path and g denotes the flow direction for each node i in the road network.

In this study, the concept of the Pareto frontier and the corresponding weighting method is extended to solve the eco-routing problem with a travel time constraint. To utilize a Pareto-optimal approach to this eco-routing problem, such as by implementing a weighting method, we transform P1 into a bi-objective like optimization problem as follows:

$$\text{P2: Min } Z(x) = w_t Z_t(x) + w_c Z_c(x) \tag{7-27}$$

Subject to constraints (7-24)-(7-26).

Where,

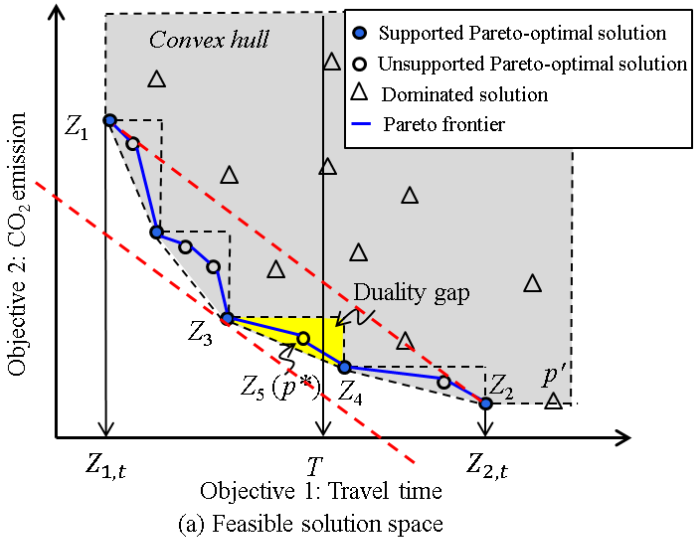
$$Z_t(x) = \sum_{ij \in A} t_{ij} x_{ij} \quad (7\text{-}28)$$

$$Z_c(x) = \sum_{ij \in A} c_{ij} x_{ij} \quad (7\text{-}29)$$

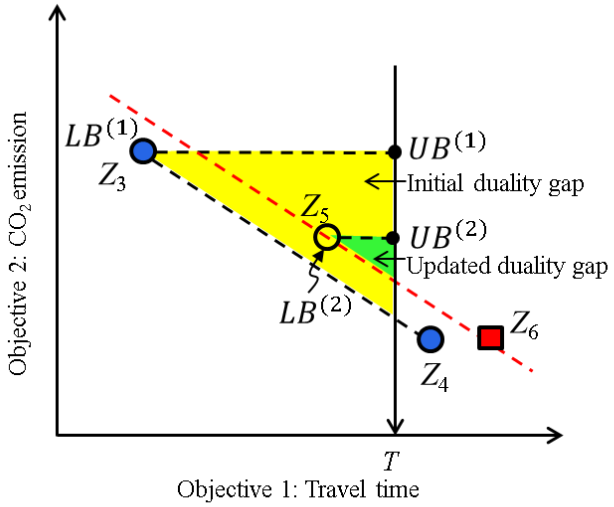
w_t: weighting parameter for the objective of travel time;

w_c: weighting parameter for the objective of CO_2 emissions.

Note that if constraint (7-24) is relaxed, solving P2 will generate a Pareto frontier (Pareto-optimal or non-dominated path set). Considering that this study aims to find an optimal path but not the path set, we design a heuristic algorithm to search the optimal solution along the Pareto frontier. The heuristic search procedure will stop when the travel time of a candidate Pareto-optimal solution reaches the specified travel time budget (T).



(a) Feasible solution space



(b) Optimal solution in duality gap

Figure 7-2 Search procedure used by the Pareto-optimal based heuristic algorithm

7.4.2 Routing approach

Prior to introducing the Pareto-optimal based heuristic approach, we first list some fundamental definitions relating to the concept of Pareto optimality.

Definition 1 (Pareto-optimal solution and Pareto frontier). Given a bi-objective shortest path problem, a path $p^* \in P$, where P is the path set from origin r to destination s, is called a Pareto-optimal solution if there is no other path $p \in P$ with $Z_t(x) \leq Z_t(x^*)$ and $Z_c(x) \leq Z_c(x^*)$ with at least one inequality being strict. An equivalent condition to the Pareto-optimal solution p^* is that there is no other feasible solution $p \in P$ that dominates p^*. The set of Pareto-optimal or non-dominated solutions is called the Pareto frontier.

Definition 2 (Supported and unsupported Pareto-optimal solutions) (Sedeno-Noda and Raith, 2015). Supported Pareto-optimal solutions (supported non-dominated solutions) are those feasible solutions that can be obtained from a weighted sum optimization problem Min $(w_t Z_t + w_c Z_c)$ for $w_t, w_c > 0$. All other Pareto-optimal solutions or non-dominated solutions are called unsupported Pareto-optimal solutions.

As illustrated in Figure 7-2(a), the supported Pareto-optimal solutions lie on the lower left boundary of the convex hull of the feasible solution set, whereas the unsupported Pareto-optimal solutions lie in the triangle areas determined by two adjacent supported Pareto-optimal solutions. The Pareto frontier comprises all the Pareto-optimal solutions including both supported and unsupported solutions. Assuming that the traveler prefers a fuel-efficient path (e.g., p^*) with the lowest CO2 emissions within the travel time constraint (T), the optimal path can be found along the Pareto frontier. A weighting method (Coutinho-Rodrigues et al., 1999) can be applied to efficiently determine the Pareto frontier, using a weighting utility function including all the objectives so as to transform a multi-objective problem into a single objective shortest-path problem. Unfortunately this method does not identify the unsupported Pareto-optimal solutions that lie in the interior of the convex hull. For example, when the travel time budget is set to T, the optimal solution should be p^* in Figure 7-2(a). However, p^* is an unsupported Pareto-optimal solution in this case, so it will not be found by the weighting method. To fill this gap, a three-step Pareto-optimal based algorithm is proposed in the following sections.

Step 1: Initialization

The algorithm gets started by searching two initial Pareto-optimal solutions with two weighting parameter sets, i.e., $(w_t, w_c) = (\varepsilon, 1-\varepsilon)$ and $(w_t, w_c) = (1-\varepsilon, \varepsilon)$, where ε is a sufficiently small number, i.e., $0 < \varepsilon \ll 1$. Note that a shortest path minimizing the CO_2

emissions ($Z_c(x)$) may be obtained when we solve the weighting program for $\varepsilon = 0$, but the corresponding shortest path could be a dominated path as the second objective is not taken into account. For example, as shown in Figure 7-2(a), p' could be found if $w_t = 0$, but p' is not the Pareto-optimal solution because it is dominated by Z_2. Therefore, we use a small ε value in the initial weighting parameter sets to avoid the risk of selecting a dominated solution at the initial iteration. Assumed that the initial Pareto-optimal solutions are Z_1 and Z_2, three situations should be considered for different settings of travel time budgets. (1) If the travel time budget is less than the travel time of Z_1, i.e., $T < Z_{1,t}$, no available paths can be found; (2) If the travel time budget is between the travel time of Z_1 and Z_2, i.e., $Z_{1,t} \le T \le Z_{2,t}$, it is easy to show that the optimal path will be located on the Pareto frontier as shown in Fig. 1(a); (3) If the travel time budget is more than the travel time of Z_2, i.e., $T > Z_{2,t}$, the optimal path is Z_2 which can be found directly by using the shortest-path algorithm with the weighting parameter set $(w_t, w_c) = (\varepsilon, 1-\varepsilon)$. Here, we put the focus on how to find the optimal path for the second situation. The two resulting solutions, i.e. Z_1 and Z_2, are recorded into a list L sorted ascending by path travel time, i.e., $L = \left\{Z_a^{(0)}, Z_b^{(0)}\right\}$, where $Z_a^{(0)} = Z_1$, $Z_b^{(0)} = Z_2$, and $Z_{a,t}^{(0)} \le Z_{b,t}^{(0)}$. And then a heuristic algorithm approximates the true optimal solution in next steps.

Step 2: Find the initial duality gap

To ensure a rapid approach to the optimal solution, a NISE-like methodology (Coutinho-Rodrigues et al., 1999) is used to update the weighting parameters. The weighting parameters are generated from the two Pareto-optimal solutions in list L iteratively. The most widely used method for generating weighting parameters is the perpendicular approach (Xie and Waller, 2012b), which leads to a parameter vector perpendicular to the line going through the two Pareto-optimal solution points (see the red dash line in Figure 7-2(a)). Specifically, given two Pareto-optimal solutions in list L, a new parameter set is generated by using the following linear functions:

$$\begin{cases} w_t^{(n)}\left(Z_{a,t}^{(n-1)} - Z_{b,t}^{(n-1)}\right) + w_c^{(n)}\left(Z_{a,c}^{(n-1)} - Z_{b,c}^{(n-1)}\right) = 0 \\ w_t^{(n)} + w_c^{(n)} = 1 \end{cases} \tag{7-30}$$

Thus,

$$w_t^{(n)} = \frac{Z_{b,c}^{(n-1)} - Z_{a,c}^{(n-1)}}{\left(Z_{b,c}^{(n-1)} - Z_{a,c}^{(n-1)}\right) + \left(Z_{a,t}^{(n-1)} - Z_{b,t}^{(n-1)}\right)} \tag{7-31}$$

$$w_c^{(n)} = \frac{Z_{a,t}^{(n-1)} - Z_{b,t}^{(n-1)}}{\left(Z_{b,c}^{(n-1)} - Z_{a,c}^{(n-1)}\right) + \left(Z_{a,t}^{(n-1)} - Z_{b,t}^{(n-1)}\right)} \tag{7-32}$$

where,

$Z_{a,t}^{(n-1)}$: Path travel time of $Z_a^{(n-1)}$ in list L;

$Z_{b,t}^{(n-1)}$: Path travel time of $Z_b^{(n-1)}$ in list L;

$Z_{a,c}^{(n-1)}$: Path CO_2 emission of $Z_a^{(n-1)}$ in list L;

$Z_{b,c}^{(n-1)}$: Path CO_2 emission of $Z_b^{(n-1)}$ in list L;

n: Iteration number.

