A Proactive User Optimum-Oriented Route Guidance System Incorporating Individual Users' Route Choice Preferences

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Bingrong Sun

August 2018

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: ______________________

This Dissertation has been read and approved by the examining committee:

Advisor: Byungkyu Park

Committee Member: T. Donna Chen

Committee Member: John Miller

Committee Member: Hongning Wang

Committee Member: Andrew Mondschein

Committee Member: ______________________

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

August 2018

ABSTRACT

The route guidance system is one of the most effective ways of reducing traffic congestion. Existing route guidance systems are mostly reactive and self-interested, and simplify drivers' route choice preferences by assuming drivers only pursue the shortest travel time/distance. These features keep the route guidance system from adequately accommodating drivers' heterogeneous route choice preferences, and also from proactively avoiding congestion. Researchers designed more advanced route guidance systems that can optimize transportation efficiency or proactively avoid congestion, but drivers' heterogeneous route choice preferences have not been fully incorporated. Because of the lack of considering drivers' preferences and possible consequent reactions to the guidance, simplifying drivers' preferences in route guidance systems may lead to the discrepancy between the expected and actually generated traffic conditions, and might also undermine the performance of traveler information related intelligent transportation system strategies. Meanwhile, emerging information technologies applied in the transportation domain make it possible to collect drivers' behavior data even at the individual level, which provides great resources for analyzing drivers' preference heterogeneity. Therefore, this research proposes a proactive user optimum-oriented route guidance system that incorporates individual users' preferences in order to achieve users' better satisfaction and transportation system performance improvement. Individual drivers' route choice preferences can be captured from his/her historical preference data, then are adequately considered and coordinated by incorporating individual route choice models into the process of searching for user optimal conditions. Then, routes recommendations can be generated for each user based on the user optimal condition in which no user can improve his/her experience by changing to another route.

In order to make the route guidance system accurately capture, predict thus fully consider each driver's route choice preference, several approaches were firstly explored to establish individual route choice models, including traditional discrete choice model, mixed logit model, support vector machine and multi-task linear model adaptation (MT-LinAdapt). Three stated preference datasets collected from 102 participants as well as three synthetic datasets were used to evaluate and compare the performance of different approaches. The evaluation showed that MT-LinAdapt has the highest prediction accuracy which is up to 8% and 18% higher than other approaches when there is adequate and inadequate historical preference data, respectively. Additionally, it has implementation feasibility advantages: (1) does not require segmentation criteria (e.g., sociodemographic information) to distinguish drivers'

heterogeneous preferences; (2) also works well with limited amount of individual preference data and very heterogeneous preference data; and (3) can be updated in real time as individuals' preference data accumulates. Therefore, MT-LinAdapt is recommended to establish the individual-level route choice preference models in the application of route guidance systems.

The framework of a proactive user optimum-oriented route guidance system was proposed which contains two components: (1) established individual route choice models and (2) incorporated individuals' route choice preferences in searching for user optimum conditions. Such user optimum conditions are used as guidance information. With a commonly used Sioux Falls network and user population whose preferences were synthesized from surveyed participants, the proposed route guidance system at both perfect and imperfect market penetration rates was compared to existing route guidance strategies including travel time based real-time guidance and travel time based User Equilibrium (UE) guidance. An evaluation platform which is made of a traffic simulation module (DTAlite) and a route choice module (Matlab) was established and utilized to conduct the evaluation. The proposed route guidance system demonstrated advantageous performance in aspects of users' satisfaction (up to 22% more satisfied users), system mobility and sustainability (up to 10% of travel time reduction and up to 42% of delay reduction), and future traffic conditions estimation (up to 70% links having more accurate volume estimation). At imperfect market penetration rates, users of the proposed route guidance system interact with those drivers who use real-time guidance system and who take habitual routes. The generated performance improvement gradually increases as the market penetration rate increases. In addition, the proposed route guidance system has the potential to be extended for additional traffic control and management strategies so that further system performance improvement could be achieved, such as personalized incentive scheme.

The proposed route guidance system framework and the evaluation results extend the existing literature and have broader impacts on the following aspects: (1) Established individual route choice models to capture individual drivers' route choice preferences; (2) Proposed a proactive user optimum-oriented route guidance system for system performance and users' satisfaction improvements; (3) Prepared the foundation for designing personalized traffic control and management strategies that have great potential to further improve transportation system performance.

ACKNOWLEDGMENT

This dissertation would not be possible without many people’s help. I would like to take this chance to express my appreciation and thankfulness to them, for guiding me, encouraging me and helping me go through such a long journal and finally achieved one of the biggest achievements in my life.

First of all, I would like to thank my advisor, Dr. Byungkyu (Brian) Park. I am really grateful for having him as my doctoral advisor. He supported and encouraged me to pursue something I'm really interested in and gave me help and guidance all along the way in my Ph.D. study. Without all of these, I would not be able to make my way through the difficult time while I prepared this dissertation, and I really appreciate all of these.

I would also like to thank all my committee members, Dr. Donna Chen, Dr. John Miller, Dr. Honing Wang, Dr. Brian Smith and Dr. Andrew Mondschein for their guidance and suggestions to this dissertation. I enjoyed and learned a lot from all the discussion with them. Their valuable suggestions improved the quality of this dissertation.

My appreciation also goes to all professors who taught me and guided me before, especially Dr. Shengchuan Zhao and Dr. Ying-en Ge in my undergraduate and master study, and Dr. Michael Fountain here at UVa. Their broad knowledge and patient mentorship help me find the path that I am interested in. I would not have gone such far without their help and guidance.

I also want to thank all my friends for being always supportive and patient around me. I value all the laughter and tears we had, all the adventures we went for and even all the weights we gained together. Thank you for being there and make my Ph.D. life such colorful.

At last, I want to thank my dear families. Their unconditional love supports me and encourages me to fearlessly explore my life and purse what I am passionate about. Love you all!

TABLE OF CONTENT

CHAPTER 1. INTRODUCTION

1.1 Background

The route guidance system is an important part of the Advanced Traveler Information System (ATIS). Drivers use route guidance systems in daily life to find the routes leading them to their destinations or check real time traffic conditions so that congested areas can be avoided. With the information provided by route guidance systems, drivers can make informed route choice decisions and transportation network's efficiency can be improved. It is recognized that the route guidance system is one of the most effective ways to reduce traffic congestion (Han et al. 2016).

More and more drivers are using route guidance systems nowadays. As of 2017, Google Maps app ranks the fifth of the most popular smart phone apps in the United States and it was accessed by 56% of the mobile users which is 56% of 224.3 million smart phone users ("Mobile Apps U.S. Smartphone Audience Reach 2017 | Statistic"). Route guidance systems not only provide most recent traffic condition information but also provide users with some customized options, such as allowing drivers to select among options of showing the route with shortest travel time, shortest distance and without toll in Google Maps. The route guidance systems can also store users' home, workplace and other favorite places. With more and more users participating in as well as more information that can be collected with route guidance systems, route guidance systems have large potential to influence drivers' travel behaviors and consequently the transportation system performance.

1.2 Research Motivation

Route guidance systems bring much conveniences to drivers' driving activities. As its user group is becoming larger and larger and it can collect more and more demand-side data, there are several motivations to make the route guidance systems to better serve users' needs as well as improve transportation system performance.

Simplified behavior component in existing route guidance systems could deteriorate users' satisfaction. Existing route guidance systems simplify drivers' route choice preferences in route guidance process by recommending routes to users based on single criterion, such as shortest travel time or distance. However, drivers' route choice preferences can be very heterogeneous among the population. Different departure time, trip purposes, driving experiences, risk attitudes, and etc. can all bring different route choice preferences (Amirgholy et al. 2017; Liu, He, and Recker 2007). The limited options in existing route guidance systems cannot fully accommodate drivers' preference heterogeneity, therefore recommending suboptimal routes to users may deteriorate users' satisfaction.

Existing route guidance systems are reactive and self-interested. Most of existing route guidance systems recommend routes to users based on either historical traffic conditions or the real-time traffic conditions. Though it can quickly react to the changing traffic conditions, it does not consider how the recommended routes are going to affect the future traffic conditions. It works more like an alert system which warns drivers' about the congestion already happened instead of guiding drivers to proactively prevent the congestion from happening (Liang and Wakahara 2014). In addition, the route recommendations are usually self-interested. It neither considers if other users' behaviors are going to affect the quality of the recommendations made to the subjective user nor considers if the recommendation is going to influence other users. The characteristics of reactive and self-interested can undermine the route guidance systems' service reliability and consequently influence users' satisfaction and transportation system performance. Therefore, a route guidance system that can consider users' possible reactions and proactively prevent congestion is more ideal than reactive and self-interested ones.

Emerging information technologies provide an opportunity to design and implement a route guidance system that can further improve system performance and users' satisfaction. Route guidance systems nowadays not only have large user population but also can collect much users' route choice preference data, even at the individual level. It provides an opportunity to capture individual driver's route choice preference and predict their possible route choice decisions. It can help route guidance systems better capture the heterogeneous route choice preferences among user population and understand the needs of specific users when using route guidance. By knowing users' possible reactions in different scenarios, a new route guidance system can be designed and implemented to overcome limitations of existing route guidance systems so that system performance and users' satisfaction can be both improved.

1.3 Research Objectives and Scope

With the behavior related data that could be collected from route guidance systems, the main research goal is to propose a route guidance system that can further improve transportation system performance and users' satisfaction by considering individual drivers' route choice preferences in designing routing strategy. The route guidance system should be able to address the existing route guidance systems' limitations in terms of inadequately considering users' preference heterogeneity and the reactive and self-interested features. In other words, the desired route guidance system should consider individuals' route choice preferences, be proactive and also consider the impacts of recommendations on traffic conditions. To achieve this goal, three objectives are identified as follows.

Capture drivers' route choice preferences at the individual level. Different modeling approaches can be explored to find the most suitable one for the application of a route guidance system that considers drivers' preferences at the individual level. The capability of accurately predicting drivers' route choice decisions is the major measurement. The modeling approach should be able to handle the characteristics of the data that can be collected from route guidance systems and also have certain implementation feasibility, such as the capability of being updated in real time and working well when there is limited amount of preference data.

Design a proactive user optimum-oriented route guidance system. The route guidance system should consider all individual driver's route choice preferences when generating route recommendations. It also should consider the impacts of possible recommendations on the traffic conditions and consequently the impacts on other users' route choice decisions.

Evaluate and quantify the performance of the proposed route guidance system in terms of transportation system performance and users' satisfaction. Proper measurement should be selected to represent the system performance and users' satisfaction. The proposed route guidance system should be evaluated against representative existing routing strategies. As it may take time to implement the proposed route guidance system in practice, the performance at different imperfect penetration rates need to be evaluated as well.

It should be noted that this research only considers the route choice aspect of travelers' behavior. Though departure time choice and mode choice are also important aspects of travelers' behaviors, this research assumes drivers all have fixed preferred departure time with driving as the travel mode.

1.4 Dissertation Organization

In the following chapters of this dissertation, Chapter 2 first explores the capabilities of traditional mixed logit model combined with the Bayes rule in estimating individual drivers' route choice preference. Chapter 3 explores the capability of an advanced sentiment analysis approach, MT-LinAdapt in route guidance system application from the perspectives of both prediction accuracy as well as implementation feasibilities. In Chapter 4, a proactive user optimum-oriented route guidance system is proposed by capturing individual user's route choice preferences and integrating individual's preferences in the process of searching for user optimum conditions. The proposed route guidance system was evaluated against existing routing strategies and the evaluation results were analyzed. In Chapter 5, some possible extended applications of the proposed route guidance system (e.g., personalized incentive scheme) that can further improve system performance are discussed as well as some possible practical issues in the implementation of the proposed route guidance system. At last, conclusions are made in Chapter 6 as well as possible future research.

CHAPTER 2: Applying Mixed Logit Model and Bayes Rule for Drivers' Route Choice Preferences Modeling at the Individual Level

Bingrong Sun, B. Brian Park

Department of Civil and Environmental Engineering, University of Virginia, 351 McCormick Rd
Charlottesville, Virginia, United States, 22904

ABSTRACT

Personalized Advanced Traveler Information System (ATIS) services such as personalized route recommendation and trip planning require knowing drivers' route choice preferences at the individual level. To distinguish drivers' heterogeneous route choice preferences, existing modeling approaches typically build a model which includes segmentation criteria (e.g., sociodemographic characteristics) with an assumption that drivers belonging to the same group have similar route choice preference. Meanwhile, random parameter models such as mixed logit (ML) model are also widely used to capture drivers' route choice preferences heterogeneity. However, segmentation criteria are not easy to obtain in practice and only using the preference distributions in the ML model cannot accurately tell the preference of a specific driver. Thus, this paper explores the capability of using ML models combined with Bayes rule to capture individual drivers' route choice preferences. With a stated preference dataset in which each of 44 participants went through 81 route choice scenarios containing multiple route attributes, a ML model that considers coefficients' correlation was established. Individual participant's route choice preference was obtained with Bayes rule on the basis of the estimated coefficients' distributions. The finally obtained individual drivers' route choice models were evaluated against a multinomial logit model regarding the capability of correctly predicting each participant's choices. The results showed that the ML model combined with Bayes rule has an average prediction accuracy of 89% which is nearly 20% higher than the multinomial logit model. The ML model combined with Bayes rule has the potential to be used for modeling individual-level drivers' route choice preferences for personalized ATIS services.

Keywords: Individual Route Choice Model; Mixed Logit Model; Bayes Rule; Personalized Route Recommendation

INTRODUCTION

Advanced Traveler Information System (ATIS) applications such as personalized route guidance and personal trip planning require the knowledge of individual driver's route choice preference (Liu et al., 2014; Nadi and Delavar, 2011; Pahlavani and Delavar, 2014). In these applications, a route is recommended to a particular driver based on his/her own route choice preference in order to improve driver's satisfaction. In addition, drivers are more willing to comply with the routes that are recommended based on his/her preference. This can lead to increased drivers' compliances and consequently has a great potential for traffic engineers to improve transportation network performance (Bifulco et al., 2007; Domenico et al., 2015; Paz and Peeta, 2009).

However, drivers' route choice preference varies person by person in many aspects. Drivers perceive route information differently (Tawfik et al., 2010; Parthasarathi et al., 2013; Peruch et al., 1989; Matthews, 1981). Drivers value the same attribute differently (Liu et al., 2007; Perk et al., 2011; Rogers et al., 1999; Rogers and Langley, 1998) and even have different preferences at different time of day (Amirgholy et al., 2017; Liu et al., 2007). As pointed by Lida et al., (1992) "route choice behavior varies depending on the individual." Therefore, for better satisfaction of personalized ATIS users as well as the potential of utilizing ATIS to improve system performance, it is desirable to have a modeling approach that can capture drivers' heterogeneous route choice preference and accurately predict drivers' preferred routes at the individual level.

To model drivers' route choice different preferences, the most common approach is to use a logit-type model that incorporates segmentation criteria to differentiate drivers' preferences. Sociodemographic characteristics are frequently used as the segmentation criteria (Abdel-Aty et al., 1994; Li et al., 2005; Shiftan et al., 2011). Assumptions are made that drivers with same sociodemographic characteristics (such as age, gender and income) have a similar preference (Ben-Aakiva and Lerman, 1985; Prato, 2009). Other criteria such as driving pattern can also be used to categorize drivers into different groups (Tawfik and Rakha, 2013; Peeta and Yu, 2004). Therefore, drivers are basically divided into different groups based on certain segmentation criteria and drivers belonging to the same group behave similarly. However, in practice, the information regarding above mentioned segmentation criteria are difficult to obtain because of privacy issues. The information includes age, gender, income, driving experiences and other sociodemographic characteristics which are typically used as segmentation criteria (Abdel-Aty et al., 1994; Li et al., 2005; Shiftan et al., 2011). In some widely used route guidance services such as

Google Maps service, it requires users' age and gender information when users register Google account, but users have an option of "Rather not say." Another popular route guidance service Waze also does not require socioeconomic information. Therefore, using sociodemographic information as segmentation criteria to distinguish drivers' different route choice preferences is not easy in the application of providing personalized route planning or guidance. Given various characteristics in drivers' population, it is also difficult to select proper criteria to segment drivers into groups (Hensher and Greene, 2003; Peeta and Yu, 2004). In addition, the preference heterogeneity within groups is not considered.

In the family of discrete choice model, the mixed logit model is widely used to capture drivers' preference heterogeneity (Train, 2009; Hess and Train, 2017; Sarrias and Daziano, 2017; Bansal et al., 2017). The mixed logit model assumes people's preferences following certain distributions among the population. Instead of using a single utility function for all individuals (i.e., a multinomial logit model), the mixed logit model allows the parameters associated with attributes in the utility function following certain distributions among population. Thus, it is considered to be more realistic than fixed parameter models (Greene and Hensher, 2003). Given this nature, mixed logit models have been applied to capture heterogeneous preferences in different areas, such as predicting customers' preferences and designing targeted policy to consumers (Revelt and Train, 1998; Train, 1998), estimating drivers' route choice preference variation (Ben-Elia and Shiftan, 2010; Han et al., 2001; Liu et al., 2004; Tian et al., 2012), estimating pedestrian's exists choice preferences (Haghani et al., 2015) and railway passengers' preferences in purchasing tickets (Hetrakul and Cirillo, 2013).

In the ATIS applications such as personalized route guidance or personal trip planning, using the mixed logit model to obtain the preference distribution is still not enough, because these applications require knowing the preference of a specific driver. When only the preference distribution is used for describing drivers' route choice preferences, it describes drivers' preference pattern as a whole. Taste parameters are randomly assigned to each individual driver from the distribution. The resulted preferences among drivers match the preference distribution pattern, but a specific driver's preference may not be accurately captured when look at each individual driver. Fortunately, the mixed logit model is also capable of capturing preference at the individual level when combined with the Bayes rule, in a sense that each individual can have his/her own route choice model. Revelt and Train (1998) used a mixed logit model to analyze the tastes distribution as well as individual preferences of residential customers' choice among energy suppliers, but to the best of authors' knowledge, the mixed logit model explicitly considering individual preference has not been used in transportation domain. This is probably because traditional drivers' route choice behavior study mainly focus on drivers' general preference and behavior patterns in order to understand the behavior impacts of drivers as a group. With emerging information communication technologies in ATIS, new transportation services such as personalized route recommendation and trip planning bring the needs of utilizing each specific driver's preference. Therefore, given the potential that mixed logit models have in capturing preferences at the individual level, this paper aims at evaluating the capability of the mixed logit model combined with Bayes rule in modeling and predicting individual drivers' route choice preferences.

In the rest of this paper, the mixed logit model together with Bayes rule are introduced in Methodology section, followed by the description of a stated preference dataset that was used for developing the mixed logit model in Survey and Data section. The mixed logit model combined with Bayes rule is applied to the dataset in Model Implementation section and the results are interpreted in Results Analysis section. Finally, conclusions and future research are discussed.

METHODOLOGY

With a mixed logit model, the utility of choosing alternative route *i* for driver *n* is (Train, 2009):

$$U_{ni} = \boldsymbol{\beta}'_{\boldsymbol{n}}\boldsymbol{X}_{\boldsymbol{ni}} + \varepsilon_{ni} \quad (1)$$

Where $\boldsymbol{\beta}_{\boldsymbol{n}}$ is vector of parameters associated with each variable for user *n*. $\boldsymbol{X}_{\boldsymbol{ni}}$ is a vector containing the alternative *i*'s variables and drivers' sociodemographic characteristics if available. In the application of this research, only alternative option's attributes are included in order to represent the situation in practice that sociodemographic data is not easy to obtain because of privacy concerns. ε_{ni} is the unobserved extreme random value for driver *n* choosing alternative *i*. It is typically assumed to follow an identically and independently distribution across drivers and alternatives in the multinomial logit model, but the mixed logit model allows for a relaxation of this assumption. Also, different from the multinomial logit model where drivers have the same form of utility function, drivers' utility function parameters follow certain distribution among population in the mixed logit (ML) model.

Then, the logit probability L_{ni} for user *n* choosing alternative *i* can be written as follows (Hensher and Greene, 2003):

$$L_{ni} = \frac{\exp(\boldsymbol{\beta}'_{\boldsymbol{n}}\boldsymbol{X}_{\boldsymbol{ni}})}{\sum_j \exp(\boldsymbol{\beta}'_{\boldsymbol{n}}\boldsymbol{X}_{\boldsymbol{nj}})} \quad (2)$$

Since the mixed logit (ML) model considers heterogeneous preferences among the population, β here follows a distribution. Commonly used distributions include normal, lognormal, uniform and triangular distribution (Hensher and Greene, 2003). Assuming density function of β is noted as $f\,(\beta|\theta)$, θ represents the parameters of the distribution. For example, if β follows a normal distribution $N(u,\sigma^2)$, θ represents the mean u and the standard deviation σ of the normal distribution. Assuming all parameters in $\boldsymbol{\beta_n}$ follow normal distributions, then $\boldsymbol{\beta_n}$ can be written as (Train, 2009):

$$\boldsymbol{\beta_n} = \boldsymbol{\mu_n} + \boldsymbol{L\eta} \quad (3)$$

$\boldsymbol{\eta}$ is a standard normal distribution. **L** is the Choleski factor of covariance matrix $\mathbf{\Omega}$ and it can be defined as a lower-triangular matrix such that $\mathbf{LL}'= \mathbf{\Omega}$. When the parameters in $\boldsymbol{\beta_n}$ are independent from each other, the off-diagonal elements in **L** are all 0.

Given the distribution density function $f\,(\beta|\theta)$, the choosing probability becomes (Train, 2009):

$$P_{ni} = \int L_{ni} f(\beta|\theta) d\beta \qquad (4)$$

The above equation is the form of a mixed Logit model, as it is a mixture of the logit function evaluated at different values of β with $f\,(\beta)$ as mixing distribution. The underlying distribution θ can be estimated. One of the most common ways to estimate the parameters of underlying distribution is the simulation method (Train, 2009). Equation (4) can be approximated with simulation as:

$$\hat{P}_{ni} = \frac{1}{R}\sum_{r=1}^{R} L_{ni}(\beta^r) \qquad (5)$$

Given the distribution as $f\,(\beta|\theta)$, β^r is the β value obtained by a random draw from the distribution for r^{th} time. β^r is plugged into Equation (5) to calculate the simulated likelihood. This process is conducted R times and the average likelihood is the simulated likelihood for driver n choosing alternative i. There are different ways to conduct the random draws so that the estimation can be efficient and accurate. Commonly used ways to perform random draws include pseudo-random draws (Sándor and Train, 2004), quasi-random draws (Sándor and Train, 2004), Halton draws (Train, 2009) and scrambled Halton draws (Bhat, 2003). Therefore, the simulated log-likelihood of all drivers' choices can be written as (Train, 2009):

$$Simulated\ log\ likelihood = \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \hat{P}_{nj} \qquad (6)$$

Where d_{nj} equals to 1 when driver *n* chose alternative *j*. Otherwise, d_{nj} equals 0. The estimated θ value gives the distribution that drivers' preferences follow and it represents the heterogeneous preferences among drivers.

Once parameters' distributions are obtained, a specific individual's preference can be estimated with Bayes rule. The choice probability for driver *n*'s series of choices is (Train, 2009):

$$P(y_n|x_n, \beta) = \prod_{t}^{T} L_{nt}(y_{nt}|\beta_n) \qquad (7)$$

Since only the distribution of β is known, the equation above can be written as (Train, 2009):

$$P(y_n|x_n, \theta) = \int P(y_n|X_n, \beta) f(\beta|\theta) d\beta \qquad (8)$$

By Bayes' rule, we can have β's conditional distribution

$$h(\beta|y_n, x_n, \theta) = \frac{P(y_n|x_n, \beta) f(\beta|\theta)}{P(y_n|x_n, \theta)} \qquad (9)$$

Therefore, the expected individual taste $\bar{\beta}_n$ can be written as (Train, 2009):

$$\bar{\beta}_n = \int \beta * h(\beta|y_n, x_n, \theta) d\beta$$

$$= \frac{\int \beta * P(y_n|x_n, \beta) f(\beta|\theta) d\beta}{P(y_n|x_n, \theta)}$$

$$= \frac{\int \beta * P(y_n|x_n, \beta) f(\beta|\theta) d\beta}{\int P(y_n|x_n, \beta) f(\beta|\theta) d\beta} \qquad (10)$$

However, the equation above does not have a closed form. A simulation approach has to be used to obtain its value. Random draws can be made from the distribution $f(\beta|\theta)$ and the simulated taste β can be the weighted sum of values obtained (Train, 2009), as shown in Equation (11).

$$\hat{\beta}_n = \sum_r w^r \beta^r \qquad (11)$$

In which

$$w^r = \frac{P(y_n|x_n, \beta^r)}{\sum_r P(y_n|x_n, \beta^r)} \qquad (12)$$

Thus, individual driver *n*'s specific preference can be estimated with combining the mixed logit model and Bayes rule.

SURVEY AND DATA

In order to develop the mixed logit model and to test its performance in predicting individual driver's route choice preference, a stated preference survey was designed and conducted. In order to generate realistic survey scenarios, the basic procedure used for the survey design is to check the traffic information on a popular navigation system (Google Maps) and include adequate variations for survey questions. Then, origin and destination (OD) pairs of three distance levels were chosen from the local area. For each OD pair, two alternative routes suggested by Google Maps were included in the survey to form binary route choice scenarios. Based on researchers' findings (Jan et al., 2000), several route attributes that are considered as belonging to major influencing factors were included in the survey, including *Distance*, *Travel Time*, *the Number of Controlled Intersections* and *Pedestrian Level*. The information about these attributes was checked on Google Maps within a day and across days of week to have realistic values as well as adequate variations, except *Pedestrian Level* which was checked manually through field visit. Eventually, all route attributes have three levels of values except that Pedestrian Level that was set to have two levels of "High" and "Low." Taguchi design in Minitab (Simpson et al., 2001) was used to develop the experiment design. In total, 81 questions were generated. All the questions were shuffled before showing to participants. The trip purpose of the scenarios was casual trip. A total of 44 participants who are mostly undergraduate students at the University of Virginia took part in the survey. The linkage of the online questionnaire was sent to all participants who volunteered for the surveys. The participants were informed with the time the survey would take and were asked only to participate in if they would take their time to answer the questions to the best of their ability. To ensure the quality of participants' answers, the survey includes dominant questions in which one route is better than the other in all aspects. All participants passed the answer quality screening.

MODEL IMPLEMENTATION

To test the performance of the mixed logit model, especially its capability of capturing individual driver's route choice preference, 80% of every participant's data was randomly selected to build the model and the rest of the data was reserved for testing. The dataset was standardized by each route attribute. To be more specific, each column contains the values of a route attribute. Each value in a column was subtracted from the column mean and divided by the column standard deviation. The dataset after standardization was used for model training and testing later. Before estimating the mixed logit model, distribution types of coefficients and the number of random draws need to be determined. Each of them was discussed below.

Select Distribution Types

As explained in the Methodology section, coefficients of utility function are assumed to follow certain distributions. The types of distribution need to be defined before the model estimation. Common distributions that were used in transportation include normal distribution, lognormal, uniform and triangular (Hensher and Greene, 2003). Instead of assuming distribution types arbitrarily, Hensher and Greence (2003) suggested to use empirical distributions that were observed from the data. The process is possibly to obtain the coefficient estimate for each sampled individual and then plot all individuals' coefficient estimates so that a distribution shape can be observed. However, it is very likely that some individuals may not have significant coefficient estimates because there might not be enough amount of or enough variations in individual's data. Therefore, this research adopted the *Q+1* model method suggested by Hensher and Greene (2003) to determine the distribution types. *Q* is the number of surveyed individuals. Here, *Q* equals 44. A multinomial logit model ***M*** was built firstly based on all individuals' data. Then the data of each individual was removed respectively and the rest *Q-1* individuals' data was used to build a multinomial logit model, called $\boldsymbol{M_i}$. Therefore, there were one model developed from all individuals' data and *Q* models developed from partial individuals' data. For a participant *i*'s particular coefficient, the difference between its numerical values in model ***M*** and in model $\boldsymbol{M_i}$ is considered as the impact of this participant *i*'s preference on the whole group's preference. The distribution of these differences across individuals shows the distribution of individuals' heterogeneous preferences. Kernel density estimator was used to plot the distribution of the difference

(Hensher and Greene, 2003). The bandwidth of the density estimation is determined by h=1.06 $\sigma N^{-1/5}$ (Scott and Sain, 2005). The shape was observed to obtain the underlying distribution for each coefficient in the mixed logit model. The shapes of each coefficient's density plot are shown in Figure 1.

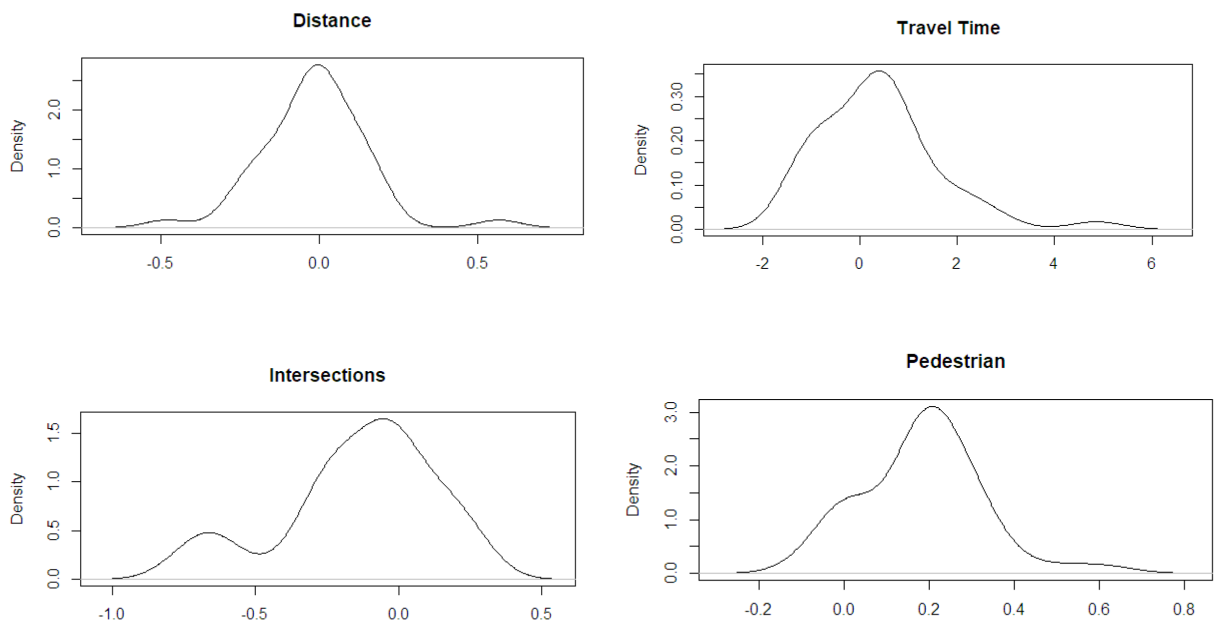


Figure 1 Density plots of each coefficent's distribution among participants

As shown in Figure 1, the coefficient distributions of "Distance" and "Pedestrian" are very close to the shape of normal distribution. The distribution shapes of "Travel Time" and "Number of Intersections" are between normal distribution and log-normal distribution. Both of the normal and log-normal distributions are commonly used in mixed logit model estimations. The normal distribution has flexiable properties and is widely understood, but the estimated parameters could have either positive or negative signs (Department for Transport UK, 2014). The log-normal distribution can restrict the parameter signs to be either positive or negative when modelors have a sense about the signs of parameters (Hensher and Greene, 2003), but it also has several disadvantages, including difficult to find the starting values in the estimation (Train, 2009; Han et al., 2001), considerably difficult to converge (Li et al., 2010) and having very long tails on one end which represents unreasonable values (Nahuelhual et al., 2004). As mentioned in (Hensher and Greene, 2003), all types of distribution have their own advantages and limitations. On the other hand, the application of personalized route guidance system requires accurate predictions regarding which route the driver would like to take. The model performance in terms of fitness and prediction accuracy are the measurements personalized route guidance system care more about. With the dataset in this research, the staring value is very difficult to find when using log-normal distribution. Therefore, the coefficients of "Travel time" and "Number of Intersections" were also assumed to follow normal distribution in the estimation process.

Select the Numbers of Random Draws

Halton intelligent draw was used to conduct random draws in the simulation process of estimating coefficients' distributions (Train, 2009). The number of random draws, namely R value in equation (5), can have impacts on model estimations. A small number of random draws may not be enough to accurately estimate the model. A large number of random draws can decrease the estimation errors but make the model estimation process inefficient. There is no standard number of random draws that works for every situation (Hensher and Greene, 2003). Therefore, a range of different numbers of draws were tested until the estimated model is stable. In this paper, numbers of draws were tested including 50, 100 and 1000. With the dataset in this research, standard errors of estimated parameters did not decrease much when increasing the numbers of draws. Therefore, the number of draws in equation (5) was set to be 100.

Model Establishment

After determining the distribution types and the number of random draws, gmnl package in software R (Sarrias and Daziano, 2017) was used to conduct the estimation. The utility function has four route attributes including "Distance", "Travel Time", "Number of Controlled Intersections" and "Pedestrian Level." The utility function coefficients of all route attributes are assumed to follow normal distribution. Therefore, $\boldsymbol{\beta_n}$ in Equation (1) contains four elements and each of the elements follows a normal distribution $N(u,\sigma^2)$ with u as the mean and σ as the standard deviation. In addition, it is very likely that participants' preferences regarding different route attributes are correlated. The correlation among route attributes is also considered. Therefore, the expected estimation results should include the mean and standard deviation of each coefficient and the Choleski factor of coefficients' covariance matrix $\boldsymbol{\Omega}$ as shown in Equation (3).

RESULT ANALYSIS

This section first evaluates the estimated mixed logit model at the aggregate level. Then selected participants' estimated preferences are discussed at the individual level.

Model Performance at the Aggregate Level

The estimated mixed logit (ML) model is shown in Table 1, as well as a multinomial logit (MNL) model which was included for comparison. The MNL model is also known as a fixed parameter model. Therefore, the MNL model only has estimated coefficient for each variable. On the other hand, the mixed logit model estimates the distribution of preference regarding each route attribute. Therefore, the estimation includes the estimated means and Choleski factors of coefficients' variance-covariance matrix.

As shown in Table 1, the estimated coefficients in MNL and coefficients' means in the mixed logit (ML) model all have negative signs, which are as expected. It is easy to understand that the route with higher values of these attributes (e.g., longer travel time) is usually less favored by drivers, thus drivers' utility of choosing certain route is reduced when the values of these route attributes increase.

Among the variables considered, "Distance" is not significant in both MNL and ML models with p-values of 0.163 and 0.139, but the standard deviation of "Distance" is significant in the estimation of ML model. That means participants have various preferences regarding "Distance" and their preferences balanced out when they are seen as a group. Therefore, the overall preference regarding "Distance" is not significantly different from zero in the MNL model. This is not reflected in the MNL as it does not consider preference heterogeneity.

The magnitude of "Travel Time" coefficient is the largest among all coefficients in both the ML model and the MNL model. That means "Travel Time" has the largest impact on participants' decisions among all route attributes. It is noted that the stated preference data was normalized before the model estimation. Following the same logic, the second and third most influential factors are "Number of Intersections" and "Pedestrian Level."

Table 1 Model Estimation Results of the Mixed Logit Model and the Multinomial Logit Model

	MNL			Mixed Logit		
Coefficients	Estimates	Std.error	p-value	Estimates	Std.error	p-value
Distance.mean	-0.308	0.221	0.163	-0.555	0.375	0.139
Travel Time.mean	-5.156	0.270	0.000	-10.437	1.091	0.000
Number of Intersections.mean	-1.225	0.052	0.000	-2.561	0.233	0.000
Pedestrian Level	-0.746	0.036	0.000	-1.510	0.179	0.000
sd.dis.dis	-	-	-	-1.554	0.261	0.000
sd.dis.tt	-	-	-	4.401	0.951	0.000
sd.dis.ped	-	-	-	1.376	0.161	0.000
sd.dis.inter	-	-	-	0.502	0.105	0.000
sd.tt.tt	-	-	-	-6.504	1.040	0.000
sd.tt.ped	-	-	-	0.321	0.182	0.077
sd.tt.inter	-	-	-	0.608	0.137	0.000
sd.ped.ped	-	-	-	1.336	0.216	0.000
sd.ped.inter	-	-	-	0.114	0.216	0.599
sd.inter.inter	-	-	-	0.903	0.132	0.000
Log likelihood	-1131.5			-728.79		
Pseudo-R^2	0.4180			0.6252		

Table 2 shows the variance-covariance matrix of coefficients estimates. The diagonal entries are the variances of coefficients. Based on the magnitude of variance, “Travel Time” has the largest variance. It indicates that participants’ preference regarding “Travel Time” varies most among four route attributes. Then, participants’ preferences regarding “Pedestrian Level” is the second most heterogeneous and has larger variance than those of “Distance” and “Number of Intersections.” The significant values of parameters’ variance show that participants do have various preferences regarding the same route attributes.