The weighting single objective shortest-path problem specified by the updated $w_t^{(n)}$ and $w_c^{(n)}$ can be solved iteratively. A heuristic strategy is taken to identify the two candidate supported Pareto-optimal solutions that are closest to the true optimal solution. That is, the new supported Pareto-optimal solution is added to list L and meanwhile the previous solution, which has a larger absolute difference from the travel time budget, is removed from list L iteratively. For example, as shown in Figure 7-2(a), Z_3 is obtained by solving the weighting shortest path problem with updated $w_t^{(1)}$ and $w_c^{(1)}$. And then Z_3 is added to L, replacing Z_1 because $|Z_{1,t} - T| > |Z_{2,t} - T|$. Thus, the items in L are updated, i.e., $Z_a^{(1)} = Z_3$, and $Z_b^{(1)} = Z_2$ at the first iteration. The iteration procedure is ended if no new supported Pareto-optimal solutions can be found. For example, since there are no other new supported Pareto-optimal solutions that can be found by updating the weighting parameters generated by Z_3 and Z_4, the parameter generation procedure is stopped. As shown in Figure 7-2(a), the optimal solution Z_5 should be located in the yellow triangle region, which is called the duality gap, though Z_3 is a potential optimal solution. The duality gap can be determined using the two adjacent supported Pareto-optimal solutions in list L. It must be noted that Z_3 may not be the optimal solution because the heuristic algorithm only identifies those supported Pareto-optimal solutions that lie on the boundary of the convex hull, but it cannot find unsupported Pareto-optimal solutions, e.g., Z_5. For this reason, we present a procedure to identify unsupported Pareto-optimal solutions in the duality gap and determine the optimal solution in the next step.

Step 3: Update the duality gap and determine the optimal solution

An efficient algorithm for closing the duality gap was introduced by Current et al. (1990) for the bi-objective routing problem, in which the upper bound is updated by checking the trade-off of two objective values according to the traveler's preference

interactively. This inspires us to develop a heuristic approach to close the duality gap for the constrained eco-routing problem. In our algorithm, the upper bound and lower bound are updated according to the updated solution in the duality gap, but the weighting parameters do not need to update. And the duality gap is further limited to a smaller triangle region bounded by the travel time budget as shown in Figure 7-2(b). Considering that the optimal solution may locate in the duality gap, a k-shortest path algorithm (Yen, 1971) is applied to search for potential optimal solutions until there are no further new unsupported Pareto-optimal solutions to be found in the duality gap. Specifically, the k-shortest path searching procedure stops when the upper bound (UB) of the duality gap has been reached. At each iteration, the upper bound is determined by the two candidate solutions in list $L = \left\{Z_a^{(n)}, Z_b^{(n)}\right\}$ and the travel time budget (T):

$$UB^{(n)} = w_t^{(n)} T + w_c^{(n)} Z_{a,c}^{(n)} \tag{7-33}$$

Any k-shortest path algorithm for a directed network can be used in this study. The objective function for the k-shortest path algorithm is as follows:

$$\text{Min } Z(x) = \sum_{ij \in A}\left(w_t^{(n)} t_{ij} + w_c^{(n)} c_{ij}\right) x_{ij} \tag{7-34}$$

Subject to constraints (7-24)-(7-26).

Here, we use Yen's well-known algorithm (Yen, 1971) to find the unsupported Pareto-optimal solution. The UB is updated when a new unsupported Pareto-optimal solution is found in the duality gap by the k-shortest path algorithm. For example, at the first iteration, $Z_a^{(n)}$ in list L is updated to Z_5 and the UB is updated to $UB^{(1)}$ when Z_5 is found. Correspondingly, the lower bound (LB) can be defined as the weighted sum cost of the current unsupported Pareto-optimal solution, i.e., $Z_a^{(n)}$.

$$LB^{(n)} = w_t^{(n)} Z_{a,t}^{(n)} + w_c^{(n)} Z_{a,c}^{(n)} \tag{7-35}$$

Note that in using the k-shortest path searching procedure, it is possible that solutions outside the duality gap might be found, e.g., Z_6. These solutions are not added to list L and the UB is not updated; instead the next kth shortest path is calculated until a feasible solution is found in the duality gap.

7.4.3 Solution algorithm

The solution algorithm is based on the discussion presented above, which is presented below in pseudo code as Algorithm 7-1.

Algorithm 7-1 Pareto-optimal based heuristic algorithm

1	**Step 1**: Initialization
2	Initialize $\boldsymbol{L} \coloneqq \{Z_a, Z_b\}$, $Z_a \coloneqq null$, $Z_b \coloneqq null$;
3	Set travel time budget: $T \coloneqq t_0$;
4	Set $w_t := 1$, $w_c := \varepsilon$;
5	Find the minimal travel time solution: $Z_1(Z_{1,t}, Z_{1,c})$;
6	Set $w_t := \varepsilon$, $w_c := 1$;
7	Find the minimal CO_2 emission solution: $Z_2(Z_{2,t}, Z_{2,c})$;
8	Update $\boldsymbol{L}$, $Z_a \coloneqq Z_1$, $Z_b \coloneqq Z_2$;
9	**If** $(T < Z_{1,t})$
10	Return $null$;
11	**End if**
12	**If** $(Z_{1,t} \le T \le Z_{2,t})$
13	Go to **Step 2**;
14	**End if**
15	**If** $(T > Z_{2,t})$
16	Return Z_2;
17	**End if**
18	**Step 2**: Find the initial duality gap
19	**While** $(Z_a \neq Z_b)$
20	Update weighting parameters w_t and w_c using Eqs. (7-31) and (7-32);
21	Find the supported Pareto-optimal solution for Min $Z(x) = w_t Z_t(x) + w_c Z_c(x)$: Z';
22	**If** ($Z' = Z_a$ or $Z' = Z_b$)
23	Break;
24	**Else**
25	**If** $(\lvert Z_{a,t} - T \rvert > \lvert Z_{a,t} - T \rvert)$
26	Update $\boldsymbol{L}$, $Z_a \coloneqq Z'$;
27	**Else**
28	Update $\boldsymbol{L}$, $Z_b \coloneqq Z'$;
29	**End if**
30	**End if**
31	**End while**
32	**Step 3**: Update the duality gap and determine the optimal solution
33	Initialize $UB := w_t T + w_c Z_{a,c}$ and $LB := w_t Z_{a,t} + w_c Z_{a,c}$; $k := 1$;
34	**while** $(LB < UB)$;
35	Find the k^{th} best solution for Min $Z^{(k)}(x) = \sum_{ij \in A}(w_t t_{ij} + w_c c_{ij}) x_{ij}$;
36	**If** $(Z^{(k)} = null)$;
37	Break;

38	**End if**
39	Update $LB := w_t Z_t^{(k)} + w_c Z_c^{(k)}$;
40	**If** ($Z_{a,t}^{(k)} < T$ and $Z_{a,c}^{(k)} < Z_a$)
41	Update $\boldsymbol{L}$, $Z_a := Z^{(k)}$;
42	Update UB, $UB := w_t T + w_c Z_{a,c}$;
43	**End if**
44	$k := k + 1$
45	**End while**
46	Return the optimal solution: Z_a;

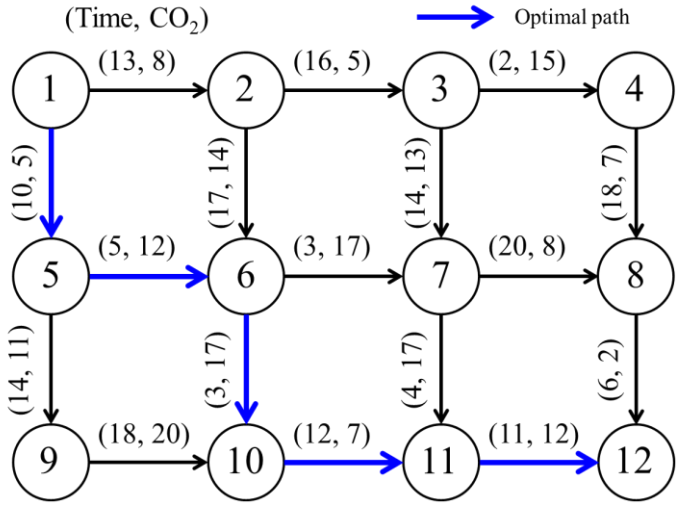


Figure 7-3 Illustration of the tested network and the optimal path

7.4.4 An illustrative example for the proposed eco-routing approach

An example of an eco-routing problem with a travel time constraint is shown here to illustrate the effectiveness of the proposed search procedure. The tested network and the link attributes are shown in Figure 7-3. The origin and destination are set to node 1 and node 12, respectively. The objective is to find the most eco-friendly path within the travel time budget (e.g., T=42). The solutions in outcome space are shown in Figure 7-4, and the iteration results are shown in Table 7-3.

In the first step, given the initial weighting parameter sets, ($w_t = 0.999, w_c = 0.001$) and ($w_t = 0.001, w_c = 0.999$), the first two supported Pareto-optimal solutions, i.e., Z_1 and Z_2 shown in Figure 7-4, are identified by solving the weighted sum shortest path problem. Then, the travel times of the two solutions are checked against the specified travel time budget. Since the travel time budget is between the travel times of these two Pareto-solutions, i.e., 33<T<69, Z_1 and Z_2 are added to list L, and the calculation procedure goes to step 2.