Table 2 also shows the covariance between each pair of coefficients. The coefficients of “Distance” and “Travel Time” have negative covariance of -6.841. That means participants who have large magnitude of “Distance” coefficient tend to have a small magnitude of “Travel Time” coefficient. That is saying a participant who strongly prefers a route with shorter distance tend not to be motivated by shorter travel time. Similar negative covariance also exists between the coefficients of “Distance” and “Number of Intersections”, “Distance” and “Pedestrian Level.” The coefficient of “Travel Time” is negatively correlated with the coefficient of “Number of Intersection” and is positively correlated with the coefficient of “Pedestrian Level.” This means participants who strongly prefer shorter travel time tend to have larger weight on “Pedestrian Level” and tend not to be easily motivated by fewer number of intersections. At last, the coefficient of “Number of Intersections” is positively correlated with the coefficient of “Pedestrian Level.” This indicates participants who like the route with fewer number of intersections also is more easily motivated by lower pedestrian level.

Table 2 The Variance and Covariance Matrix of Estimated Coefficients in the Mixed Logit Model

	Distance	Travel Time	Number of Intersections	Pedestrian Level
Distance	2.416			
Travel Time	-6.841	61.669		
Number of Intersections	-0.781	-1.741	1.450	
Pedestrian Level	-2.138	3.968	1.038	3.779

By considering participants’ various preferences regarding different route attributes and the correlation among attributes’ coefficients, the ML model has better overall fitness to the survey data than the MNL model. As shown in Table 1, the log likelihood of the ML model (-728.79) is much higher than that of the MNL model (-1131.5). The Pseudo-R^2 of the ML model is also higher than that of the MNL model by around 20%.

Model’s Performance at the Individual Level

The Mixed logit model can estimate individual’s specific preference by following equations (7) to (12). Estimated preferences of four selected participants were shown in Table 3 to demonstrate how participants’ preferences vary. As shown in Table 3, Participant 1 cares about travel time most and the coefficient of “Travel Time” has the largest magnitude among four coefficients in his/her utility function. The “Number of Intersection” is the second important factor that affects Participant 1’s route choice decision. The most important factor for both Participant 2 and Participant 3 is also “Travel Time”, but the second most important factor is “Pedestrian Level” for Participant 2 and “Distance” for Participant 3. As to Participant 4, the most important factor affecting his/her route choice decision becomes the number of intersections and his/her preference regarding other three route attributes are very similar. Since the distribution type of all parameters are assumed to be positive, certain percentage of participants got positive signs in the estimation results, though the signs for attributes such as “Distance” and “Travel Time” are usually negative. The percentages of positive signs for four route attributes are: 34% for “Distance”, 8% for “Travel Time”, 1% for “Number of Controlled Intersections” and 21% for “Pedestrian Level”. For all parameters, majority of participants’ estimated preferences has negative signs.

Table 3 Utility Function Coefficients of Selected Participants’ Individual Models

Coefficient	Distance	Travel Time	Number of Intersections	Pedestrian Level
Participant 1	-0.224	-14.974	-2.018	-0.218
Participant 2	-0.382	-4.758	-2.883	-3.602
Participant 3	-3.405	-5.279	-0.344	-0.972
Participant 4	-1.516	-1.475	-2.025	-1.353

When looking at the individual level, it can be found that participants’ preferences could be very different from one another. Furthermore, the survey data was collected from a group of participants who have relatively homogeneous sociodemographic characteristics (i.e., undergraduate students have similar age, education level and driving experiences). Using a MNL model that includes common sociodemographic characteristics as segmentation criteria may categorize them into the same group and assumes they have similar preferences. Thus, using the MNL model may not be able to capture the heterogeneous preference among drivers. One the other hand, the individual route choice models obtained with the mixed logit model can capture the preference heterogeneity. Therefore, it is necessary to build individual’s route choice model in the applications such as personalized route recommendation and trip planning which requires the knowledge of drivers’ route choice preferences at the individual level.

Prediction Accuracy at the Individual Level

The applications of personalized route recommendation or trip planning aims at providing better personalized services that he/she would like. Therefore, models’ capabilities of correctly predicting individual driver’s decision or preference are worthy of investigation. Coefficients at the individual level were obtained with the mixed logit model from last section. Each participant has his/her own utility function. This section evaluates these individual route choice models’ capability in correctly predicting individual participant’s route choice decisions. As noted, 80% of each individual participant’s data was used to establish models and the rest 20% was reserved to test models’ performances on each participant. A MNL model was included for comparison. The prediction accuracy in terms of the percentage of models' predictions that match participant's actual choices is used as the measurement for evaluation.

Table 4 summarizes the prediction accuracies of individual route choice models and MNL models on each participant. As shown in Table 4, the prediction accuracy of the individual model is much higher than that of MNL model on most of participants. The average prediction accuracy of the individual models is 89.8%, compared to the prediction accuracy of MNL model which is 73.4%. The individual route choice model and MNL model's prediction accuracies on each participant were compared with a paired t test. The prediction accuracy of the individual route choice model is significantly higher than that of the MNL model (p value: $3*10^{-8}$).

Table 4 Prediction Accuracy of Mixed Logit Model and Multinomial Logit Model on Each Participant

Participant	MNL	Mixed Logit	Participant	MNL	Mixed Logit
1	64.7%	76.5%	23	64.7%	88.2%
2	76.5%	82.4%	24	76.5%	82.4%
3	58.8%	82.4%	25	88.2%	94.1%
4	88.2%	82.4%	26	47.1%	94.1%
5	100.0%	100.0%	27	82.4%	88.2%
6	82.4%	94.1%	28	58.8%	100.0%
7	94.1%	82.4%	29	52.9%	100.0%
8	70.6%	82.4%	30	52.9%	100.0%
9	76.5%	100.0%	31	82.4%	94.1%
10	76.5%	94.1%	32	88.2%	88.2%
11	58.8%	76.5%	33	100.0%	100.0%
12	88.2%	94.1%	34	58.8%	70.6%
13	88.2%	88.2%	35	82.4%	100.0%
14	70.6%	82.4%	36	70.6%	82.4%
15	94.1%	94.1%	37	100.0%	94.1%
16	52.9%	76.5%	38	82.4%	88.2%
17	64.7%	82.4%	39	82.4%	82.4%
18	82.4%	100.0%	40	70.6%	94.1%
19	47.1%	100.0%	41	70.6%	94.1%
20	76.5%	94.1%	42	64.7%	100.0%
21	58.8%	94.1%	43	76.5%	82.4%
22	64.7%	76.5%	44	41.2%	100.0%
Average Prediction Accuracy		MNL: 73.4%			ML: 89.8%

Given the natures of two types of model, the individual route choice models obtained with ML model are expected to have better performances. As analyzed in Table 3, participants' route choice preferences do vary. Facing the same route choice scenario, participants could have different route choice decisions. The individual models obtained with ML model considers the heterogeneous preferences thus has advantageous performance.

CONCLUSIONS

Emerging technologies in the intelligent transportation systems bring up the need to understand driver's preference at the individual level, such as personalized route recommendations or trip planning. This paper applied the mixed logit model together with Bayes rule in capturing and predicting drivers' route choice preferences at the individual level. In addition to estimate the distribution of the heterogeneous route choice preferences, this paper obtained the specific participant's route choice preference at the individual level.

A stated preference dataset was used to demonstrate how to estimate the individual route choice preference parameters under the mixed logit model and the Bayes rule. The estimation results show that participants do have heterogeneous preferences under the same route attribute, even though participants have very similar sociodemographic characteristics (i.e. age, education level and driving experiences). Their preference regarding "Travel Time" varies most among four route attributes. The results also show that participants' preferences regarding different route attributes have either positive or negative correlations as shown in Table 2. At the aggregate level, the mixed logit model has better overall fitness (Pseudo-R^2: 0.6252) than a regular multinomial logit model (Pseudo-R^2: 0.4180). At the individual level, the individual route choice model can help researchers and engineers better understand and predict the route choice preference of a specific individual driver. The evaluation results show that the individual route choice models have 20% higher prediction accuracy than that of a regular multinomial logit model.

With the results of this paper, we demonstrated the mixed logit model's capabilities in capturing and predicting individual's route choice preferences, and provides a new perspective of modeling individual drivers' route choice behaviors. Recent papers also utilized the mixed logit model to capture intra-consumer heterogeneous preferences (Ben-Akiva et al., 2015; Hess and Rose, 2009) and proposed a mixed heuristic model to capture individual-level preferences (Gonzalez-Valdes and Raveau, 2018). These can be explored in the future research for personalized ATIS services. In addition, as the applications of personalized route recommendation and trip planning still have other challenges in practice, such as the need to update the preference model in real time and some users having limited preference data for building model, more modeling approaches will be explored to better serve the purposes of personalized route recommendation and trip planning.

REFERENCES

Abdel-Aty, M., Kitamura, R., Jovanis, P., Vaughn, K., 1994. INVESTIGATION OF CRITERIA INFLUENCING ROUTE CHOICE: INITIAL ANALYSIS USING REVEALED AND STATED PREFERENCE DATA.

Amirgholy, M., Golshani, N., Schneider, C., Gonzales, E.J., Gao, H.O., 2017. An advanced traveler navigation system adapted to route choice preferences of the individual users. International Journal of Transportation Science and Technology, Special Issue on Urban Spatiotemporal Behavior and Network Assignment 6, 240–254. https://doi.org/10.1016/j.ijtst.2017.10.001

Bansal, P., Daziano, R.A., Achtnicht, M., 2017. Extending the logit-mixed logit model for a combination of random and fixed parameters. Journal of Choice Modelling. https://doi.org/10.1016/j.jocm.2017.10.001

Ben-Aakiva, M., Lerman, S.R., 1985. DISCRETE CHOICE ANALYSIS: THEORY AND APPLICATION TO TRAVEL DEMAND.

Ben-Akiva, M., McFadden, D., Train, K., 2015. Foundations of stated preference elicitation consumer behavior and choice-based conjoint analysis.

Ben-Elia, E., Shiftan, Y., 2010. Which road do I take? A learning-based model of route-choice behavior with real-time information. Transportation Research Part A: Policy and Practice 44, 249–264. https://doi.org/10.1016/j.tra.2010.01.007

Bhat, C.R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B: Methodological 37, 837–855. https://doi.org/10.1016/S0191-2615(02)00090-5

Bifulco, G.N., Simonelli, F., Pace, R. di, 2007. Endogenous Driver Compliance and Network Performances under ATIS, in: 2007 IEEE Intelligent Transportation Systems Conference. Presented at the 2007 IEEE Intelligent Transportation Systems Conference, pp. 1028–1033. https://doi.org/10.1109/ITSC.2007.4357722

Department for Transport UK, 2014. Supplementary Guidance Mixed Logit Models. Transport Appriaisal and Strategic Modeling Division, Department for Transport.

Domenico, M.D., Lima, A., González, M.C., Arenas, A., 2015. Personalized routing for multitudes in smart cities. EPJ Data Science 4, 1. https://doi.org/10.1140/epjds/s13688-015-0038-0

Gonzalez-Valdes, F., Raveau, S., 2018. Identifying the presence of heterogeneous discrete choice heuristics at an individual level. Journal of Choice Modelling 28, 28–40. https://doi.org/10.1016/j.jocm.2018.05.001

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological 37, 681–698. https://doi.org/10.1016/S0191-2615(02)00046-2

Haghani, M., Sarvi, M., Shahhoseini, Z., 2015. Accommodating taste heterogeneity and desired substitution pattern in exit choices of pedestrian crowd evacuees using a mixed nested logit model. Journal of Choice Modelling 16, 58–68. https://doi.org/10.1016/j.jocm.2015.09.006

Han, B., Algers, S., Engelson, L., 2001. Accommodating drivers' taste variation and repeated choice correlation in route choice modeling by using the mixed logit model, in: 80th Annual Meeting of the Transportation Research Board.

Hensher, D.A., Greene, W.H., 2003. The Mixed Logit model: The state of practice. Transportation 30, 133–176. https://doi.org/10.1023/A:1022558715350

Hess, S., Rose, J.M., 2009. Allowing for intra-respondent variations in coefficients estimated on repeated choice data. Transportation Research Part B: Methodological 43, 708–719. https://doi.org/10.1016/j.trb.2009.01.007

Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. Journal of Choice Modelling 23, 1–8. https://doi.org/10.1016/j.jocm.2017.03.001

Hetrakul, P., Cirillo, C., 2013. Accommodating taste heterogeneity in railway passenger choice models based on internet booking data. Journal of Choice Modelling 6, 1–16. https://doi.org/10.1016/j.jocm.2013.04.003

Iida, Y., Akiyama, T., Uchida, T., 1992. Experimental analysis of dynamic route choice behavior. Transportation Research Part B: Methodological 26, 17–32. https://doi.org/10.1016/0191-2615(92)90017-Q

Jan, O., Horowitz, A., Peng, Z.-R., 2000. Using Global Positioning System Data to Understand Variations in Path Choice. Transportation Research Record: Journal of the Transportation Research Board 1725, 37–44. https://doi.org/10.3141/1725-06

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. https://doi.org/10.3141/1926-19

Li, H., Huang, H., Liu, J., 2010. Parameter Estimation of the Mixed Logit Model and Its Application. Journal of Transportation Systems Engineering and Information Technology 10, 73–78. https://doi.org/10.1016/S1570-6672(09)60065-9

Liu, H.X., He, X., Recker, W., 2007. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. Transportation Research Part B: Methodological 41, 448–461. https://doi.org/10.1016/j.trb.2006.07.002

Liu, H.X., Recker, W., Chen, A., 2004. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. Transportation Research Part A: Policy and Practice 38, 435–453. https://doi.org/10.1016/j.tra.2004.03.003

Liu, L., Xu, J., Liao, S.S., Chen, H., 2014. A real-time personalized route recommendation system for self-drive tourists based on vehicle to vehicle communication. Expert Systems with Applications 41, 3409–3417. https://doi.org/10.1016/j.eswa.2013.11.035

Nadi, S., Delavar, M.R., 2011. Multi-criteria, personalized route planning using quantifier-guided ordered weighted averaging operators. International Journal of Applied Earth Observation and Geoinformation 13, 322–335. https://doi.org/10.1016/j.jag.2011.01.003

Nahuelhual, L., Loureiro, M.L., Loomis, J., 2004. Using Random Parameters to Account for Heterogeneous Preferences in Contingent Valuation of Public Open Space. Journal of Agricultural and Resource Economics 29, 537–552.

Pahlavani, P., Delavar, M.R., 2014. Multi-criteria route planning based on a driver's preferences in multi-criteria route selection. Transportation Research Part C: Emerging Technologies 40, 14–35. https://doi.org/10.1016/j.trc.2014.01.001

Paz, A., Peeta, S., 2009. Behavior-consistent real-time traffic routing under information provision. Transportation Research Part C: Emerging Technologies 17, 642–661. https://doi.org/10.1016/j.trc.2009.05.006

Peeta, S., Yu, J.W., 2004. Adaptability of a hybrid route choice model to incorporating driver behavior dynamics under information provision. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 34, 243–256. https://doi.org/10.1109/TSMCA.2003.822272

Perk, V.A., DeSalvo, J.S., Rodrigues, T.A., Versoza, N.M., Bovino, S.C., 2011. Improving Value of Travel Time Savings Estimation for More Effective Transportation Project Evaluation (No. BDK85 911-21). Florida Department of Transportation, Tampa, Florida.

Prato, C.G., 2009. Route choice modeling: past, present and future research directions. Journal of Choice Modelling 2, 65–100. https://doi.org/10.1016/S1755-5345(13)70005-8

Revelt, D., Train, K., 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. The Review of Economics and Statistics 80, 647–657. https://doi.org/10.1162/003465398557735

Rogers, S., Fiechter, C., Langley, P., 1999. A Route Advice Agent that Models Driver Preferences. ResearchGate.

Rogers, S., Langley, P., 1998. Personalized Driving Route Recommendations. ResearchGate.

Sándor, Z., Train, K., 2004. Quasi-random simulation of discrete choice models. Transportation Research Part B: Methodological 38, 313–327. https://doi.org/10.1016/S0191-2615(03)00014-6

Sarrias, M., Daziano, R.A., 2017. Individual-specific point and interval conditional estimates of latent class logit parameters. Journal of Choice Modelling. https://doi.org/10.1016/j.jocm.2017.10.004

Scott, D.W., Sain, S.R., 2005. 9 - Multidimensional Density Estimation, in: Rao, C.R., Wegman, E.J., Solka, J.L. (Eds.), Handbook of Statistics, Data Mining and Data Visualization. Elsevier, pp. 229–261. https://doi.org/10.1016/S0169-7161(04)24009-3

Shiftan, Y., Bekhor, S., Albert, G., 2011. Route choice behaviour with pre-trip travel time information. IET Intelligent Transport Systems 5, 183–189. https://doi.org/10.1049/iet-its.2010.0062

Simpson, T.W., Poplinski, J.D., Koch, P.N., Allen, J.K., 2001. Metamodels for computer-based engineering design: survey and recommendations. Engineering with computers 17, 129–150.

Tawfik, A., Rakha, H., 2013. Latent Class Choice Model of Heterogeneous Drivers' Route Choice Behavior Based on Learning in a Real-World Experiment. Transportation Research Record: Journal of the Transportation Research Board 2334, 84–94. https://doi.org/10.3141/2334-09

Tian, H., Gao, S., Fisher, D.L., Post, B., 2012. Mixed-Logit Latent-Class Model of Strategic Route Choice Behavior with Real-Time Information. Presented at the Transportation Research Board Annual Meeting, Washington D.C, pp. 12–2867.

Train, K.E., 2009. Discrete Choice Methods with Simulation, 2 edition. ed. Cambridge University Press, Cambridge ; New York.

Train, K.E. (University of C., 1998. Recreation demand models with taste differences over people. Land economics (USA).

CHAPTER 3: Modeling Individual Drivers' Route Choice Preferences for Personalized Route Recommendation

Bingrong Sun[a], Lin Gong[b], B. Brian Park[a], Hongning Wang[b], Jia Hu[c,d]

[a] Department of Civil and Environmental Engineering, University of Virginia, United States

[b] Department of Computer Science, University of Virginia, United States

[c] Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, China

[d] Institute for Advanced Study, Tongji University, China

ABSTRACT

Current route guidance systems have simplified assumptions about drivers' route choice preferences and cannot adequately accommodate drivers' heterogeneous route choice preferences. Main challenges that route guidance systems do not consider drivers' heterogeneous preferences include: (i) difficulty in acquiring exogenous criteria (e.g., sociodemographic information) that can be used to differentiate drivers' preferences; (ii) difficulty in capturing preference of individuals with limited preference data; and (iii) difficulty in updating route choice models in real time as data accumulates. To address these, this paper introduces a Multi-Task Linear Classification Model Adaption (MT-LinAdapt) model that captures drivers' common aspects of route choice preferences and yet adapts to each driver's own preference. In MT-LinAdapt, an aggregate level model is updated to capture the common aspects of drivers' route choice preference, while individual drivers' models are simultaneously adapted from the aggregate model to capture each driver's own preference. MT-LinAdapt was evaluated against three existing route choice models including an aggregate model, an individual model and a mixed logit model. With three stated preference datasets and three synthetic datasets, MT-LinAdapt's performance was compared to three existing models in two representative scenarios of route guidance applications: (i) when users have adequate historical preference data, and (ii) when users have limited historical preference data. The evaluation results showed that, with survey datasets, MT-LinAdapt achieved up to 8% higher prediction accuracy in the adequate data scenario and up to 18% higher prediction accuracy in the limited data scenario than the existing models. The advantages of MT-LinAdapt are even greater when users' route choice preferences become more heterogeneous.

Keywords: Individuals' Route Choice Preferences; Route Recommendation; Multi-Task Linear Classification Model Adaptation; Mixed Logit; Support Vector Machine; Route Guidance System.

INTRODUCTION

The route guidance system is an important part of Advanced Traveler Information System (ATIS) (Ben-Elia et al., 2013). A user can obtain information about possible routes connecting his/her origin and destination by using a route guidance system and makes an informed decision regarding which route to take. Routes are usually recommended based on a single criterion such as the shortest travel time or the shortest distance without fully considering the driver's own preference. Existing route guidance systems, such as Google Maps and Waze, typically offer limited options for users to customize his/her own preferences. For example, options of avoiding highways, tolls, ferry or dirt roads are available. However, many studies have shown that drivers do not make route choices solely based on a single criterion such as travel time or distance (Ben-Akiva et al., 1984; Lima et al., 2016), and drivers also have heterogeneous route choice preferences (Li et al., 2016). For example, drivers perceive information differently (Tawfik et al., 2010; Parthasarathi et al., 2013; Peruch et al., 1989; Matthews, 1981), value route attributes differently (Feng et al., August 11, 2; Liu et al., 2007; Perk et al., 2011; Rogers et al., 1999; Rogers and Langley, 1998) or have various preferences at different time (Amirgholy et al., 2017; Liu et al., 2007). In other words, drivers' heterogeneous route choice preferences were not adequately considered in current route guidance systems. The reason that route guidance systems do not consider individual driver's route choice preference is due to some challenges in practice. For example, it is difficult for route guidance systems to obtain sociodemographic information or other exogenous criteria that existing modeling approaches typically require to differentiate drivers' preferences. Also, it is difficult to capture the preferences of users who have a limited amount of preference data, and it is difficult for route guidance systems to update captured preference efficiently when more preference data accumulates, and so on. Therefore, this paper focuses on the problem of how to tackle these challenges and consider each individual driver's route choice preference for route recommendations in route guidance systems.

It is important to consider each individual user's route choice preference in a route guidance system. The first reason is to improve users' satisfactions by providing route recommendations that users would like to take. Amirgholy et al. (Amirgholy et al., 2017) found that, in around 60% of trips that drivers use route guidance service, drivers either do not like the suggested routes before the trips were started or change to other routes during their trips. Some drivers also mentioned the experience of ignoring suggested turns until the route guidance system re-calculated a route that avoids a certain area in the city (The user's preference was to avoid that area) ("How does the navigation system choose a route?," n.d.). If each individual's specific route choice preference can be properly considered in the route guidance system, these phenomena could be largely mitigated. The other even more important reason is to improve the transportation system performance. Considering individual driver's route choice preference can increase driver's compliance with route guidance and thus improve network performance, especially when recommendations are made for improving the system performance. Several studies have shown that driver's compliance rate to recommendations has significant impacts on network modeling accuracy (Wang et al., 2017) and a high compliance rate is a requirement for improving road network performance (Bifulco et al., 2007; Domenico et al., 2015; Paz and Peeta, 2009). A higher compliance rate means drivers should be more likely to take the suggested routes. In other words, the route guidance system has to consider users' route choice preferences in order to achieve a higher compliance rate and better system performance, especially when the guidance was made for improving the network performance, such as giving guidance for avoiding traffic congestion (Bazzan and Klügl, 2005), achieving a system optimum goal (Klein et al., 2018), reducing system-wide delay (Ma et al., 2016), efficiently reallocating system capacity (Adler and Blue, 2002) and even providing personalized sustainable travel incentives (Azevedo et al., 2018).

Many researchers have studied drivers' route choice preferences. To capture and describe drivers' route choice preferences, researchers developed various route choice models. Given the focus of this paper is how to consider individual driver's route choice preference and provide better route guidance service that every individual user may like, existing route choice models are divided into three types in this paper according to the data used for establishing models. The three types of models are aggregate route choice models which are based on preference data at the aggregate level, individual route choice models which are based on individual user's own preference data, and mixed logit models that combine both aggregate and individual levels data. Each of them is described in this section.

Aggregate Route Choice Models

Aggregate models are established based on route choice preference data from a group of representative drivers. Data collected from all drivers are put together to build a model that can be applied to everyone.

It usually comes with an assumption that drivers who have the same sociodemographic characteristics would share the same route choice preference (Ben-Akiva and Lerman, 1985). Therefore, sociodemographic characteristics such as age, gender, income, etc. are also included in the model as criteria to differentiate drivers' route choice preferences. Following this concept, aggregate models are established with different modeling approaches, including discrete choice models and machine learning methods. Discrete choice model family calibrates drivers' utility functions and calculate an individual driver's probability of choosing alternative routes, such as multinomial logit models (Ben-Akiva and Lerman, 1985; Prato, 2009). Machine learning methods treat a route choice decision as a classification problem in the sense of classifying a route into the category of choosing or not choosing. Different machine learning methods that have been investigated by researchers for route choice modeling include neural network (Yang et al., 1993), hybrid route choice model (Peeta and Yu, 2005), support vector machine (Lee and Li, 2016; Sun and Park, 2017), decision tree (Park et al., 2007) etc. Some of these machine learning techniques were compared with traditional discrete choice models in traveler behavior study and sometimes showed better performances (Yamamoto et al., 2002; Zhang and Xie, 2008). Another variation of aggregate route choice models is the multi-class route choice model. It usually first divides drivers into different groups based on certain criteria (for example learning and choice evolution pattern) and then build a model for each group (Peeta and Yu, 2004; Tawfik and Rakha, 2013). Drivers within the same group share the same route choice preference. Therefore, with aggregate route choice models, drivers with same sociodemographic characteristics or belonging to the same class have the same possibilities of choosing alternative routes.

The application of route guidance provides route recommendations to individual drivers. Using sociodemographic information to differentiate drivers' preferences make aggregate models lack sufficient preference heterogeneity to accurately predict each user's route choice preference, as drivers who have the same sociodemographic characteristics can still have different route choice preferences. In addition, users' sociodemographic information is required when applying aggregate models to predict users' route choice decisions. However, it is not easy to obtain some sociodemographic information in route guidance process due to privacy concerns, for example, age, income level, education, household structure, profession, number of cars in family, etc., which are usually included in aggregate route choice models (Jan et al., 2000). Aggregate models divide drivers into different classes with different preferences based on any other exogenous criteria such as learning patterns. However, it is difficult to pick proper criteria for making segmentations (Hensher and Greene, 2003; Peeta and Yu, 2004) and the heterogeneity existing within a class is still difficult to determine.

Individual Route Choice Models

Individual models are established based on individual driver's own route choice preference data. With emerging information and communication technologies, it is possible for route guidance systems to collect route choice preference data at an individual driver's level. The data is usually collected from each single driver by observing his/her route choice behaviors from multiple either stated or realistic route choice scenarios (Mahmassani et al., 2013). Route choice models can be established for each user based on his/her own preference data. Therefore, it does not require segmenting drivers into different groups based on either sociodemographic characteristics or other exogenous criteria.

The individual route choice models are most commonly used in the personalized route guidance (Nadi and Delavar, 2011; Pahlavani et al., 2012; Pahlavani and Delavar, 2014; Liu et al., 2014). Rogers et al. (1999; 1998) collected stated preference route choice data from 24 participants and used the method of differential perceptron to capture individual participant's preference. Park et al. (2007) generated route choice data from simulation and used decision tree to model individual "driver's" adaptive route choice preference. Nadi and Delavar (2011) conducted a survey among 32 tourists and used a pairwise comparison method and an ordered weighted averaging method to incorporate their different route choice preferences. In Pahlavani and Delavar' research (2014), a participant selected the criteria, rating scale and then rated a sample route set to get the initial training data. The data was used to train a linear neuro-fuzzy model to learn the driver's route selection decisions.

Individual route choice models can capture heterogeneous route choice preference to the maximum extent. However, in order to build a valid or meaningful individual route choice model, it requires a certain amount of data. The data amounts that were used by researchers for building individual models have values of 675 accumulated trips (Park et al., 2007) and 232 driver-rated virtual routes (Pahlavani and Delavar, 2014). In real life, it might take a certain length of time to get this amount of data. In practice, drivers may give up using a new route guidance system after several times of unsatisfied experiences, so the prediction accuracy of the route guidance system should be good when only several trips' (for example, 5 to 6 trips) preference data is available. In addition, an individual route choice model is built based on subject driver's historical preference data. When new trip scenarios are not covered by historical data, it is very likely that the model does not work well in the new scenarios. In reality, drivers' preference could vary with different trip purposes, different departure time, and different distances, etc. It is impossible to collect the data that cover all possible trip characteristics of any particular driver may face.

Mixed Logit Models

Mixed logit models belong to discrete choice model family. Mixed logit models treat each utility function coefficient as a random parameter following certain distribution and thus account for the heterogeneous preferences in population. They utilize the data from all sampled drivers to estimate distribution parameters at the aggregate level and also utilize a particular driver's own data to adapt to his/her taste. Mixed logit models have been used in modeling drivers' route choice behaviors and are considered more realistic because of their capabilities of considering heterogeneous preferences (Ben-Elia and Shiftan, 2010; Greene and Hensher, 2003; Han et al., 2001; Liu et al., 2004; Razo and Gao, 2013; Tian et al., 2012). Recently, Amirgholy et al., (2017) proposed an advanced travelers navigation system with a dynamic mixed logit model which can adapt to individual driver's preference, but the performance of their proposed model was not demonstrated yet. Also, existing literature using mixed logit models for route choice modeling did not explore their performance at individual driver's level, which is essential to the application of route guidance. Meanwhile, mixed logit models were used in other domains to analyze preferences at the individual level (Revelt and Train, 1998; Train, 1998). Therefore, mixed logit models have potential to be used for capturing route choice preference at the individual level. However, mixed logit models also have some unideal features for solving the research question of this paper, such as they are usually solved with simulation methods (Han et al., 2001; Revelt and Train, 2001). Simulation method has the vulnerabilities such as the sacrifice of model accuracy and being inefficient to calculate. The estimated results could be affected by the number of random draws used in the simulation and the way to conduct random draws (Train, 2009).

As discussed above, to capture individual drivers' route choice preferences in route guidance systems, three types of existing modeling approaches have their own limitations in tackling the challenges of existing route guidance systems. Aggregate route choice models have assumptions that drivers' preferences can be distinguished with sociodemographic characteristics or other criteria and require those data when applying the models. The assumption and requirement limit aggregate models' capabilities of incorporating user's heterogeneous preferences. Individual route choice models require a certain amount of and enough coverage of a user's historical preference data. A model established based on personal historical data may not work well in new scenarios. As to mixed logit models, using the simulation method for model estimation could be inefficient or damage models' accuracies.

Research Objective

When looking at drivers' route choice preferences in the real world, drivers share some homogeneous aspects of route choice preferences (e.g., all drivers like the route with shorter travel time) and meanwhile each driver has his/her own emphasis (e.g., some drivers prefer routes with less cost while others like more expensive but more reliable routes). For applications of route guidance which need to provide recommendations to each individual user, a possible approach is to capture the homogenous part of drivers' route choice preferences and adapt to individual's route choice preference at the same time. Meanwhile, any changes or new contributions from individual's preference should be updated to aggregate homogeneous part of preferences. Following this concept, this paper introduces Multi-Task Linear Classification Model Adaptation (MT-LinAdapt), a route choice modeling approach that was designed to identify the homogeneous route choice preference for all drivers and also capture the heterogeneous route choice preference existing among individuals. Thus, the model is expected to overcome the limitations of existing route choice models in the application of route guidance and provide better route recommendations to each individual user based on his/her own preference.

MT-LinAdapt is expected to have following featured capabilities:

- Does not need exogenous criteria such as sociodemographic characteristics to differentiate drivers' heterogeneous preferences;
- Works well when an individual driver has a limited amount of preference data;
- Capable of being updated in real time as additional data accumulates;
- Allowing drivers have own unique preference in route guidance systems;
- Compatible with any linear classification models.

The rest of the paper is organized as follows. The MT-LinAdapt model is firstly introduced in the Methodology section. Then, three groups of data collected from stated preference surveys are discussed in the Survey and Data section. In the Model Comparison section, two comparison scenarios are set up and three commonly used existing route choice models are selected for comparison. The implementation of MT-LinAdapt and selected existing models are discussed. Then, the performance of the MT-

LinAdapt model against selected existing models is evaluated in the Results Analysis section. At last, some discussion and conclusions are made.

METHODOLOGY

This paper introduces the MT-LinAdapt model which is expected to tackle the challenges of considering individual driver's preference in route guidance systems. MT-LinAdapt roots in social psychology theories and treats the formation of sentiment as a social norm (Gong et al., 2016). With the social norm theory, individual's opinion or decision usually is largely affected by other society members' opinions. Thus, members of the society have common criteria used to make decisions or form opinions. Meanwhile, each member has his/her own preference that is different from other member's. Members influence each other and the social norm of the whole society tends to shift or evolve. Based on how social norm forms and evolves, MT-LinAdapt tries to minimize the error rates of sentiment classification at the individual level and the aggregate level together by defining it as a joint optimization problem.

It is generally understood that drivers' route choice preferences follow a kind of social norm. Drivers tend to have some common criteria to choose one route over the others, while each individual driver has his/her own emphasis that is different from other drivers. Given a group of drivers, MT-LinAdapt can identify the homogeneous route choice preference across all drivers (for example, all drivers like the route with shorter travel time), then capture the heterogeneous route choice preference existing among individuals (for example, some drivers prefer routes with less cost while others like more expensive but more reliable routes). Instead of requiring drivers' sociodemographic characteristics or other criteria to differentiate their route choice preferences, the MT-LinAdapt model adapts the aggregate route choice preference to an individual level so that individuals' preference can be captured. Meanwhile, the change of individual drivers' preference could lead to drivers' aggregate preferences shifting and evolving. These shifting and evolving can also be captured by MT-LinAdapt.

Therefore, there are two adaptation processes: the adaptation from the aggregate preference to individual preference, and the adaptation of the aggregate preference. A linear classification model typically has a form of $y = sign(\boldsymbol{w}\boldsymbol{X} + c)$ in which $\boldsymbol{w}$ is a weight vector, $\boldsymbol{X}$ is the feature vector, c is the intercept and y is the predicted classification label. $\boldsymbol{X}$ contains the attributes of alternative routes in route choice scenario. $\boldsymbol{w}$ indicates how important each route attribute is. At the aggregate level of MT-LinAdapt, all drivers share the same weight vector, $\boldsymbol{w_s}$**,** which is a *k* dimensional vector. *k* is the number of route attributes that affects drivers' route choice decisions.

When adapting the aggregate preference into the individual level, all individuals' weight vectors can be obtained by:

$$\boldsymbol{W} = [\boldsymbol{w_1}, \boldsymbol{w_2}, \dots, \boldsymbol{w_i}, \dots, \boldsymbol{w_n}] = \begin{bmatrix} w_{11} & w_{21} & \dots & w_{i1} & \dots & w_{n1} \\ w_{12} & w_{22} & \dots & w_{i2} & \dots & w_{n2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{1k} & w_{2k} & \dots & w_{ik} & \dots & w_{nk} \end{bmatrix} = [\boldsymbol{w_s}, \boldsymbol{w_s}, \dots, \boldsymbol{w_s}] \circ \boldsymbol{A_u} + \boldsymbol{B_u}$$

$$= \begin{bmatrix} w_{s1} & w_{s1} & \cdots & w_{s1} \\ w_{s2} & w_{s2} & \cdots & w_{s2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{sk} & w_{sk} & \cdots & w_{sk} \end{bmatrix} \circ \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{n1} \\ b_{12} & b_{22} & \cdots & b_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1k} & b_{2k} & \cdots & b_{nk} \end{bmatrix} \qquad (1)$$

W is the matrix in which column *i* represents Driver *i*'s weight vector. Each weight vector contains *k* elements corresponding to the weights for *k* route attributes. Each driver's weight vector is obtained by scaling and shifting the aggregate preference, $\boldsymbol{w_s}$. Based on individual driver's preference data, the scaling and shifting operations for different drivers are different. $\boldsymbol{a_i}$ and $\boldsymbol{b_i}$ in Matrix $\mathbf{A_u}$ and $\mathbf{B_u}$ represent the specific adaptation operations based on driver *i*'s preference data. ∘ represents the operation to calculate the entry-wise product of two matrices.

Since the aggregate preference evolves when individual drivers' preference changes, $\boldsymbol{w_s}$ also adapts to capture the preferences changing in preference data. Therefore, the similar adaptation process can be conducted to $\boldsymbol{w_s}$ as well.

$$\boldsymbol{w_s} = \boldsymbol{w_0} \circ \boldsymbol{A_s} + \boldsymbol{B_s} = \begin{bmatrix} w_{01} \\ w_{02} \\ \vdots \\ w_{0k} \end{bmatrix} \circ \begin{bmatrix} a_{s1} \\ a_{s2} \\ \vdots \\ a_{sk} \end{bmatrix} + \begin{bmatrix} b_{s1} \\ b_{s2} \\ \vdots \\ b_{sk} \end{bmatrix} \qquad (2)$$

$\boldsymbol{w_0}$ is a *k* dimensional vector representing a prior weight vector which can be obtained by building a model based on a dataset consists of a small portion from every individual's data. Any linear classification model can be incorporated into Equation (1) and Equation (2). Depending on the specific classification model, the problem becomes to find the $\boldsymbol{A_u}$, $\boldsymbol{B_u}$, $\boldsymbol{A_s}$ and $\boldsymbol{B_s}$ that can minimize the prediction errors at both aggregate and individual levels.

To demonstrate how MT-LinAdapt works, this paper adopts logistic regression as the linear classification model with a binary route choice scenario to show how $\boldsymbol{A_u}$, $\boldsymbol{B_u}$, $\boldsymbol{A_s}$ and $\boldsymbol{B_s}$ can be obtained. It is noted that MT-LinAdapt can incorporate other linear classification models and can be extended to scenarios with multiple alternatives. When using individual drivers' weights in Logistic Regression, the probability of choosing alternative *1* for driver *i* (*i=1,2,…,n*) in scenario *j* is:

$$P_{ij}\left(y=1\middle|\boldsymbol{x_j}\right)=\frac{\exp\left(\boldsymbol{w_i x_{j1}}\right)}{\exp\left(\boldsymbol{w_i x_{j1}}\right)+\exp\left(\boldsymbol{w_i x_{j0}}\right)}$$

$$=\frac{\exp((\boldsymbol{a_i}\circ\boldsymbol{w_s}+\boldsymbol{b_i})\boldsymbol{x_{j1}})}{\exp((\boldsymbol{a_i}\circ\boldsymbol{w_s}+\boldsymbol{b_i})\boldsymbol{x_{j1}})+\exp((\boldsymbol{a_i}\circ\boldsymbol{w_s}+\boldsymbol{b_i})\boldsymbol{x_{j0}})}$$

$$=\frac{\exp((\boldsymbol{a_i}(\boldsymbol{w_0}\circ\boldsymbol{A_s}+\boldsymbol{B_s})+\boldsymbol{b_i})\boldsymbol{x_{j1}})}{\exp((\boldsymbol{a_i}(\boldsymbol{w_0}\circ\boldsymbol{A_s}+\boldsymbol{B_s})+\boldsymbol{b_i})\boldsymbol{x_{j1}})+\exp((\boldsymbol{a_i}(\boldsymbol{w_0}\circ\boldsymbol{A_s}+\boldsymbol{B_s})+\boldsymbol{b_i})\boldsymbol{x_{j0}})} \tag{3}$$

$\boldsymbol{X_j}$ includes route attributes that driver *i* experienced in route choice scenario *j*. $\boldsymbol{x_{jm}}$ is the route attributes of alternative *m* (*m=0* or *1*). Therefore, $\boldsymbol{A_u}$, $\boldsymbol{B_u}$, $\boldsymbol{A_s}$ and $\boldsymbol{B_s}$ can be retrieved by maximizing log-likelihood. The log-Likelihood function for driver *i* with all scenarios that he/she experienced is:

$$L_i(\boldsymbol{a_i},\boldsymbol{A_s},\boldsymbol{b_i},\boldsymbol{B_s})=\sum_{j=1}^{J}[y_j logP_{ij}\left(y_j=1\middle|x_j\right)+\left(1-y_j\right)logP_{ij}\left(y_j=0\middle|x_j\right)] \tag{4}$$

in which y_j is the user's choice in scenario *j*. As the MT-LinAdapt model tries to fit each individual driver's preference, it can be very sensitive to individual's historical data. This could lead to overfitting when a particular driver has very limited data (for example, 1 or 2 observations). In other words, the model can fit very limited data well but fails to capture this driver's general preference. To avoid this overfitting issue, regularization terms are added to both the individual level (5a) and the aggregate level (5b), as shown below.

$$R(\boldsymbol{a_i},\boldsymbol{b_i})=\frac{1}{2}\eta_1(\boldsymbol{a_i}-\boldsymbol{I})^T(\boldsymbol{a_i}-\boldsymbol{I})+\frac{1}{2}\eta_2\boldsymbol{b_i}^T\boldsymbol{b_i} \tag{5a}$$

$$R(\boldsymbol{A_s},\boldsymbol{B_s})=\frac{1}{2}\eta_3(\boldsymbol{a_s}-\boldsymbol{I})^T(\boldsymbol{a_s}-\boldsymbol{I})+\frac{1}{2}\eta_4\boldsymbol{b_s}^T\boldsymbol{b_s} \tag{5b}$$

The regularization terms are added to the log-likelihood function as penalties. They penalize the log-likelihood function when $\boldsymbol{A_u}$, $\boldsymbol{B_u}$, $\boldsymbol{A_s}$ and $\boldsymbol{B_s}$ deviate too much from keeping weights unchanged, in other words, scaling weight vectors by 1 and shifting weight vectors by 0. Therefore, taking *N* drivers' preference data together, the objective function is:

$$\max L(\boldsymbol{A_n}, \boldsymbol{B_n}, \boldsymbol{A_s}, \boldsymbol{B_s}) = \sum_{i=1}^{N}[L_i(\boldsymbol{a_i}, \boldsymbol{b_i}) - R(\boldsymbol{a_i}, \boldsymbol{b_i})] - R(\boldsymbol{A_s}, \boldsymbol{B_s}) \qquad (6)$$

Which can be efficiently solved by a gradient-based optimizer. The parameters η_1, η_2, η_3 and η_4 need to be tuned to make the model work the best. This problem can be viewed as a joint maximization problem. The problem is converted to find the $\boldsymbol{A_u}$, $\boldsymbol{B_u}$, $\boldsymbol{A_s}$ and $\boldsymbol{B_s}$ that maximize the log-likelihood function.

Equation (6) can be used to estimate MT-LinAdapt with all users' data together. In practice, as more preference data accumulates, route guidance systems need to update MT-LinAdapt in real time so that users' preference could be accurately captured. Considering the tremendous data that might be handled by route guidance systems, retraining the whole model may not be an efficient option. MT-LinAdapt can be updated in real time when more data accumulates. Equations (7) to (10) demonstrate how to update feature k's weight at both individual and aggregate level. For example, a driver i has a new data observation ($\boldsymbol{x_j}$,y_j). His/her weight for feature k can be updated with Equations (7) and (8). The impacts of this individual's new observation on aggregate weight can be obtained with Equations (9) and (10).

$$\frac{\partial L(\boldsymbol{A_u}, \boldsymbol{A_s}, \boldsymbol{B_u}, \boldsymbol{B_s})}{\partial a_{ik}} = \Delta_{ij}(a_{sk}w_{0k} + b_{sk})x_{jk} \qquad (7)$$

$$\frac{\partial L(\boldsymbol{A_u}, \boldsymbol{A_s}, \boldsymbol{B_u}, \boldsymbol{B_s})}{\partial b_{ik}} = \Delta_{ij}x_{jk} \qquad (8)$$

$$\frac{\partial L(\boldsymbol{A_u}, \boldsymbol{A_s}, \boldsymbol{B_u}, \boldsymbol{B_s})}{\partial a_{sk}} = \Delta_{ij}a_{ik}w_{0k}x_{jk} \qquad (9)$$

$$\frac{\partial L(\boldsymbol{A_u}, \boldsymbol{A_s}, \boldsymbol{B_u}, \boldsymbol{B_s})}{\partial b_{sk}} = \Delta_{ij}a_{ik}x_{jk} \qquad (10)$$

in which $\Delta_{ij} = y_j - P_i(y_j = 1|\boldsymbol{x_j})$. With Equations (7) to (10), MT-LinAdapt can efficiently capture drivers' evolving route choice preferences.

SURVEY AND DATA

Ideally, for route choice behavior study, drivers' historical preferences data should be collected from real route guidance systems including recommended route sets, route attributes and drivers' final route choice decisions. A driver's final route choice decision could be obtained by checking his/her travel profile which is available in some route guidance service providers, such as Google Maps or Waze. The driver's final choice can be captured by analyzing his/her GPS trajectory, as proven by several researchers (Derevitskiy et al., 2016; Herring, 2010). However, the recommended route sets and associated route attributes of each alternative are practically impossible to be accessed by researchers. It imposes difficulties of using revealed preference data for the analysis at the current stage. To obtain the route choice preference data that can represent drivers' historical preference data when using route guidance systems, stated preference surveys were designed to collect data. Once there are promising results, more efforts can be put later into collecting revealed preference data while using real world route guidance systems.

Stated Preference Datasets

We have collected the route choice preference data through three stated preference surveys. By showing participants recommended routes with associated information and recording their choices in the questionnaires, basic data components that can be accumulated in route guidance systems were obtained. Three surveys were designed and conducted by following the same design procedure. In order to generate realistic survey scenarios, the basic procedure used for survey design was to check the traffic information on a current popular navigation system (Google Maps) and include adequate variations for survey questions. Then, origin and destination (OD) pairs at three distance levels were chosen from local areas. Each OD pair has two to three recommended alternative routes in Google Maps. The questions in all surveys were binary route choice scenarios. The details are discussed as follows.

In the survey Dataset 1, route attributes considered include *Distance*, *Travel Time*, *Possible Longest Travel Time*, *Number of Controlled Intersections* and *Fuel Cost*. The information about these attributes was checked on Google Maps within a day and across days of the week to have realistic values as well as adequate variations, except *Fuel Cost* was calculated based on Equation (11). The fuel efficiency of Ford vehicles ranging between 22 and 32 miles per gallon was used. Based on this, three levels of fuel efficiencies, 22, 27 and 32 miles per gallon, were used to calculate fuel costs.

$$Fuel\ cost = \frac{Distance}{Fuel\ effciency} * Fuel\ price \qquad (11)$$

Taguchi design in Minitab (14) was used to make an experiment design including eight variables with each having three levels. The survey contained three scenarios like this. In total, there were 81 questions for three scenarios. Scrutinizing efforts were made to eliminate dominated questions in which one route was absolutely superior to the other in every aspect. Finally, 74 questions were left to form the final questionnaire. All the questions were shuffled before showing to participants. The trip purpose of this survey was for a casual trip.

Dataset 2 and Dataset 3 were collected from another two stated preference surveys which were designed following the same procedure, except that they had different route attributes in survey questions. Route attributes considered in Dataset 2 include *Distance, Travel Time, Number of Controlled Intersections* and *Pedestrian Level*. Information about the first three attributes was obtained in the same way as the first survey. *Pedestrian Level* had two levels: high and low. Taguchi design generated 81 survey questions. The trip purpose of this survey was assumed to be a casual trip. In Dataset 3, route attributes included *Distance, Travel Time, Possible Longest Travel Time* and *Number of Controlled Intersections*. Following the same procedure, information of these attributes was obtained. After applied the experiment design, 81 survey questions were generated and scrutinized in order to remove the dominant questions. Eventually, 63 survey questions were kept in the final survey questionnaire. The trip purpose was assumed to be commute trips.

Three datasets in total were collected from 102 participants. As noted, three datasets were collected from three different groups of participants with different trip purposes. Therefore, preference reflected from each dataset could be different. Dataset 1 was collected from 28 participants who were mostly undergraduate students at the University of Virginia. The participants were invited to sit in a driving simulator and were shown with the information about two routes. They were asked to choose the routes that they prefer and their choices were recorded. Each subject went through all 74 survey questions (except one participant only finished 60 questions). The stated preference surveys of Dataset 2 and Dataset 3 were conducted as online surveys. The linkages of online questionnaires were sent to participants who volunteered for the surveys. The participants were informed with the time the survey would take and were asked only to participate if they would take their time and answer the questions to the best of their ability. 44 and 30 participants took part in each survey, respectively. Most of the participants were students of the University of Virginia. In the surveys, efforts were made to ensure answers quality such as including repeated questions or dominant questions.

Synthetic Datasets

In reality, heterogeneity of drivers' route choice preferences varies. The most homogeneous situation is that every driver has the same preference, namely, they care about the same route attributes with the same importance when making route choice decisions. The most heterogeneous situation is that everyone values different route attributes with different importance when making route choice decisions. Drivers' preferences in the real world would be in between of these two extreme cases. The most homogeneous case, however, is not likely to happen and would be easy to model. On the other hand, the most heterogeneous case is not likely but could happen. Given most survey participants are college students in this research, the route choice preferences contained in the survey data could be less heterogeneous than real world with general drivers. Thus, synthetic datasets were made based on the survey questions to generate datasets with more heterogeneous preferences.

To generate a group of synthetic drivers with more heterogeneous preferences than stated preference datasets, each synthetic driver was assumed to only care one or a combination of route attributes (the number of route attributes considered could be *1, 2, 3… k*, where *k* is the maximum number of attributes that a survey contains). For example, a synthetic driver was assumed to only care the attribute of *Travel Time*, then his/her responses to the survey questions would be the routes with shorter travel time. Another synthetic driver who cares the attribute combination of *Travel Time* and *Fuel Cost* would choose the route with shorter travel time and cheaper fuel cost. When criteria have conflicts, namely one route has shorter travel time but the other has lower cost, this synthetic driver would take one of the routes randomly.

Synthetic drivers with synthetic preferences went through three survey questionnaires and generated three sets of data. Given the number of attributes considered in each dataset and the possible combinations of attributes that synthetic drivers care about, synthetic Dataset 1 has 30 synthetic drivers. Synthetic Dataset 2 and Synthetic Dataset 3 have 14 synthetic drivers. Their responses were also used in later analysis to represent more heterogeneous cases than stated preference datasets based on college students.

MODEL COMPARISONS

Comparison Scenarios

Two representative application scenarios of route guidance systems were set up to evaluate the MT-LinAdapt model. To consider individual drivers' route choice preferences in route guidance systems, these two scenarios represent the possible challenges that route guidance systems facing.

Scenario 1: Users have adequate historical preference data

This scenario represents a challenge that route guidance systems are facing in practice, namely long term users have adequate historical data but their sociodemographic information is not available possibly due to privacy concerns. In this scenario, the historical data covers a particular user's preference regarding most of possible trips with different levels of route attributes (such as distance, travel time, and etc.), different time of day, different trip purposes, and so on.

Scenario 2: Users have limited historical preference data

This scenario represents a challenge that there is a very limited amount of historical data for an individual user, for example, a new user or a tourist just starts using the route guidance system or the route guidance system just starts operation. It is difficult for the route guidance system to learn this particular driver's preference with a limited amount of data, for instance, just using the route guidance system for 5 to 6 trips. How to generate good route recommendations with the limited amount of data at this point is very important, because it is essential to guarantee users' satisfaction so that the system could keep its users.

Select Commonly Used Existing Models for Comparison

To have a better understanding regarding MT-LinAdapt model's performance, three representative existing models were selected to compare with it. Based on existing types of route choice models as reviewed in the Introduction section, an aggregate model, an individual model and a mixed logit model were selected for comparison. One of commonly used machine learning methods, Support Vector Machine (SVM), was chosen here to build the aggregate and the individual route choice models, as researchers have used it for many traveler behavior analysis such as route choice and mode choice modeling (Lee and Li, 2016; Sun and Park, 2017; Zhang and Xie, 2008). SVM demonstrated good performance in these applications. The concept of SVM is to map the data points into high dimensional space and find a hyperplane which can separate the points belonging to different categories. The estimation of SVM model is to maximize the distance of all data points to the separation plane. Readers could refer to several literatures (Steinwart and Christmann, 2008; Zhang and Xie, 2008) for detailed objective functions and constraints. A mixed logit model was also selected for comparison. It allows users' preferences regarding a route attribute following a certain distribution. With all users' data, parameters of the preference distribution can be estimated (e.g., with an assumption of normal distribution, the mean and the standard deviation are parameters to be estimated). Based on the estimated distributions, individual's expected preference regarding certain route attribute can be obtained with Bayes' rule. For a detailed explanation, readers can refer to Chapter 6 and Chapter 11 of *Discrete Choice Method with Simulation* (Train, 2009).

Data Preparation

Each dataset was divided into training data and testing data. The training data was used for establishing models and the testing data was used for evaluating the performance of the established models. Training data and testing data were developed from each of the six datasets. 20% of each individual participant's data was randomly selected as testing data. All established models are to be tested on each participant's testing data to see models' performance on each individual user. The rest 80% data of each individual was used to form the training data for different types of route choice models.

For aggregate models, all participant's training data in each dataset was put together as the training dataset. Each route choice observation was considered as an independent data point. Neither sociodemographic data nor participant's identification was included. This training dataset was used for building aggregated SVM models.

For individual models, each participant's training data was used alone to build a route choice model for him/herself. This training dataset was used to build individual SVM models.

As to MT-LinAdapt and mixed logit models, their training data was formed with aggregate models' training data with an additional column indicating participants' identity number.

All datasets were standardized by each route attribute. To be more specific, in each dataset, each column contains the values of a route attribute. Each value in a column was subtracted from the column mean and divided by the column standard deviation. The datasets after standardization were used for later model training and testing.

The random divisions between training and testing data were conducted fifty times to avoid data divisions' impacts on model performance. For each time of data division, models established based on training data were tested on each individual's testing data. The performance measurement used here is prediction accuracy which is defined as the percentage that model's predicted choices match a participant's actual choices. With fifty times of data division, each model has fifty prediction accuracies for each individual. The average prediction accuracy for each individual was used for final model comparisons.

For Scenario 1 comparison, all training data which represents adequate amount data was used for building models. For Scenario 2 comparison, each individual participant's testing data was still kept for testing models' performance, but the training data was randomly divided into ten groups with each group having only 10% of training data. Then, the MT-LinAdapt model, as well as other commonly used existing models, were established with gradually increased percentages of data, namely 10%, 20%, 30%...100% of training data. Therefore, the minimum amount of data that was used for the model development was 10% of the training dataset. That is equivalent to around 5 to 6 data points per participant in our datasets. In practice, route guidance systems could either obtain users' preferences by analyzing user's first 5 to 6 trips or asking each user 5 to 6 stated preferences questions when users sign up for the service.

Model Implementation

The application details of MT-LinAdapt and three commonly used existing models were elaborated in this section.

Support Vector Machine (SVM)

The Sklearn package with Python was used for training SVM. Since SVM models with linear kernel function have decent performance in travelers' behavior study (Zhang and Xie, 2008; Lee and Li, 2016; Sun and Park, 2017), the linear kernel function was adopted for the SVM model in this research.

SVM has a penalty parameter that needs to be determined with cross validation. The range explored in the cross validation is a geometric sequence from 10^{-5} to 10^{5} by a factor of 10, which is a commonly used range for penalty parameter in SVM (Ben-Hur and Weston, 2010). The training data was further split into five groups. Each group was used as validation data once on all possible values. This random split was conducted five times. The value with highest average performance on validation data was selected to be the penalty parameter value.

SVM was used to establish both aggregate models and individual models, therefore it was applied to training datasets prepared for aggregate models and individual models, respectively. Two types of models were tested on each individual user's testing data to obtain the prediction accuracies of the aggregate models and the individual models.

Mixed Logit Model

Software R with the gmnl package (Sarrias and Daziano, 2017) was used to estimate the mixed logit model. The package was chosen because it can output random parameters' distributions in terms of means and standard deviations among sampled data, as well as model coefficients at the individual level. The utility model's coefficients for each individual participant were used to predict this participant's choices in his/her testing data.

Two settings for the mixed logit model should be determined before model estimation, including the parameters' distributions and the number of draws used in the simulation. Each of them was discussed in detail as follows.

Each random parameter follows an underlying distribution among the population. The type of the distribution should be defined before estimating mixed logit models. Common distributions that were used in transportation include normal distribution, lognormal, uniform and triangular (Hensher and Greene, 2003). Hensher and Greence (2003) suggested using empirical distributions that were observed from data. The process is to possibly obtain the parameter estimate for each sampled individual and then plot everyone's parameter estimates so that a distribution shape could be observed. Kernel density estimator was used to plot each parameter's distribution in the dataset. The shape was observed to obtain the underlying distribution for each parameter among participants. Following the same procedure (Hensher and Greene, 2003), final settings for most parameters were normal distributions, except "Pedestrian Level" in Dataset 2 and "Fuel Cost" in Dataset 1 using uniform distributions. One thing should be noted is that the shape of "Numbers of Controlled Intersections" in Dataset 1 looks more like a lognormal distribution. However, the starting values in the estimation process were difficult to find, which appears to be a common issue for lognormal distribution in mixed logit models (Train, 2009; Han et al., 2001), so the parameter of “Number of Controlled Intersections” was also assumed to follow a normal distribution.

The number of random draws used for calculating simulation probabilities could influence model accuracies. As discussed by many researchers (Hensher and Greene, 2003; Revelt and Train, 2001; Han et al., 2001), random draws were made to calculate the simulated likelihood. A large number of random draws can take a long time and make the estimation not efficient. A small number of random draws may not be enough to estimate accurate preference. Therefore, a sufficient number of random draws of 1000 was made. Halton draw was used in the estimation (Train, 2009).

With the observed coefficients’ distributions and the number of random draws, the mixed logit models were established for each dataset and associated individual route choice preferences were generated as well. With estimated individual’s utility function, his/her route choices in the testing dataset were predicted and compared with their actual choices.

MT-LinAdapt Model

The MT-LinAdapt model described in the Methodology section was coded with Java. In the training process, four parameters η_1, η_2, η_3 and η_4 in Equation (5) need to be determined with cross validation. The ranges of these parameters are the same, namely from 0.1 to 1 with the step of 0.1. To reduce the efforts of cross validation process, four parameters were divided into two groups and parameters in the same group were adjusted together. The combination with the best performance was used for final model building. With the parameters selected, the MT-LinAdapt models were established based on training data and tested on each individual user’s testing data.

RESULTS ANALYSIS

Two possible challenging scenarios of route guidance applications were set up in the Comparison Scenario section. The performances of MT-LinAdapt and the selected three commonly used existing models are discussed in this section.

Scenario 1: Users Have Adequate Historical Data

In reality, when drivers have been using route guidance systems for certain length of time, each of them could have adequate historical data accumulated, but sociodemographic information which is used for differentiating drivers’ preferences is difficult to obtain. In this situation, the performance of MT-

LinAdapt was compared with the performance of three selected existing models, including aggregate models, individual models and mixed logit models. Prediction accuracies of MT-LinAdapt and other models on testing data were compared using paired T test. Table 1 summarizes all models' mean prediction accuracies as well as hypothesis test results between MT-LinAdapt and each selected existing model.

Table 1 Models' Prediction Accuracies and Paired T Test Results between MT-LinAdapt and Other Models

Dataset	MT-LinAdapt	Aggregate Model	Individual Model	Mixed Logit Model
Dataset 1	78.8%	70.9% ***	76.0% ***	78.1%
Dataset 2	90.6%	82.0% ***	89.5% ***	90.3% .
Dataset 3	87.7%	83.8% ***	86.6% ***	87.1% **
Synthetic Dataset 1	81.3%	65.2% ***	80.1% ***	79.3% **
Synthetic Dataset 2	90.3%	73.7% ***	88.8% ***	85.1% **
Synthetic Dataset 3	83.3%	64.2% ***	83.2%	80.8% *

Note: significance level code, 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The model performance of MT-LinAdapt models was first compared with aggregate models, as shown in the second and third columns of Table 1. MT-LinAdapt has significantly better performance than aggregate models in all datasets. The prediction accuracy of MT-LinAdapt is 4% to 19% higher than aggregate models. MT-LinAdapt models' advantage of higher prediction accuracy is greater in synthetic datasets than surveyed datasets. That is because synthetic datasets contain more heterogeneous preferences. When preferences become more heterogeneous, drivers share little common aspect of route choice preference. It becomes more difficult for aggregate models to find drivers' preference commonality, therefore, aggregate models fail to fit anyone's preference very well. On the contrary, MT-LinAdapt adapts aggregate preference to the individual level, thus shows even higher prediction accuracies than aggregate models when users have more heterogeneous preferences.

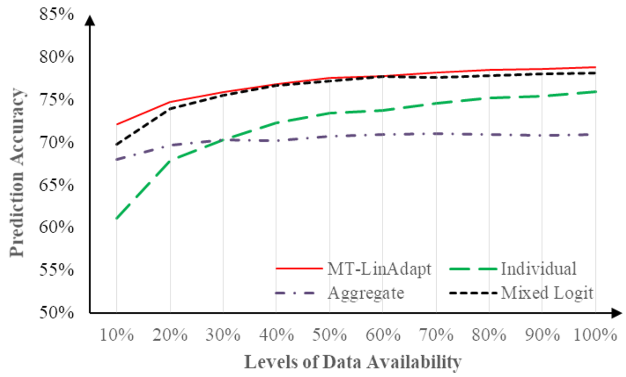
As shown in the second and forth columns of Table 1, when compared to individual route choice models, MT-LinAdapt has significantly better performance than individual models in most of the datasets except for Synthetic Dataset 3. Actually, the advantage of MT-LinAdapt is not that big in all datasets. MT-LinAdapt mostly has only 1% higher prediction accuracy than individual models. In Synthetic Dataset 3, MT-LinAdapt basically has similar performance to the individual route choice model. Because both models consider drivers' route choice preferences at the individual level, when a driver has enough historical data reflecting his/her preference, both of two models can capture the preference very well. At this point, individual route choice models also work better than aggregate models. In practice, making personalized route recommendation service can make users more likely to be satisfied than making route recommendations based on a simple unified rule.

MT-LinAdapt has significantly better performance than mixed logit models in most cases except for Dataset 1, as shown in the second and fifth columns of Table 1. The advantages are minor in surveyed datasets (less than 1% higher than mixed logit models) and statistically significant in synthetic datasets (2% to 5% higher than mixed logit models). Mixed logit models first estimate parameters' distributions at the aggregate level then generate the individual level with Bayes rule. The estimated distribution has impacts on the quality of individual user's model. When users' preferences are as heterogeneous as synthetic datasets, it becomes more difficult to accurately estimate user's preference into a pre-defined distribution type. The parameters (i.e. means and standard deviations) of distribution have higher standard errors in synthetic datasets than in surveyed datasets in the estimation. Thus, the advantages of MT-LinAdapt are greater in synthetic datasets than in surveyed datasets.

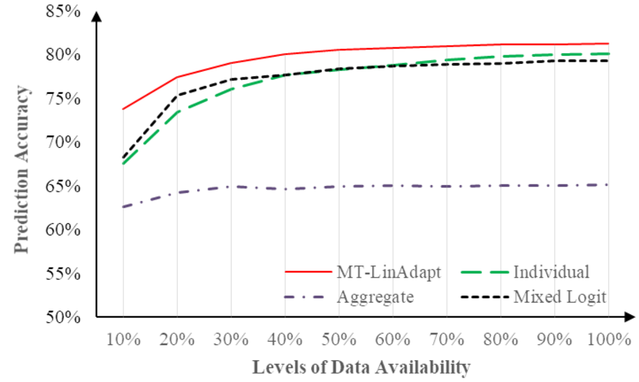
Scenario 2: Users Have Limited Historical Data

In applications of route guidance systems, there is a situation where a user does not have adequate historical data (e.g., a user or tourist just starts using the route guidance system). To test the performance of MT-LinAdapt in this scenario, this section tested models' prediction accuracies at different levels of data availability. The route choice models built at various data availabilities were tested on each individual's testing data. Thus, for each of six datasets, each model has a set of prediction accuracies at different levels of data availability. To avoid the influences of data divisions on models' performance, the described process was conducted fifty times and the average prediction accuracy on each participant was calculated. Then, the average prediction accuracy across users was used for model comparison, as shown in Figure 1.

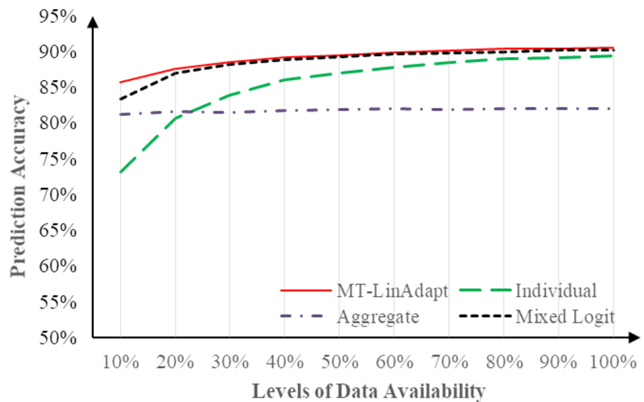
In Figure 1, the horizontal axis gives levels of data availability and the vertical axis shows the average prediction accuracy across users/synthetic users. Some observations can be obtained from the plots.



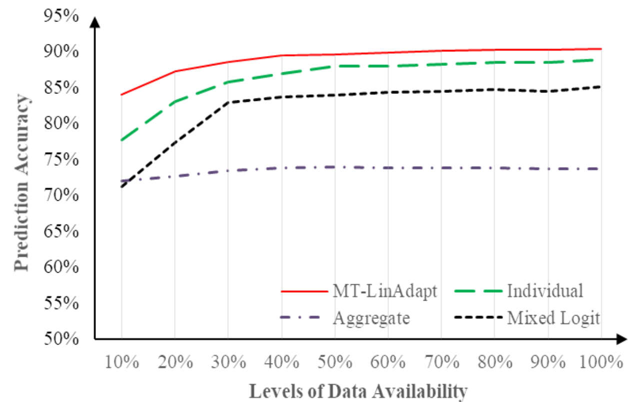
(a) Dataset 1



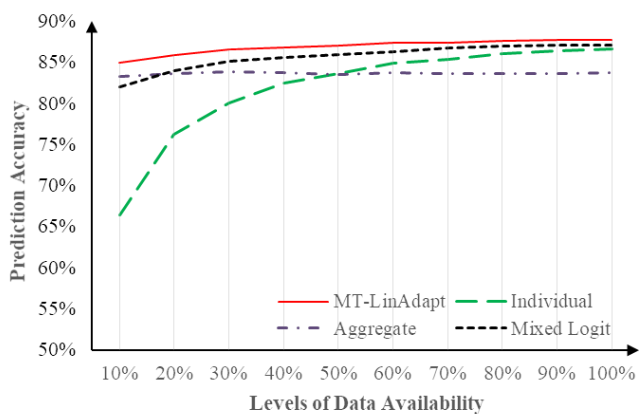
(b) Synthetic Dataset 1



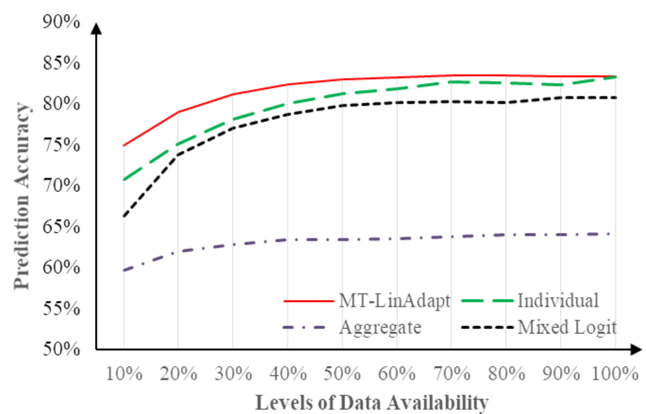
(c) Dataset 2



(d) Synthetic Dataset 2



(e) Dataset 3



(f) Synthetic Dataset 3

Figure 1 Models Performances at Different Levels of Data Availability

MT-LinAdapt has the highest prediction accuracies at all levels of data availability in all datasets. Its performance advantage is especially greater at low levels of data availability. Table 2 shows how much better MT-LinAdapt's performance is at low data availabilities by calculating the prediction accuracy difference between MT-LinAdapt and other models. In Table 2, the numbers are the specific differences of prediction accuracies between MT-LinAdapt has and other models. The relative magnitude of better performance is indicated with shaded areas for better visualization. Take data availability of 10% (5 to 6 observations) as an example, the prediction accuracy of MT-LinAdapt can be 1.6% to 15.3% higher than aggregate models, 4.2% to 18.5% higher than individual models, and 2.3% to 12.7% higher than mixed logit models. MT-LinAdapt's advantage is especially greater at low data availabilities is due to its nature that drivers' preferences are adapted at both aggregate and individual levels. For a particular user, when testing data contains new scenarios that the user never experienced before, MT-LinAdapt can predict the user's preference by combining the common preference (in other words, social norm) which was formed with other users who experienced this scenario before. Therefore, when a user has only limited data either because he/she is a new user or the system just starts operation, MT-LinAdapt can still perform well.

Table 2 The Prediction Accuracy Differences between MT-LinAdapt and Selected Existing Models at Low Data Availability Levels

Existing Models	Dataset	Prediction Accuracy Difference at Low Data Availibity Levels		
		10%	20%	30%
Aggregate model	Dataset 1	4.19%	5.08%	5.58%
	Dataset 2	4.55%	6.08%	6.98%
	Dataset 3	1.60%	2.29%	2.70%
	Synthetic Dataset 1	11.17%	13.20%	14.15%
	Synthetic Dataset 2	11.92%	14.61%	15.21%
	Synthetic Dataset 3	15.27%	17.06%	18.37%
Individual model	Dataset 1	11.07%	6.82%	5.56%
	Dataset 2	12.50%	6.99%	4.55%
	Dataset 3	18.46%	9.70%	6.48%
	Synthetic Dataset 1	6.26%	4.03%	3.06%
	Synthetic Dataset 2	6.27%	4.20%	2.83%
	Synthetic Dataset 3	4.20%	4.01%	3.00%
Mixed Logit model	Dataset 1	2.34%	0.83%	0.33%
	Dataset 2	2.39%	0.66%	0.35%
	Dataset 3	2.97%	1.92%	1.38%
	Synthetic Dataset 1	5.59%	2.03%	1.97%
	Synthetic Dataset 2	12.67%	9.96%	5.71%
	Synthetic Dataset 3	8.70%	5.31%	4.12%

Individual models do not perform well at low data availabilities, but their prediction accuracies gradually increase as more data is available. When all training data was used for model establishment, individual models have similar performance as MT-LinAdapt, as shown in Table 1. Individual route choice models only use a driver's own data for building models. The advantage of this nature is that obtained models can avoid the impacts of other users' preference so that the model can precisely capture this particular user's preference. However, the disadvantage of this characteristic is requiring an adequate amount of data. As each user may only experience limited scenarios in terms of amount as well as the variety, the individual model established based on his/her own historical data may not work well in some new scenarios. Thus, it is easy to understand that individual route choice models have relatively low prediction accuracies at low data availabilities (such as 10%, 20% and 30%), but has better performance when more data is available.

Aggregate models' performance does not improve much as more data is available. As shown in Figures 1, the increases of prediction accuracy mostly are within 2% for aggregate models, while the increase of individual models' performance can reach as high as 20% when gradually increasing data availability. That is because aggregate models can cover most of the possible scenarios at low data availabilities (such as 10%, 20% and 30%) by combining all users' experienced scenarios together. Therefore, when a particular user faces an entirely new scenario, aggregate models can still work well. However, aggregate models were built under the impacts of all users. This nature makes aggregate models shifted by groups of users as a whole but may not capture an individual user's preference precisely. Therefore, as more data is available, aggregate models' prediction accuracies do not increase much.

The general trend of mixed logit models' performance can be discussed by datasets. In surveyed datasets, the performance of mixed logit models is similar to that of MT-LinAdapt, better than both individual models and aggregate models. In synthetic datasets, the performance of mixed logit models is generally worse than both MT-LinAdapt and individual models, and better than aggregate models. When compared to aggregate models, mixed logit models can further adapt to individual tastes on the basis of preference distribution at the aggregate level. Thus, mixed logit models can have higher prediction accuracies than aggregate models in most cases and at least similar performance at low levels of data availability (such as 10%). When mixed logit models are compared to individual models, individual preference obtained with mixed logit models were generated from estimated parameter distributions. When it is difficult to accurately estimate coefficients' distributions with limited data (5 to 6 trip observations) or with very heterogeneous preference data in which preference data cannot fit into pre-defined distribution types, mixed logit models may not work as well as individual models, as individual models can specifically capture a particular user's preference. At last, the concepts of mixed logit models and the MT-LinAdapt model are very similar. Both of them adapt the aggregate preference into each individual's taste, but mixed logit models generate individual tastes based on estimated coefficient distributions while MT-LinAdapt considers preference at the aggregate and individual levels together. In addition, it is difficult for mixed logit models to establish a meaningful model out of data when everyone has only a limited amount of data. Among 50 times of data divisions at 10%, 20% and 30% of data availabilities, there are many cases in which all parameters are not significant. When it is difficult to fit users' preference into the pre-assumed type of parameter distributions because of low data availability or large preference heterogeneity, mixed logit models' capability of generating precise individual tastes is undermined. Therefore, MT-LinAdapt has 2.3% to 12.8% higher prediction accuracies than mixed logit models at data availability of 10%. The general trend of mixed logit models is much lower than that of MT-LinAdapt model in synthetic datasets.