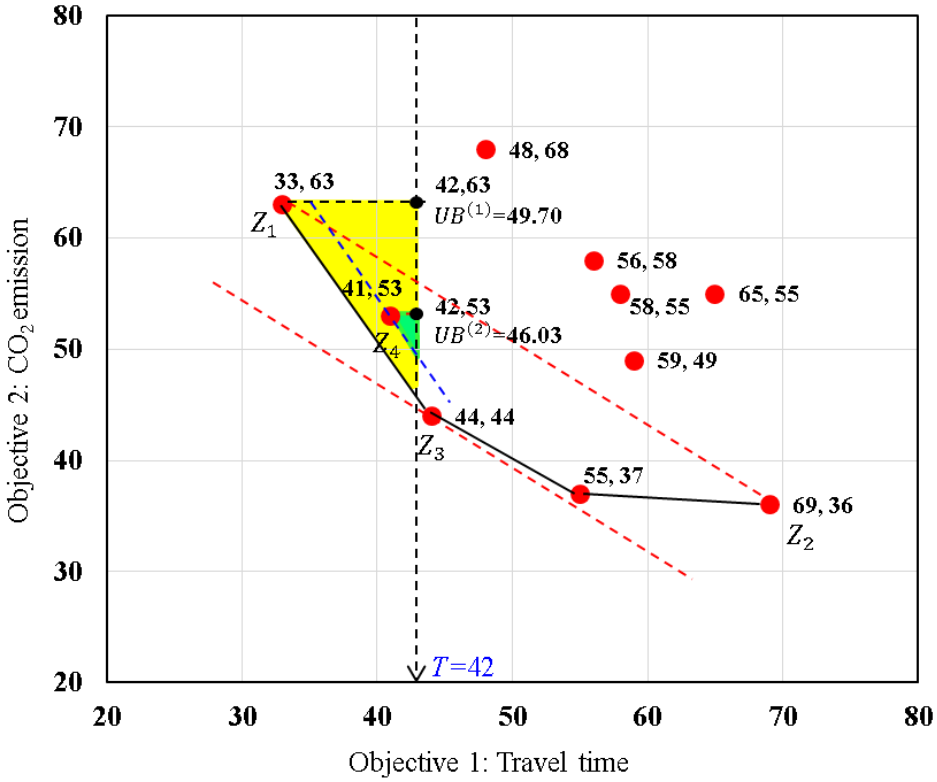


Figure 7-4 Solutions in outcome space

Table 7-3 Solution process and iteration results for illustrative eco-routing example

Step	Iteration	w_t	w_c	LB	UB	List $\boldsymbol{L}$		Objective value			
						Z_a	Z_b	$Z_{a,t}$	$Z_{a,c}$	$Z_{b,t}$	$Z_{b,c}$
1	1	1.000	0.010	null	null	1-5-6-7-11-12	null	33	63	null	null
1	2	0.010	1.000	null	null	1-5-6-7-11-12	1-2-3-7-8-12	33	63	69	36
2	1	0.429	0.571	null	null	1-5-6-7-11-12	1-5-6-7-8-12	33	63	44	44
2	2	0.633	0.367	null	null	1-5-6-7-11-12	1-5-6-7-8-12	33	63	44	44
3	1	0.633	0.367	44.00	49.70	1-5-6-10-11-12	1-5-6-7-8-12	41	53	44	44
3	2	0.633	0.367	45.40	46.03	1-5-6-10-11-12	1-5-6-7-8-12	41	53	44	44

In step 2, the weighting parameters are updated using Eqs. (7-31) and (7-32), i.e., $w_t = 0.43$ and $w_c = 0.57$. Then, the next supported Pareto-optimal solution is found, i.e., Z_3 shown in Figure 7-4, using the shortest-path algorithm. Comparing the travel time budget with the travel times of the two solutions in list L, i.e., Z_1 and Z_2 shown in Figure 7-4, it is found that the travel time of Z_1 is closer than that of Z_2. Thus, Z_2 is removed and Z_3 is added to list L. The weighting parameters are updated iteratively using the objective values of the two candidate solutions in list L, i.e., Z_1 and Z_3. In iteration 2, because no other solutions can be found using the updated weighting parameters, i.e., $w_t = 0.63$ and $w_c = 0.37$, the duality gap can be determined by Z_1 and Z_3, and the calculation procedure goes to step 3.

In step 3, we find the optimal solution inside the duality gap. First, the UB value of the weighted sum cost of the k-shortest path is set using Eq. (7-33), i.e. UB=0.633*42+0.367*63=49.7. The LB value is set as the weighted sum cost of one of the solutions in list L, i.e., LB=0.633*33+0.367*63=44. Then, the k-shortest path algorithm searches iteratively for the kth best path constrained by UB and LB. If the kth best path falls inside the duality gap, i.e., Z_4 shown in Figure 7-4, the first solution in list L is replaced by the kth best path, i.e., $Z_a = Z_4$. And then UB and LB are updated by the new solutions in list L, i.e., UB=0.633*42+0.367*53=46.03, B=0.633*41+0.367*53=45.4. The k-shortest path algorithm is called again with the updated UB and LB. Because there are no more solutions to be found in the updated duality gap, the algorithm stops and the optimal solution, i.e., Z_4, is found.

7.5 Numerical experiment and discussion

7.5.1 Comparison of CO_2 emission models

The estimation performance is evaluated by using the following statistical metrics, namely, the mean absolute percentage error (MAPE) and squared correlation coefficient (r^2). The smaller the values of MAPE, the closer are the estimated values to the observed values. r^2 provides an index of the correlation between the predicted values and the observed values. A larger value closing to one represents a better predictor.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{7-36}$$

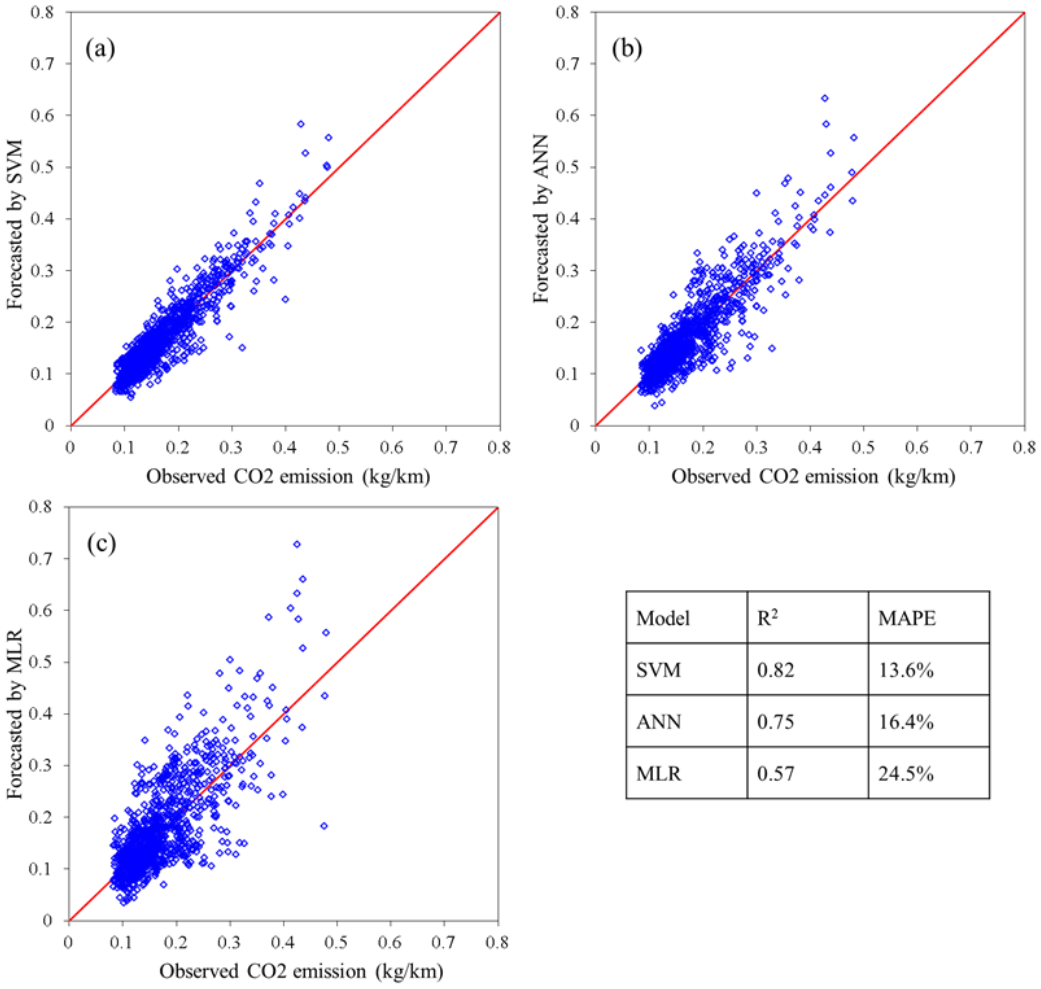
$$r^2 = \frac{\left(n\sum_{i=1}^{n}\hat{y}_i y_i - \sum_{i=1}^{n}\hat{y}_i\sum_{i=1}^{n}y_i\right)^2}{\left(l\sum_{i=1}^{n}(\hat{y}_i)^2-\left(\sum_{i=1}^{n}\hat{y}_i\right)^2\right)\left(l\sum_{i=1}^{n}(y_i)^2-\left(\sum_{i=1}^{n}y_i\right)^2\right)} \tag{7-37}$$

where *n* is the total number of samples, y_i and $\hat{y}_i$ represent the actual and the estimated outputs, respectively.

To demonstrate the advantage of the proposed SVM model, the multiple linear regression (MLR) model and artificial neural network (ANN) model, are used for performance comparison. The MLR model can be expressed as follows:

$$\hat{y}_{MLR} = w_0 + \sum_{i=1}^{m} w_i x_i \tag{7-38}$$

where $(x_1, x_2, \ldots, x_m)$ denotes the input vector, $\{w_0, w_1, \ldots, w_m\}$ are the model coefficients to be estimated, $\hat{y}_{MLR}$ denotes the CO_2 emission.



Model	R^2	MAPE
SVM	0.82	13.6%
ANN	0.75	16.4%
MLR	0.57	24.5%

Figure 7-5 Model performance comparison

Table 7-4 Car models

model	Samples	Displacement (cc)	Ideal fuel consumption rate (L/100km)	Weight (kg)
Corolla Fielder-1NZ	8791	1496	6.2	1280
Corolla Fielder-2ZR	615	1797	6.7	1330
IQ	3188	996	5.7	860
ISIS-2ZR	1646	1797	6.9	1560
ISIS-3ZR	410	1986	7.9	1560
MarkX	26	2499	9.1	1520
Noah	9864	1986	8	1500
Premio	2688	1986	7.7	1440
RAV4	3378	2362	6.5	1150
Vitz-1KR	4241	996	4.6	980
Vitz-1NZ	4376	1496	5.8	1050
Vitz-2SZ	4960	1296	5.6	1030
Voxy	20443	1986	7	1500
WISH	5430	1797	6.6	1355

The classical backpropagation neural network will be used to establish the ANN model (Cortez et al., 2009). The structure of ANN model can be written as follows:

$$\hat{y}_{ANN} = w_{o,0} + \sum_{j=m+1}^{o-1} f\left(\sum_{i=1}^{m} x_i w_{j,i} + w_{j,0}\right) w_{o,i} \quad (7\text{-}39)$$

$$f(x) = \frac{1}{e^{-x}+1} \quad (7\text{-}40)$$

where $w_{j,i}$ denotes the weight of the connection from node j to i, and o is the output node. The number of hidden nodes (H) is determined by a regularization method (Hastie et al., 2001).