The comparisons between the MT-LinAdapt model and three selected existing models showed that MT-LinAdapt has significantly better performance than aggregate models, individual models and mixed logit models at different levels of data availability. With datasets collected in this research, when a user has adequate historical data, the prediction accuracy of MT-LinAdapt can be 4% to 8% higher than that of aggregate models, 1% to 3% higher than that of individual models, and 1% to 5% higher than that of mixed logit models. When a user only has limited amount of data (5 to 6 observations), the prediction accuracy of MT-LinAdapt can be 1.6% to 4.5% higher than that of aggregate models, 11% to 18% higher than that of individual models, and 2.3% to 2.9% higher than that of mixed logit models. When users' preferences are more heterogeneous than the collected datasets, for example, as heterogeneous as synthetic datasets, the benefits in terms of more accurate predictions can be even larger. Given the observed preference collected in this paper were mostly from college students, the real-world preferences are expected to be more heterogeneous with general driver population and different trip purposes, attributes, etc. Therefore, in the application of route guidance systems, MT-LinAdapt is expected to predict users' route choice decisions more accurately than the selected commonly used existing models.

DISCUSSIONS

Drivers not only have some shared common aspects of route choice preferences but also have their own tastes. The MT-LinAdapt model is able to capture common preferences at the aggregate level and adapt to each individual user's specific preference. With the purpose of considering each individual driver's preference in route guidance systems, the MT-LinAdapt model can overcome the difficulties that route guidance systems are facing. For example, MT-LinAdapt does not require sociodemographic information or other exogenous criteria for distinguishing drivers' preferences. Also, MT-LinAdapt does not require a large amount of data for accurately capturing a user's preference, can be easily solved with a gradient based optimizer and can be updated efficiently in real time as data accumulates.

MT-LinAdapt was compared against aggregated SVM models, individual SVM models and mixed logit models with three stated preference datasets and three synthetic datasets. MT-LinAdapt demonstrates up to 8% and 18% higher prediction accuracies than commonly used existing models in two representative route guidance scenarios: (1) users have adequate historical preference data (as shown in Table 1) and (2) users have a limited amount of historical preference data (as shown in Figure 1). Therefore, MT-LinAdapt can work well for long-term users who have adequate historical data and also a new user who has a limited amount of data (for example, 5 to 6 trips in this paper).

Among selected commonly used existing route choice models, aggregate models work well when each user have very limited data and all users share similar route choice preferences, but aggregate models' performances are compromised when users have very heterogeneous preferences. Individual models work well when every user has adequate data and users' preferences are very heterogeneous, but they do not work well in the situation that users have a limited amount of data. Mixed logit models have similar performances with MT-LinAdapt in some of the comparisons except the cases with a limited amount of data and very heterogeneous user preferences. Furthermore, mixed logit models do not have closed form solutions and are usually solved with simulation method. That means the estimation accuracy could be influenced by how the simulation is conducted and the model can be inefficient to implement the model in practice.

In addition to the two scenarios of route guidance tested in this paper, MT-LinAdapt also has another advantage which would work well with route guidance applications. MT-LinAdapt has the potential to allow users to select different route attributes for their own interests. It is likely that drivers select different attributes they care for in the real world and this leads to the observations in the format of sparse matrices. By grouping route attributes, MT-LinAdapt can be adjusted to update preferences at both the aggregate level and the individual level for all attributes simultaneously (Gong et al., 2016). At last, the whole route guidance system with the MT-LinAdapt model can also be used as Automated Vehicles' (AVs) routing systems, as AV's passengers do not need to operate vehicles and it is important for AVs to understand its passengers' preferences.

CONCLUSIONS

In order to consider each individual driver's route choice preference when making route recommendations in route guidance systems, we introduced the Multi-Task Linear Classification Model Adaptation (MT-LinAdapt) and demonstrated its capability of accommodating drivers' heterogeneous preferences. The MT-LinAdapt model captures the common aspects of drivers' route choice preferences at the aggregate level and adapts to individual's specific preference simultaneously. The model does not require personal sociodemographic information (e.g., age, gender, income, etc.) or other criteria to differentiate drivers' different route choice preferences. In addition, it can be easily solved with a gradient-based optimizer. With three surveyed datasets and three synthetic datasets, the evaluation of the MT-LinAdapt model shows that it outperforms representative existing models by 0.1% to 19% in different datasets with various data availabilities. A detailed investigation of the evaluation results reveals the following key findings:

- When each user has a limited amount of historical preference data (5 to 6 trips), the prediction accuracy of MT-LinAdapt is 1.6% to 15.7% higher than aggregate models, 4.2% to 12.5% higher than individual models, and 2.3% to 12.7% higher than mixed logit models. It demonstrates that MT-LinAdapt can better learn drivers' preference when they have a limited amount of data than commonly used existing models;
- When each user has adequate historical preference data (50 to 64 trips), the prediction accuracy of MT-LinAdapt is 3.9% to 19.2% higher than aggregate models, 0.1% to 2.8% higher than individual models, and 0.3% to 5.2% higher than mixed logit models. It means MT-LinAdapt also has advantages over commonly used existing models when drivers have an adequate amount of data.

- The advantages of MT-LinAdapt compared to three commonly used existing models are much greater in datasets with more heterogeneous preferences.
- MT-LinAdapt has the capability of updating models in real time as behavior data accumulates and the flexibility of allowing each driver to select his/her own route attributes based on his/her interests.

These advantages of MT-LinAdapt could help route guidance system consider each individual driver's specific preference when making route recommendations in practice, and consequently improve users' satisfaction, increase users' compliances with the guidance system and potentially achieve better road network performance. The higher compliance rate is expected to benefit road network performance especially when routes were recommended for achieving a system performance goal. A better compliance rate also means knowing which route drivers are more likely to take in advance. That can also provide a perspective of utilizing disaggregate route choice decisions to calibrate network assignment models and help predict and estimate future traffic conditions (Ben-Akiva et al., 2007, 2015). These network level impacts of considering individual route choice preferences in route guidance systems would be evaluated in a future research.

REFERENCES

Adler, J.L., Blue, V.J., 2002. A cooperative multi-agent transportation management and route guidance system. Transportation Research Part C: Emerging Technologies 10, 433–454. https://doi.org/10.1016/S0968-090X(02)00030-X

Amirgholy, M., Golshani, N., Schneider, C., Gonzales, E.J., Gao, H.O., 2017. An advanced traveler navigation system adapted to route choice preferences of the individual users. International Journal of Transportation Science and Technology, Special Issue on Urban Spatiotemporal Behavior and Network Assignment 6, 240–254. https://doi.org/10.1016/j.ijtst.2017.10.001

Azevedo, C.L., Seshadri, R., Gao, S., Atasoy, B., Akkinepally, A.P., Christofa, E., Zhao, F., Trancik, J., Ben-Akiva, M., 2018. Tripod: Sustainable Travel Incentives with Prediction, Optimization, and Personalization. Presented at the Transportation Research Board 97th Annual MeetingTransportation Research Board.

Bazzan, A.L.C., Klügl, F., 2005. Case studies on the Braess Paradox: Simulating route recommendation and learning in abstract and microscopic models. Transportation Research Part C: Emerging Technologies, Agents in Traffic and Transportation: Exploring Autonomy in Logistics, Management, Simulation, and Cooperative Driving 13, 299–319. https://doi.org/10.1016/j.trc.2005.07.003

Ben-Akiva, M., Bergman, M.J., Daly, A.J., Ramaswamy, R., 1984. Modeling inter-urban route choice behaviour, in: Proceedings of the 9th International Symposium on Transportation and Traffic Theory. VNU Science Press Utrecht, The Netherlands, pp. 299–330.

Ben-Akiva, M., Bottom, J., Gao, S., Koutsopoulos, H.N., Wen, Y., 2007. Towards Disaggregate Dynamic Travel Forecasting Models. Tsinghua Science & Technology 12, 115–130. https://doi.org/10.1016/S1007-0214(07)70019-6

Ben-Akiva, M., Gao, S., Lu, L., Wen, Y., 2015. DTA2012 Symposium: Combining Disaggregate Route Choice Estimation with Aggregate Calibration of a Dynamic Traffic Assignment Model. Netw Spat Econ 15, 559–581. https://doi.org/10.1007/s11067-014-9232-z

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. 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. https://doi.org/10.1016/j.trc.2012.07.001

Ben-Elia, E., Shiftan, Y., 2010. Which road do I take? A learning-based model of route-choice behavior with real-time information. Transportation Research Part A: Policy and Practice 44, 249–264. https://doi.org/10.1016/j.tra.2010.01.007

Ben-Hur, A., Weston, J., 2010. A User's Guide to Support Vector Machines, in: Data Mining Techniques for the Life Sciences, Methods in Molecular Biology. Humana Press, pp. 223–239. https://doi.org/10.1007/978-1-60327-241-4_13

Bifulco, G.N., Simonelli, F., Pace, R. di, 2007. Endogenous Driver Compliance and Network Performances under ATIS, in: 2007 IEEE Intelligent Transportation Systems Conference. Presented at the 2007 IEEE Intelligent Transportation Systems Conference, pp. 1028–1033. https://doi.org/10.1109/ITSC.2007.4357722

Derevitskiy, I., Voloshin, D., Mednikov, L., Karbovskii, V., 2016. Traffic Estimation on Full Graph of Transport Network Using GPS Data of Bus Movements. Procedia Computer Science, 5th International Young Scientist Conference on Computational Science, YSC 2016, 26-28 October 2016, Krakow, Poland 101, 207–216. https://doi.org/10.1016/j.procs.2016.11.025

Domenico, M.D., Lima, A., González, M.C., Arenas, A., 2015. Personalized routing for multitudes in smart cities. EPJ Data Science 4, 1. https://doi.org/10.1140/epjds/s13688-015-0038-0

Feng, T., Arentze, T., Timmermans, H.J.P., August 11, 2. Capturing preference heterogeneity of truck drivers' route choice behavior with context effects using a latent class model. European Journal of Transport and Infrastructure Research 13.

Gong, L., Al Boni, M., Wang, H., 2016. Modeling social norms evolution for personalized sentiment classification, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. pp. 855–865.

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological 37, 681–698. https://doi.org/10.1016/S0191-2615(02)00046-2

Han, B., Algers, S., Engelson, L., 2001. Accommodating drivers' taste variation and repeated choice correlation in route choice modeling by using the mixed logit model, in: 80th Annual Meeting of the Transportation Research Board.

Hensher, D.A., Greene, W.H., 2003. The Mixed Logit model: The state of practice. Transportation 30, 133–176. https://doi.org/10.1023/A:1022558715350

Herring, R.J., 2010. Real-Time Traffic Modeling and Estimation with Streaming Probe Data using Machine Learning. eScholarship.

How does the navigation system choose a route? [WWW Document], n.d. . Tesla Motors Club. URL https://teslamotorsclub.com/tmc/threads/how-does-the-navigation-system-choose-a-route.67394/ (accessed 11.19.17).

Jan, O., Horowitz, A., Peng, Z.-R., 2000. Using Global Positioning System Data to Understand Variations in Path Choice. Transportation Research Record: Journal of the Transportation Research Board 1725, 37–44. https://doi.org/10.3141/1725-06

Klein, I., Levy, N., Ben-Elia, E., 2018. An agent-based model of the emergence of cooperation and a fair and stable system optimum using ATIS on a simple road network. Transportation Research Part C: Emerging Technologies 86, 183–201. https://doi.org/10.1016/j.trc.2017.11.007

Lee, E.W.M., Li, M.C.W., 2016. Intelligent Route Choice Model for Passengers' Movement in Subway Stations, in: International Symposium on Neural Networks. Springer, pp. 385–392.

Li, D., Miwa, T., Morikawa, T., Liu, P., 2016. Incorporating observed and unobserved heterogeneity in route choice analysis with sampled choice sets. Transportation Research Part C: Emerging Technologies 67, 31–46. https://doi.org/10.1016/j.trc.2016.02.002

Lima, A., Stanojevic, R., Papagiannaki, D., Rodriguez, P., González, M.C., 2016. Understanding individual routing behaviour. Journal of The Royal Society Interface 13, 20160021. https://doi.org/10.1098/rsif.2016.0021

Liu, H.X., He, X., Recker, W., 2007. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. Transportation Research Part B: Methodological 41, 448–461. https://doi.org/10.1016/j.trb.2006.07.002

Liu, H.X., Recker, W., Chen, A., 2004. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. Transportation Research Part A: Policy and Practice 38, 435–453. https://doi.org/10.1016/j.tra.2004.03.003

Liu, L., Xu, J., Liao, S.S., Chen, H., 2014. A real-time personalized route recommendation system for self-drive tourists based on vehicle to vehicle communication. Expert Systems with Applications 41, 3409–3417. https://doi.org/10.1016/j.eswa.2013.11.035

Ma, J., Smith, B.L., Zhou, X., 2016. Personalized real-time traffic information provision: Agent-based optimization model and solution framework. Transportation Research Part C: Emerging Technologies 64, 164–182. https://doi.org/10.1016/j.trc.2015.03.004

Mahmassani, H.S., Koppelman, F., Frei, C., Frei, A., Haas, R., 2013. Synthesis of Traveler Choice Research: Improving Modeling Accuracy for Better Transportation Decisionmaking (No. No. FHWA-HRT-13-022).

Matthews, M.H., 1981. Children’s Perception of Urban Distance. Area 13, 333–343.

Nadi, S., Delavar, M.R., 2011. Multi-criteria, personalized route planning using quantifier-guided ordered weighted averaging operators. International Journal of Applied Earth Observation and Geoinformation 13, 322–335. https://doi.org/10.1016/j.jag.2011.01.003

Pahlavani, P., Delavar, M.R., 2014. Multi-criteria route planning based on a driver’s preferences in multi-criteria route selection. Transportation Research Part C: Emerging Technologies 40, 14–35. https://doi.org/10.1016/j.trc.2014.01.001

Pahlavani, P., Delavar, M.R., Frank, A.U., 2012. Using a modified invasive weed optimization algorithm for a personalized urban multi-criteria path optimization problem. International Journal of Applied Earth Observation and Geoinformation 18, 313–328. https://doi.org/10.1016/j.jag.2012.03.004

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, 159–166. https://doi.org/10.1049/iet-its:20060074

Parthasarathi, P., Levinson, D., Hochmair, H., 2013. Network Structure and Travel Time Perception. PLOS ONE 8, e77718. https://doi.org/10.1371/journal.pone.0077718

Paz, A., Peeta, S., 2009. Behavior-consistent real-time traffic routing under information provision. Transportation Research Part C: Emerging Technologies 17, 642–661. https://doi.org/10.1016/j.trc.2009.05.006

Peeta, S., Yu, J.W., 2005. A Hybrid Model for Driver Route Choice Incorporating En-Route Attributes and Real-Time Information Effects. Netw Spat Econ 5, 21–40. https://doi.org/10.1007/s11067-005-6660-9

Peeta, S., Yu, J.W., 2004. Adaptability of a hybrid route choice model to incorporating driver behavior dynamics under information provision. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 34, 243–256. https://doi.org/10.1109/TSMCA.2003.822272

Perk, V.A., DeSalvo, J.S., Rodrigues, T.A., Versoza, N.M., Bovino, S.C., 2011. Improving Value of Travel Time Savings Estimation for More Effective Transportation Project Evaluation (No. BDK85 911-21). Florida Department of Transportation, Tampa, Florida.

Peruch, P., Giraudo, M.-D., Garling, T., 1989. Distance cognition by taxi drivers and the general public. Journal of Environmental Psychology 9, 233–239. https://doi.org/10.1016/S0272-4944(89)80037-4

Prato, C.G., 2009. Route choice modeling: past, present and future research directions. Journal of Choice Modelling 2, 65–100. https://doi.org/10.1016/S1755-5345(13)70005-8

Razo, M., Gao, S., 2013. A rank-dependent expected utility model for strategic route choice with stated preference data. Transportation Research Part C: Emerging Technologies, Selected papers from the Seventh Triennial Symposium on Transportation Analysis (TRISTAN VII) 27, 117–130. https://doi.org/10.1016/j.trc.2011.08.009

Revelt, D., Train, K., 2001. Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier (No. 0012001), Econometrics. EconWPA.

Revelt, D., Train, K., 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. The Review of Economics and Statistics 80, 647–657. https://doi.org/10.1162/003465398557735

Rogers, S., Fiechter, C., Langley, P., 1999. A Route Advice Agent that Models Driver Preferences. ResearchGate.

Rogers, S., Langley, P., 1998. Personalized Driving Route Recommendations. ResearchGate.

Sarrias, M., Daziano, R., 2017. Multinomial Logit Models with Continuous and Discrete Individual Heterogeneity in R: The gmnl Package | Sarrias | Journal of Statistical Software 79. https://doi.org/10.18637/jss.v079.i02

Steinwart, I., Christmann, A., 2008. Support Vector Machines. Springer Science & Business Media.

Sun, B., Park, B.B., 2017. Route choice modeling with Support Vector Machine. Transportation Research Procedia, World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016 25, 1811–1819. https://doi.org/10.1016/j.trpro.2017.05.151

Tawfik, A., Rakha, H., 2013. Latent Class Choice Model of Heterogeneous Drivers' Route Choice Behavior Based on Learning in a Real-World Experiment. Transportation Research Record: Journal of the Transportation Research Board 2334, 84–94. https://doi.org/10.3141/2334-09

Tawfik, A.M., Rakha, H.A., Miller, S.D., 2010. Driver route choice behavior: Experiences, perceptions, and choices, in: Intelligent Vehicles Symposium (IV), 2010 IEEE. IEEE, pp. 1195–1200.

Tian, H., Gao, S., Fisher, D.L., Post, B., 2012. Mixed-Logit Latent-Class Model of Strategic Route Choice Behavior with Real-Time Information. Presented at the Transportation Research Board Annual Meeting, Washington D.C, pp. 12–2867.

Train, K.E., 2009. Discrete Choice Methods with Simulation, 2 edition. ed. Cambridge University Press, Cambridge ; New York.

Train, K.E. (University of C., 1998. Recreation demand models with taste differences over people. Land economics (USA).

Wang, J., Lv, J., Wang, C., Zhang, Z., 2017. Dynamic Route Choice Prediction Model Based on Connected Vehicle Guidance Characteristics [WWW Document]. Journal of Advanced Transportation. https://doi.org/10.1155/2017/6905431

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, 59–66. https://doi.org/10.3141/1807-08

Yang, H., Kitamura, R., Jovanis, P.P., Vaughn, K.M., Abdel-Aty, M.A., 1993. Exploration of route choice behavior with advanced traveler information using neural network concepts. Transportation 20, 199–223. https://doi.org/10.1007/BF01307059

Zhang, Y., Xie, Y., 2008. Travel Mode Choice Modeling with Support Vector Machines. Transportation Research Record: Journal of the Transportation Research Board 2076, 141–150. https://doi.org/10.3141/2076-16

CHAPTER 4. A Proactive User Optimum-Oriented Route Guidance System Incorporating Individual Users' Route Choice Preferences

Bingrong Sun, B. Brian Park

Department of Civil and Environmental Engineering, University of Virginia, 351 McCormick Rd
Charlottesville, Virginia, United States, 22904

ABSTRACT

The route guidance system has been an effective way of mitigating traffic congestion. Existing route guidance systems in practice and literature tend to simplify drivers' heterogeneous route choice preferences when designing route guidance strategies. This could undermine the performance of the designed route guidance strategies and also deteriorate users' satisfaction. The emerging information technologies provide an opportunity of analyzing drivers' route choice preferences heterogeneity by collecting drivers' preference data at the individual level. Therefore, this paper proposes a proactive user optimum-oriented route guidance system which (1) establishes individual route choice models with each user's historical preference data and (2) incorporates individuals' route choice preferences in searching for user optimum conditions. Such user optimum conditions are provided to users as guidance information. An evaluation platform which contains a Traffic Simulation module (DTALite) and Decision Module (Matlab) was set up for evaluation. With the Sioux Falls network and user population whose preferences were synthesized from surveyed participants, the proposed route guidance system at both perfect and imperfect market penetration rates was compared to existing route guidance strategies including travel time based real time guidance and User Equilibrium conditions based guidance. The proposed route guidance system demonstrated advantageous performance in aspects of users' satisfaction (e.g., up to 22% more satisfied users), system mobility and sustainability (e.g., up to 10% of travel time reduction and up to 42% of delay reduction), and future traffic conditions estimation (e.g., up to 70% links having more accurate volume estimation). At imperfect market penetration rates, the performance improvement gradually increases as the market penetration rate increases.

Keywords: Route Guidance System; Proactive Guidance; User Optimum; Individual' Route Choice Preference

I. INTRODUCTION

Route guidance systems have been an important part of Advanced Traveler Information System (ATIS). Drivers can use route guidance systems to find the route leading them to their destinations or check real time traffic conditions so that congested areas can be avoided. With the information provided by route guidance systems, drivers can make informed route choice decisions thus transportation network's efficiency can be improved. Route guidance systems in practice such as Google Maps and Waze usually find the route with shortest distance or travel time and recommend them to drivers. In addition to these, researchers proposed different route guidance strategies to further utilize route guidance systems in order to improve transportation system efficiency (Angelelli et al. 2016; Paz and Peeta 2009a; Liang and Wakahara 2014; Du, Han, and Chen 2015). Most of routing strategies in the literature can be characterized as either reactive or proactive, centralized or decentralized, and simple criterion based or multiple criteria (behavior consistent) based.

A. Reactive versus Proactive Route Guidance

Route guidance systems generate route recommendations usually based on certain traffic conditions. Depending on the type of traffic conditions used to generate route recommendations, route guidance systems can be divided into reactive route guidance and proactive route guidance.

The reactive route guidance systems generate route recommendations based on historical or real-time traffic conditions. Such historical and real-time traffic conditions could be link travel time that is updated every certain periods such as 3 seconds in (Deflorio 2003), 30 seconds in (Khaled, Ardeshir, and Raman 2003), the combination of historical travel time and the updated travel time information (Khaled, Ardeshir, and Raman 2003), inflow and outflow on each link in previous 300 seconds (He, Guan, and Ma 2013), or the histogram of number of vehicles on each lane in previous 15 seconds (El-Sayed, Thandavarayan, and Hawas 2017). These traffic condition indictors can be obtained by loop detectors, roadside sensors, probe vehicles and other traffic monitoring methods. The route guidance systems react to the general traffic conditions in the past or what is happening in the network. In other words, the reactive route guidance system only responses to the traffic conditions that already happened. Reactive route guidance systems using real-time traffic conditions are probably very close to the current commonly used route guidance services. It can quickly react to the traffic conditions so that congested areas can be avoided when users request for route guidance services. Though it can quickly react to the changing traffic conditions, it does not consider how the recommended routes are going to affect the future traffic conditions. One extreme case is when a large amount of route guidance requests get the recommended routes that share certain portion of routes. The shared road links would become heavily congested and the traffic conditions when trips are actually made will be different from what was shown in the route guidance systems. As pointed out by Liang and Wakahara (2014), reactive route guidance system is more like an alert system which warns drivers about the congestion already happened instead of guiding drivers to proactively prevent the congestion from happening.

Proactive route guidance systems generate route recommendations based on predicted traffic conditions. Some literatures also refers them as iterative, predictive or anticipatory route guidance systems (He, Guan, and Ma 2013; Dong, Mahmassani, and Lu 2006; Liang and Wakahara 2014). This type of route guidance system can be further divided into two categories: user optimum oriented and system optimum oriented. Both of them aim at improving the performance of route guidance system but target on different objectives.

The User Optimum oriented route guidance system focuses on drivers' benefits and is implemented following the Wardrop's first principle. The Wardrop's first principle defines that in a stable equilibrium, all drivers cannot further reduce his/her travel cost by switching to another route (Wardrop 1952). These user optimum (also treated interchangeably with user equilibrium) conditions are obtained and then used to generate route recommendations. If all drivers comply with the recommended routes, user optimum conditions can be reached. Travel time is frequently used to represent drivers' cost because it is recognized as one of the most important explanatory variable for understanding people's travel decisions (A. Small 2012). Iterative process is often used to solve for the user optimum conditions. Researchers have demonstrated that user optimum oriented route guidance can have good system efficiency (Deflorio 2003; Giglio and Sacco 2014; He, Guan, and Ma 2013), but it is also widely recognized system optimum-oriented route guidance system has better system efficiency than user optimum oriented route guidance system (Roughgarden and Tardos 2000; Han et al. 2016). Also, existing user equilibrium based route guidance systems mostly simplify drivers' route choice preference such as assuming travel time is the only route attribute that drivers care when making route choice decisions. Drivers may not follow the route recommendations made with a simplified preference assumption by the route guidance system.

When system performance is of concern especially to transportation authorities, system optimum-oriented route guidance system is also widely discussed. At system optimum conditions, drivers make route choice decisions so that system performance indicators such as total travel time or total delay are minimized (Ferris and Ruszczyński 2000; Lafortune et al. 1993). However, it is widely recognized that routing drivers based on system optimum conditions brings unfairness to drivers (Jahn et al. 2005; Du, Han, and Chen 2015). Some users have to take longer detours than the route they preferred so that the system can achieve its optimum goal. It is unfair and also not practical because drivers are not likely to comply with the recommendations thus eventually the network cannot reach the system optimum conditions. That is why researchers have recognized that system optimum is not suitable for the application of route guidance (He, Guan, and Ma 2013). In order to address the unfairness, researchers included some user inconveniences constraints in the settings of system optimum problem. To measure the inconveniences of a route, Angelelli et al. (2016) used the ratio of the distance difference between the route and the shortest route to the distance of the shortest route. The ratio has to be less than certain threshold for drivers to accept a particular route. Han et al. (2016) formulated a bi-level problem to consider the users' preferences at the lower level and the network efficiency at the upper level. Paz and Peeta (2009b) proposed behavior consistent route guidance system which finds the overlap of system desired and users preferred routes. When including users' constraints in the searching for system optimum, the unfairness can be reduced meanwhile system efficiency can be improved. Still, similar as both system optimum and user optimum oriented route guidance systems, simplified behavior representations were adopted in these research, such as only assuming travel time is the only influencing factor (Angelelli et al. 2016; Han et al. 2016) or using an aggregate model to represent drivers' behaviors (Paz and Peeta 2009a). This could also undermine the performance of these route guidance systems that target on system optimum with users' inconveniences constraints.

B. Centralized versus Decentralized Route Guidance

Depending on the operation scheme, route guidance systems can be divided into centralized and decentralized systems. Centralized route guidance system comprises a control center and its system users. The requests for route guidance services and the route recommendation information are all communicated directly between the center and each single user (Hawas and El-Sayed 2015). The centralized route guidance system can coordinate all system users and utilize all possible information to process and generate route guidance information. That also means it requires efficient communication devices. On the other hand, decentralized route guidance systems can handle the traffic conditions and route guidance information locally, either through communication with surrounding users or the roadside facilities (Du, Han, and Chen 2015; Hawas and El-Sayed 2015). It avoids the necessity of large amount of data communication, but it may not be able to find the system wide optimum conditions (Hawas and El-Sayed 2015). In addition, as pointed by Hawas and El-Sayed (2015), the decentralized system may suffer vehicle cycling issues, namely algorithms do not result in a fixed point solution. The vehicle cycling issues can be mitigated by allowing local controllers to share information, but that also involves certain information communication and processing.

C. Single Criterion versus Multiple Criteria (Behavior Consistent)

Depending on the criteria used to generate route recommendations, route guidance systems can be divided into single criterion based guidance and multiple route attributes/behavior consistent route guidance. Multiple criteria based guidance systems are equivalent to behavior consistent route guidance, because a behavior consistent route guidance usually considers multiple criteria. For the single criterion guidance system, the most frequently used criterion is travel time as it is one of the most important explanatory variable for understanding people's travel decisions (A. Small 2012). Routes with the shortest travel time are recommended to the users. Other commonly used criteria including distance (Jahn et al. 2005) and input and output flow (He, Guan, and Ma 2013). The single criterion represents simplified drivers' route choice behaviors that are considered in route guidance systems. On the other hand, some route guidance systems include multiple route attributes that drivers may care when making route choice decisions. Wahle et al. (2001) included dynamic arc costs (i.e., travel time and traffic density) and static arc cost (i.e., road type and route length) for generating route recommendations. Khaled et al. (2003) designed a route guidance system that considers route length, capacity, free flow travel time, current travel time information, previous travel time experienced, etc. Paz and Peeta (2009b) included several if-then rules concerning different aspects of drivers' route choice decisions. Alder et al. (2005) included 9 objectives as decision evaluation criteria. Multiple route attributes are usually included in the form of linear weighted utility function (Adler et al. 2005) or fuzzy logic (Paz and Peeta 2009a; Wahle et al. 2001). Given the fact that drivers usually make route choice decisions based on more than one route attribute, it is natural to think that route guidance systems including multiple criterion is better, from the perspective of realistically representing drivers' behaviors.

Most of the route guidance systems in practice or in literature combines one or multiple features described above. Having different features can bring route guidance systems different performances in terms of improving transportation system efficiency and users' satisfaction. Shared by all different types of route guidance systems discussed above, one common issue that they are suffering is the simplified drivers' preferences in the process of generating route recommendations. Most researchers agreed that the proactive route guidance system can bring more benefits as long as the prediction is accurate and reliable (Liang and Wakahara 2014; Dong, Mahmassani, and Lu 2006). One of the factors that causes any inconsistency between predicted and actual traffic condition is the incorrect driver behavior modeling (Paz and Peeta 2009a). Regardless of user optimum or system optimum based route recommendations, if drivers' behaviors are not properly modeled and consequently drivers do not follow the recommended routes, the resulted traffic conditions could far deviated from the predicted traffic conditions which targeted on any optimum conditions. As Paz and Peeta (2009a) pointed out, "an incorrect prediction of the drivers' likely reactions to the information strategies can results in the generation of erroneous information strategies, negatively impacting network performance." Many other researchers (Schofer, Koppelman, and Charlton 1997; Giglio and Sacco 2014; Khaled, Ardeshir, and Raman 2003) also pointed out the necessity of considering drivers' behavior representation in route guidance. From users' perspective, Schofer et al. (1997) conducted a survey among 100 drivers and asked for their experiences of using route guidance systems. In the survey, participants thought that route guidance systems should learn drivers' preferences and drivers were even willing to let the computer have learning mode and teach the computer to learn their preferences. Participants also liked to have their own route choice preference and criteria that can change at different time.

Researchers have tried to include more realistic route choice models in designing route guidance systems, such as Khaled et al. (2003) used a neural network model to represent drivers' route choice preferences in route guidance, Paz and Peeta (2009a) used a fuzzy multinomial logit model which incorporates several if-then rules for representing drivers' behaviors, Wahle et al. (2001) adopted fuzzy decision model to describe drivers' route choice decisions. These route choice models are mostly developed from the aggregate data collected from a group of drivers, such as 150 samples from 15 drivers (Khaled, Ardeshir, and Raman 2003) or assumed preference distributions (Paz and Peeta 2009a). However, drivers' preferences are very heterogeneous in terms of information perception, decision rules and cared route attributes (A. M. Tawfik, Rakha, and Miller 2010; Parthasarathi, Levinson, and Hochmair 2013; Feng, Arentze, and Timmermans 2013; Liu, He, and Recker 2007; Amirgholy et al. 2017). Fortunately, with the emerging information technologies applied in transportation domain such as GPS, it is possible to collect route choice preferences even at the individual level. Establishing drivers' route choice models at the individual level can capture drivers' heterogeneous preference to the maximum extent. Paz and Peeta (2009a) also mentioned that the capability of tracking individual drivers in route guidance system can provide robust models to describe individual drivers' route choice preferences. Therefore, this research discusses the framework of incorporating individuals' route choice preferences when designing route guidance systems.

D. Research Scope

The route guidance system framework in this research adopted the scheme of the centralized, proactive user optimum-oriented with considering multiple route attributes. Since the focus of this research is to represent drivers' route choice preferences realistically in route guidance systems, the scheme of multiple route attributes allow drivers to include different route attributes in their own models. Also, a centralized scheme allows the route guidance system to consider all aspects of the system details together. Though centralized system requires large communication bandwidth, with the emerging information technologies such as 5G, it is promising that future communication devices and infrastructure can handle a city wide route inquiry at the same time. Proactive user optimum-oriented scheme is chosen instead of system optimum oriented. Though some researchers concluded that the system efficiency at the user optimum conditions is not as good as that of system optimum conditions, this research aims at increasing drivers' compliance behavior by incorporating individual route choice preferences in generating route recommendations. Guaranteeing all drivers would like to comply with the recommendations means the traffic conditions at user optimum. Therefore, if drivers' compliances are guaranteed, traffic control and management strategies can be implemented and are more likely to be effective so that system performances can be improved. Instead of implementing system optimum oriented guidance with user constraints, improving system performance on the basis of users' satisfaction is the approach this research adopts.

Thus, this paper proposes a user optimum-oriented proactive route guidance system that incorporates individual drivers' route choice preferences. In the proposed route guidance system, drivers' route choice preferences can be learned from his/her historical trips and stored in the system in the form of individual route choice models. When a driver requests route guidance service for a trip, his/her request and preference are considered with other drivers' requests in the same time interval. Then, user optimum conditions that incorporate every individual driver's particular route choice preference can be obtained and used as the guidance information. The rest of this paper discusses the proposed route guidance system in details. Section II introduces the framework and major components of the proposed route guidance system. Section III evaluates the proposed routing system with an example network and compares the proposed route guidance system with several existing route guidance strategies. At last, some conclusions and future research are described in Section IV.

II. ROUTE GUIDANCE SYSTEM FRAMEWORK

The proposed proactive user optimum-oriented route guidance system has two major components, namely how to capture drivers' route choices at the individual level and how to incorporate the individuals' route choice preferences into the process of searching for user optimum conditions. Each of the major components is discussed in this section.

A. Individual Route Choice Model

Capturing individual driver's route choice preference is very critical to the proposed route guidance system, because it determines if the recommended routes are users' preferred ones thus comply with the recommendations or not. Traditionally, probably because limited individual driver's preference data is available, researchers typically establish a model for a group of drivers and differentiate drivers' preference by incorporating segmenting variables, such as sociodemographic information (e.g., age, gender, income, etc.) and driving patterns (A. Tawfik and Rakha 2013). This approach has an assumption that drivers who belong to the same segmented group have similar route choice preference. As more information communication technologies such as GPS are applied in transportation applications, more drivers' preference data even at the individual level becomes available. Researchers have explored different methods to model individual driver's route choice preference, such as the method of differential perceptron (Rogers and Langley 1998), decision tree (Park et al. 2007), ordered weighted averaging method (Nadi and Delavar 2011) and neuro-fuzzy model (Pahlavani and Delavar 2014).

A general form of an individual driver *i*'s route choice model in the proposed route guidance system can be written in Equation (1).

$$Y = f(\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{S}) \qquad (1)$$

In which

Y: the selected route among given route set. When there are *n* alternative routes which are labeled as *1,2,…,n,* then the value of *Y* could be set to represent each route.

Table 1 Inputs of Individual Route Choice Model

Types	Variables	Brief Description	Examples
I	***X***	Routes attributes	Travel Time, Reliability, Number of Controlled Intersections, Distance, etc.
II	***T***	Trip related	Departure Time, Trip Purpose, Weather, People Who Travel with, etc.
III	***S***	Driver's states	Driver's Physical States and Emotional States

X: information of route attributes on alternative routes. In a particular trip scenario, there are several alternative routes in the route set for the driver to choose. The information of these route attributes on different routes is included in ***X*** and is shown to the drivers. Drivers see the information and make route choice decisions based on his/her perceptions, knowledge and experiences. Different drivers can include different route attributes in his/her own model. The attributes that are included in the model can be either determined by drivers based on what they care or can be learned by the route choice model from the driver's historical trips.

T: the trip related other information that can influence route choice decisions. Though these information is not directly related to alternative routes, drivers make route choice decisions under the influence of them. For example, drivers who make trips at different times of day may choose different routes. Therefore, it is necessary to include the impacts of these influencing factors in the individual route choice models. The information of these variables can be inferred by the route guidance system from the historical trips. For example, the variable of trip purpose can be inferred from the destination or the land use of the destinations. Also, when drivers connected their digital calendar with the route guidance system (for example, Waze allows users to connect Google Calendar and Facebook events), the route guidance system can also tell the trip purposes from driver's digital calendar.

S: the drivers' states including drivers' physical states and emotional states, for example, tired, angry, etc. Dia (2002) mentioned that individual route choice model should involve some emotional elements so that drivers' beliefs, motives and impulsive actions can be modeled. Some researchers (Abdić et al. 2016; Fridman et al. 2018) have studied using in-vehicle camera to analyze drivers' facial expression so that the drivers' emotional states (e.g., frustration) can be detected. Therefore, when drivers are in different emotional states, different routes can be recommended to drivers. These functions of route guidance system probably require certain levels of vehicle automation, but $\boldsymbol{S}$ is included in the individual route choice models to incorporate future extended functions.

Depending on the specific modeling approaches used for building individual route choice models, the data of decision variables can be organized in different formats. In the proposed route guidance system, machine learning methods are used to capture each individual driver's route choice preference, given they were widely used to handle large amount data efficiently (Gong, Al Boni, and Wang 2016) and have more general assumptions (Xiong et al. 2016) such as without assuming random utility maximization.

B. User Optimum Conditions Generation with Individuals' Preferences

As described in Wardrop's first principle, user optimum (also treated interchangeably with User Equilibrium) conditions are defined as the flow pattern in which no user can unilaterally reduce their travel cost (Wardrop 1952). When user equilibrium conditions exist at all time intervals, the flow pattern is called dynamic user equilibrium (DUE) (Carey and Ge 2012). The goal of the proposed route guidance system is to find the user optimum conditions that incorporate individuals' route choice preferences and provide guidance based on user optimum conditions.

In order to consider individual drivers' route choice preferences in the process of searching for user optimum conditions, agent based traffic assignment is used (Nagel and Flötteröd 2009). Agent based modeling has been used in transportation research for route guidance system study (Adler and Blue 2002; Adler et al. 2005; Arokhlo et al. 2011; Dia 2002). It has the flexibility of allowing agents have different behavior rules so that the behaviors of different transportation system stakeholders can be represented. As discussed in Nagel and Flötteröd (2009), flow can be discreted into "appropriate number of travelers for every OD pair and every time slot, and distribute them across the time slot." Therefore, assuming a time-dependent OD matrix is available, the number of agents traveling between different OD pairs in all time slots can be generated. Each of the agent has his/her route choice decision rules which represent a driver's route choice preference in the route guidance system. Then, dynamic user optimum (also treated interchangeably with dynamic user equilibrium) conditions can be solved by using agent based traffic assignment. Following iterative procedure is used to obtain user optimum conditions. It should be noted that in this research, the travel demand which is in the format of time dependent origin-destination matrix is assumed to be known. Drivers are assumed to be willing to share their preference data with the route guidance system. These assumption will be discussed in the Discussion section about its reasonableness in implementation.

Iterative Process

Assume the road network is made of a set of nodes, **N** and a sets of arcs **A**. Each node represents an intersection and each arc represents a road section in the real road network. The nodes are numbered 1,…, *N*. A route can be denoted as a sequence of nodes (i_1, i_2, i_3..., i_k). The set $\{\boldsymbol{X_a^n}\}$ represent the traffic conditions on all arcs $\in\boldsymbol{A}$ in iteration *n*. The set $\{\boldsymbol{C^n}\}$ represent users' route choice decisions in iteration *n*.

Step 0: Initialization. Perform a network loading based on historical traffic pattern $\{\boldsymbol{X_a^0}\}$. Check each agent's route choice decision based on the traffic condition at its departure time, *t*. This generates all agents' route choice decisions $\{\boldsymbol{C^0}\}$.

Step 1: Iterations:

(1) Network loading: Load agents on the network based on their departure time and selected routes. Updated traffic conditions $\{\boldsymbol{X_a^n}\}$ can be obtained.

(2) Update choices: Check agents' route choice decisions based on traffic conditions $\{\boldsymbol{X_a^n}\}$ at its departure time, *t*, using their individual route choice models. This generates all agents' route choice $\{\boldsymbol{C^n}\}$.

Step 4: Convergence criterion: Compare agents' choices $\{\boldsymbol{C^n}\}$ and $\{\boldsymbol{C^{n-1}}\}$ to see how many agents' choices are different in consecutive iterations. If the difference is less than *m*, stop. Otherwise, go to Step 1. *m* is a predefined threshold and can be defined as, for example, *m=1%*demand*.

Some details of the procedures above are discussed in the rest part of this section, including route set generation, traffic information processing, and the application of individual route choice models.

Route set generation

A route set consists of several alternative routes that drivers can select a route from. The route set can be different by OD pairs, by departure time and even by drivers. Researchers have explored different methods of generating route sets. There are mainly four types of three generation methods (C. G. Prato

2009): deterministic shortest path-based method, stochastic shortest path-based methods, and constrained enumeration methods.

Deterministic shortest path method treats the network conditions as deterministic values and find the best *K* paths based on selected criteria (e.g., travel time, distance) by repeatedly searching the network. The widely used K-shortest path algorithm, link elimination approach, link penalty approaches and labeling approach all belong to this category (C. G. Prato 2009). Instead of assuming the network conditions are deterministic, stochastic shortest path-based methods assume drivers perceive traffic conditions with errors thus drivers' perceived link conditions following certain probability distributions (C. G. Prato 2009). The path cost is calculated from the summation of link costs which are obtained by making random draws from the probability distributions. The probability distributions could be different on different links and also vary by drivers due to the fact that drivers may perceive traffic conditions differently. Constrained enumeration methods assume that driver generate the route set based on behavioral rules instead of the minimum cost. One representative algorithm in this category is the branch and bound algorithm (C. Prato and Bekhor 2006). By setting up thresholds for different decision criteria, a certain link is considered to be included or excluded to form a route when searching for routes in the network. Since this research aims at realistically representing drivers' route choice behaviors and branch-and-bound algorithm is considered to be able to reflect behavior assumptions (C. G. Prato 2009), the branch-and-bound algorithm (C. Prato and Bekhor 2006) is selected to generate route set in the traffic assignment process in this research. The details of the branch and bound algorithm is discussed below.

The branch and bound algorithm (C. Prato and Bekhor 2006) constructs a connection tree between the origin and the destination by processing a sequence of road links. While going through each link, certain bounded criteria are set up to avoid unrealistic routes as well as include as many heterogeneous alternative routes as possible.

For a given OD pair, a connection tree is constructed to connect the origin node, *o*, and the destination node, *d*. Starting from the node *o*, all the nodes that locate at the downstream of the node *o* are considered as the candidate nodes that the current tree would connect to. Only the nodes that satisfy the pre-defined criteria are included in the connection tree. To be more general, when the connection tree reaches an intermediate node, *x* and the downstream node of *x* is *y*, following criteria (C. Prato and Bekhor 2006) are considered to determine if the connection tree will reach node *y* or not.

(1) Directional constraint

$$\delta_D D^*(x,d) > D^*(y,d) \qquad (2)$$

δ_D is the adjusting factor which defines the tolerance level of the directional criterion. $D^*(a, b)$ represents the shortest distance between nodes *a* and *b* in the network. This constraint makes sure that the next node of the route should be able to bring the route towards the destination. It can deviate a little from the direction towards the destination but the deviation magnitude is limited by δ_D.

(2) Travel time constraint

$$\delta_T T^*(o,y) > T(o,y) \qquad (3)$$

δ_T is the adjusting factor which defines the tolerance level of the travel time criterion. $T^*(a,b)$ represents the shortest travel time between nodes *a* and *b* in the network. $T(a, b)$ represents the travel time between nodes *a* and *b* if following the route that is under consideration. This constraint makes sure that the travel time would not increase too much if the connection tree includes node *y*.

(3) Distance constraint

$$\delta_{DD} D^*(a,b) > D(a,b), \quad \forall\, node \subseteq \{o,a,\ldots,x,y\} \qquad (4)$$

δ_{DD} is the adjusting factor which defines the tolerance level of the distance criterion. $D(a, b)$ represents the distance between node *a* and node *b* if following the route that is under consideration. $\{o,a,\ldots,x,y\}$ is the node set that contains all the nodes in existing route and the node under consideration. This constrain requires the distance between any two nodes on this route should not be larger than δ_{DD} times of the shortest distance between them. The constraint rejects the route that have unrealistic detour.

(4) Loop constraints

$$y \notin \{o, a, \ldots, x\} \qquad (5)$$

The nodes in the bracket are all the existing nodes which are already included in the route. The node under consideration should not be the same as any of these existing nodes. This constraint avoids the route that contains loop which is not very likely to be taken by drivers in real life.

(5) Intersection constraints

$$nInter\,(o, y) < NC \qquad (6)$$

nInter(o, y) represents the number of controlled intersections if following the route that is under consideration. If it excesses certain predefined amount, *NC*, the route is rejected.

When a node satisfying all five constraints described above, the node can be included in the connection tree and its downstream nodes will be evaluated in the next step until the destination is reached. The tolerant factors in Equation (2) to (6) should be calibrated with the observed data.

After the initial route sets are generated, similarity check needs to be conducted to avoid the situations that routes with too much overlaps are included in the route set at the same time. For the initial route set of an OD pair, a route was randomly selected to be the first route included in the final route set. Then, all other routes in the initial route set was compared to every route in the final route set based on equation (7). The overlapping ratio was calculated as (C. Prato and Bekhor 2006):

$$\frac{LS_{i,j}}{L_j} < r \qquad (7)$$

Where: $LS_{i,j}$ represents the number of shared links between route i and route j. Route i is the candidate route in the initial route set and Route j is the route in the final route set. L_j is the number links containing in the Route j. r is the defined overlapping ratio threshold. It can also be calibrated with observed field data.

Traffic Information Processing

In the traffic assignment, traffic conditions $\{\boldsymbol{X_a^n}\}$ in iteration n should be processed in order to obtain the route conditions for the assignment in the next iteration. This process is the traffic information processing in the route guidance system. Depending on the route attributes that route guidance users care or select in the route guidance system settings, the route information can be categorized into two types: fixed route attributes and time-dependent attributes.

Fixed route attributes refer to the route attributes that do not change with time, such as route distance, the number of controlled intersections, etc. The information regarding the fixed route attributes can be simply calculated as the summation of the link attributes' values. To be more specific, assume the traffic conditions $\{\boldsymbol{X_a^n}\}$ on acr $(i, j) \in \boldsymbol{A}$ are available from the assignment results of the last iteration. A route can be denoted as a sequence of nodes $(i_1, i_2, i_3..., i_k)$. Then, the fixed route attribute can be collocated with Equation (8):

$$x_r = \sum_{n=1}^{k-1} x_{i_n, i_{n+1}} \qquad (8)$$

In which x_r is the fixed attribute of Route r and $x_{i_n, i_{n+1}}$ represents the attribute related link condition on the link connecting node i_n and node i_{n+1} which are the *nth* and *n+1th* nodes along driver *i*'s Route *r*.

The time-dependent attributes refer the route attributes that their values vary at different time. Route attributes such as travel time, possible longest travel time, fuel cost and etc. belong to this category because their values vary by time. With the time-dependent link conditions, the value of time-dependent route attribute at time *t* can be calculated as:

$$x_r(t) = \sum_{n=1}^{k-1} x_{i_n, i_{n+1}}(t_{i_n}^*) \qquad (9)$$

in which $t_{i_n}^*$ is the time that the agent enters the link staring with node i_n. $t_{i_n}^* = \sum_{m=1}^{n-1} tt_{i_m, i_{m+1}}$ and $t_{i_1}^* = t$. Basically, the route attribute of route *r* at time *t* equals the summation of the time dependent link conditions. As the link conditions keep changing, the link conditions could be different at different entering time. Therefore, travel time on previous links is needed to calculate the entering time on link *n*.

With Equation (8) and (9), information of the two types of route attributes can be obtained by processing the assignment results from the previous iteration. The information is fed into individual drivers' route choice model to predict their updated route choices $\{\boldsymbol{C_i^n}\}$.

Utilization of the Individual Route Choice Model

In traffic assignment process, individual route choice models are used in each iteration to predict drivers' preferred routes. As discussed in the II A section Equation (1), three types of model input can be obtained and fed into individual route choice models to obtain agents' predicted route choices.

One thing that needs to be noted is the size discrepancy between the route set and the number of alternatives considered in the individual route choice models. It is possible that a user's individual route choice model is estimated based on preference data of scenarios with *m* alternative routes, but the application scenario contains *m'* alternative routes in which *m≠m'*. In behavior study, when the number of choices increases, the decision task becomes more difficult and takes longer time to do reasoning (Churchland, Kiani, and Shadlen 2008), so the number of alternative routes shown to drivers should not be too many. The number of alternative routes that are showed to users in Google Maps is typically 2 or 3. Also, a route choice model is typically built when there are alternative routes, so *m* is more than 1.

Therefore, the value of *m* could be either 2 or 3 and the value of *m'* varies depending on different route set generation methods. Some situations in which $m \neq m'$ are discussed below in terms of how to apply individual route choices in these situations.

- When *m*=2, *m'*<2

m'<2 means that there is only one alternative in the application of individual route choice models, namely *m'*=1. Therefore, there is no need to apply route choice model because there is only one available route connecting the origin and destination of the trip.

- When m=2, m'>2

In this situation, the pairwise comparison is used to select the final choice. In behavior study, one of the decision strategies that researchers used to describe people's decision-making behavior is pairwise comparison (Pfeiffer 2012). As described by Pfeiffer (2012), "…compare alternatives sequentially in pair. They eliminate the weaker alternative of that pair and keep the stronger alternative for forming a new pair." Therefore, the pairwise comparison was adopted to in this situation. Two routes can be randomly drawn from *m'* alternative routes in the route set. Individual route choice model which was established with the data collected from binary route choice scenarios can be applied to tell which route the user is likely to prefer and marks it as *r**. Then, another route is randomly drawn from *m'-2* routes left in the route set and is evaluated against the *r** with individual route choice model. The winning route is updated as *r**. The process can be repeated until there is no route left in the route set. The final *r** is the route decision that predicted by the individual route choice model.

- When *m*=3, m'>3

In this situation, the individual route choice model can be re-estimated to handle two alternative routes at a time, namely re-estimate a route choice model in which *m*=2. Then, the rules in situations of *m*=2 can be applied. In order to estimate a route choice model in which *m*=2 from the preference data that contains 3 alternatives, 3 routes in preference data can be break down into two pairs of route scenarios. Assume three routes are labeled as [*a, b, c*] in preference data and the user selected Route *a*. Then, a training sample containing three alternatives can be break into two training samples containing two alternatives, namely [*a, b*] and [*a, c*] with Route *a* as the chosen route. A binary individual route choice model can be established from the preference data that was generated from further breaking down the original preference data. Then, the rule in situation m=2 can be applied to handle the situations in which m'>3.

- When m=3, m'<3

Based on the previous discussion, m'=1 or m'=2 in this situation. When m'=1, that means there is only one available route in the application scenario. Therefore, individual route choice model is not needed and the only available route is the selected one. When m'=2, the similar process in situation “m=3, m'>3” can be applied, namely re-estimate a binary individual route choice model with newly structured preference data.

C. Applying Route Guidance System in Rolling Scheme along Time Horizon

In the framework of the proposed route guidance system, only pre-trip route guidance is considered. That means the user receives the recommended route before the trip starts and stays on the routes until the trip finishes. In reality, it is possible for the route guidance systems regenerate routes for users based on the changing traffic conditions. The proposed route guidance system does not directly consider the en-route guidance and assumes it can be handled as a new route guidance request of which the origin is the user’s current location and the destination is the same with the pre-trip guidance request.

D. Framework

Two major components of the proposed proactive user optimum-oriented guidance system were discussed in previous sections. To have a whole idea of the proposed route guidance system, the framework in Figure 1 shows how the proposed route guidance system works. As shown in Figure 1, the widely used route guidance systems can collect individual users’ route choice preference data. The collected individual-level preference data is used to establish individual route choice models with machine learning methods for every participated driver. Then, the route guidance system can search for the user optimum conditions with the knowledge of users’ possible route choices. Finally, the obtained user optimum conditions are used to generate route guidance information and route recommendations.

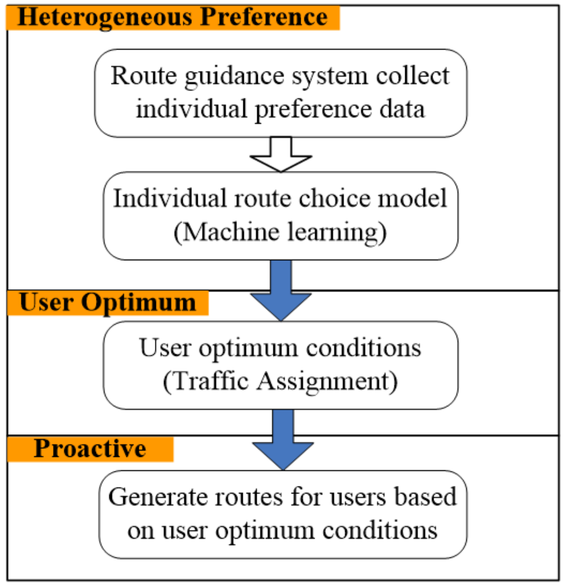


Figure 1 The Framework of the Proactive User Optimum-Oriented Route Guidance System

To have a clear view of how the user optimum conditions as well as the traffic information are obtained, the framework in Figure 2 shows the calculation for one evaluation time interval, *t*. All the notations in Figure 2 are summarized in Table 2. When evaluation time interval *t* starts, some of the drivers who departured in previous time intervals may still be in the network. These demand will be considered together with the new demand departing within current time interval *t*. The initial data input for interval *t* includes two parts: users' actual route choices in time interval *t-j* and the historical traffic conditions $\{\boldsymbol{X}^{0}_{\mathrm{r,t}}\}$ for time interval *t*. The iterative process starts with the first iteration *n*=1. Every user's route choice when seeing the historical traffic conditions $\{\boldsymbol{X}^{0}_{\mathrm{r,t}}\}$ is predicted with his/her individual route choice model. When *n*≠1 in the iterative process, the traffic conditions $\{\boldsymbol{X}^{\mathrm{n}}_{\mathrm{r,t}}\}$ is used to predict every user's possible route choices. If user *i* belongs to $\boldsymbol{D_t}$ whose departure time is within current time interval *t*, or if user *i* belongs to $\boldsymbol{D_{t-j}}$ who departured within previous time intervals and have not finished his/her trip yet, then user *i*'s estimated route choice preference is used to predict the route he/she would prefer, $C^{n}_{i,t}$. For users who belong to $\boldsymbol{D_{t-j}}$, there are two different situations. For users who belong to $\boldsymbol{D_{t-j}}$ as well as $\boldsymbol{D_R}$ who would like to have regenerated routes during their trips, their requests of regenerating recommended route update their origins and destinations. The new origins are their locations at the beginning of time interval *t*. Their destinations are still their original destinations. On the other hand, if user *i* departure within a previous time interval but belongs to the group of drivers $\boldsymbol{D_{NR}}$ who do not like to have re-generated recommended routes, then his/her original route choices $C^{*}_{i,t-j}$ is kept for this time interval. Once all users' choices $\{\boldsymbol{C}\}_{\mathrm{t}}^{\mathrm{n}}$ are obtained for any iteration *n*, they are compared with users' choices in previous iteration, $\{\boldsymbol{C}\}_{\mathrm{t}}^{\mathrm{n}-1}$. If there are more than $s*$, the pre-defined number of users having different route choices in two consecutive iterations, the converging criterion is not met and the system conducts another time of network loading with route choices $\{\boldsymbol{C}\}_{\mathrm{t}}^{\mathrm{n}}$. The generated traffic conditions go back to the input block and another iteration starts. If there are less than $s*$ users having different route choices in two consecutive iterations, the generated traffic conditions $\{\boldsymbol{X}_{\mathrm{r}}\}_{\mathrm{t}}^{\mathrm{n}}$ is recognized as user optimum conditions and the choice set $\{\boldsymbol{C}\}_{\mathrm{t}}^{\mathrm{n}}$ is considered as user optimum route choices. Therefore, $\{\boldsymbol{X}_{\mathrm{r}}\}_{\mathrm{t}}^{\mathrm{n}}$ and $\{\boldsymbol{C}\}_{\mathrm{t}}^{\mathrm{n}}$ are determined to be the predicted traffic conditions $\{\boldsymbol{X}'_{\mathrm{r}}\}_{\mathrm{t}}$ and route recommendations $\{\boldsymbol{C}'\}_{\mathrm{t}}$. Both of them are broadcasted to users. Since there might be discrepancy between the predicted drivers' route choice preference and drivers' true preferences. When users see the information, they make decisions based on their own true preferences. Their actual decisions $\{\boldsymbol{C*}\}_{\mathrm{t}}$ generate the actual traffic conditions $\{\boldsymbol{X}_{\boldsymbol{r}}^{*}\}_{\mathrm{t}}$. The process showed in Figure 2 can be applied to each time interval in a rolling scheme.

Table 2 Notations in Figure 2

Notations	Meaning
$\{\boldsymbol{C*}\}_{\mathrm{t-j}}$	A set of routes that were chosen by users who departure in previous interval and have not finished trips yet.
$\boldsymbol{X}^{n}_{r,t}$	The traffic conditions of alternative routes in iteration n for evaluation interval t.
$\boldsymbol{X}^{0}_{r,t}$	The traffic conditions of alternative routes in historical traffic conditions.
D_t	The users whose departure time is within current time interval t.
D_{t-j}	The users whose departure time is within previous evaluation intervals but whose trips have not been finished yet.
D_R	The users who choose to update their recommended routes in trips.
D_{NR}	The users who choose not to update their recommended routes in trips.
$C*_{i,t}$	The route that user *i* finally took in evaluation interval t.
$C^{n}_{i,t}$	The route that the route guidance system predicted for user *i* based on his/her individual route choice model in iteration *n*.
$\{\boldsymbol{C}\}_{\mathrm{t}}^{\mathrm{n}}$	All users' routes that were estimated by the route guidance system in iteration n.
$\{\boldsymbol{C}'\}_{\mathrm{t}}$	All users' routes that were estimated by the route guidance system after the iterative process stops.
$\{\boldsymbol{X}_{r}'\}_{\mathrm{t}}$	The traffic conditions of alternative routes after the iterative process stops.
$\{\boldsymbol{C*}\}_{\mathrm{t}}$	The routes that were finally chosen by users who departure in evaluation time interval *t*.
$\{\boldsymbol{X}_{\mathrm{r}}\boldsymbol{*}\}_{\mathrm{t}}$	The traffic conditions of alternative routes generated by the $\{\boldsymbol{C*}\}_{\mathrm{t}}$
$s*$	The converge criterion defined in the route guidance system.

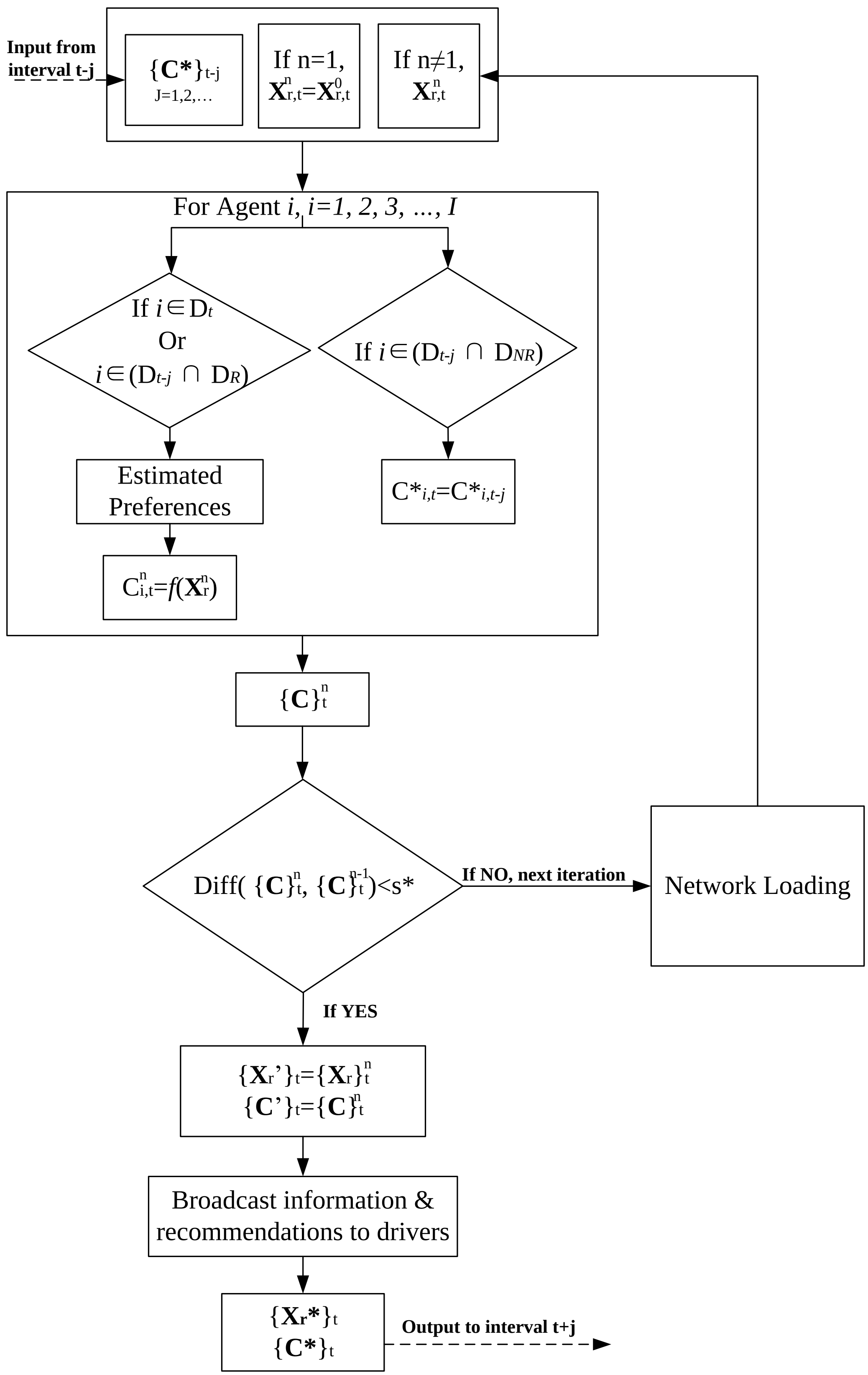


Figure 2 The Flow Chart of the Proposed Route Guidance System in Interval *t*

From a user's perspective, he/she can choose to either update recommended route or not in user settings of the route guidance system. When the user requests a route guidance service, his/her OD information together with all other drivers' OD information are gathered and processed by the route guidance system. A recommended route is returned to the user and the user starts the trip by following the recommended route. If the trip ends before the current evaluation interval ends, there are not additional steps to do. If the trip has not been finished when the current evaluation interval ends, a new route would be recommended to the user in the next evaluation interval depending on whether the user chooses to update his/her recommended route in trips or not. If the user belongs to the group of drivers who choose to update the route, the user's trip is considered as another new request with his/her current location as the origin and original destination as the destination. Otherwise, the user can keep using the same recommended route as in the previous evaluation interval. The process can be repeated until the user reaches the destination.

III. EVALUATION AND COMPARISON

In order to evaluate the performance of the proposed route guidance system, this section sets up a scenario to compare the proposed route guidance system with several existing route guidance strategies. The details in the comparison including scenario set up, an evaluation platform establishment, individual route choice models' application, selected existing route guidance strategies and performance evaluation, are discussed in this section.

A. Test-bed Set Up

A test-bed was set up to demonstrate how the proposed proactive user-optimum oriented route guidance system works and how its performance compared to other routing strategies. Sioux Falls network is widely used in traffic analysis and is also adopted in this research.

The Sioux Fall network has 24 traffic analysis zones and 76 road links. Initially, Sioux Falls network was used by LeBlanc (1975) for static traffic assignment. The traffic demand which was given in an origin-destination matrix and road links features was also used for static traffic assignment. In order to test the proposed route guidance system in dynamic traffic assignment manner, some adjustments were made. The link setting including capacity, speed limit and number of lanes were adjusted based on Chakirov's work (Chakirov 2014). The author selected the network characteristics in terms of capacity, road type, speed limit, etc. by matching the real-world network characteristics to the network structure used in previous static analysis. The initial OD matrix of Sioux Falls network gives the daily traffic demand among all 24 traffic analysis zones and used 10% to represent hourly traffic demand. It is equivalent to 10% peak hour volume. Since traffic control and management strategies focus more on congested traffic conditions in urban transportation system, therefore, the 15% peak hour volume index is set here and 15% of daily demand is the hourly demand used in this evaluation. The total demand of 54045 drivers were distributed in one hour evaluation interval which was further break into four 15-minute intervals having the demand share of 20%, 30%, 30% and 20%. Agents depart from their origins in a constant rate within each 15-minute interval, but the departure rate within different intervals varies. Given most of the trips in morning peak are commute trips which typically have fixed desired arriving time, every agent is assumed to have fixed departure time in this research. When departure time choice preference data is available, departure time choice behavior can also be considered. The network structure and the hourly OD matrix are included in the Appendix.

B. Individual Route Choice Model Development

Generating synthetic preferences

Assuming each agent represents a unique driver who is using route guidance system in the road network, each agent has his/her own route choice preference. In reality, when drivers participate in the proposed route guidance system, each driver's route choice preference can be estimated by the route guidance system. In this evaluation, agents' route choice preferences were synthetized from a group of participants' preferences which were observed from a stated preference survey.

A stated preference survey was conducted among 28 participants who are mostly students at the University of Virginia. The participants were invited to sit in the driving simulator and were shown 76 binary route choice scenarios. Based on the information of five attributes including *Distance*, *Travel Time*, *Possible Longest Travel Time*, *Fuel Cost* and *Number of Controlled Intersections*, participants were asked to choose a route to go based on the information. Their answers were recorded and analyzed to obtain individual route choice preference. The details of experiment design and survey implementation was discussed in Chapter 3 of this dissertation.

Participant i's preference is denoted as a weight vector, $\boldsymbol{w^i}$=[w_1^i, w_2^i, ..., w_n^i]^T, in which w_n^i represents participant i's weight regarding route attribute n. Assume the weight of all drivers' route choice preference regarding a certain route attribute, $\boldsymbol{w_n}$, follows a normal distribution as shown in Equation (10). The parameters of the distribution, u_n and s_n are the mean and standard deviation calculated from observed participants' preferences following Equation (11) and Equation (12). Then, agent i's preference in terms of the weight regarding the route attribute *n* can be randomly drawn from the normal distribution, $N(u_n, s_n)$. The synthetic preference of an agent can be written as $\boldsymbol{w^{i'}}$ as given in Equation (13). These randomly generated weights are considered as the "true" preference for the evaluation purpose in this research. It is noted that driver's true preference is by no means available to the route guidance system. That is to say, route guidance systems provide services and information based on their estimated drivers' preferences, but drivers react to the services and information based on their true preferences. To have an idea about the performance of route guidance systems, the assumed "true" preference is not involved in designing routing strategies but used in the final evaluation to simulate drivers' actual responses to the information in practice. Synthetic drivers with synthetic preferences went through a questionnaire that was used in the stated preference survey and indicated their preferred routes in answers. As given in Equation (14), when synthetic agent *i* sees the route choice scenario $\boldsymbol{x_j}$ in the questionnaire, its answer is $\boldsymbol{y_j}$. ε_j is added in the generation of synthetic agents' decisions to represent the unobserved influencing factors and such error was added by drawing from a pre-defined normal distribution. These questionnaires can be seen as the observations that route guidance systems can learn drivers' preferences from. In reality, these observations can be either made by observing drivers' historical trips or asking driver's stated preference with survey questions. With these observations, individual route choice models can be established. Driver i's estimated preference is in noted as in Equation (15).

$$\boldsymbol{w_n} \sim N(u_n, s_n) \tag{10}$$

$$u_n = \frac{1}{I}\sum_{i=1}^{I} w_n^i \tag{11}$$

$$s_n = \sqrt{\frac{\sum_{i=1}^{I}(w_n^i - u_n)}{I-1}} \tag{12}$$

$$\boldsymbol{w^{i'}} = \begin{pmatrix} w_1^{i'} \\ w_2^{i'} \\ \vdots \\ w_n^{i'} \end{pmatrix} \tag{13}$$

$$y_j^{i'} = \boldsymbol{w^{i'}}\boldsymbol{x_j} + \varepsilon_j \tag{14}$$

$$\widehat{\boldsymbol{w}}^{i'} = \begin{pmatrix} \widehat{w}_1^{i'} \\ \widehat{w}_2^{i'} \\ \vdots \\ \widehat{w}_n^{i'} \end{pmatrix} \quad (15)$$

Building Individual Route Choice Models

With the preference data collected from synthetic agents, individual driver's route choice model can be established, namely building each agent a route choice model. Following the structure of the individual route choice model described in Section II A, a machine learning technique of support vector machine (SVM) with a linear kernel function was adopted to demonstrate how individual route choice models can be established. In reality, there are situations where individual driver does not have adequate historical preference data, for example, he/she is a new user and just starts using the system or the route guidance system just starts operation. To handle these situations, modeling approaches with more advanced features such as Multi-task linear model adaptation, can be used to establish individual route choice models. In the scenario of this research, each agent went through all 76 questions in the survey questionnaire. It is considered as having enough amount as well as variation of historical preference data. Therefore, a simple machine learning method is used here for demonstration.

The concept of SVM is to map the data points into high dimensional space and find a hyperplane which can separate the points belonging to different categories. The hyperplane can be represented as follows:

$$\boldsymbol{\omega X} + b = 0 \quad (16)$$

In which $\boldsymbol{\omega}$ is the vector of hyperplane's slopes, $\boldsymbol{X}$ is the vector of alternative's features and b is the intercept. To find this hyperplane, $\boldsymbol{\omega}$ and b can be obtained by solving the following optimization problem (Steinwart and Christmann 2008).

$$\text{Min } \frac{1}{2}\boldsymbol{\omega}\boldsymbol{\omega}^{\boldsymbol{T}} + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{n}} \epsilon_{\mathrm{i}} \quad (17)$$

$$s.t.\, y_i(\boldsymbol{\omega X} + b) \geq 1 - \epsilon_i, i = 1,2 \dots, n$$

$$\epsilon_i \geq 0, i = 1,2 \dots, n$$

Solving the optimization problem in Equation (17) results in maximizing the distances between data points to the hyperplane. ϵ_i is the slack variable which adjusts the margin between data points and hyperplane. It gives some tolerance for the data points that cannot be linearly separated. C is a hyper parameter which puts penalty on these data points. The optimization problem can be converted to a dual problem and solved with Lagrange method.

Therefore, the problem of choosing which route is converted to classifying the scenarios into the category of choosing route *j* (*j=1, 2, …, J)*. We can find each driver a hyper plane which is represented with $\widehat{\boldsymbol{w}}^{\boldsymbol{i}'}\boldsymbol{X} + \mathrm{b} = 0$ in which $\widehat{\boldsymbol{w}}^{\boldsymbol{i}'}$ is the estimated weight vector of driver *i*, as noted in Equation (15). $\boldsymbol{X}$ contains all the information in Equation (1), namely including route attributes information, trip related information, and driver's characteristics. In this demonstration, only route attribute information was include. *b* is the intercept of the hyper plane, which can be estimated with SVM. In this research, since the route choice scenarios in the survey are all binary route choice scenarios, a SVM for binary classification was used. When there are multiple alternative routes, multi-class classification SVM can be adopted (Hsu and Lin 2002).

For all agents, the data of each agent was used to establish his/her own route choice model following Equation (16) to (17). In order to apply SVM models, the penalty parameter needs to be determined with cross validation. The range explored in the cross validation is a geometric sequence from 10^{-5} to 10^{5} by a factor of 10, which is a commonly used range for penalty parameter in SVM (Ben-Hur and Weston 2010). A particular agent's data was split into five groups. Each group was used as validation data once on all possible values of penalty parameter. This random split was conducted five times. The value with highest average performance on validation data was selected to be the penalty parameter value. With a penalty parameter, the agent's weight $\widehat{\boldsymbol{w}}^{\boldsymbol{i}'}$ can be obtained with SVM models given in Equation (16) to (17). This process was conducted for all agents. Each agent has its "true" preference $\boldsymbol{w}^{\boldsymbol{i}'}$ and the estimated preference $\widehat{\boldsymbol{w}}^{\boldsymbol{i}'}$ based on its stated preference survey answers that were generated with "true" preference. In reality, only $\widehat{\boldsymbol{w}}^{\boldsymbol{i}'}$ is available to route guidance systems which is learned from driver's historical preference data.

C. Evaluation Platform

To evaluate the performance of the proposed route guidance system in the test-bed, an evaluation platform which is made of a Decision module and a Traffic Simulation module was set up, as shown in Figure 3. The decision module was mainly coded in Matlab and the Traffic Simulation module was made based on DTAlite. DTAlite is a lightweight simulation based Dynamic Traffic Assignment agent-based mesoscopic simulation tool (Zhou and Taylor 2014). It utilizes parallel computing thus can efficiently finish network loading process.