Figure 7-5 shows the goodness-of-fit for the vehicle CO_2 emission predictions by the three models. The vehicle models are shown in Table 7-4. The link-based sample size includes 70056 records. 90% of the samples are used for training and 10% are used for testing. It can be seen from the r^2 that the prediction of the SVM model is much closer to the observed target than the other two models. The MAPEs are 13.6%, 16.4%, and 24.5%, respectively. The advantage of SVM over ANN is likely due to the differences in the training procedure. The training of SVM ensures an optimum fit, while ANN training might fall into a local optimum. On the other hand, the objective function of SVM endows a linear penalty to out-of-range errors. Thus, SVM is expected to be less sensitive to outliers and such effect will lead to a higher accuracy for low error tolerances. MLR model is the worst one since it cannot capture the nonlinear features of the explanatory variables. Therefore, it can be concluded that the proposed SVM structure can provide a better result.

7.5.2 Sensitivity analysis for CO_2 emission model

To reveal how the selected factors influence the CO_2 emission, this study proposes the following simple methodology for sensitivity analysis. Once a SVM structure has been trained on a large set of input variables, calculate an average value for each input variable. Then, holding all variables at their average values but one each time, vary the one input over its entire range and analyze the variability produced in the outputs.

Here, sensitivities of the four explanatory variables are examined. Figure 7-6(a) portrays a bowl-shape variation in the CO_2 emission. The optimal CO_2 emission appears to occur at the average speed of approximate 16.9m/s (61km/h). This finding indicates that an increase in average travel speed from 1m/s to 16.9m/s could result in a decrease of 78% in the CO_2 emission, while a decrease in average travel speed from 16.9m/s to 30m/s could also result in an increase of 34% in the CO_2 emission. It indicates that driving too slowly or too fast will lead to high CO_2 emission. Figure 7-6 (b) illustrates the relationship between

the acceleration and CO_2 emission. The CO_2 emission increases slowly when the acceleration is negative, while it increases dramatically when the acceleration is between 0 and 0.5m/s^2. Interestingly, the CO_2 emission does not always increase as the acceleration increases. It is found that CO_2 emission decreases as the acceleration increases from 0.5m/s^2 to 2m/s^2. A possible reason is that the fuel efficiency during acceleration improves as the revolution per minute (RPM) increases, which results in fewer fuel consumption and CO_2 emission. Figure 7-6 (c) shows how the road gradient influences the CO_2 emission. As expected, the CO_2 emission approximately follows a linear relationship with the road gradient. It indicates that the CO_2 emission increases by 2% as a result of 1 degree increase in road gradient. Figure 7-6 (d) illustrates the trend of the CO_2 emission with vehicle displacement. It is found that the vehicle with lower displacement will result in lower CO_2 emission.

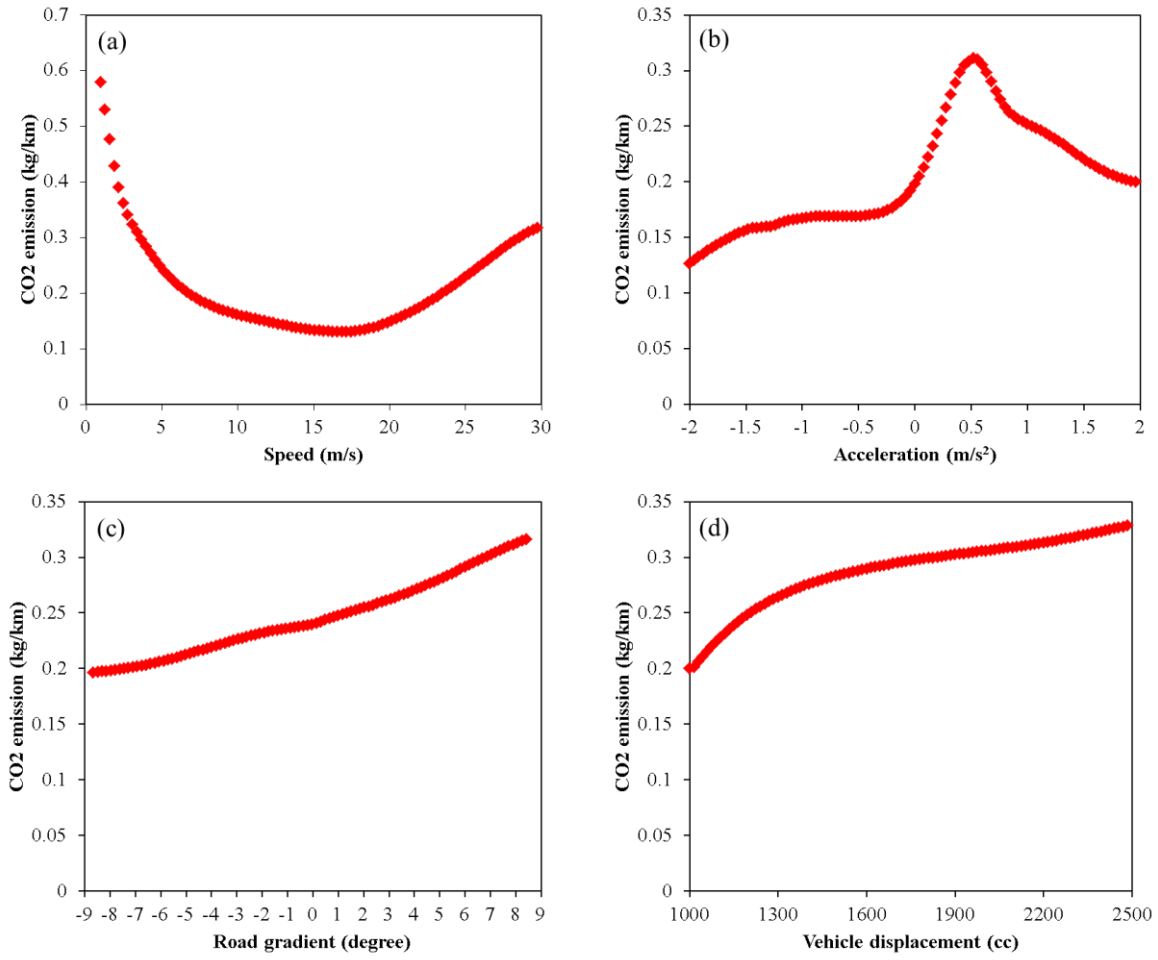


Figure 7-6 Sensitivity analysis on explanatory variables

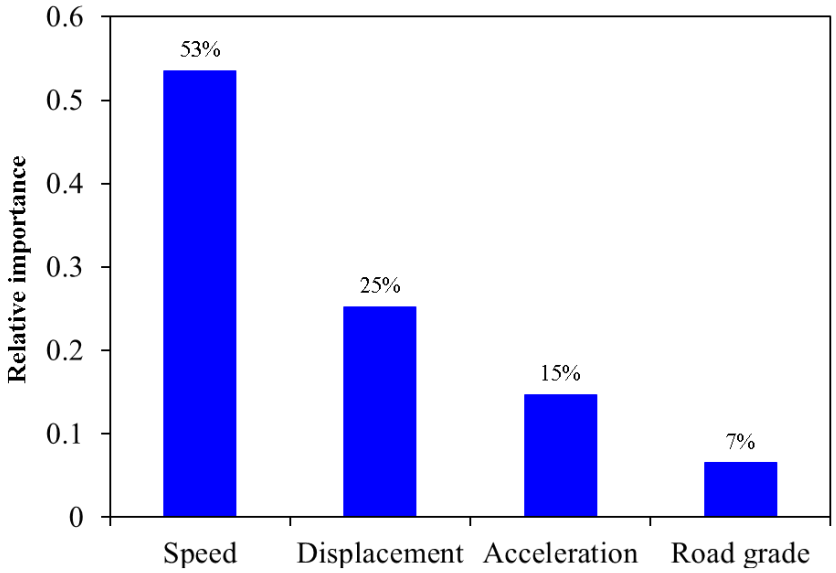


Figure 7-7 Relative importance for each explanatory variable

To better understand the relative importance of each explanatory variable to the trip fuel consumption, we conduct the elasticity analysis after training SVM. Elasticity measures the percentage change of a dependent variable to a percentage change in an explanatory variable. Accordingly, the elasticity can be written as follows:

$$E_a = \frac{df(\mathbf{x})}{f(\mathbf{x})} / \frac{dx_a}{x_a} \tag{7-41}$$

where E_a denotes the elasticity for variable x_a, a denotes the subscript for each explanatory variable, $f(\boldsymbol{x})$ is the regression model by SVM. The output of Eq. (7-41) can be obtained by numerical solution after training SVM.

Keeping all input variables at their mean values except x_a which changes through its whole range with $i \in \{1,\ldots, n\}$ intervals, the point elasticity (E_a^i) for each input x_a^i can be written as follows:

$$E_a^i = \frac{y_a^{i+1} - y_a^i}{x_a^{i+1} - x_a^i} \frac{x_a^i}{y_a^i} \tag{7-42}$$

If an explanatory variable ($x_a \in \{x_a^1,\ldots, x_a^n\}$) is relevant, then it should produce a larger sum of absolute point elasticity ($\left|E_a^i\right|$). Thus, its relative importance (R_a) can be represented by:

$$R_a = \frac{\sum_{i=1}^{n} |E_a^i|}{\sum_{a=1}^{m} \sum_{i=1}^{n} |E_a^i|} \tag{7-43}$$

where *n* is the number of samples, *m* is the number of explanatory variables.

The relative importance for explanatory variables in SVM models are shown in Figure 7-7. It demonstrates that the average speed plays the most important role in vehicle CO_2 emission, while the road gradient has slight impact. The average speed and vehicle displacement occupy 78% relative importance to the model. It indicates that a vehicle with lower displacement running in around 60km/h is likely to save fuel consumption and reduce the CO_2 emission. This ranking helps travelers determine an economic and eco-friendly path.