The Decision module designs different route guidance strategies, generates route sets for drivers, realistically mimics drivers' route choice behavior, and processes and generates route guidance information. The input of this module is mainly traffic conditions either historical traffic conditions or the traffic conditions from the previous iteration. The output of the Decision module is the route choice of each agent.

In the Traffic Simulation module, the network configuration including information of road network structure, capacity, number of lanes, etc. is set up. The network loading process is conducted in DTAlite. The traffic flow model used in traffic assignment is Newell's simplified kinematic wave model. It simulates the demand in terms of agents' movements from origins to destinations. The input of the traffic simulation module is the route choice decisions of every single agent and the output is the time-dependent traffic conditions in terms of time dependent link performance such as volume, speed, density, queue length, etc. These time-dependent traffic conditions are utilized by the decision module again to conduct the calculation for the next iteration.

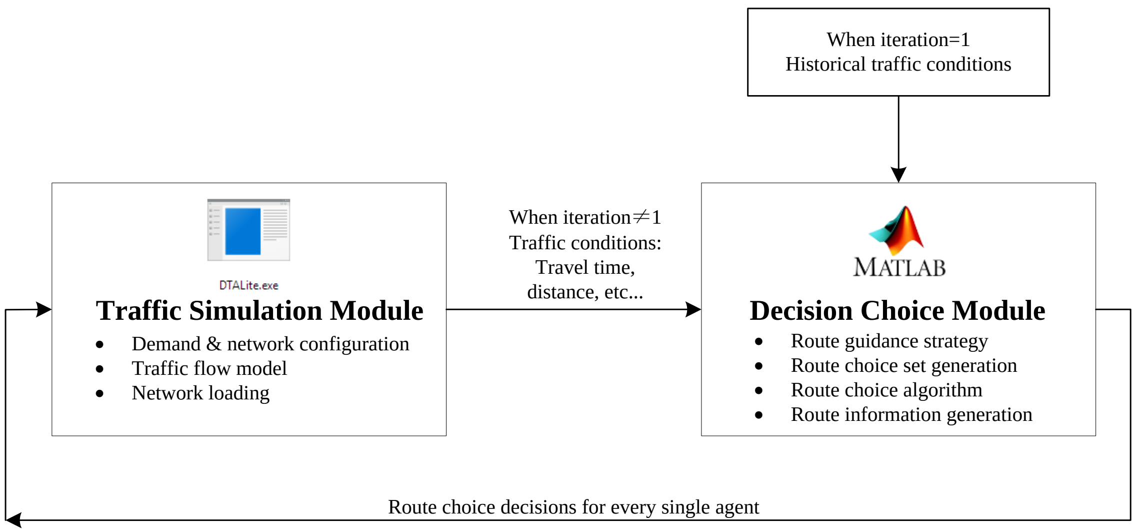


Figure 3 The Framework of the Evaluation Platform Comprise a Dynamic Traffic Simulation Module and a Decision Choice Module

Route set generation

In the traffic assignment process, the route set is pre-defined. As described in II B section, the branch and bound algorithm is used to generate the route set. Branch and bound algorithm constructs a connection tree between the origin and destination by processing a sequence of road links. While going through each link, certain bounded criteria are set up to avoid the unrealistic routes as well as include as many heterogeneous alternative routes as possible. As explained in the Section II A, the branch and bound algorithm needs to determine the values of adjusting factors and threshold values for some criteria. These values should be calibrated with observed data on specific road network in practice. In this research, starting with the values used in (C. Prato and Bekhor 2006), adjustments were made to generate a fairly reasonable route set. Then, some OD pairs were randomly selected and compared with the suggested route shown on Google Maps. The tolerance values that generated a route set with similar routes recommended by Google Maps were eventually used. The tolerance values used for generating the final route set are as: $\delta_D = 1.2$, $\delta_{\mathrm{T}} = 1.25$, $\delta_{\mathrm{D}} = 1.2$, $N = 10$. With the tolerance values set up, route sets were generated for each OD pair.

After the initial route sets were generated, similarity check was conducted to avoid the situations that overlapping routes are included in the route set at the same time. In this research, different levels of similarity threshold were set for routes with different amounts of links. For the route having less than 6 links, r in Equation (7) was set to be 0.8. That means a route containing 5 links is accepted if it has less than 3 overlapping links with all existing routes in the final route set. For routes having equal to or more than 6 links, r was set to be 0.5. Therefore, the initial route set was narrowed down by eliminating overlapping routes.

For the 24 traffic analysis zones in Sioux Falls network, there are 552 OD pairs. Route set was generated for each OD pair using the branch and bound algorithm described above. After the similarity check, the number of alternative routes in the final route sets for each OD pair ranges from one to ten. The composition of different numbers of alternative routes were shown in Figure 4.

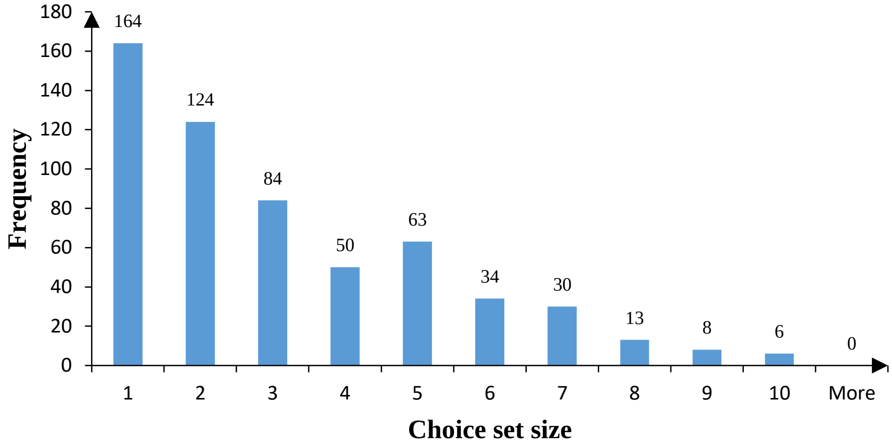


Figure 4 The Frequency of OD Pairs Having Certain Choice Set Size

Traffic information processing

Drivers make decisions based on the route information provided to them. The information of all routes in the route sets was obtained by processing traffic conditions. As explained in Section II B, in the process of searching for user optimum conditions, drivers' decisions in iteration *n* are made based on the traffic conditions in iteration *n-1*. The information of route attributes that are considered in the drivers' personalized route choice model is gathered from the traffic conditions generated in the last iteration. When the traffic assignment is in the first iteration, there are no traffic conditions from the previous iteration that can be used to process traffic information. Therefore, the historical traffic conditions are used to process the traffic information in the first iteration.

Historical traffic condition generation

In order to more realistically represent the traffic conditions, the starting point of the iteration process is not from the status of free flow conditions. Instead, historical traffic conditions are used. Historical traffic conditions represents the typical traffic conditions in the past. Drivers adjust their route choices based on their experiences with the network and finally have their typical route choices which lead to typical traffic conditions. The weighted average of traffic conditions in several iterations before reaching equilibrium is usually used to represent typical historical traffic conditions (Yang, Koutsopoulos, and Ben-Akiva 2000). Therefore, travel time based traffic assignment was conducted and average conditions (time dependent link travel time) of 10 iterations before reaching equilibrium was used as historical traffic conditions.

Calculating route attributes from traffic assignment results

The network is made of a set of nodes, **N** and a sets of arcs **A**. The nodes are numbered 1,..., N. The acr (i,j) ∈**A** has time dependent conditions associated with it, as shown below.

Length of the link: $d_{i,j}$;

Travel time on the link at time *t*: $tt_{i,j}(t)$;

Possible longest travel time on the link at time *t*: $pltt_{i,j}(t)$;

Fuel cost on the link at time *t*: $fc_{i,j}(t)$.

Most the above information can be obtained from the output of the traffic simulation module which outputs the traffic conditions very one minute. In order to calculate the fuel cost, following equation is used. The gas price was assumed to be $3 per gallon.

$$fc_{i,j}(t) = \frac{d_{i,j}}{fe_{i,j}(t)} * 3 \qquad (18)$$

in which the fuel efficiency (mile/gallon) was assumed to be three levels based on the speed (mile/hour):

$$fe_{i,j}(t) = \begin{cases} 22 & \frac{d_{i,j}}{tt_{i,j}(t)} \leq 25 \\ 27 & 25 < \frac{d_{i,j}}{tt_{i,j}(t)} < 40 \\ 32 & \frac{d_{i,j}}{tt_{i,j}(t)} > 40 \end{cases} \qquad (19)$$

With the link attributes, route attributes can be calculated accordingly. A route can be denoted as a sequence of nodes (i_1, i_2, i_3..., i_k). The information of all route attributes including distance, travel time, possible longest travel time, fuel cost and number of intersections that are included in the users' route choice models can be calculated as follows.

- Distance of route *r*

$$d_r = \sum_{n=1}^{k-1} d_{i_n, i_{n+1}} \qquad (20)$$

- Travel time of route *r* at departure time *t*

$$tt_r(t) = \begin{cases} \sum_{n=1}^{h-1} tt_{i_n, i_{n+1}}(t_{i_n}^*) & when\ h = 0\ or\ in\ 1st\ iteration \\ \frac{1}{h}\sum_{i=1}^{h} tt_i^r & when\ h > 0 \end{cases} \qquad (21)$$

in which $t^*_{i_n}$ is the time that enters the link staring with node i_n. $t^*_{i_n} = \sum_{m=1}^{n-1} tt_{i_m,i_{m+1}}$ and $t^*_{i_1} = t$. h is the total number of agents who traveled on route r in the previous iteration. tt^r_i is the travel time that experienced by the agent i in the previous iteration on route r.

- Possible longest travel time of route r

$$pltt_r(t) = \begin{cases} tt_r(t) & when\ h = 0\ or\ in\ 1st\ iteration \\ max(tt^r_1, tt^r_2, \dots, tt^r_h) & when\ h > 0 \end{cases} \quad (22)$$

- Fuel cost of route r at time t

$$fc_r(t) = \sum_{n=1}^{k-1} fc_{i_n,i_{n+1}}(t^*_{i_n}) \quad (23)$$

- Number of controlled intersection of route r

$$inter_r = k - 1 \quad (24)$$

With the process described above, the information of all five route attributes for any alternative route can be obtained. Then, these information can be used as the input of individual route choice models for predicting drivers' route choice decisions.

Applying Individual Route Choice Models

In the traffic assignment, after the route information was obtained from the traffic information processing, the traffic conditions on alternative routes in route set are available to agents. Route conditions can be plugged into each drivers' estimated route choice preference $\hat{\boldsymbol{w}}^{\boldsymbol{i}'}$ to know the route that they are likely to take. In the route sets, there are more than two alternative routes for some origin and destination pairs. On the other hand, the route choice models developed in this research is for binary route choice scenario. In reality, when there are more than two or three alternatives, one decision strategy that people usually use is to conduct pairwise comparison to determine the final selection (Pfeiffer 2012). Therefore, when there are more than two routes, agents use pair comparison to select the route to go, as discussion in Section II B.

D. Comparison Route Guidance System with Different Routing Strategies

The proposed route guidance system was evaluated against three existing route guidance strategies, including travel time based real-time route guidance with different information updated intervals, and using travel time based user equilibrium condition as guiding information. Each of them is discussed below.

Travel time based real-time route guidance

Real-time route guidance strategy is very similar to the route guidance systems in the practice nowadays. Users can get the real-time traffic conditions as the information when they departure. Route guidance system operators (such as Google Maps) keep monitoring and collecting the traffic conditions on the network and provide the real-time traffic conditions to their users. This can help the drivers avoid the congested area and save travel time.

To represent the travel time based real-time route guidance system, the evaluation time period of [*0*, *T*] can be discretized in to *M* intervals with length of σ. The time intervals can be indexed as *1, 2, …, M*. When agent *i* departure at time *t* and *t* falls into time interval *m*, its decisions can be calculated with its individual route choice model as follows:

$$Y_{i,m} = f(\boldsymbol{X_{m-1}}) \qquad m = 1,2,\dots,M \tag{25}$$

in which $\boldsymbol{X_{m-1}} = [d^r, tt^r_{m-1}, pltt^r_{m-1}, fc^r_{m-1}, inter^r]$. When *m*=1, the traffic conditions in interval *m* in the historical traffic conditions were used as information and provided to drivers. $\boldsymbol{X_0} = [d^r_{his}, tt^r_{m,his}, pltt^r_{m,his}, fc^r_{m,his}, inter^r_{his}]$.

The Equation (25) describes how agent *i* uses the traffic conditions that happened recently and are collected by the route guidance system operators to make a route choice decision. Depending on how often the information is updated, users who departure in a time interval get the information about traffic conditions in the previous iteration. When the length of time interval σ is small enough, the information is nearly real-time. It is easy to understand that the smaller the value of σ is, the faster users can react to the congestion in the network. As implemented in the practice, based on the collected traffic conditions, route with shortest travel time is recommended to the users. Because of data collection and processing efforts, route guidance system operators may adopt different levels of information update frequency. In this evaluation, real-time route guidance system with σ=1 minute and σ=10 minutes are tested.

Travel time based user equilibrium condition as information

Using user equilibrium condition as information to provide route guidance service is one of the commonly used route guidance strategies proposed in the literature and have been proved to have good performance (Deflorio 2003; Giglio and Sacco 2014; He, Guan, and Ma 2013). The concept of dynamic user equilibrium (DUE) is defined as the flow pattern in which the actual route travel time is the same on all used routes and is less than or equal to the travel time on non-used route, at any time (Carey and Ge 2012). Therefore, at any time interval *m*, following conditions should be met.

$$tt_m^r = \begin{cases} = tt_m^{r\,*} & if\ f_m^r > 0 \\ \geq tt_m^{r\,*} & if\ f_m^r = 0 \end{cases} \quad where\ tt_m^{r\,*} = min(tt_m^r | r \in J) \qquad (26)$$

In which tt_m^r is the travel time on route *r* in time interval *m*. $tt_m^{r\,*}$ is the shortest travel time among all alternative routes in time interval *m*. f_m^r is the traffic flow or the number of agents choosing route *r* in time interval *m*.

In the algorithms of finding the DUE conditions, method of successive average (MSA) is one of the widely used algorithms (Sheffi 1985). With MSA, the route flow is adjusted within iterations with following procedures (Sheffi 1985).

Step 0: Initialization. Perform a network loading based on the historical travel times $\{\boldsymbol{tt_m^{r,0}}\}$. A traffic pattern $\{\boldsymbol{f_m^{r,1}}\}$ is generated. Set iteration number, *n*=1.

Step 1: Update. Based on the traffic pattern $\{\boldsymbol{f_m^{r,n}}\}$, the travel time information on all routes $\{\boldsymbol{tt_m^{r,n}}\}$ can be obtained.

Step 2: Direction finding. Perform a network loading based on the current travel time conditions $\{\boldsymbol{tt_m^{r,n}}\}$. This yields an auxiliary traffic pattern $\{\boldsymbol{y_m^{r,n}}\}$.

Step 3: Move. Find the new flow pattern using $f_m^{r,n+1} = f_m^{r,n} + \alpha_n (y_m^{r,n} - f_m^{r,n})$ in which α_n ($0<\alpha_n<1$) is a fraction determines how much flow is moved. $\alpha_n = \frac{1}{n}$ is used here.

Step 4: Convergence criterion. If convergence criterion is reached, stop. If not, set *n=n+1* and repeat Step 1 to Step 4.

Relative gap which is widely used as the convergence criterion is selected to determine whether the iterative process should stop (Chiu et al. 2011). The relative gap function is defined as in Equation (27).

$$rel_{gap} = \frac{\sum_m \sum_{s\in S} (\sum_{r\in R_S} f_m^r tt_m^r) - \sum_m \sum_{s\in S} D_{s,m} tt_{s,m}^*}{\sum_m \sum_{s\in S} D_{s,m} tt_{s,m}^*} \qquad (27)$$

The relative gap is basically measuring how far the current flow pattern deviates from the flow pattern in which everyone is using the shortest travel time route. *m* represent any time interval. *s* is a specific OD pair. *r* represents one of all *R* alternative routes between this OD pair *s*. f_m^r represents the number of drivers who choose route r in time interval m and tt_m^r is the associated travel time on route r in time interval m. $\boldsymbol{D_{s,m}}$ is the total demand traveling between OD pair s in time interval m and $tt_{s,m}^*$ is the shortest travel time between OD pair s in time interval m. In this evaluation, the relative gap criterion is set to be 0.01%. When the relative gap is less than 0.01%, the iterative process is terminated.

In traditional UE assignment, traffic demand is treated as aggregate traffic flow. The flow pattern is adjusted also in an aggregate manner in the process of MSA. However, traffic demand is made of many single individual drivers and each of them is represented by a single agent in the evaluation platform. Since travel time based UE assignment assumes drivers have simplified preference in terms of preferring route with shorter travel time, drivers are not considered as being different from each other. Therefore, in MSA process, when certain fraction of traffic flow should be switched among alternative routes, associated amount of agents were randomly selected and switched accordingly.

E. Performance Evaluation at Perfect Market Penetration Rate

Evaluation flow chart

The proposed route guidance system was evaluated against travel time based real-time route guidance system and travel time based UE condition as information for guidance. The performance is evaluated in three aspects including users' satisfaction, system performance, and the capability of estimating future traffic conditions.

From the users' perspective, users' satisfaction is evaluated by checking if the route recommended by the route guidance system match the users' own choices (represented by the "true preference"). In the proposed route guidance system, routes are recommended based on estimated route choice preferences. While in other route guidance systems, users' route choice preferences are simplified to be preferring the route with shortest travel time. The routes are recommended based on this simplified preference. In the evaluation, the number of agents whose truly preferred route match the recommended routes is used as measurement of users' satisfaction.

For transportation authorities, the ultimate goal is to improve the transportation system performance. The performance can be measured in aspects of mobility and sustainability. Therefore, the system total travel time, total delay, energy consumption and greenhouse gas emission are used to measure the system performance of different route guidance systems.

By knowing the possible routes that each driver may take, it is likely to estimate the future traffic conditions even before the trips are made. This is one of the major advantages of the proposed proactive user-optimum oriented route guidance system. Accurately estimating future traffic conditions can not only provide the drivers with more reliable traffic information but also allow traffic engineers to implement traffic control and management strategies to avoid the possible congestion. The accuracy of future traffic condition prediction is measure as the discrepancy between the predicted link volumes estimated by the route guidance systems and the actual link volumes generated with users' "true preferences".

With the purpose of evaluating the route guidance systems in three aspects mentioned above, the evaluation flow chart for each type of route guidance system is discussed below.

Figure 5 shows the evaluation flow chart of the proposed proactive user-optimum oriented route guidance system. The system learns drivers' route choice preference from the historical preference data and has every individual's estimated preference. Then, based on the estimated preferences, user optimum conditions can be found and used as the route guidance information. The predicted traffic conditions are also generated in the process of searching for the user-optimum conditions. The generated route guidance information is fed into agents' true preferences which are unknown to the system to obtain the actual traffic conditions. This represents the situation that traffic information is broadcasted to the public and drivers react to the traffic information based on their own preferences in reality. The consequences of drivers' reaction are the actual performance of the proposed route guidance system. Therefore, users' actual route choices are compared to the recommended routes in the generated guidance information to evaluate users' satisfaction. The network conditions in actual traffic conditions are used to evaluate the system performance. The discrepancy between actual traffic conditions and predicted traffic conditions can show how well the route guidance system estimates future traffic conditions.

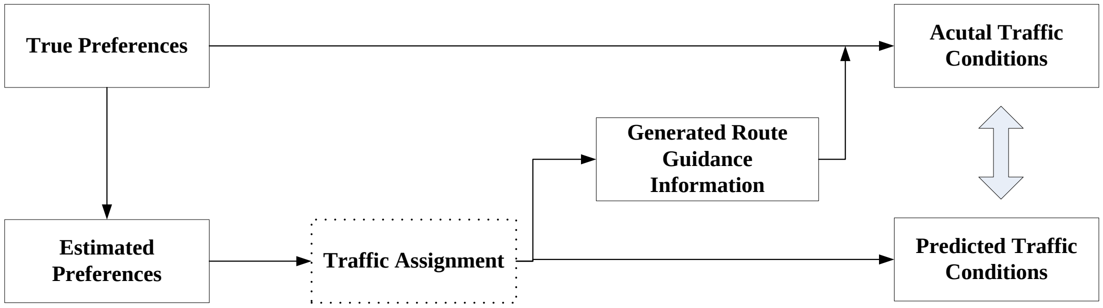


Figure 5 The Evaluation Flow Chart of the Proposed Proactive Route Guidance System

Figure 6 shows the evaluation flow chart of travel time based real-time-guidance system. As explain in Section III D, the real time route guidance system keep monitoring the traffic conditions and send the real-time traffic information to users. The system assumes all users prefer the route with shortest travel time. The assumed preference of demand $\boldsymbol{D}$ in time interval m is represented by $F'_{D,m}$. Therefore, $F'_{D,m}(\boldsymbol{X_{m-1}})$ represents the predicted route choice of demand $\boldsymbol{D}$ in time interval m when they obtain the traffic conditions $\boldsymbol{X_{m-1}}$ that is collected from most recent short time interval m-1 by the route guidance system. Accordingly, $F_{D,m}(\boldsymbol{X_{m-1}})$ represents users' actual choices based on their true preferences when they see information $\boldsymbol{X_{m-1}}$. The discrepancy between $F'_{D,m}(\boldsymbol{X_{m-1}})$ and $F_{D,m}(\boldsymbol{X_{m-1}})$ is used to evaluate users' satisfaction. The generated actual traffic condition $\boldsymbol{X}$ is the consequences of drivers' actual reactions and therefore is used for evaluate the system performance. Meanwhile, when the real-time route guidance system is used for estimating future traffic conditions, only assumed route choice preferences are used in the simulation process and it is assumed that all drivers follow the recommended routes with shortest travel time. The generated predicted traffic conditions $\boldsymbol{X'}$ are compare to the actual traffic conditions $\boldsymbol{X}$ to evaluate the route guidance system's capability in future traffic condition estimation.

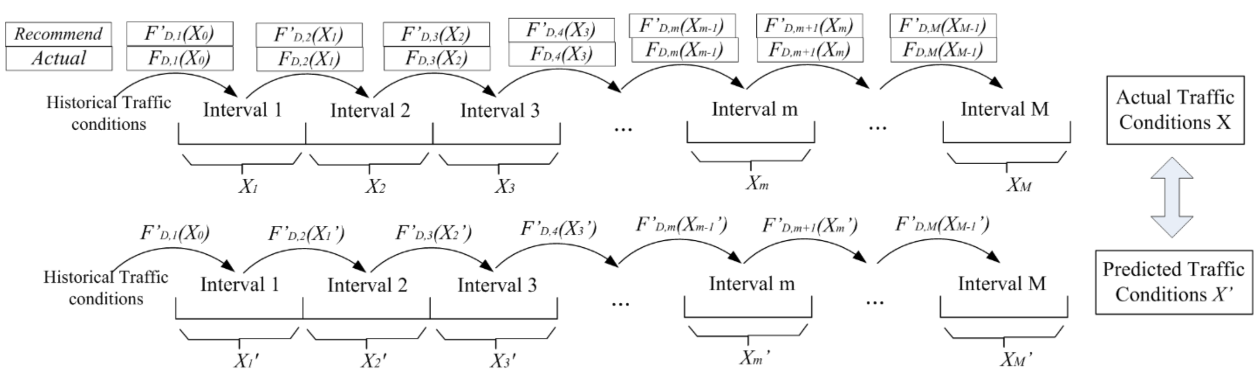


Figure 6 The Evaluation Flow Chart of the Real-time Route Gudiance System

At last, Figure 7 shows the evaluation flow chart of the route guidance system using travel time based UE conditions as information. The travel time based UE conditions are found firstly and then used as information. In reality, the information is broadcasted to the users and the users react to the information based on their own preferences. This generates the actual traffic conditions as the effects of sending travel time based UE conditions as information. The system performance of the route guidance system is evaluated based on actual traffic conditions. Travel time based UE also assumes users prefer the route with shortest travel time. Based on this preference assumption, the system recommends the route with shortest travel time to users and assume they all follow the recommendations. This generates the predicted traffic conditions. The recommended routes are compared to drivers' actual choices to reflect the users' satisfactions, and the discrepancy between the predicted and actual traffic conditions shows how well the route guidance system can estimate future traffic conditions.

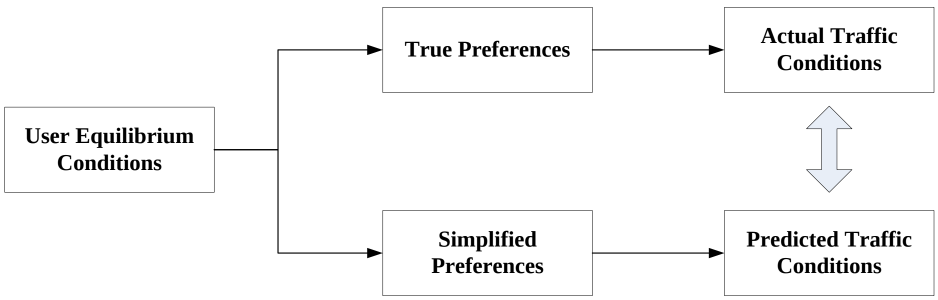


Figure 7 The Evaluation Flow Chart of Travel Time based UE Conditions as Information

With the evaluation flow charts described above, three types of route guidance systems were evaluated in three important aspects, including users' satisfaction, system performances and capability in future traffic condition estimation. Each of these three aspects is discussed below in details.

User' satisfaction

Since the ultimate purpose of route guidance systems is to better serve drivers, users' satisfaction is one of the most important performance indicators. Users' satisfaction is measured as the number of drivers whose preferred routes match with the routes recommended by the route guidance systems. Table 3 summarizes the number of satisfied drivers whose preferred routes match with the recommended routes in each route guidance system. As shown in Table 3, the proposed system has the most satisfied drivers which is around 88.64% of all users. It is because that the proposed route guidance system considers all the factors that drivers care for making route choice decisions therefore it has better prediction of which route a driver would like to take. On the other hand, the real-time-guidance system and the route guidance system using UE condition as information both assume drivers only care travel time when making route choice decisions. Therefore, the recommendations made based on the simplified preference assumptions lead to lower percentage of satisfied users. Among these two types of route guidance systems, the real-time-guidance system has better performance than the route guidance system using UE condition as information. It is because the real-time-guidance system can capture the changing traffic conditions and help the drivers know the current traffic conditions before the trips are made. Even though it simplifies the route choice preference to be only caring travel time, travel time is still the most important attribute for many drivers. Therefore, it performs better than the guidance system using UE conditions as information which does not adjust their recommendations based on the real-time traffic conditions.

Table 3 The Amount of Users Who Are Satisfied with Route Recommendations

Route Guidance System	Number of satisfied users	Percentage of satisfied users
Real-time 1min	38,096	70.49%
Real-time 10 min	37,935	70.19%
UE as information	35,963	66.54%
Proposed	47,908	88.64%

System mobility and sustainability

The mobility and sustainability impacts of the proposed route guidance system are evaluated against other three route guidance systems in this section. System total travel time and total delay are used to measure the system mobility impacts. The system total travel time is defined as the summation of all agents' experienced travel time. The system total delay is measured as number of vehicles times the delayed travel time. It is calculated by summing the multiplication of link delay and the link volume on all links. The sustainability impacts of route guidance systems include energy consumption and emission.

As shown in Table 4, the proposed route guidance system has the minimum system total travel time and total delay among all routing strategies. The sustainability impacts show a similar trend. Though the proposed route guidance system targets on the user-optimum conditions which is not necessarily associates with better system performance, the total travel time and total delay was reduced because travel time is still one of the most important attributes that many drivers considered. Therefore, the target of user-optimum has some overlap with the goal of minimizing system total travel time. In addition, because drivers' route choice decisions can be more accurately predicted with the proposed route guidance system, it is easier to know the possible congested routes in traffic assignment process and drivers can avoid those areas in the process of searching for user-optimum conditions. Therefore, it avoids the situation that drivers are navigated to the congested routes and consequently reduces the delay and travel time.

Among other route guidance systems, the results show similar trend as in the analysis of users' satisfaction. The real-time-guidance system has better system performance than the guidance system using UE conditions as information in the aspect of both mobility and sustainability. It is because the latter does not adjust its recommendations based on the changing traffic conditions. On the other hand, the real-time-guidance system captures the real-time traffic conditions and makes recommendation accordingly. It is easy to understand that real-time-guidance system with 1 minute information update interval has better performance than the one with 10 minutes information update interval, because drivers can react to the changing traffic conditions faster when the information is updated more frequently.

Table 4 The Mobility and Sustainability Impacts of Different Routing Strategies

Route Guidance System	Total Travel Time (hours)	Total Delay (veh*hours)	Total Energy (10^9KJ)	CO2 (ton)	NOX (kg)	CO (kg)	HC (kg)
Real-time 1min	26,983.1	9,831.9	2.81	201.9	146.2	1594.8	51.0
Real-time 10 min	28,039.6	10,809.5	2.87	206.0	146.6	1611.3	52.1
UE as information	30,760.5	13,857.4	2.97	213.5	148.0	1628.2	55.4
Proposed	24,719.3	7,860.4	2.65	190.8	141.3	1525.3	47.8

Future traffic conditions estimation

Volume is one of the most important factors used in designing and implementing traffic control and management strategies, such as traffic signal control. Therefore, the link volume difference between predicted and actual traffic conditions (namely the $\{X_r^*\}_t$ as shown in Figure 2) is used to evaluate the capability of each route guidance system in estimating future traffic conditions.

Figure 8 shows the number of links having different levels of volume discrepancy in each route guidance system. There are 76 links in the Sioux Falls network. For the proposed route guidance system, as shown in the Figure 8, 74 links have link volume discrepancy less than 10% and only 2 links have discrepancy within the range of [10%, 20%]. Other route guidance systems have nearly 20 out of 76 links having volume discrepancy between [10%, 20%], around 15 to 30 links having discrepancy between [20%, 50%] and 3 to 6 links having discrepancy between [50%, 100%]. More links in the ranges with lower discrepancy means more accurate estimation of future traffic conditions. Therefore, the proposed route guidance system has more accurate volume estimation than other route guidance systems. This advantage of the proposed route guidance system is the result of better knowing drivers' route choice preferences. Thus, it is easier to predict the routes that drivers are likely to take and consequently predict traffic conditions. As traffic volume plays an important role in designing and implementing traffic control and management strategies, an accurate estimation can help the traffic engineers implement a design that is suitable for the traffic conditions and avoid delays caused by an impropriate design. Therefore, the proposed route guidance system in which most of the links have less than 10% volume discrepancy can help traffic engineers better estimate future traffic conditions thus better design traffic control and management strategies. This is also consistent with Paz and Peeta's results (Paz and Peeta 2009a), namely considering drivers' possible reaction in route guidance can reduce the traffic condition estimation error.

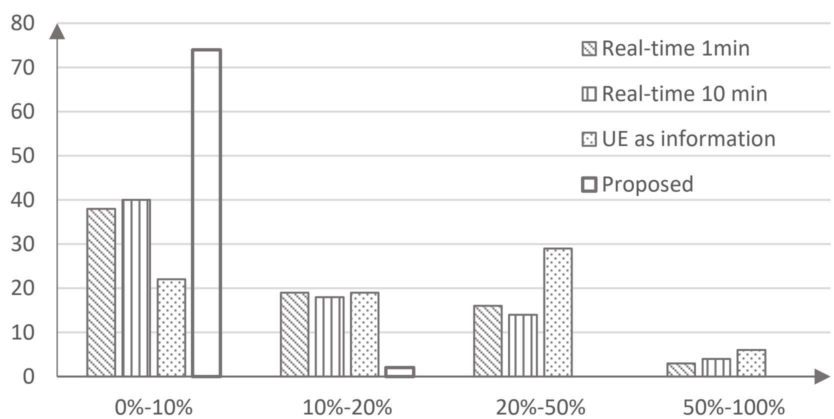


Figure 8 The Number of Links Having Different Volume Estimation Accuracy in Different Route Guidance Strategies

F. Performance Evaluation at Imperfect Market Penetration Rates (MPRs)

Evaluation flow chart

In reality, it takes time for a certain route guidance systems to reach a full MPR. When a full MPR is not reached, drivers who do not participate in the proposed route guidance either use current route guidance system or do not use route guidance services at all. Therefore, the interaction among drivers who use different routing strategies could generate certain traffic patterns. Whether the proposed route guidance system is going to bring benefits to both users and transportation system at imperfect market penetration rates is worthy of investigation.

To evaluate the performance of the proposed route guidance system at imperfect MPRs, it is assumed that there are three types of drivers on the road network, namely members, real-time-guidance users and habitual drivers.

- Members: the users of the proposed route guidance system. They receive route guidance information based on the predicted user optimum conditions.
- Real-time-guidance users: the drivers who use real-time-guidance systems. The operators of the real-time-guidance systems keep collecting traffic conditions on the network in real time and send the traffic information to drivers as well as the recommended route with shortest travel time.
- Habitual drivers: the drivers who do not use any route guidance services. They choose the route to go based on the historical traffic conditions according to their own preferences.

Therefore, there are two types of route guidance system in addition to three types of drivers with different routing strategies.

- User optimum-oriented guidance: its operators need to generate the route recommendations for the members. The operators know its members' estimated preference but do not know other drivers' in the network who use real time information and who habitually use one route. Therefore, the operators assume real-time users always follow the shortest travel time routes based on their estimate real-time traffic conditions and the habitual users always choose the shortest distance routes. Then, operators can generate the user optimum conditions with assuming that different groups of drivers have different behavior rules and send the user optimum conditions as information for guidance.
- Real time information guidance: its operators keep monitoring the traffic conditions on the road network and send the real time information to drivers.

The interaction among three types of drivers and two types of route guidance systems can be shown with the framework in Figure 9.

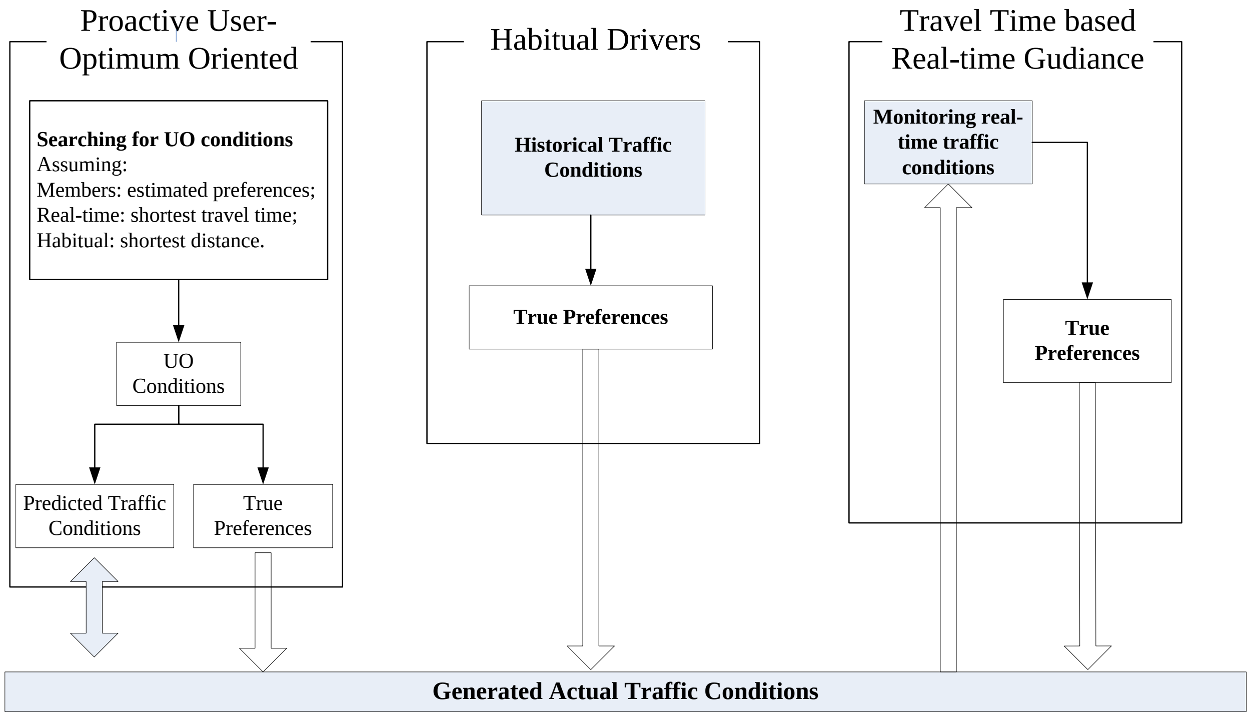


Figure 9 The Interaction among Three Types of Drivers and Two Types of Route Guidance System

Based on a survey conducted by Amirgholy et al. (2017), drivers use route guidance system in 36% of their trips. Therefore, among the drivers who are not members, the ratio between real-time-guidance users and habitual drivers is 40% versus 60%. When the market penetration rate of the proactive user-optimum oriented route guidance system is *p*, the market share of the real-time-guidance users is (1-*p*)*0.4 and the market share of the habitual drivers is (1-*p*)*0.6. The market penetration rate of the proposed route guidance system is gradually increased with 10% of incremental adjustment from 0% to 100%. Agents belonging to different types of users follow different behavior rules. Their route choice decisions interact with each other and generate different traffic patterns under different market penetration rates.

In order to better evaluate the performance of the proposed route guidance system, a route guidance strategies using travel time based user equilibrium is implemented at different market penetration rates as well. Using travel time based user equilibrium conditions as route guidance information is commonly used in the literature and has been proved to have good system performance (Deflorio 2003; Giglio and Sacco 2014). Its performance was compared with the proposed route guidance. In the UE condition route guidance system, the operator of the route guidance system assumes all drivers in the system only consider travel time as the only component of travel cost. Then, travel time based user equilibrium conditions can be found and provided to the participated drivers.

The performance of different route guidance strategies are again evaluated in three aspects including users' satisfaction, system performance and the capability to accurately estimate traffic conditions. The performance of two types of route guidance strategies are discussed in each of three aspects below.

Users' satisfaction

Users' satisfaction is measured as the number of drivers whose preferred route match the recommended routes. The evaluation here uses the amount of unsatisfied drivers as measurement here. Table 5 summarizes the number and the percentage of unsatisfied drivers when different route guidance system is under investigation. As explained in the evaluation follow chart Figure 9, travel time based real-time guidance system coexists in the system in addition to the route guidance system under investigation (either the proposed one or the one using UE conditions as information). The number of unsatisfied drivers is calculated as the summation of all unsatisfied drivers from both route guidance systems, namely the real-time guidance and system under investigation. As shown in Table 4, when the MPR of the route guidance system under investigation increases, the number of drivers using guidance systems also changes. When the proposed proactive user-optimum oriented system is under investigation, the actual number of unsatisfied users increases because the total number of guided users increases, but the actual percentage of unsatisfied drivers deceases from 19.4% to 11.4%. That means the proposed route guidance system can increase users' satisfaction when its MPR gradually increases. On the other hand, when the route guidance system using UE condition as information is under investigation, the percentage of unsatisfied driver increases as its MPR increases. That means the route guidance system actually deteriorate users' satisfaction. Therefore, the proposed route guidance system is expected to bring benefits to users even at imperfect market penetration rates.

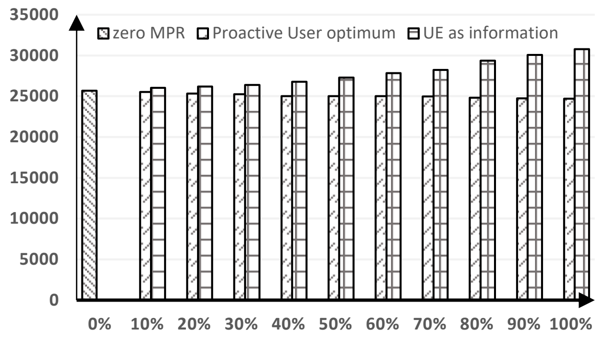
Table 5 Users' Satisfaction when different routing strategies are at imperfect market penetration rates

Market share of system under investigation	Number of guidance users	Proactive user-optimum oriented		System using UE conditions as information	
		Unsatisfied users	%	Unsatisfied Users	%
10%	24,861	4,829	19.4%	5,994	24.1%
20%	28,103	4,982	17.7%	7,421	26.4%
30%	31,346	5,158	16.5%	8,777	28.0%
40%	34,589	5,227	15.1%	10,068	29.1%
50%	37,832	5,347	14.1%	11,369	30.1%
60%	41,074	5,543	13.5%	12,754	31.1%
70%	44,317	5,683	12.8%	14,070	31.7%
80%	47,560	5,960	12.5%	15,459	32.5%
90%	50,802	6,020	11.8%	16,774	33.0%
100%	54,045	6,137	11.4%	18,082	33.5%

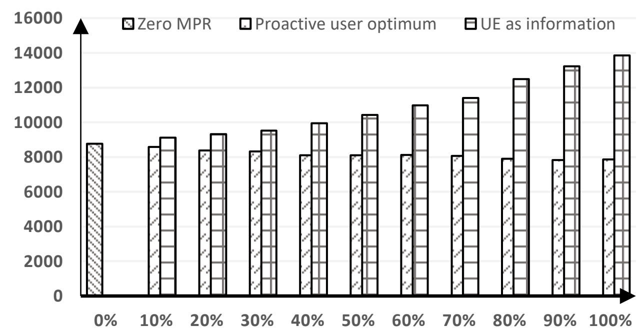
System mobility and sustainability

The system total travel time is calculated as the summation of all agents' travel time. The system total delay is measured as number of vehicles times the delayed travel time. It is calculated by summing the multiplication of link delay and the link volume on all links. The total travel time and total system delay under different route guidance routing strategies at various MPRs are shown in Figure 10.

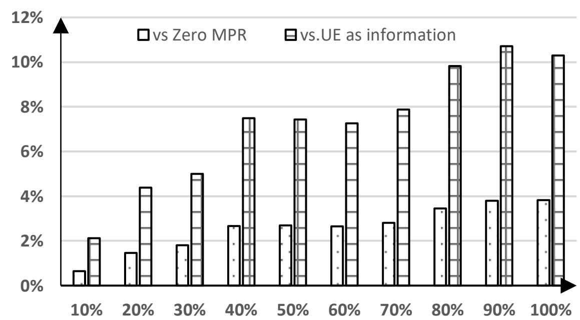
As shown in Figure 10 (a) and (b), compare to the zero MPR case, gradually increasing the MPR of proactive user optimum route guidance system can decrease the system total travel time and total delay. On the contrary, increasing MPR of the route guidance system using UE conditions as information generates more system total travel time and total delay. To have a better understanding of the magnitude of total travel time and delay changes, Figure 10 (c) and (d) show the percentage reduction of using the proposed route guidance system when compared to the zero MPR scenario and scenarios of using UE conditions as information at various MPRs. As shown in Figure 10 (c), when the MPR of the proposed route guidance system increases from 10% to 100%, the total travel time can be reduced from 0.65% to 3.82% compared to the zero MPR case. The reduction of total delay can reach 2.12% to 10.29% when the MPR increases from 10% of 100%. When compared to the scenario of using UE conditions as information, the reduction percentages are even larger. As shown in Figure 10 (c) and (d), the proposed route guidance system has 1.96% less total travel time and 5.96% less total delay than the route guidance system using UE conditions as information at the MPR of 10%. As the MPR increases, the total travel time and total delay can be 19.64% and 43.28% less. Therefore, the proposed proactive user-optimum oriented route guidance system can reduce the total travel time and total delay even at imperfect MPRs. The reduction amount increases as the MPR increases.



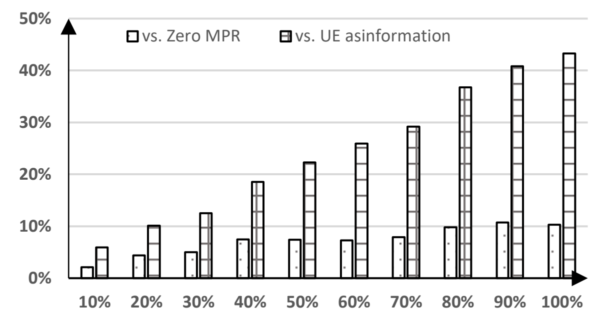
(a) Total travel time at various MPRs (hours)



(b) Total delay at various MPRs (vehicle*hours)



(c) Total travel time reduction (%)



(d) Total delay reduction (%)

Figure 10 The Mobility Impacts of the Proposed Route Guidance and UE as Information Route Guidance at Different MPRs

In addition to the mobility benefits, the sustainability benefits are also evaluated in terms of the energy consumption and the emission. DTALite combines the MOVES Lite and generates the vehicles' energy consumption and by converting mesoscopic vehicles' movement into vehicle trajectory second by second (Zhou et al. 2015). The emission rates adopted in this research were the default rates came with DTAlite and can also be found here (https://github.com/xzhou99/dtalite_software_release). Table 6 summarizes the energy consumption and the emissions of CO2, NOx, CO and HC. From Table 6, the sustainability impacts of two different routing strategies show similar trend as their mobility impacts. Compared to the zero MPR case, the proposed route guidance system gradually decreases the energy consumption and all kinds of emissions as MPR increases. As to the route guidance system using UE conditions as information, the system performance in terms of energy consumption and emission amount get worse as its MPR increases. Therefore, the results show that the proposed proactive user-optimum oriented route guidance can also have suitability improvement to the transportation system even at imperfect MPRs and the benefits increases as the MPR increases.

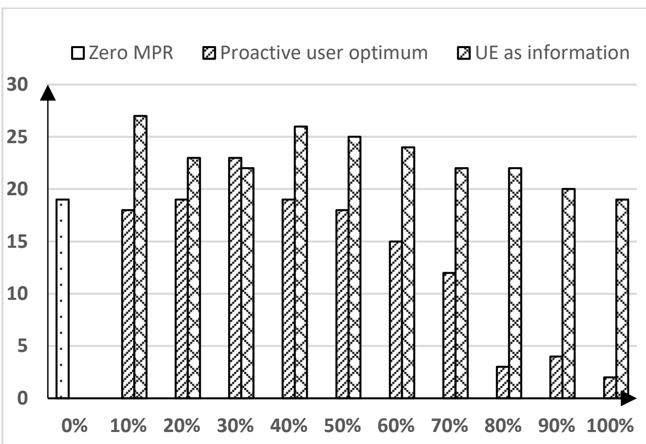
Table 6 The Sustainability Impact Comparison between the Proposed Route Guidance and Using UE as Information-route Guidance

Member	Energy (KJ) 10^9		CO2 (ton)		NOX (kg)		CO (kg)		HC (kg)	
0%	2.72		195		143		1550		49	
	Proposed	UE as info	Proposed	UE as infor	Proposed	UE as infor	Proposed	UE as infor	Proposed	UE as infor
10%	2.71	2.74	194.5	196.7	142.5	143.4	1543.8	1556.5	48.9	49.6
20%	2.69	2.75	193.6	197.4	142.4	143.7	1539.6	1560.4	48.6	49.8
30%	2.69	2.76	193.3	198.2	142.3	143.9	1537.9	1564.1	48.4	50.0
40%	2.68	2.78	192.5	199.7	142.1	144.3	1535.5	1570.5	48.2	50.5
50%	2.68	2.81	192.5	201.7	141.9	145.0	1535.8	1581.8	48.1	51.1
60%	2.68	2.84	192.5	204.0	142.0	145.7	1534.8	1592.2	48.1	51.8
70%	2.67	2.86	192.0	205.5	141.8	146.3	1531.9	1600.9	48.1	52.4
80%	2.66	2.91	191.3	209.1	141.6	147.2	1528.1	1613.3	47.9	53.8
90%	2.66	2.94	190.9	211.5	141.4	147.7	1525.7	1623.1	47.7	54.7
100%	2.65	2.97	190.8	213.5	141.3	148.0	1525.3	1628.2	47.8	55.4

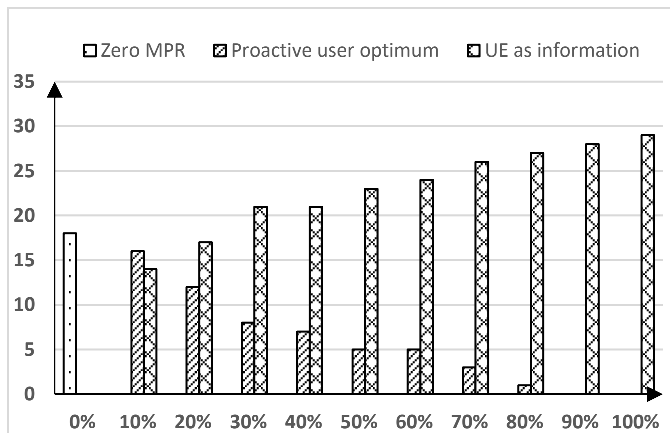
Future traffic conditions estimation

To measure the capability of estimating future traffic conditions, the link volume discrepancy between predicted traffic conditions and actual traffic conditions is used.

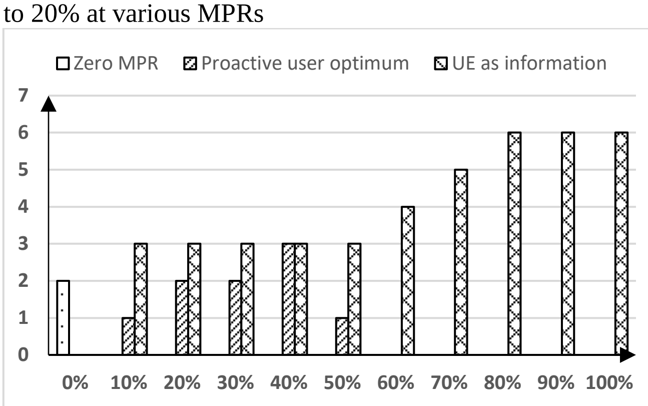
The numbers of links having different levels of volume discrepancy are summarized in Figure 11. Zero MPR case represents the scenario that there are only real-time-guidance users and habitual drivers in the road network. The other cases include either certain MPR of the proposed route guidance system or the route guidance system using UE conditions as information. In the Sioux Falls network, there are 76 links in total. In the zero MPR case, there are 19 links on which the volume discrepancy is within [10%, 20%], 18 links on which the discrepancy is within [20%, 50%], 2 links between [50%, 100%] and 1 link between [100%, 200%]. When the MPR of the proposed route guidance system increases from 10% to 100%, the number of links having different levels of discrepancy is largely reduced. As shown in Figure 11, the number of links having volume discrepancy between [100%, 200%] becomes zero at MPR of 40% (as shown in (d)), the number of links having discrepancy between [50%, 100%] becomes zero at MPR of 50% (as shown in (c)). When the MPR increases to 80%, the number of links having discrepancy [20%, 50%] becomes zero. When the MPR reaches 100%, there are only less than 5 links having volume discrepancy between [10%, 20%]. Therefore, as the MPR increases, the proposed route guidance system has more accurate prediction of link volume. With 100% of MPRs, 74 out of 76 links having link volume discrepancy less than 10%. On the other hand, the route guidance system using UE conditions as information generally makes the estimation less accurate as its MPR increases. As shown in Figure 11, the number of links having volume discrepancy in [100%, 200%] becomes zero when the MPR reaches 80%. In other discrepancy levels, the numbers of links having discrepancy in ranges of [20%, 50%] and [50%, 100%] are higher than those of zero MPR cases and actually increases as the MPR increases. In the discrepancy range of [10%, 20%], the number of links having this level of discrepancy increases first and then decreases. Generally, the route guidance system using UE conditions as information does not help with the traffic condition estimation.



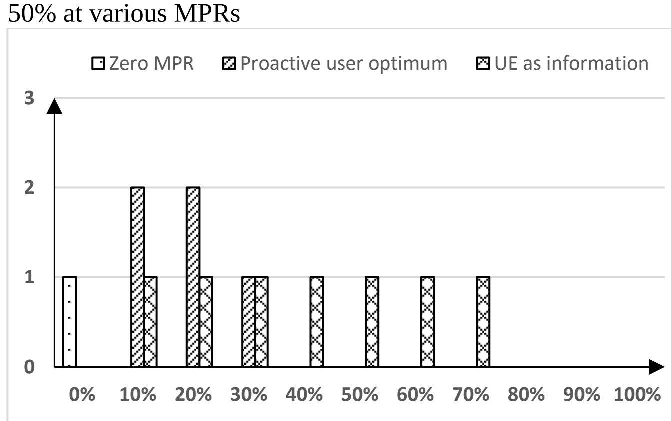
(a) Number of links having volume differences within 10% to 20% at various MPRs



(b) Number of links having volume differences within 20% to 50% at various MPRs



(c) Number of links having volume differences within 50% to 100% at various MPRs



(d) Number of links having volume differences within 100% to 200% at various MPRs

Figure 11 The Number of Links Having Different Levels of Volume Estimation Error at Different MPRs When Using Different Routing Strategies

The evaluation results in this section showed that the proposed proactive user optimum-oriented route guidance system have good performance in aspects of users' satisfaction, system performance in terms of mobility and sustainability, and the capabilities of accurately predict future traffic conditions. When not all drivers participate in the proposed route guidance system, it still has performance improvement at imperfect MPRs.

IV. CONCLUSIONS

The route guidance system has been an important component of ATIS and has been recognized as one of the most effective ways to mitigate congestion. This paper proposed a proactive user optimum-oriented route guidance system which contains two major components: the machine learning method based individual route choice models and the traffic assignment process that incorporates individuals' preferences in the process of searching for user optimum. The user optimum conditions are then used to generate route recommendations. The proposed route guidance system was evaluated against existing routing strategies at both perfect and imperfect market penetration rates with a Sioux Falls network example. Their performances were compared with each other in users' satisfaction, system mobility and sustainability, and future traffic condition estimation.

The evaluation results show that the proposed route guidance system has superiors performance in all three evaluated aspects (as shown in Table 3, Table 4 and Figure 8), when compared with travel time based real-time route guidance with different information updated intervals, and using travel time based user equilibrium condition as guiding information. When the proposed route guidance system was evaluated at imperfect market penetrates, the performance gradually increases as the market penetrate increase (as shown in Table 5 & 6, and Figure 10 &11). The advantageous performance is especially significant when compared to implementing the routing strategy using UE conditions as information.

By incorporating individual drivers' route choice preference when designing route guidance strategy, users' satisfaction can be improved. System mobility and sustainability can also be improved as long as drivers cares efficiency indicators when they make route choice decisions. In addition, by knowing the possible route selections of drivers, future traffic conditions can be estimated with a decent prediction accuracy so that traffic control and management strategies can be implemented in advanced. Therefore, the proposed route guidance system demonstrate its capability in system performance and users' satisfaction improvement. In future research, the factors that are going to influence the performance of the proposed route guidance system are going to be investigated, such as prediction accuracy of the individual route choice models and preference distribution among demand.

V. REFERENCES

A. Small, Kenneth. 2012. “Valuation of Travel Time.” *Economics of Transportation* 1 (December): 2–14. https://doi.org/10.1016/j.ecotra.2012.09.002.

Abdić, Irman, Lex Fridman, Daniel McDuff, Erik Marchi, Bryan Reimer, and Björn Schuller. 2016. “Driver Frustration Detection from Audio and Video in the Wild.” In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 1354–1360. IJCAI’16. New York, New York, USA: AAAI Press. http://dl.acm.org/citation.cfm?id=3060621.3060809.

Adler, Jeffrey L, and Victor J Blue. 2002. “A Cooperative Multi-Agent Transportation Management and Route Guidance System.” *Transportation Research Part C: Emerging Technologies* 10 (5–6): 433–54. https://doi.org/10.1016/S0968-090X(02)00030-X.

Adler, Jeffrey L., Goutam Satapathy, Vikram Manikonda, Betty Bowles, and Victor J. Blue. 2005. “A Multi-Agent Approach to Cooperative Traffic Management and Route Guidance.” *Transportation Research Part B: Methodological* 39 (4): 297–318. https://doi.org/10.1016/j.trb.2004.03.005.

Amirgholy, Mahyar, Nima Golshani, Craig Schneider, Eric J. Gonzales, and H. Oliver Gao. 2017. “An Advanced Traveler Navigation System Adapted to Route Choice Preferences of the Individual Users.” *International Journal of Transportation Science and Technology*, Special Issue on Urban Spatiotemporal Behavior and Network Assignment, 6 (4): 240–54. https://doi.org/10.1016/j.ijtst.2017.10.001.

Angelelli, E., I. Arsik, V. Morandi, M. Savelsbergh, and M. G. Speranza. 2016. “Proactive Route Guidance to Avoid Congestion.” *Transportation Research Part B: Methodological* 94 (December): 1–21. https://doi.org/10.1016/j.trb.2016.08.015.

Arokhlo, Mortaza Zolfpour, Ali Selamat, Siti Zaiton Mohd Hashim, and Md Hafiz Selamat. 2011. “Route Guidance System Based on Self Adaptive Multiagent Algorithm.” In *Computational Collective Intelligence. Technologies and Applications*, 90–99. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23938-0_10.

Ben-Hur, Asa, and Jason Weston. 2010. “A User’s Guide to Support Vector Machines.” In *Data Mining Techniques for the Life Sciences*, 223–39. Methods in Molecular Biology. Humana Press. https://doi.org/10.1007/978-1-60327-241-4_13.

Carey, Malachy, and Y. E. Ge. 2012. “Comparison of Methods for Path Flow Reassignment for Dynamic User Equilibrium.” *Networks and Spatial Economics* 12 (3): 337–76. https://doi.org/10.1007/s11067-011-9159-6.

Chakirov, Artem. 2014. “Enriched Sioux Falls Scenario with Dynamic and Disaggregate Demand.” *Arbeitsberichte Verkehrs- Und Raumplanung* 978. http://e-citations.ethbib.ethz.ch/view/pub:127267.

Chiu, Yi-Chang, Jon Bottom, Michael Mahut, Alex Paz, Ramachandran Balakrishna, Travis Waller, and Jim Hicks. 2011. “Dynamic Traffic Assignment: A Primer.” *Transportation Research E-Circular*, no. E-C153 (June). https://trid.trb.org/view.aspx?id=1112932.

Churchland, Anne K., Roozbeh Kiani, and Michael N. Shadlen. 2008. “Decision-Making with Multiple Alternatives.” *Nature Neuroscience* 11 (6): 693–702. https://doi.org/10.1038/nn.2123.

Deflorio, Francesco Paolo. 2003. “Evaluation of a Reactive Dynamic Route Guidance Strategy.” *Transportation Research Part C: Emerging Technologies*, World Congress on Intelligent Transport Systems, 11 (5): 375–88. https://doi.org/10.1016/S0968-090X(03)00031-7.

Dia, Hussein. 2002. “An Agent-Based Approach to Modelling Driver Route Choice Behaviour under the Influence of Real-Time Information.” *Transportation Research Part C: Emerging Technologies* 10 (5): 331–49. https://doi.org/10.1016/S0968-090X(02)00025-6.

Dong, Jing, Hani Mahmassani, and Chung-Cheng Lu. 2006. “How Reliable Is This Route?: Predictive Travel Time and Reliability for Anticipatory Traveler Information Systems.” *Transportation Research Record: Journal of the Transportation Research Board* 1980 (January): 117–25. https://doi.org/10.3141/1980-17.

Du, Lili, Lanshan Han, and Shuwei Chen. 2015. “Coordinated Online In-Vehicle Routing Balancing User Optimality and System Optimality through Information Perturbation.” *Transportation Research Part B: Methodological* 79 (September): 121–33. https://doi.org/10.1016/j.trb.2015.05.020.

El-Sayed, Hesham, Gokulnath Thandavarayan, and Yasser Hawas. 2017. “A Cost Effective Route Guidance Method for Urban Areas Using Histograms.” Research article. Wireless Communications and Mobile Computing. 2017. https://doi.org/10.1155/2017/4507352.

Feng, Tao, Theo Arentze, and H.J.P. Timmermans. 2013. “Capturing Preference Heterogeneity of Truck Drivers’ Route Choice Behavior with Context Effects Using a Latent Class Model.” *European Journal of Transport and Infrastructure Research* 13 (August).

Ferris, Michael C., and Andrzej Ruszczyński. 2000. “Robust Path Choice in Networks with Failures.” *Networks* 35 (3): 181–94. https://doi.org/10.1002/(SICI)1097-0037(200005)35:3<181::AID-NET2>3.0.CO;2-Y.

Fridman, Lex, Bryan Reimer, Bruce Mehler, and William T. Freeman. 2018. “Cognitive Load Estimation in the Wild.” In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 652:1–652:9. CHI ’18. New York, NY, USA: ACM. https://doi.org/10.1145/3173574.3174226.

Giglio, Davide, and Nicola Sacco. 2014. “A Control System for the Individual Route Guidance in Traffic Flow Networks.” *IFAC Proceedings Volumes*, 19th IFAC World Congress, 47 (3): 948–56. https://doi.org/10.3182/20140824-6-ZA-1003.02303.

Gong, Lin, Mohammad Al Boni, and Hongning Wang. 2016. “Modeling Social Norms Evolution for Personalized Sentiment Classification.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1:855–865. http://www.cs.virginia.edu/~lg5bt/files/acl2016.pdf.

Han, Linghui, Huijun Sun, David Z. W. Wang, Chengjuan Zhu, and Jianjun Wu. 2016. “The Combination of Continuous Network Design and Route Guidance.” *Computers & Operations Research* 73 (September): 92–103. https://doi.org/10.1016/j.cor.2016.03.012.

Hawas, Yaser E., and Hesham El-Sayed. 2015. “Autonomous Real Time Route Guidance in Inter-Vehicular Communication Urban Networks.” *Vehicular Communications* 2 (1): 36–46. https://doi.org/10.1016/j.vehcom.2015.01.001.

He, Zhengbing, Wei Guan, and Shoufeng Ma. 2013. “A Traffic-Condition-Based Route Guidance Strategy for a Single Destination Road Network.” *Transportation Research Part C: Emerging Technologies* 32 (July): 89–102. https://doi.org/10.1016/j.trc.2013.04.005.

Hsu, Chih-Wei, and Chih-Jen Lin. 2002. “A Comparison of Methods for Multiclass Support Vector Machines.” *IEEE Transactions on Neural Networks* 13 (2): 415–25. https://doi.org/10.1109/72.991427.

Jahn, Olaf, Rolf H. Möhring, Andreas S. Schulz, and Nicolás E. Stier-Moses. 2005. “System-Optimal Routing of Traffic Flows with User Constraints in Networks with Congestion.” *Operations Research* 53 (4): 600–616. https://doi.org/10.1287/opre.1040.0197.

Khaled, Hamad, Faghri Ardeshir, and Nanda Raman. 2003. “A Behavioral Component Analysis of Route Guidance Systems Using Neural Networks.” *Computer-Aided Civil and Infrastructure Engineering* 18 (6): 440–53. https://doi.org/10.1111/1467-8667.00329.

Lafortune, Stéphane, Raja Sengupta, David E. Kaufman, and Robert L. Smith. 1993. “Dynamic System-Optimal Traffic Assignment Using a State Space Model.” *Transportation Research Part B: Methodological* 27 (6): 451–72. https://doi.org/10.1016/0191-2615(93)90017-5.

LEBLANC, LARRY J. 1975. “An Algorithm for the Discrete Network Design Problem.” *Transportation Science* 9 (3): 183–99.

Liang, Zilu, and Yasushi Wakahara. 2014. “Real-Time Urban Traffic Amount Prediction Models for Dynamic Route Guidance Systems.” *EURASIP Journal on Wireless Communications and Networking* 2014 (May): 85. https://doi.org/10.1186/1687-1499-2014-85.

Liu, Henry X., Xiaozheng He, and Will Recker. 2007. “Estimation of the Time-Dependency of Values of Travel Time and Its Reliability from Loop Detector Data.” *Transportation Research Part B: Methodological* 41 (4): 448–61. https://doi.org/10.1016/j.trb.2006.07.002.

Nadi, S., and M. R. Delavar. 2011. “Multi-Criteria, Personalized Route Planning Using Quantifier-Guided Ordered Weighted Averaging Operators.” *International Journal of Applied Earth Observation and Geoinformation* 13 (3): 322–35. https://doi.org/10.1016/j.jag.2011.01.003.

Nagel, Kai, and Gunnar Flötteröd. 2009. “Agent-Based Traffic Assignment: Going from Trips to Behavioral Travelers.” In . https://infoscience.epfl.ch/record/152368?of=HB.

Pahlavani, Parham, and Mahmoud R. Delavar. 2014. “Multi-Criteria Route Planning Based on a Driver’s Preferences in Multi-Criteria Route Selection.” *Transportation Research Part C: Emerging Technologies* 40 (March): 14–35. https://doi.org/10.1016/j.trc.2014.01.001.

Park, K., M. Bell, I. Kaparias, and K. Bogenberger. 2007. “Learning User Preferences of Route Choice Behaviour for Adaptive Route Guidance.” *IET Intelligent Transport Systems* 1 (2): 159–66. https://doi.org/10.1049/iet-its:20060074.

Parthasarathi, Pavithra, David Levinson, and Hartwig Hochmair. 2013. “Network Structure and Travel Time Perception.” *PLOS ONE* 8 (10): e77718. https://doi.org/10.1371/journal.pone.0077718.

Paz, Alexander, and Srinivas Peeta. 2009a. “On-Line Calibration of Behavior Parameters for Behavior-Consistent Route Guidance.” *Transportation Research Part B: Methodological* 43 (4): 403–21. https://doi.org/10.1016/j.trb.2008.07.007.

———. 2009b. “Behavior-Consistent Real-Time Traffic Routing under Information Provision.” *Transportation Research Part C: Emerging Technologies* 17 (6): 642–61. https://doi.org/10.1016/j.trc.2009.05.006.

Pfeiffer, Jella. 2012. “Fundamentals on Decision-Making Behavior.” In *Interactive Decision Aids in E-Commerce*, 15–45. Contributions to Management Science. Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-2769-9_2.

Prato, Carlo, and Shlomo Bekhor. 2006. “Applying Branch-and-Bound Technique to Route Choice Set Generation.” *Transportation Research Record: Journal of the Transportation Research Board* 1985 (January): 19–28. https://doi.org/10.3141/1985-03.

Prato, Carlo Giacomo. 2009. “Route Choice Modeling: Past, Present and Future Research Directions.” *Journal of Choice Modelling* 2 (1): 65–100. https://doi.org/10.1016/S1755-5345(13)70005-8.

Rogers, Seth, and Pat Langley. 1998. “Personalized Driving Route Recommendations.” *ResearchGate*, July. https://www.researchgate.net/publication/2741171_Personalized_Driving_Route_Recommendations.

Roughgarden, T., and E. Tardos. 2000. “How Bad Is Selfish Routing?” In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, 93–102. https://doi.org/10.1109/SFCS.2000.892069.

Schofer, Joseph, Frank Koppelman, and William Charlton. 1997. “Perspectives on Driver Preferences for Dynamic Route Guidance Systems.” *Transportation Research Record: Journal of the Transportation Research Board* 1588 (January): 26–31. https://doi.org/10.3141/1588-04.

Sheffi, Yosef. 1985. *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*. Englewood Cliffs, N.J: Prentice Hall.

Steinwart, Ingo, and Andreas Christmann. 2008. *Support Vector Machines*. Springer Science & Business Media.

Tawfik, Aly M., Hesham A. Rakha, and Shadeequa D. Miller. 2010. “Driver Route Choice Behavior: Experiences, Perceptions, and Choices.” In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, 1195–1200. IEEE.

Tawfik, Aly, and Hesham Rakha. 2013. “Latent Class Choice Model of Heterogeneous Drivers’ Route Choice Behavior Based on Learning in a Real-World Experiment.” *Transportation Research Record: Journal of the Transportation Research Board* 2334 (December): 84–94. https://doi.org/10.3141/2334-09.

Wahle, J., O. Annen, Ch. Schuster, L. Neubert, and M. Schreckenberg. 2001. “A Dynamic Route Guidance System Based on Real Traffic Data.” *European Journal of Operational Research*, Artificial Intelligence on Transportation Systems and Science, 131 (2): 302–8. https://doi.org/10.1016/S0377-2217(00)00130-2.

Wardrop, J G. 1952. “Road Paper. Some Theoretical Aspects of Road Traffic Research.” *Proceedings of the Institution of Civil Engineers* 1 (3): 325–62. https://doi.org/10.1680/ipeds.1952.11259.

Xiong, Chenfeng, Xiqun Chen, Xiang He, Xi Lin, and Lei Zhang. 2016. “Agent-Based En-Route Diversion: Dynamic Behavioral Responses and Network Performance Represented by Macroscopic Fundamental Diagrams.” *Transportation Research Part C: Emerging Technologies* 64 (March): 148–63. https://doi.org/10.1016/j.trc.2015.04.008.

Yang, Qi, Haris Koutsopoulos, and Moshe Ben-Akiva. 2000. “Simulation Laboratory for Evaluating Dynamic Traffic Management Systems.” *Transportation Research Record: Journal of the Transportation Research Board* 1710 (January): 122–30. https://doi.org/10.3141/1710-14.

Zhou, Xuesong, Shams Tanvir, Hao Lei, Jeffrey Taylor, Bin Liu, Nagui M. Rouphail, and H. Christopher Frey. 2015. “Integrating a Simplified Emission Estimation Model and Mesoscopic Dynamic Traffic Simulator to Efficiently Evaluate Emission Impacts of Traffic Management Strategies." *Transportation Research Part D: Transport and Environment* 37 (June): 123–36. https://doi.org/10.1016/j.trd.2015.04.013.

Zhou, Xuesong, and Jeffrey Taylor. 2014. "DTALite: A Queue-Based Mesoscopic Traffic Simulator for Fast Model Evaluation and Calibration." *Cogent Engineering* 1 (1): 961345. https://doi.org/10.1080/23311916.2014.961345.

CHAPTER 5. DISCUSSION

5.1 Further Extension of the Proposed Route Guidance System

The proposed proactive user optimum-oriented route guidance system demonstrated its capabilities in improving system performance, users' satisfaction and traffic conditions estimations. As already discussed in Chapter 4 Section I A, some researchers (Han et al. 2016; Roughgarden and Tardos 2000) found that the route guidance based on user optimum conditions do not have as good system performance as the route guidance system based on system optimum conditions. However, the natures of the proposed route guidance system make it possible to be extended so that the system performance can be further improved on the basis of user optimum conditions. This section discusses how these natures can help further extend the proposed route guidance system and demonstrates the potential of the proposed route guidance system in system performance improvement.

Personalized incentives scheme

One of the important characteristics that the proposed route guidance system has is the knowledge of the individual drivers' route choice preferences. Knowing a single driver's preference means that the route guidance system operators are able to motivate the driver to make certain decisions that operators would like him/her to choose. The motivation can be made towards different goals such as improving system performance goal, avoiding certain areas of the network for special events, and etc.

Using economic methods to motivate drivers' behaviors has been used in both theoretical discussion and practice, such as congestion pricing and public transit benefits (Luten 2004). Different amounts of financial penalty or benefits are provided to the drivers to push drivers away from or pull drivers back to the behavior decisions that the system desires them to make. London implemented cordon congestion charging since 2003 (Bhatt, Higgins, and Berg 2008). All the vehicles entering the center area of London metropolitan need to pay a fixed amount charge. The amount of charge is typically the same for all drivers except some various price for certain types of vehicles or drivers (for example, vehicles with renewable energy and residents who live inside of the charging area can pay less) (Luten 2004). Cities of Stockholm and Singapore implemented similar kinds of pricing schemes (Bhatt, Higgins, and Berg 2008). The pricing amount is determined by the possible responses of the driver population. For example, estimate drivers' responses based on demand elasticity in order to achieve desired level of services, or analyze delays and drivers' value of time to find a cut-off charging level that can make certain amount of drivers change to other travel options (Bank 2015). However, asking for the same amount of charge from all drivers is recognized as not efficient from the perspective of microeconomic theory (Button 2004). As explained in (Button 2004), the effects of congestion and pricing can vary by person. Take the value of time as an example, when the price is determined as a cut-off value from the value of time distribution, drivers with higher value of time are not going to change their travel behaviors because the price is not high enough to make them change. On the other hand, drivers with lower value of travel time make changes but the price is actually much higher than the price that can make drivers change travel behavior. The pricing scheme with unified price is called second-best pricing and is recognized as inefficient (Button 2004; Bank 2015). On the other hand, it is more ideal to charge each drivers with the exact required amount that can change his/her behavior decisions. Such pricing scheme with individual-based price is called first-best pricing (Button 2004). It is recognized that first-best pricing is more efficient than second-best pricing from the perspective of microeconomic. However, because of the incapability to know individual driver's preference, only second-best pricing is widely implemented in practice, as in London and Stockholm.

With the proposed route guidance system, individual drivers' route choice preferences are available. Therefore, the exact required amount of either pricing or incentives for changing a driver's route choice decisions can be calculated. An example of providing personalized incentives for improving system performance is used here to demonstrate the potential of the proposed route guidance system. Providing incentives to make drivers change to system desired routes works in a similar way as congestion pricing. Assuming the funding used to provide the incentives is either from the system operators or transportation authorities, the required funding amount is the less the better. In other words, we do not want to give more incentives than it is necessary to make a driver change travel behavior decisions, and also still want to make sure the provided incentives are enough to make a driver change behavior decisions.