7.5.3 Performance analysis of eco-routing in a real-world network

In this study, we collect data from probe vehicles fitted with GPS and an on-board diagnostics (OBD) device. Link information such as link length and link travel time can be extracted after map-matching (Miwa et al., 2012). Link CO_2 emissions can be obtained from the OBD reader. Actually, the OBD reader is able to provide comprehensive controller area network (CAN) bus data, including fuel consumption and emissions, instantaneously. However, considering the applicability and robustness of the routing model for a navigation system, we aggregate the emissions data at the link level. A real-world network with 4072 nodes and 12,877 links in Toyota city, Japan, is used to verify the eco-routing procedure and evaluate the likely environment benefits of implementing it. The trip records used for verification are from 153 GPS probe vehicles over a period of 10 months.

In Figure 7-8, a randomly selected real-world OD pair in the central area of Toyota city is extracted to demonstrate the performance of the proposed eco-routing strategy. For this demonstration, the travel time budget is set as 1.2 times the observed travel time. That means we allow the eco-routing model to find an eco-friendly path with 20% extra travel time buffer relative to the observed path. As shown in Table 7-5, different routing strategies and their associated path performances are compared with the proposed eco-routing strategy. The shortest distance path takes a direct route through the central urban area with a high density of intersections, which may lead to great fluctuations in speed. The observed path selected by the driver has similar features to the shortest distance path. Even though the shortest path and the observed path are shorter, the high density of intersections and high value of coefficient of variance (COV) of link average speed will result in higher CO_2 emissions as compared to the proposed eco-friendly path. On the other hand, the least travel time path detours onto expressways to profit from their higher average travel speed. As compared with the shortest distance path, the travel time saving is 10.26% even though the distance traveled is 1.16 times longer. However, this path

generates the most CO_2 emissions because it is the longest among the four paths. The eco-friendly path, determined as proposed in this study, offers significantly better performance in regard to CO_2 emissions at the cost of a very small increase in travel time and a minor detour. The eco-friendly path reduces CO_2 emissions by 13.02%, 10.24%, and 18.92% relative to the observed path, the shortest distance path, and the least travel time path, respectively. In particular, in comparison with the observed path as selected empirically by the driver, the eco-friendly path offers a significant advantage in both travel time (reduced by 16.95%) as well as CO_2 emissions (reduced by 13.02%), though the distance traveled is a little longer (by 0.3km). We note that the saving in CO_2 emissions compared with the least travel time path naturally comes at the expense of increased travel time, since the objective of the routing algorithm is no longer travel time. This is consistent with previous studies (Ahn and Rakha, 2008; Yao and Song, 2013; Boriboonsomsin, 2012).

Figure 7-8 Case study for eco-routing

Table 7-5 Path performance comparison

Performance indices	OP	DP	TP	EP	Relative difference (%)		
					EP vs. OP	EP vs. DP	EP vs. TP
Trip distance (km)	8.71	8.00	9.33	9.01	3.42	12.58	-3.43
Trip average speed (km/h)	17.23	20.52	26.66	21.45	24.53	4.54	-19.52
Intersection density (km^{-1})	8.15	8.75	7.50	5.88	-27.82	-32.75	-21.60
COV of link average speed	0.63	0.74	0.58	0.43	-31.62	-42.16	-26.21
Travel time (h)	0.51	0.39	0.35	0.42	-16.95	7.69	20.00
CO_2 emission (kg)	1.33	1.29	1.43	1.16	-13.02	-10.24	-18.92

Note: OP = Observed path; DP = Shortest distance path; TP = Least travel time path; EP = Eco-friendly path; COV = Coefficient of variance

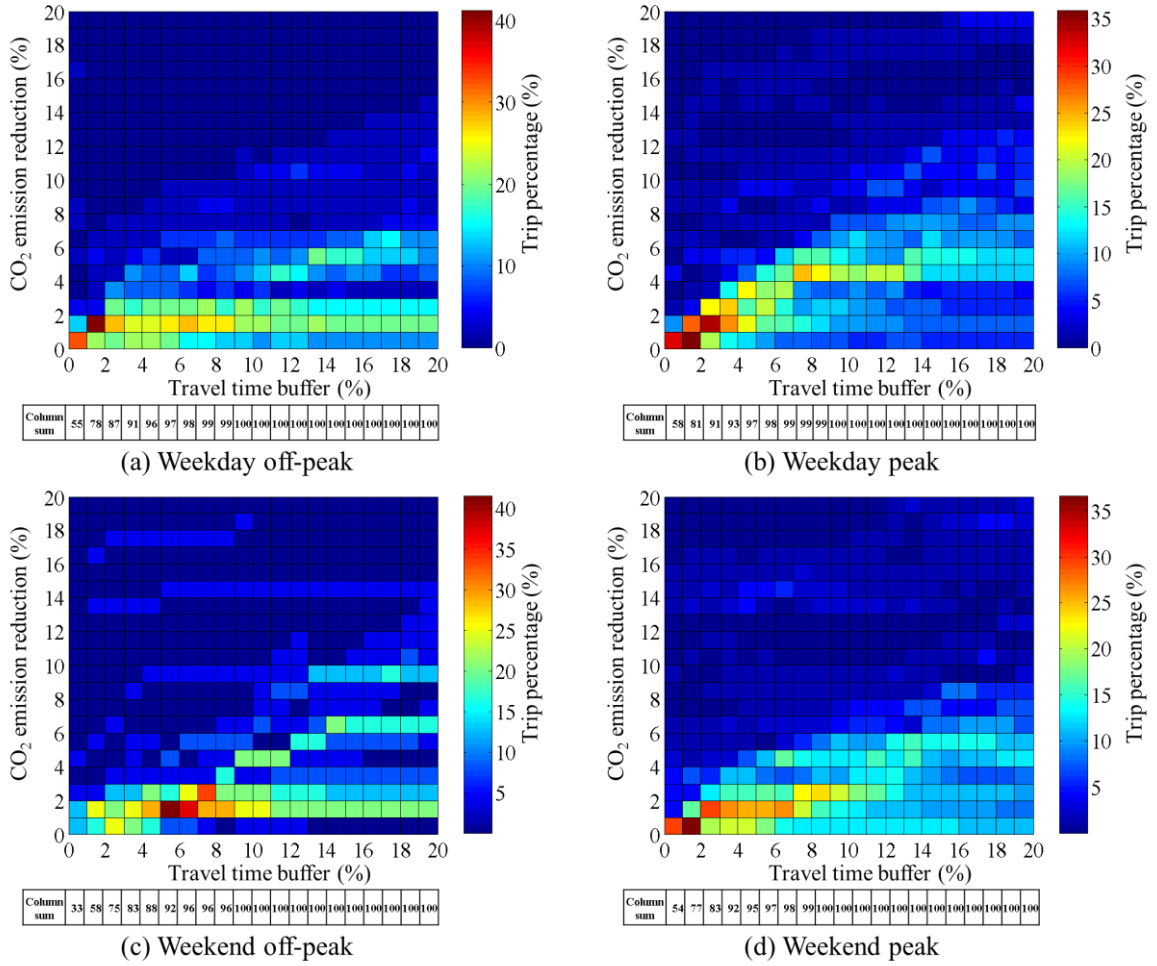


Figure 7-9 Impact of travel time buffer on the percentage of trips with CO_2 emission reduction

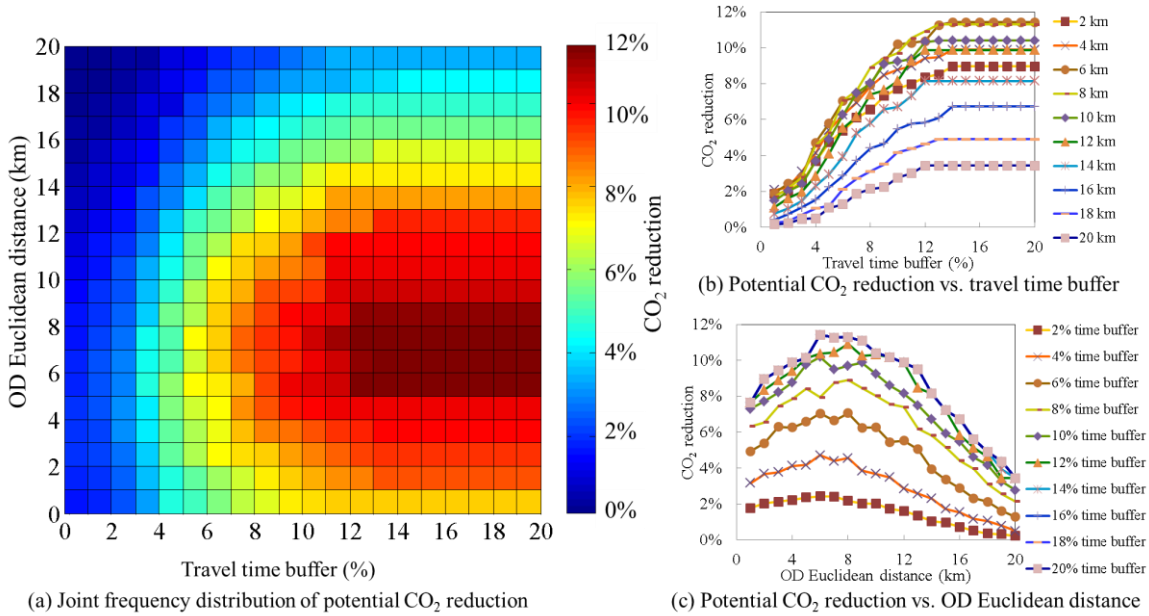


Figure 7-10 Impact of travel time buffer and OD distance on CO_2 reduction

7.5.4 Sensitivity analysis of potential CO_2 emission reduction

Because the benefit tradeoff between CO_2 emissions and travel time budget is important both to individual travelers and to the provision of routing guidance, the network-wide reduction in CO_2 emissions that can potentially be realized with various settings of travel time budget is analyzed in this section. The concept of a travel time buffer is used, defined as the percentage increment over the least travel time. OD pairs collected from 7989 real-world trip records are used to conduct this sensitivity analysis. For each trip, the eco-routing model calculates eco-friendly paths under various travel time buffers, and then the potential reduction in CO_2 emissions is calculated by comparing these with the least travel time path.

Figure 7-9 illustrates how many percentage of trips for which CO_2 emissions are reduced by a certain amount under different conditions. The x-axis denotes the percentage increment in travel time budget (the travel time buffer). The y-axis denotes the percentage reduction in CO_2 emissions compared to the least travel time path. The color of each cell denotes the percentage of trips represented by that cell. The single row table below each sub-graph denotes the cumulative percentage of trips with that buffer value for which CO_2 emissions are reduced. This shows that the percentage of trips with reduced CO_2 emissions at the cost of a very small increase in travel time. However, the degree of the CO_2 reduction varies by trips. For example, some trips can achieve a reduction of 15% while others only reach 5%, even when the travel time buffer is increased to 20% or more. This indicates that eco-routing does not always provide a significant CO_2 reduction as compared

to traditional routing based on travel time. Interestingly, if the travel time buffer is increased to 10%, almost all trips see some degree of CO_2 emissions reduction. This suggests that a relatively small increase in travel time enables network-wide CO_2 emissions to be effectively reduced by the eco-routing strategy. On the other hand, the cumulative percentage of the trips for which CO_2 emissions are reduced on weekdays increases faster with increasing travel time buffer than on weekends. This indicates that the overall reduction in CO_2 emissions on weekdays is more sensitive. Further, the cumulative percentage of trips for which CO_2 emissions are reduced in peak hours increases faster with increasing buffer than in off-peak hours. This suggests that the eco-routing strategy has a significant potential to reduce CO_2 emissions during peak hours.

Figure 7-10(a) shows the joint percentage distribution of potential CO_2 reduction for different OD distances and travel time buffers. The cell color denotes the percentage CO_2 emissions reduction comparing to the path with least travel time. In general, it is found that the eco-friendly path enables to reduce CO_2 emissions by an average of 11% for OD distances between 6km and 9km when the travel time buffer is greater than 10%. Figure 7-10(b) shows the curves of potential CO_2 reduction as the travel time buffer increases. There is a significant rise in CO_2 emissions reduction as the travel time buffer increases from 1% to 12%. However, with further rises in travel time buffer, the CO_2 reduction remains relatively stable. That is, few paths with lower CO_2 emissions can be found once the travel time buffer increases beyond a certain threshold (e.g., approximately 12% in the studied network). This indicates that a travel time buffer of 12% is appropriate for the eco-routing strategy, since this provides the greatest CO_2 emissions reduction for the least cost in travel time. On the other hand, compared to trips with longer OD distance, shorter trips have the potential for a larger CO_2 reduction percentage. For example, for trips with 6km OD distance the potential reduction in average CO_2 emissions is from 2% to 11%, while for trips with 20km OD distance the potential reduction is only from 0.2% to 3%. Figure 7-10(c) shows the trend of potential CO_2 reduction percentage as OD Euclidean increases. The curves are shaped like a mountain for all buffer values. The CO_2 reduction rises to a peak for OD distances up to 8km, then falls for OD distances greater than 8km. Trips with larger OD distances are not as sensitive to CO_2 reduction as shorter trips; this is because a larger percentage of shortest travel time paths and eco-friendly paths may overlap due to the same choice of expressway or major road.

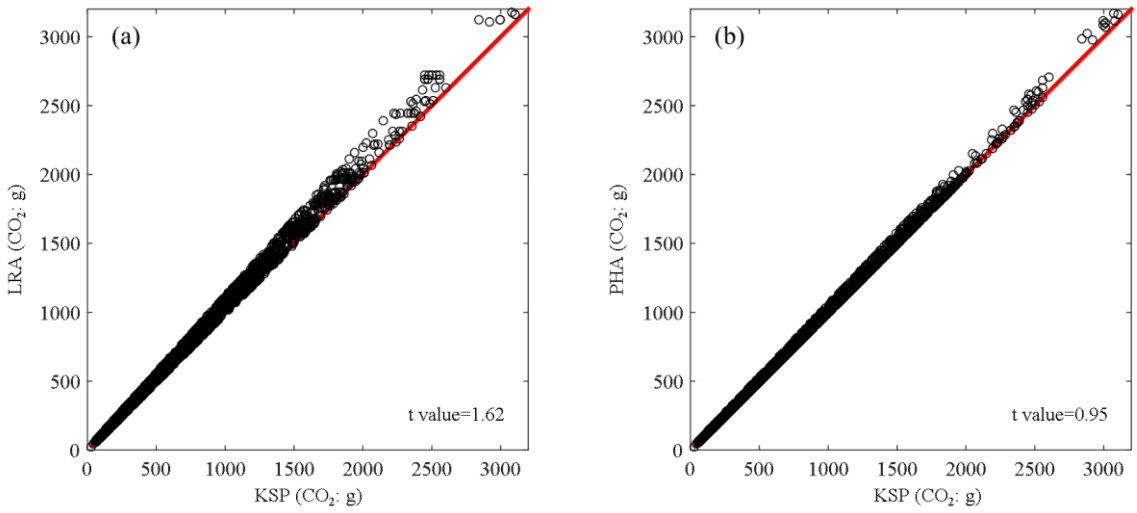


Figure 7-11 Solution quality comparison

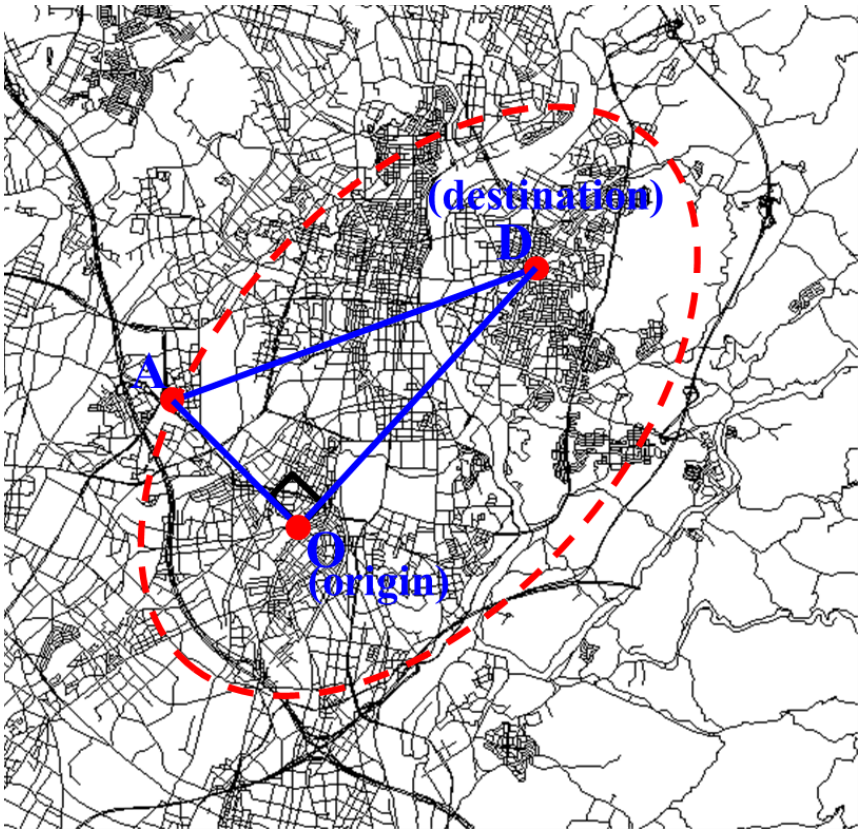


Figure 7-12 Network generation

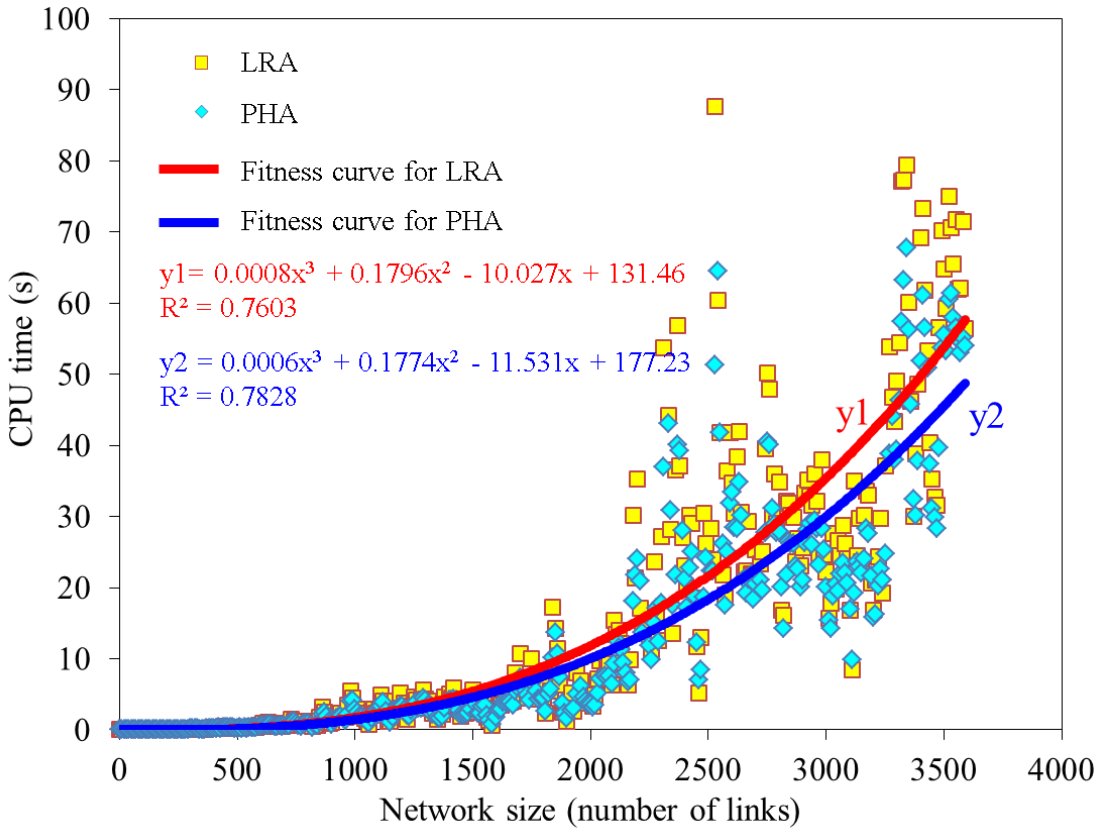


Figure 7-13 Computation time comparison

7.6 Computation efficiency analysis

A comparative evaluation analysis for the proposed Pareto-optimal based heuristic approach (PHA) against the Lagrangian relaxation approach (LRA) is presented in this section. The Lagrangian relaxation approach is a classical and widely used approach for the constrained shortest path problem (Carlyle et al., 2008; Ahuja et al., 1993). Both algorithms are programed in C# on the Windows 7 platform and run on a PC with Intel Core i7-4960X 3.6GHz CPU and 8GB RAM. We use two performance measures to evaluate the computation efficiency: (1) solution quality; (2) computation time.