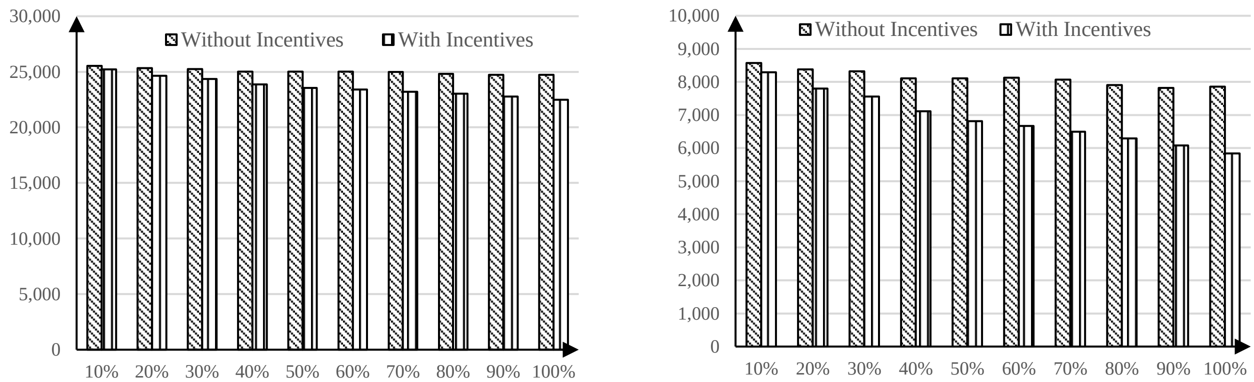
Depending on different goals, system-desired individuals' decisions can be different. Paz and Peeta (2009) explained that the ideal states of the system could be the system optimum conditions or user equilibrium conditions. In the evaluation process in Chapter 4 Section III E, the travel time based user equilibrium conditions have more efficient system performance in terms of total travel time. The reason that it cannot be reached in reality is because drivers do not comply with the route recommendations that are generated with simplified drivers' preference (i.e. only caring the route attribute of travel time). In this section, the travel time based user equilibrium conditions are set to be system-desired conditions. Drivers' route choices under the travel time based user equilibrium conditions are the system-desired individuals' choices. Each driver will be provided with a personalized incentive and motivated towards the system-desired choices.

Assuming $\boldsymbol{X_{ip}}$ and $\boldsymbol{X_{is}}$ represent the route attributes information of Driver *i*'s preferred route, *p*, and the system-desired route, *s*, when travel time based user equilibrium conditions are used to generate route recommendations. Route *p* is known because system operators can predict Driver *i*'s route choices with his/her route choice model $f(\boldsymbol{X})=\boldsymbol{\beta X}$. Therefore, the amount of the incentives can be determined with:

$$s_i = \frac{\beta_i(X_{ip} - X_{is})}{\beta_m} \qquad (1)$$

In which β_m is the weight of the route attribute that contains monetary value, for example, toll. Therefore, the difference in other aspects between two routes can be converted to monetary value through the monetary term in the route choice model. The calculated personalized incentive is added to or reduced from the monetary term in $\boldsymbol{X_{is}}$. With the updated $\boldsymbol{X_{is}}'$, iterative process is conducted again to find the new user optimum conditions. In the iterative process, s_i can be adjusted according to the difference between Route *s* and Route *p*. When the user optimum conditions are found, the amount of s_i is determined and provided to drivers. Driver *i*'s route choice decision would change to Route *s*. However, drivers may not change route choice decisions even they are provided with incentives because of the prediction accuracy of individual route choice models. Therefore, in the evaluation process, the travel time based user equilibrium conditions together with personalized incentives are fed into drivers' true preferences. The generated traffic conditions are used to evaluate the performance of the personalized incentive scheme.

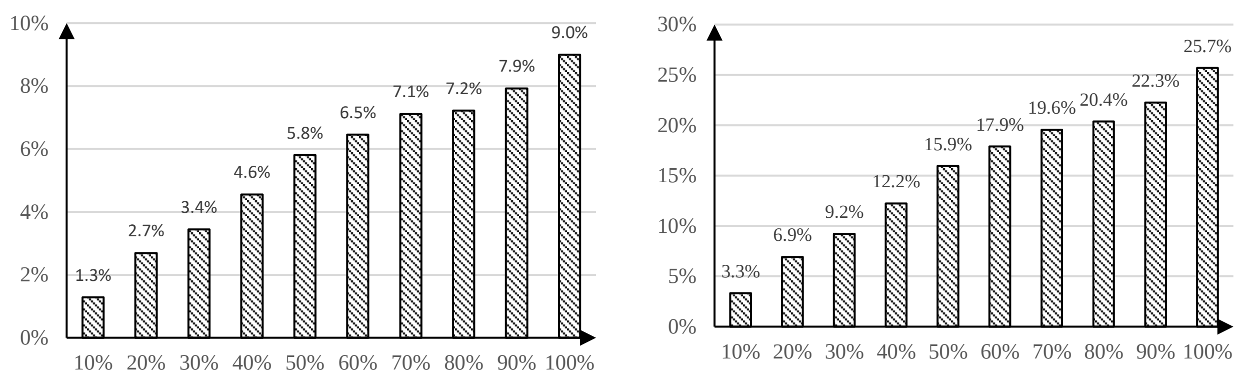
The same Sioux Falls network and demand settings were used to evaluate the personalized incentive scheme. The weight of "Fuel Cost" is used as β_m term in Equation (1). It should be noted that fuel cost is typically very small for a trip in urban areas. The impacts of fuel cost on drivers' decisions are not as huge as the monetary term such as high way tolls. Because of the limitation of data availability, the fuel cost is used here, but more effective monetary term can be obtained in practice. The total travel time and total system delay are summarized in Table 1. Figure 1 (a) and (b) show the total travel time and total delay at different market penetration rates with and without personalized incentives. With the personalized incentive scheme based on the proposed route guidance system, both of total system travel time and system delay are further reduced when compared to only implementing the proactive route guidance system. There is a clear trend that the reduction amount increases as the market penetration rate increases. Figure 2 shows the reduction percentages of both system total travel time and delay. When market penetration rate increases from 10% to 100%, the total travel time reduction can be ranging from 1.3% to 9.0% and the total delay reduction can be ranging from 3.3% to 25.7%. Therefore, the potential of system performance improvement brought by the personalized incentive scheme is quite promising.



(a) System total travel time (unit: hours) at different MPRs with and without personalized incentives

(b) System total delay (unit: vehicle*hour) at different MPRs with and without personalized incentives

Figure 1 System Total Travel Time and Delay at Different MPRs with and without Personalize Incentives



(a) The total travel time reductions with personalized incentives at different MPRs

(b) The total delay reductions with personalized incentives at different MPRs

Figure 2 System Total Travel Time and Delay Reduction at Different MPRs with and without Personalize Incentives

This example demonstrated the system performance improvement of the personalized incentive scheme that is designed on the basis of the proposed route guidance system. However, a careful incentive scheme design should be made to be practical and cost-effective for implementation. In this demonstration, the required amount of total incentives increases from 1 million to 10 million when the number of participated drivers increases from 10% to 100%. For an hour of evaluation interval, the required amount of incentives may not be practical even though the system performance improvement is significant. The required amount of the total incentives could be determined by different factors, such as the accuracy of individual route choice models, drivers' preference characteristics, network conditions and etc. The impacts of these influencing factors should be carefully evaluated before implementation. On the other hand, how to find the most critical drivers who can bring largest system performance improvement so that minimum amount of incentive is required is also a question worthy of future investigation.

Volume based traffic control and management

The proposed proactive user-optimum oriented route guidance system has another advantage in terms of accurate traffic condition estimation, as demonstrated in Chapter 4 Section E and F. An accurate future volume prediction can help design volume based traffic control and management strategies. For example, designing traffic signal timing plans. Because of the consideration of drivers' possible reactions, system performance can benefit from the traffic signal timing plan that is designed based on the volume predicted by the proposed route guidance system.

Traffic volume is an important input for designing a signal timing plan. Proactive route guidance system can help estimate future traffic volumes, therefore, the signal timing plan can be designed to proactively accommodate the possible traffic arrival patterns. Though current actuated signal timing can detect the coming traffic patterns by installing loop detectors at the upstream of intersection approaches, it can only detect the immediate coming traffic. The installation and maintenance of loop detectors not only require monetary cost and labor efforts, but also could influence ongoing traffic and cause more delay. The proposed route guidance system estimate future volumes by predicting drivers' possible behavior choices and does not require additional efforts. Therefore, with relatively accurate volume estimations, signal timing plans can be better designed to accommodate the traffic patterns. In addition, signal timing plans can influence the traffic conditions which in turn influence the route attributes that drivers use for making route choice decisions. Similar as the personalized incentive scheme, signal timing plans can be made and adjusted in the iterative process to reach the system optimum conditions or any other system desired goals. In this way, the signal timing plan is considered as an influencing factor that can shape drivers' route choice decisions in addition to just better accommodate coming traffic patterns locally.

Automated vehicles (AVs)

The proposed route guidance system has the potential to be used as a routing system for automated vehicles. Automated vehicles have been believed to be the next generation of the transportation system. National Highway Traffic Safety Administration defined the vehicle automation to be 5 levels (NHTSA 2017). With automation level above 3, AVs passengers can set the destinations of their trips and the AVs can bring them to the destinations without passengers operating the vehicle or defining which route or turn should be taken.

In order to determine the route AVs take, AVs need routing algorithms. There are mainly two kinds of routing algorithms in existing literature for routing AVs. The first one is very similar to traditional vehicle routing, namely reactive and self-interested. One example is the work of Fiosins et al., (2011). They considered AVs' route planning and the trajectory optimization within road section together. At the route planning level, each individual vehicle uses stochastic shortest path algorithm to find their shortest path individually, without cooperation. The other kind of AV routing algorithm is, instead of finding best route for each AV independently, the routing for AVs can be optimized for achieving certain performance goal, such as minimizing passengers' travel time, mitigating the congestion caused by the rebalancing AVs when they are realigned with new passengers (Zhang, Rossi, and Pavone 2016), maximizing the number of vehicles reaching their destinations (Vitello et al. 2016), finding the shortest route for every vehicle by sharing all AVs' intended taking-route (Claes, Holvoet, and Weyns 2011). Similar with existing route guidance systems, existing routing algorithms for AVs also did not incorporate drivers' route choice preference heterogeneity. Therefore, for the purpose of improving both users' satisfaction and transportation system efficiency, AVs should have a routing system that can consider both users' preference and system performance. The proposed proactive user-optimum oriented route guidance system can be transferred to AVs routing system to learn users' route choice preferences and coordinate passengers to achieve better system performance.

5.2 Consideration for Practical Applications

The implementation of the proposed proactive user optimum-oriented route guidance system in practice may have some prerequisites. Assumptions were made about these prerequisites in the Chapter 4 Section II. For example, drivers are assumed to be willing to share their preference data and the time dependent OD demand is assumed to be available. These two assumptions represent the privacy and OD data issues in the practical applications and are discussed in this section.

Users' privacy

The proposed route guidance system needs to maintain the individual route choice model for each driver. The individual route choice model is developed from the driver's historical route choice preference data. The key route attributes, time of day, destination and other information need to be extracted when drivers are making trips. Sharing this information to the route guidance system may be a concern to some drivers. The privacy issue is not only a problem in transportation domain but also draws lot of attention in other industries especially in this big data era. Sharing personal data with the service providers for service quality improvement is not new at all. Google Maps collects users' trajectory data or speed data to estimate the traffic conditions for real-time travel information. Many online shopping websites such as Amazon tracks customers' purchasing history and recommends products to customers. There is always a trade-off between the privacy and the customized service.

For the specific application of the route guidance system here, drivers' privacy concerns can be possibly mitigated. For example, the route guidance system control center only needs all drivers' individual models while all the model estimation and updates can be done locally on user's end. In this way, there is no need to share drivers' personal historical trip information with the route guidance system. But, depending on the specific modeling approach adopted, certain level of model performance may be sacrificed. For example, the MT-LinAdapt approach may require all drivers' preference data to estimate both aggregate-level model and individual-level model. Even though, more advanced modeling approaches can be developed and explored to solve the privacy concern.

Obtain OD data

Dynamic OD matrix estimation is still an important research problem in the transportation network analysis (Chang and Wu 1994; Ashok and Ben-Akiva 1993; Toledo and Kolechkina 2013). Most researchers assume that OD demand is known in their analysis. Time-dependent OD data is the key input of the proposed route guidance system. The proposed route guidance system provides an additional approach to estimate the time-dependent OD demand. On the basis of the traditional methods of estimating OD data (such as traffic census data), the proposed route guidance system can infer drivers' OD demand in advance by combining drivers' other information. For example, users can connect the route guidance system with personal calendar such as Google Calenders or Facebook Events. The departure time and OD of the drivers' trips in next planning period can be obtained from the calendar. Also, some travel demand can be analyzed from drivers' trip history to see if there is any general patterns. When the proposed route guidance system is further extend to be used as an AVs' routing system, more data can be obtained because passengers need to input OD for all trips that they make with AVs and that means AVs passengers need to share more information with the AVs.

CHAPTER 6. CONCLUSIONS AND FUTURE RESEARCH

6.1 Major Content

This research proposed a proactive user optimum-oriented route guidance system that incorporates individual drivers' route choice preferences. In order to establish and evaluate the proposed route guidance system, this research was divided into three major parts and a discussion section.

- The first part explored the possibility of using the traditional mixed logit model together with the Bayes rule to estimate drivers' route choice at the individual level. A stated preference survey with binary route choice scenarios was conducted among 44 participants. The collected data was used to estimate a mixed logit model that considers the correlation among utility function coefficients. The established mixed logit model was evaluated against a regular multinomial logit model in aspects of both overall model fitness and individual's route choice decision prediction.
- The second part explored the performance of an advanced sentiment analysis approach, Multi-task Linear Model Adaptation in the application of the route guidance system that incorporates individual drivers' preferences. Additional two stated preference surveys were conducted among 58 participants. Therefore, three observed stated preference data as well as associated synthetic datasets that have more heterogeneous preferences were used to establish and evaluate the MT-LinAdapt model. The aggregate SVM model, the individual SVM model and the mixed logit model were included for comparison. The MT-LinAdapt was evaluated against three other models in two scenarios (a) drivers have adequate amount of historical preference data, and (b) drivers have limited amount of historical preference data.
- The third part established the framework of the proposed proactive user-optimum oriented route guidance system. An evaluation platform was set up with a decision module (Matlab) and a = traffic simulation module (DTAlite). The classical Sioux Falls network with time-dependent OD as well as individual drivers' preferences were used to demonstrate and evaluate the performance of the proposed route guidance system. The proposed route guidance system was firstly evaluated against two existing routing strategies, namely generating route recommendations according to travel time based UE conditions and real-time route guidance with update intervals of 1 and 10 minutes. Then, the performance of the proposed route guidance system was evaluated when its market penetration rate increases from 0% to 100%. Drivers who do not participate in the proposed route guidance system either follow the route recommendations

generated with travel time based real time guidance or use historical traffic conditions to make route choice decisions. The performance of the proposed route guidance system as well as other routing strategies were evaluated in the aspects of system mobility, system sustainability, users' satisfaction and capabilities in accurately estimating future link volumes.

- The Discussion section firstly discussed several further extended applications of the proposed route guidance system including a personalized incentive scheme, volume based traffic control and management strategies and automated vehicle routing systems. Then two practical issues that the proposed route guidance system may encounter in the implementation were also discussed.

6.2 Key Results

With the exploration and evaluation in this research, some key findings are summarized below.

- Because of considering preference heterogeneity, the mixed logit model can provide better overall model fitness (20% higher Pseudo-R^2) and more accurately predict individual drivers' route choice decisions (20% higher average prediction accuracy) than a regular multinomial logit model. It proved the necessity of considering drivers' heterogeneous route choice preference and also demonstrated that the mixed logit model can be used to estimate individual driver's route choice preference.
- MT-LinAdapt model estimates and updates users' preferences at both the aggregate and individual levels. Therefore, it not only captures the preference heterogeneity but also have a good estimation of individual drivers' preference with limited amount of preference data. When compared to the aggregate model, the individual model and the random parameter model (i.e., the mixed logit model), MT-LinAdapt achieved up to 8% higher prediction accuracy in the adequate data scenario and up to 18% higher prediction accuracy in the limited data scenario than the existing models.
- From the perspective of implementation, MT-LinAdapt also has the advantages in following aspects: (a) it works well not only on users with adequate amount of data but also new users who have limited amount of preference data; (b) it does not require sociodemographic (difficult to obtain because of privacy issue) or any other segmentation criteria to differentia drivers' heterogeneous preferences; (c) it can update individual drivers' route choice model in real time

as preference data accumulates; and (d) it allows drivers to include different route attributes that they care in the route choice models.

- The evaluation results showed that the proposed proactive user optimum-oriented route guidance system can bring much lower total travel time (up to 10% reduction), delay (up to 42% reduction), energy consumption and emissions (CO_2, NOX, CO and HC) when compared to the routing strategies of using travel time based UE condition to generate route recommendations and the real-time route guidance system with updating intervals of 1 and 10 minutes.
- The evaluation results showed that the proposed route guidance system can increase the number of drivers who are satisfied with the route recommendations. When compared to other routing strategies of using travel time based UE conditions to generate route recommendations and the real-time route guidance systems, the proposed route guidance system can reach up to 18% to 22% more satisfied users.
- The evaluation results also showed that the proposed route guidance system can more accurately estimate future traffic conditions in terms of link volume. The volume estimation of 74 out of 76 links in the Sioux Falls network have estimation errors less than 10%, while this number for travel time based UE conditions as guidance and real-time route guidance system are only 20 to 40, and there are more links having estimation errors higher than 10% in these two routing strategies.
- In addition, the evaluation results showed that the proposed route guidance system can still improve the system mobility, system sustainability, users' satisfaction and future traffic volume estimation accuracy even at imperfect market penetration rates. These benefits of the proposed route guidance system gradually increases as its market penetrate increases.
- The results of the personalized incentive scheme in the Chapter 5.1 demonstrated that the proposed system has the potential to be extended in more applications so that system performance can be further improved (e.g., 9% more reduction of total travel time and 25.7% more reduction of delay than the route guidance system without the incentive scheme). It also demonstrated that such personalized incentive scheme should be carefully designed so that it can be practical to implement.

6.3 Contributions

With the key findings summarized above, this research mainly made three contributions to existing literature as well as transportation research and applications community. Each of the three contributions is discussed below.

- **Established individual route choice models to capture individual drivers' route choice preferences**
  In this big data era, the information technologies applied in transportation domain have the ability to collect tremendous amount of behavior related data. Meanwhile, drivers' heterogeneous route choice preferences as well as the needs of providing personalized route guidance services require capturing and predicting drivers' route choice preferences at the individual level. With the possible data that can be collected from route guidance systems, this research not only explored the traditional mixed logit model together with the Bayes rules to estimate individual drivers' route choice preferences, but also explored a sentiment analysis approach, MT-LinAdapt which suits the application of route guidance system well. Both of these two models can estimate drivers' preferences at the individual level with decent performance, and do not require segmentation criteria which are not easy to obtain in practice but used by most existing modeling approaches to differentiate drivers' preferences. In addition, MT-LinAdapt is more suitable for the route guidance application from the perspective of practical implementation. Therefore, the MT-LinAdapt model introduced by this research can be adopted by route guidance systems to estimate drivers' route choice preferences at the individual level and help with providing personalized route recommendation services.
- **Proposed a proactive user optimum-oriented route guidance system for transportation system performance and users' satisfaction improvements**
  In existing route guidance systems with features of either proactive or reactive scheme, centralized or decentralized scheme, single criterion or multiple route attributes as criteria, drivers' behaviors play an important role in determining the performance of the route guidance system. This research proposed a proactive user optimum-oriented route guidance system that can learn drivers' route choice preferences and consider these preferences in designing routing strategies By knowing each individual driver's route choice preference and coordinating drivers' possible route choices in advance, the route guidance system can improve both the system performance, users' satisfaction and even reduce the errors of future traffic volume estimation. This research offers the transportation

community and route guidance service providers a new route guidance scheme that has great potential to improve both system performance and users’ satisfaction. The proposed route guidance system also provides some implications to automated vehicles routing systems which could have even larger impacts on transportation system.

- **Prepared the foundation for designing personalized traffic control and management strategies that have great potential to further improve transportation system performance**
  Most of the existing traffic control and management strategies target on general traffic as a whole. From the perspective of drivers’ behavior, because of the heterogeneous behaviors existing among population, a plan that targets on general traffic population could be ineffective to individual drivers. This is recognized as inefficient from the perspective of microeconomic, such as the inefficiency of the second-best pricing compared to the first-best pricing. By knowing each individual driver’s route choice preference, the proposed route guidance system in this research prepared the foundation for transportation researchers and engineers to design and implement personalized traffic control and management strategies, such as the personalized incentive scheme presented in the Discussion section. With the control and management strategies that are designed for each individual, the strategies are expected to be more effective than a general strategy. Therefore, a carefully designed personalized control and management strategy has great potential to bring better transportation system performance and users’ satisfaction improvement.

6.4 Future Research

Based on the findings of this research, several topics were identified as future research areas. Each of the topics is discussed in this section.

- **Further analyze drivers' route choice preferences with real data**

The route choice preference study in this research was conducted based on stated preference data. In order to validate and further analyze drivers' route choice preferences in real life, revealed preference data can be collected for route choice preferences analysis. Such data can be collected by designing a smart phone application or a web-based route guidance system, so that the traffic conditions that drivers are making route choice decisions in route guidance systems as well as the drivers' final decisions can be collected. Meanwhile, trip related information such as trip purpose and time of data can also be collected. Then, with more resources of data, all three types of individual route choice model's inputs can be collected and used for modeling individual driver's route choice preference. Therefore, the proposed individual route choice model can be validated and further analyzed.

- **Personalized traffic control and management strategies**

One of the most important characteristics of the proposed route guidance system is the knowledge of individual drivers' route choice preferences. Based on this, more personalized traffic control and management strategies can be explored in the future. One important extension is the personalized incentive scheme that was presented in the Discussion chapter. Personalized traffic control and management strategies provide a way to effectively motivate drivers' behavior decisions towards the system-desired goal while still maintaining drivers' satisfaction. Therefore, transportation problems such as traffic congestion can be mitigated from the demand-side of the transportation system. As analyzed in the Discussion chapter, the performance of personalized traffic control and management strategies is influenced by many factors, such as drivers' preference characteristics and network topology. These factors can be analyzed in future research for designing a practical and cost-effective personalized traffic control and management strategies.

- **Incorporating other aspects of traveler behaviors and non-recurrent congestion situation**

As mentioned in Chapter 4 Section II, this research only investigated travelers' route choice preferences. Other aspects of travelers' behaviors including mode choice and departure time choice also have important impacts. In future research, behavior data related to these two aspects can be collected and analyzed. The impacts of mode choice and departure time choice on transportation network performance and how these two aspects can be incorporated into the proposed proactive user optimum-oriented route guidance systems are research questions worthy of investigation. In addition, the framework of this research only considered the recurrent congestion in the demonstration. Non-recurrent congestion especially caused by incidents will be considered in future research.

REFREENCES USED IN CHAPTER 1, 5 and 6

Amirgholy, Mahyar, Nima Golshani, Craig Schneider, Eric J. Gonzales, and H. Oliver Gao. 2017. “An Advanced Traveler Navigation System Adapted to Route Choice Preferences of the Individual Users.” *International Journal of Transportation Science and Technology*, Special Issue on Urban Spatiotemporal Behavior and Network Assignment, 6 (4): 240–54. https://doi.org/10.1016/j.ijtst.2017.10.001.

Ashok, K., and Moshe Ben-Akiva. 1993. “DYNAMIC ORIGIN-DESTINATION MATRIX ESTIMATION AND PREDICTION FOR REAL- TIME TRAFFIC MANAGEMENT SYSTEMS.” In . https://trid.trb.org/view/640422.

Bank, Asian Development. 2015. *Introduction to Congestion Charging: A Guide for Practitioners in Developing Cities*. Asian Development Bank. https://www.adb.org/publications/introduction-congestion-charging-guide-practitioners-developing-cities.

Bhatt, Kiran, Thomas Higgins, and John T. Berg. 2008. “Lessons Learned from International Experience in Congestion Pricing.”

Button, Kenneth. 2004. “The Rationale For Road Pricing: Standard Theory And Latest Advances.” *Research in Transportation Economics* 9 (1): 3–25.

Chang, Gang-Len, and Jifeng Wu. 1994. “Recursive Estimation of Time-Varying Origin-Destination Flows from Traffic Counts in Freeway Corridors.” *Transportation Research Part B: Methodological* 28 (2): 141–60. https://doi.org/10.1016/0191-2615(94)90022-1.

Claes, R., T. Holvoet, and D. Weyns. 2011. “A Decentralized Approach for Anticipatory Vehicle Routing Using Delegate Multiagent Systems.” *IEEE Transactions on Intelligent Transportation Systems* 12 (2): 364–73. https://doi.org/10.1109/TITS.2011.2105867.

Fiosins, Maksims, Jelena Fiosina, Jörg P. Müller, and Jana Görmer. 2011. “Agent-Based Integrated Decision Making for Autonomous Vehicles in Urban Traffic.” In *Advances on Practical Applications of Agents and Multiagent Systems*, edited by Yves Demazeau, Michal Pěchouček, Juan M. Corchado, and Javier Bajo Pérez, 173–78. Advances in Intelligent and Soft Computing 88. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-19875-5_22.

Han, Linghui, Huijun Sun, David Z. W. Wang, Chengjuan Zhu, and Jianjun Wu. 2016. “The Combination of Continuous Network Design and Route Guidance.” *Computers & Operations Research* 73 (September): 92–103. https://doi.org/10.1016/j.cor.2016.03.012.

Liang, Zilu, and Yasushi Wakahara. 2014. “Real-Time Urban Traffic Amount Prediction Models for Dynamic Route Guidance Systems.” *EURASIP Journal on Wireless Communications and Networking* 2014 (May): 85. https://doi.org/10.1186/1687-1499-2014-85.

Liu, Henry X., Xiaozheng He, and Will Recker. 2007. “Estimation of the Time-Dependency of Values of Travel Time and Its Reliability from Loop Detector Data.” *Transportation Research Part B: Methodological* 41 (4): 448–61. https://doi.org/10.1016/j.trb.2006.07.002.

Luten, K. 2004. *Mitigating Traffic Congestion: The Role of Demand-Side Strategies*. U.S. Department of Transportation, Federal Highway Administration.

“Mobile Apps U.S. Smartphone Audience Reach 2017 | Statistic.” n.d. Statista. Accessed June 20, 2017. https://www.statista.com/statistics/281605/reach-of-leading-us-smartphone-apps/.

NHTSA. 2017. “Automated Vehicles for Safety.” Text. NHTSA. September 7, 2017. https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety.

Paz, Alexander, and Srinivas Peeta. 2009. “On-Line Calibration of Behavior Parameters for Behavior-Consistent Route Guidance.” *Transportation Research Part B: Methodological* 43 (4): 403–21. https://doi.org/10.1016/j.trb.2008.07.007.

Roughgarden, T., and E. Tardos. 2000. “How Bad Is Selfish Routing?” In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, 93–102. https://doi.org/10.1109/SFCS.2000.892069.

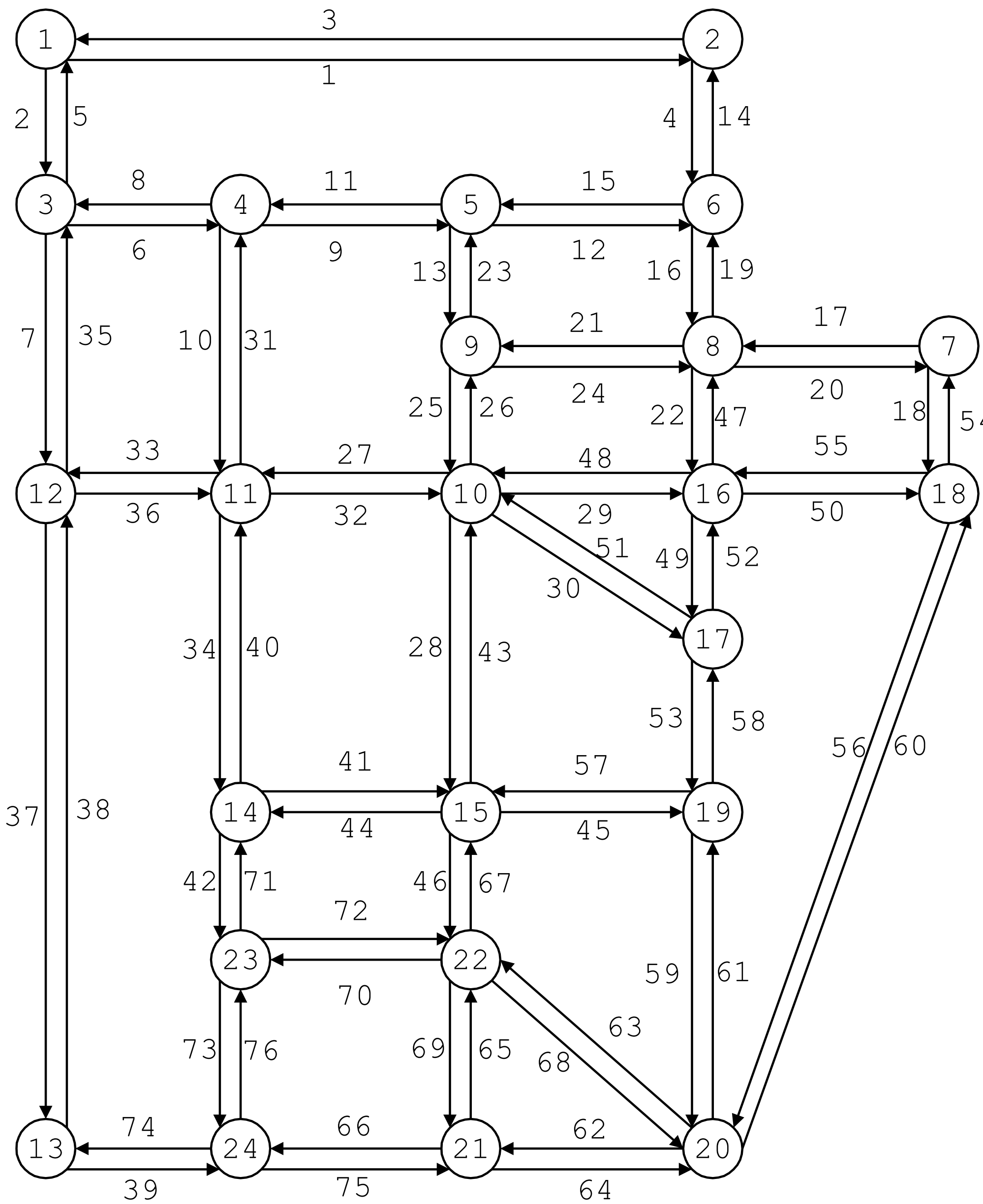
Toledo, Tomer, and Tanya Kolechkina. 2013. “Estimation of Dynamic Origin–destination Matrices Using Linear Assignment Matrix Approximations.” *IEEE Transactions on Intelligent Transportation Systems* 14 (2): 618–626.

Vitello, G., A. Alongi, V. Conti, and S. Vitabile. 2016. “A Bio-Inspired Cognitive Agent for Autonomous Urban Vehicles Routing Optimization.” *IEEE Transactions on Cognitive and Developmental Systems* PP (99): 1–1. https://doi.org/10.1109/TCDS.2016.2608500.

Zhang, Rick, Federico Rossi, and Marco Pavone. 2016. “Routing Autonomous Vehicles in Congested Transportation Networks: Structural Properties and Coordination Algorithms.” In . Robotics: Science and Systems Foundation. https://doi.org/10.15607/RSS.2016.XII.032.

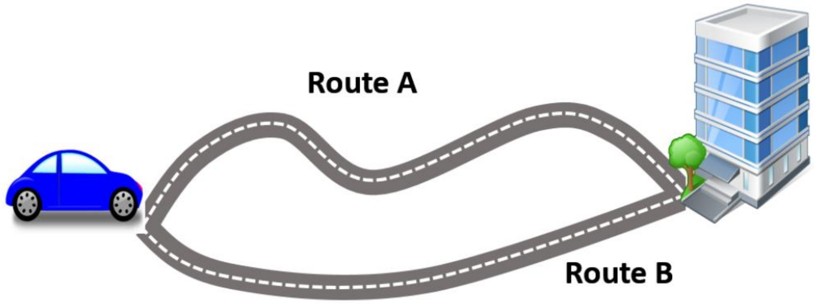
APPENDIX

Sioux Falls Network Structure



Note: The numbers in the figure are Node IDs and Link IDs.

Sampled Survey Questions



1. Which route do you prefer to take for a causal trip?

	Distance	Travel Time	Possible Longest Travel Time	Fuel Cost	No. of Controlled Intersection
Route A	8 miles	15 min	21 min	$ 0.74	10
Route B	8.8 miles	17 min	18 min	$ 0.69	9

8. Which route do you prefer to take for a causal trip?

	Distance	Travel Time	Possible Longest Travel Time	Fuel Cost	No. of Controlled Intersection
Route A	20 miles	39 min	54 min	$ 2.27	10
Route B	22 miles	42 min	50 min	$ 2.04	4

10. Which route do you prefer to take for a causal trip?

	Distance	Travel Time	Possible Longest Travel Time	Fuel Cost	No. of Controlled Intersection
Route A	20 miles	39 min	54 min	$ 2.27	20
Route B	22 miles	42 min	47 min	$ 1.72	15

Hourly Origin and Destination Matrix (Vehicles)

TOT	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
1	0	15	15	75	30	45	75	120	75	195	75	30	75	45	75	75	60	15	45	45	15	60	45	15
2	15	0	15	30	15	60	30	60	30	90	30	15	45	15	15	60	30	0	15	15	0	15	0	0
3	15	15	0	30	15	45	15	30	15	45	45	30	15	15	15	30	15	0	0	0	0	15	15	0
4	75	30	30	0	75	60	60	105	105	180	210	90	90	75	75	120	75	15	30	45	30	60	75	30
5	30	15	15	75	0	30	30	75	120	150	75	30	30	15	30	75	30	0	15	15	15	30	15	0
6	45	60	45	60	30	0	60	120	60	120	60	30	30	15	30	135	75	15	30	45	15	30	15	15
7	75	30	15	60	30	60	0	150	90	285	75	105	60	30	75	210	150	30	60	75	30	75	30	15
8	120	60	30	105	75	120	150	0	120	240	120	90	90	60	90	330	210	45	105	135	60	75	45	30
9	75	30	15	105	120	60	90	120	0	420	210	90	90	90	135	210	135	30	60	90	45	105	75	30
10	195	90	45	180	150	120	285	240	420	0	600	300	285	315	600	660	585	105	270	375	180	390	270	120
11	75	30	45	225	75	60	75	120	210	585	0	210	150	240	210	210	150	15	60	90	60	165	195	90
12	30	15	30	90	30	30	105	90	90	300	210	0	195	105	105	105	90	30	45	60	45	105	105	75
13	75	45	15	90	30	30	60	90	90	285	150	195	0	90	105	90	75	15	45	90	90	195	120	120
14	45	15	15	75	15	15	30	60	90	315	240	105	90	0	195	105	105	15	45	75	60	180	165	60
15	75	15	15	75	30	30	75	90	150	600	210	105	105	195	0	180	225	30	120	165	120	390	150	60
16	75	60	30	120	75	135	210	330	210	660	210	105	90	105	180	0	420	75	195	240	90	180	75	45
17	60	30	15	75	30	75	150	210	135	585	150	90	75	105	225	420	0	90	255	255	90	255	90	45
18	15	0	0	15	0	15	30	45	30	105	30	30	15	15	30	75	90	0	45	60	15	45	15	0
19	45	15	0	30	15	30	60	105	60	270	60	45	45	45	120	195	255	45	0	180	60	180	45	15
20	45	15	0	45	15	45	75	135	45	375	90	75	90	75	165	240	255	60	180	0	180	360	105	60
21	15	0	0	30	15	15	30	60	45	180	60	45	90	60	120	90	90	15	60	180	0	270	105	75
22	60	15	15	60	30	30	75	75	105	390	165	105	195	180	390	180	255	45	180	360	270	0	315	165
23	45	0	15	75	15	15	30	45	75	270	195	105	120	165	150	75	90	15	45	105	105	315	0	105
24	15	0	0	30	0	15	15	30	30	120	90	75	105	60	60	45	45	0	15	60	75	165	105	0