To assess the solution quality, we use the solutions generated by the k-shortest path (KSP) algorithm (Yen, 1971) as the benchmark for the PHA and LRA. The KSP algorithm sorts the solution by ascending the CO_2 emission from the minimum to the kth minimum until the first solution that is feasible to the travel time budget is obtained. Although the computation time of KSP algorithm increases dramatically with the network size increases, it can be regarded as an exact solution (Handler and Zang, 1980). We generate the paths from 1000 observed OD pairs by the three routing approaches. Figure 7-11 shows the solution quality comparison for the two tested approaches, i.e., the PHA and LRA. A 45-

degree reference line is also plotted. If the tested approach generating the same solutions as the KSP algorithm generating, the points should fall along this reference line. Through the t-test at the 95% confidence level, it is concluded that the two tested solutions and the exact solution are not significant different. However, it is found that the performance of the PHA is better than that of the LRA. This is because the PHA enables to search the unsupported Pareto-optimal solutions in the duality gap while the LRA can not find the solution in the duality gap.

To assess the computation time, a set of real-world networks is generated, whose size ranges from 95 to 3525 links. As shown in Figure 7-12, we generate elliptical shape networks from 1000 randomly observed OD pairs. The elliptical area can be defined by the right triangle (AOD) of which the length of the first right-angle side is the OD Euclidean distance and the length of the second right-angle side is the half of the OD Euclidean distance. And in this way, the network size is dependent on the OD Euclidean distance. Figure 7-13 shows the computation time of each OD pair by the PHA and LRA. Since the structure of the transportation network is various for each OD pair, the CPU time may have difference even though the network size is similar. In general, the CPU time for the PHA is slightly less than that for LRA. On the other hand, it is found that the CPU times for both the PHA and LRA can be empirically represented by a cubic polynomial function. Therefore, it can be concluded that the proposed approach is efficient and applicable to a real-world network.

7.7 Summary

This study proposes an eco-routing approach to address the problem of finding the most eco-friendly path in terms of minimum CO_2 emissions constrained by a travel time budget. The theory of Pareto-optimal optimization is introduced to solve this NP-complete routing problem. Specifically, a Pareto-optimal based heuristic approach combining the weighting method with the k-shortest path algorithm is used to find the eco-friendly path. Based on the results of a numerical experiment in which this eco-routing strategy is analyzed using data collected in Toyota city, Japan, the following key findings are obtained:

(1)The SVM based model considering the average speed, average acceleration, road gradient and vehicle displacement enables to predict the CO_2 emission, which superiors to LRM and ANN model.

(2) The sensitivity analysis for the SVM model illustrates the non-linear relationship between the CO_2 emission and the explanatory variables.

(3) The relative importance analysis indicates that the average speed and vehicle displacement occupy 78% relative importance to the SVM based model.

(4) Eco-friendly paths determined as proposed in this study offer significantly reduced CO_2 emissions at little cost in terms of increased travel time and short detours. On average, an eco-friendly path can reduce CO_2 emissions by 13.02%, 10.24%, and 18.92% relative to the observed path, the shortest distance path, and the least travel time path, respectively.

(5) Compared to the observed path as selected empirically by the driver, the eco-friendly path offers significant advantage in terms of travel time (reduced by 16.95%) and CO_2 emissions (reduced by 13.02%) though the travel distance is slightly longer (by 0.3km).

(6) In an eco-routing experiment using all the observed OD pairs, it is found that the percentage of trips in which CO_2 emissions are reduced increases as the travel time buffer is increased. Interestingly, when the travel time buffer reaches about 10%, a certain degree of CO_2 emissions reduction is achieved for almost all trips.

(7) There is a significant rise in CO_2 emissions reduction as the travel time buffer increases from 1% to 12%. And a travel time buffer of 12% is appropriate for the eco-routing strategy, since this provides the greatest CO_2 emissions reduction for the least cost in travel time.

(8) The Pareto-optimal based heuristic approach has better performance than the Lagrangian relaxation approach. Particularly, the Pareto-optimal based heuristic approach enables to search the unsupported Pareto-optimal solutions in the duality gap while the Lagrangian relaxation approach can not find the unsupported Pareto-optimal solutions.

Potential directions for future research in this area include improvement of the path finding algorithm and consideration of the stochastic characteristics of travel time and CO_2 emissions: (1) an efficient path-finding algorithm suitable for a real-time eco-routing navigation system needs to be developed; (2) considering that travel time and emissions are non-deterministic, the reliability of the eco-algorithm routing should be considered further.

7.8 References

Ahn, K., Rakha, H., Trani, A., Van Aerde, M. (2002). Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. Journal of Transportation Engineering, 128(2), 182-190.

Ahn, K., & Rakha, H. (2008). The effects of route choice decisions on vehicle energy consumption and emissions. Transportation Research Part D: Transport and Environment, 13(3), 151-167.

Barth, M., An, F., Norbeck, J., & Ross, M. (1996). Modal emissions modeling: A physical approach. Transportation Research Record: Journal of the Transportation Research Board, 1520(1), 81-88.

Barth, M., An, F., Younglove, T., Scora, G., Levine, C., Ross, M., Wenzel, T., 2000. Development of a comprehensive modal emissions model. Tech. rep., National Cooperative Highway Research Program.

Barth, M., Malcolm, C., Younglove, T., & Hill, N. (2001). Recent validation efforts for a comprehensive modal emissions model. Transportation Research Record: Journal of the Transportation Research Board, 1750(1), 13-23.

Bandeira, J., Almeida, T. G., Khattak, A. J., Rouphail, N. M., & Coelho, M. C. (2013). Generating emissions information for route selection: Experimental monitoring and routes characterization. Journal of Intelligent Transportation Systems, 17(1), 3-17.

Birol, F. (2010). World energy outlook 2010. International Energy Agency.

Boriboonsomsin, K., Barth, M. J., Zhu, W., & Vu, A. (2012). Eco-routing navigation system based on multisource historical and real-time traffic information. IEEE Transactions on Intelligent Transportation Systems, 13(4), 1694-1704.

Boriboonsomsin, K., Dean, J., & Barth, M. (2014). Examination of Attributes and Value of Ecologically Friendly Route Choices. Transportation Research Record: Journal of the Transportation Research Board, 2427(1), 13-25.

Beusen, B., Broekx, S., Denys, T., Beckx, C. (2009). Using on-board logging devices to study the longer-term impact of an eco-driving course. Transportation Research Part D: Transport and Environment, 14(7), 514-520.

Csikos, A., Varga, I., & Hangos, K. M. (2015). Modeling of the dispersion of motorway traffic emission for control purposes. *Transportation Research Part C: Emerging Technologies*.

Chang, C. C., Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology , 2(3), 27.

Current, J. R., Revelle, C. S., & Cohon, J. L. (1990). An interactive approach to identify

the best compromise solution for two objective shortest path problems. Computers & Operations Research, 17(2), 187-198.

Coutinho-Rodrigues, J. M., Clímaco, J. C. N., & Current, J. R. (1999). An interactive bi-objective shortest path approach: searching for unsupported nondominated solutions. Computers and Operations Research, 26(8), 789-798.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems, 47(4), 547-553.

Cristianini, N., Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

EPA. (2003). User's Guide to MOBILE6.1 and MOBILE6.2, Mobile Source Emission Factor Model, EPA420-R-03-010. US Environmental Protection Agency, Washington, DC.

Guo, L., Huang, S., & Sadek, A. W. (2013). An evaluation of environmental benefits of time-dependent green routing in the greater Buffalo-Niagara region. Journal of Intelligent Transportation Systems, 17(1), 18-30.

Guo, C., Yang, B., Andersen, O., Jensen, C. S., Torp, K. (2014). EcoMark 2.0: empowering eco-routing with vehicular environmental models and actual vehicle fuel consumption data. GeoInformatica, 1-33.

Hastie, T., Tibshirani, R. Friedman J. (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, NY.

JAMA, (2008). Reducing CO_2 Emissions in the Global Road Transport Sector. http://www.jama-english.jp/

Jimenez-Palacios, J. L. (1998). Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDAS remote sensing (Doctoral dissertation, Massachusetts Institute of Technology).

Kono, T., Fushiki, T., Asada, K., & Nakano, K. (2008). Fuel Consumption Analysis and Prediction Model for "Eco" Route Search. In 15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting.

Koupal, J., Michaels, H., Cumberworth, M., Bailey, C., & Brzezinski, D. (2002, April). EPA's plan for moves: A comprehensive mobile source emissions model. In

Proceedings of the 12th CRC On-Road Vehicle Emissions Workshop, San Diego, CA.

Luna, A. S., Paredes, M. L. L., de Oliveira, G. C. G., & Corrêa, S. M. (2014). Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. Atmospheric Environment, 98, 98-104.

Mensing, F., Bideaux, E., Trigui, R., Ribet, J., & Jeanneret, B. (2014). Eco-driving: An economic or ecologic driving style? Transportation Research Part C: Emerging Technologies, 38, 110-12 Minett, C. F., Daamen, W., Van Arem, B., & Kuijpers, S. (2011). Eco-routing: comparing the fuel consumption of different routes between an origin and destination using field test speed profiles and synthetic speed profiles. IEEE Forum on Integrated and Sustainable Transportation System (FISTS).

Masikos, M., Demestichas, K., Adamopoulou, E., & Theologou, M. (2013). Machine-learning methodology for energy efficient routing. IET Intelligent Transport Systems, 8(3), 255-265.

Masikos, M., Demestichas, K., Adamopoulou, E., & Theologou, M. (2015a). Mesoscopic forecasting of vehicular consumption using neural networks. Soft Computing, 19(1), 145-156.

Masikos, M., Demestichas, K., Adamopoulou, E., & Theologou, M. (2015b). Eco-friendly routing based on vehicular consumption predictions of a mesoscopic learning model. Applied Soft Computing, 28, 114-124.

Miwa, T., Kiuchi, D., Yamamoto, T., Morikawa, T., 2012. Development of map matching algorithm for low frequency probe data. Transportation Research Part C: Emerging Technologies, 22, 132-145.

Nie, Y. M., & Li, Q. (2013). An eco-routing model considering microscopic vehicle operating conditions. Transportation Research Part B: Methodological, 55, 154-170.

Ortiz-García, E. G., Salcedo-Sanz, S., Pérez-Bellido, Á. M., Portilla-Figueras, J. A., & Prieto, L. (2010). Prediction of hourly O 3 concentrations using support vector regression algorithms. Atmospheric Environment, 44(35), 4481-4488.

Pandian, S., Gokhale, S., Ghoshal, A. K. (2009). Evaluating effects of traffic and vehicle characteristics on vehicular emissions near traffic intersections. Transportation Research Part D: Transport and Environment, 14(3), 180-196.

Park, S., & Rakha, H. (2006). Energy and environmental impacts of roadway grades. Transportation Research Record: Journal of the Transportation Research Board, (1987), 148-160.

Rakha, H., Ahn, K., & Trani, A. (2004). Development of VT-Micro model for estimating hot stabilized light duty vehicle and truck emissions. Transportation Research Part D: Transport and Environment, 9(1), 49-74.

Sedeno-Noda, A., & Raith, A. (2015). A Dijkstra-like method computing all extreme supported non-dominated solutions of the biobjective shortest path problem. Computers & Operations Research, 57, 83-94.

Sivak, M., & Schoettle, B. (2012). Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. Transport Policy, 22, 96-99.

Singh, K. P., Gupta, S., & Rai, P. (2013). Identifying pollution sources and predicting urban air quality using ensemble learning methods. Atmospheric Environment, 80, 426-437.

Smola, A. J., Scholkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14(3), 199-222.

Tan, Z., & Gao, H. O. (2015). Traffic control for air quality management and congestion mitigation in complex urban vehicular tunnels. *Transportation Research Part C: Emerging Technologies*, *58*, 13-28.

Vapnik, V. N. (1999). An overview of statistical learning theory. IEEE Transactions on Neural Networks, 10(5), 988-999.

Vapnik, V. (2000). The nature of statistical learning theory. Springer.

Xie, C., & Waller, S. T. (2012). Optimal routing with multiple objectives: efficient algorithm and application to the hazardous materials transportation problem. Computer−Aided Civil and Infrastructure Engineering, 27(2), 77-94.

Yao, E., & Song, Y. (2013). Study on eco-route planning algorithm and environmental impact assessment. Journal of Intelligent Transportation Systems, 17(1), 42-53.

Yen, J. Y. (1971). Finding the k shortest loopless paths in a network. Management Science, 17(11), 712-716.

Yeganeh, B., Motlagh, M. S. P., Rashidi, Y., & Kamalan, H. (2012). Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model. Atmospheric Environment, 55, 357-365.

Chapter 8

CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

To mitigate the uncertain delay and vehicle emission, this study investigated the route search problem considering travel time reliability and CO_2 emission. The travel time distribution, CO2 emission model and the corresponding route search methodologies are discussed. The major achievements and conclusions are summarized as follows:

In chapter 3, the data collection method for travel time and CO_2 emission (fuel consumption) is introduced. Specifically, Link travel time and path travel time distributions are characterized using empirical probe vehicle data. Several classical distributions (normal, lognormal, truncated normal, and truncated lognormal) are subjected to K-S test, A-D test and χ^2 test. It is found that the truncated lognormal distribution reasonably expresses the distribution of link travel time for about 90% of the links. To make the computation of the α-reliable paths tractable, we assume that the normal distribution can be used to approximate the distribution of path travel times as composed of truncated lognormal link travel times. And this assumption is justified by a Monte-Carlo simulation.

In chapter 4, the α-reliable path problem in a stochastic network with correlated and truncated lognormal link travel times is addressed. The Lagrangian relaxation approach is applied to solve the nonlinear and non-additive problem. The availability of such reliable paths in a navigation system application would help travelers plan their travel time budgets

with a given on-time arrival probability efficiently. The proposed α-reliable path-finding algorithm was applied to a large-scale real-world network in Toyota city, Japan. The calculation performance of the proposed method was shown to be accurate within a reasonable computation time.

In chapter 5, the traveler's risk-averse preference prediction for α-reliable shortest path problem in stochastic network is investigated. A novel data collection methodology for travelers' risk-averse preferences is introduced. The observed lower bound or upper bound of the risk-averse preference is collected by using the theory of stochastic dominance. Ordered probit model is applied to learn and predict the travelers' risk preferences by considering variously individual properties (gender, age) and pre-trip information (OD distance, departure time, day of week). The parameter estimation results show that the ordered probit model enables to explain how the explanatory variables influence the level of risk-averse preference.

In chapter 6, as a parallel study of risk-averse navigation, a two-stage path finding procedure is developed to find an experientially reliable path. Firstly, the hyperpath concept helps to determine a set of potential optimal paths. Secondly, the links in the graph of hyperpath are penalized based on the choice probability and the degree of familiarity, and then the A* algorithm is used to find the optimal path. There are two advantages to this routing method. First, any reasonable detours are taken into account and the link choice probability for each outgoing link from an attractive node is estimated according to the expected delay, which guarantees the potential optimal paths are included. Second, the solution provides not only the hyperpath with the recommended link choice probability, but also a shortest path with high degree of familiarity. This can better help travelers plan their trips and avoid the risk of uncertain travel time.

In chapter 7, an eco-routing approach to address the problem of finding the most eco-friendly path in terms of minimum CO_2 emissions constrained by a travel time budget is proposed. The SVM based model considering the average speed, average acceleration, road gradient and vehicle displacement enables to predict the CO_2 emission, which superiors to LRM and ANN model. The theory of Pareto-optimal optimization is introduced to solve this NP-complete routing problem. Specifically, a Pareto-optimal based heuristic approach combining the weighting method with the k-shortest path algorithm is used to find the eco-friendly path. Eco-friendly path offers significantly reduced CO_2 emissions at little cost in terms of increased travel time and short detours. On average, an eco-friendly path can reduce CO_2 emissions by 13.02%, 10.24%, and 18.92% relative to the observed path, the shortest distance path, and the least travel time path, respectively. Compared to the observed path as selected empirically by the driver, the eco-friendly path

offers significant advantage in terms of travel time (reduced by 16.95%) and CO_2 emissions (reduced by 13.02%) though the travel distance is slightly longer (by 0.3km). he average reduction in CO_2 emissions achieved by the eco-friendly path reaches a maximum of around 11% for trip OD distances between 6km to 9km and when the travel time buffer is around 10%. This indicates that setting a travel time buffer of 10% is appropriate for this eco-routing model, because this results in the greatest reduction in CO_2 emissions for the least cost in terms of travel time.

8.2 Future work

There is significant benefit to an efficient and green transportation system by applying the navigation methods proposed in this study. Based on this study, there are still many aspects need to be further discussed. Potential directions for future research are listed as follows:

(1)The α-reliable path finding could be extended to a time-varying stochastic network. A more accurate approximation method for path travel time distribution estimation could be developed by considering the skewness characteristic. In addition to spatial link travel time correlation, temporal correlation should be considered in the finding algorithm.

(2)To improve the estimation of the traveler's risk-averse preference, more factors such as traffic condition, trip purpose and weather condition are needed to consider.

(3)Considering that travel time and emissions are non-deterministic, a time-dependent routing method needs to be extended.

(4)A more intelligent navigation system that considers the balance of travel time, travel time reliability and emission will be a promising direction for further research.

Publication list

Journal Articles

Zeng, W., Miwa, T., Wakita, Y., Morikawa, T. Application of Lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times. ***Transportation Research Part C: Emerging Technologies***, 56, 309-334, July, 2015.

Zeng, W., Miwa, T., Morikawa, T. Learning traveler's risk preference to travel time reliability using GPS probe data, ***Journal of Highway and Transportation Research and Development,*** 32, 50-58, April, 2015.

Zeng, W., Miwa, T. and Morikawa, T. Exploring trip fuel consumption by machine learning from GPS and CAN bus data, ***Journal of the Eastern Asia Society for Transportation Studies***, Vol. 11, 906-921, 2015. (Best paper award in EASTS 2015 conference).

Zeng, W., Miwa, T., Morikawa, T. Application of hyperpath strategy and driving experience to risk-averse navigation. ***IET Intelligent Transport Systems*** (accepted).

Conference Articles

Zeng, W., Miwa, T., Morikawa, T. Finding a Reliable Shortest Path for Risk-averse Navigation Based on Historical Probe Vehicle Paths. ***Transportation Research Board 94th Annual Meeting*** (No. 15-1773), January, 2015.

Zeng, W., Miwa, T. and Morikawa, T. Application of machine learning to explore traveler's risk-averse preference to travel time reliability from large-scale GPS data, ***Proceedings of 14th ITS Asia Pacific Forum***, NanJing, China, April 2015.

Zeng, W., Miwa, T., Morikawa, T. Application of machine learning and heuristic k-shortest path algorithm to eco-routing problem with travel time constraint. ***Transportation Research Board 95th Annual Meeting*** (No.16-1421), January, 2016.

Zeng, W., Miwa, T., Morikawa, T. A reliable routing for risk-averse navigation, *Proceedings of Infrastructure Planning*, Vol. 50, CD-ROM, Tottori University, November, 2014.

Submitted Articles

Zeng, W., Miwa, T., Morikawa, T. Prediction of vehicle CO2 emission and its application to eco-routing navigation. ***Transportation Research Part C: Emerging Technologies*** (under reviewed).

Zeng, W., Miwa, T., Morikawa, T. Application of machine learning and heuristic k-shortest path algorithm to eco-routing problem with travel time constraint. ***Transportation Research Part D*** (under review).

Zeng, W., Miwa, T., Morikawa, T. A Lagrangian relaxation based heuristic algorithm for eco-routing problem with travel time constraint. ***Transportation Research Part D*** (under review).