A human-centric machine learning based personalized route choice prediction in navigation systems

Sun, Bingrong; Gong, Lin; Shim, Jisup; Jang, Kitae; Park, B. Brian; Wang, Hongning; Hu, Jia

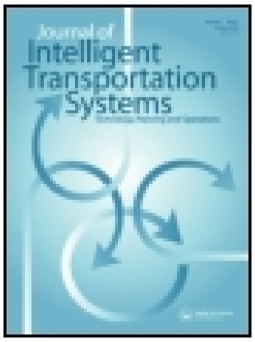
DOI
10.1080/15472450.2022.2069499

Publication date
2022

Document Version
Final published version

Published in
Journal of Intelligent Transportation Systems: technology, planning, and operations

Citation (APA)
Sun, B., Gong, L., Shim, J., Jang, K., Park, B. B., Wang, H., & Hu, J. (2022). A human-centric machine learning based personalized route choice prediction in navigation systems. *Journal of Intelligent Transportation Systems: technology, planning, and operations*, *27*(4), 523-535. https://doi.org/10.1080/15472450.2022.2069499



Journal of Intelligent Transportation Systems
Technology, Planning, and Operations

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gits20

A human-centric machine learning based personalized route choice prediction in navigation systems

Bingrong Sun, Lin Gong, Jisup Shim, Kitae Jang, B. Brian Park, Hongning Wang & Jia Hu

JOURNAL OF INTELLIGENT TRANSPORTATION SYSTEMS
https://doi.org/10.1080/15472450.2022.2069499

A human-centric machine learning based personalized route choice prediction in navigation systems

Bingrong Sun[a,b], Lin Gong[c], Jisup Shim[d*], Kitae Jang[e], B. Brian Park[a,b], Hongning Wang[c], and Jia Hu[f,g]

[a]Department of Engineering Systems and Environment, University of Virginia, VA, USA; [b]Link Lab, University of Virginia, USA; [c]Department of Computer Science, University of Virginia, USA; [d]Department of Transport & Planning, Delft University of Technology, Delft, The Netherlands; [e]Graduate School of Mobility, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea; [f]Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, China; [g]Institute for Advanced Study, Tongji University, China

ABSTRACT

Real-world route navigation data indicate that nontrivial portion of drivers do not prefer the system-recommended best routes. Current navigation systems have simplified assumptions about drivers' route choice preferences and do not adequately accommodate drivers' heterogeneous route choice preferences, mainly because of: (i) difficulty in acquiring exogenous criteria (e.g., sociodemographic information) that are typically used to differentiate drivers' preferences in behavioral modeling; and (ii) difficulty in capturing preference of individuals due to limited preference data at the individual level. To address these, this paper introduced a human-centric machine learning technique named Multi-Task Linear Classification Model Adaption (MT-LinAdapt). It can capture drivers' common aspects of route choice preferences and yet adapts to each driver's own preference. In addition, any evolvement of individual drivers' preferences can be simultaneously integrated to update the common preference for further individual drivers' preference adaptation. This paper evaluated MT-LinAdapt against two state-of-the-art route recommendation strategies including an aggregate-level and an individual-level data-based strategies, which are categorized based on the data used for modeling. With a real-world dataset containing 30,837 drivers' navigation usage data in Daegu City, South Korea, MT-LinAdapt was compared to existing strategies for its performance at different levels of data availability, and showed at least the same performance with existing strategies when minimum preference data is available and achieves up to 7% higher prediction accuracy as more data becomes available. Higher prediction accuracies are expected to bring better user satisfaction and compliance rates which can further help with transportation system control and management strategies.

ARTICLE HISTORY
Received 12 August 2021
Revised 18 April 2022
Accepted 20 April 2022

KEYWORDS
Intelligent transportation system; machine learning; personalized choice prediction; route choice; traveler information system

Introduction

The route navigation system is an important part of Intelligent Transportation Systems (ITS). A user can obtain information about possible routes connecting his/her origin and destination by using a navigation 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 considering the driver's own preference. Existing navigation systems, such as Google Maps and Waze, typically offer limited options for users to customize his/her own preferences. It is important to consider each individual user's route choice preference in a navigation system because of following two benefits. Firstly, it can improve users' satisfactions by providing route recommendations that users would like to take. A study found that, in around 60% of trips that drivers use navigation service, drivers either do not like the suggested routes before the trips were started or change to other routes during their trips (Amirgholy et al., 2017) and users may or may not follow the information provided by the system (Liao & Chen, 2015). If individuals' preferences can be properly considered in navigation systems, these phenomena could be largely mitigated. Secondly, it can help improve transportation systems performance, especially when route recommendations

CONTACT Jia Hu jh8dn@virginia.edu Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 200092, China.
*Additional affiliation: Department of Transport & Planning, Delft University of Technology, Delft.

are made for improving the system performance. Considering individual drivers' route choice preference in route recommendations can increase drivers' compliance rate thus improve network performance. Several studies have shown that driver's compliance rate to recommendations has significant impacts on network modeling accuracy (Ma et al., 2016; J. Wang et al., 2017) and a high compliance rate is a requirement for improving road network performance (Bifulco et al., 2007; Paz & Peeta, 2009).

With emerging information and communication technologies utilized, navigation systems not only have large user population but also can collect users' route choice preference data, even at the individual user's level (Zheng Wang et al., 2021). It provides an opportunity to capture individual users' route choice preferences and understand the needs of specific users when using navigation systems. Therefore, taking advantage of such data opportunity that emerging information and communication technologies brought us, this paper investigates how to capture individual users' route choice preference in navigation systems using over 30,000 drivers' real-world route choice data.

The rest of this paper is organized as follows. Literature review section first categorized the existing route recommendation strategies from data utilization perspective and identifies their limitations in practice. Methodology section described the proposed approach that this paper offered to address the identified limitations. With real-world users' data from a route navigation system, we compared the performance of the solution introduced by this paper with two existing approaches. The final section presented our conclusions and directions for future research.

Literature review

Route choice models are the major approach that researchers and practitioners used to capture and predict drivers' route choice decisions. Given the focus of this paper is how to take advantage of the data opportunity so that individual driver's route choice preference can be considered in navigation systems, existing route choice models are divided into two types in this paper according to the data used for establishing models. The two types of models are aggregate route choice models which are based on preference data at the aggregate level, and standalone individual route choice models which are based on individual user's own preference data. We categorized the existing modeling approaches based on the possible available data in practice, in order to show how we can better utilize the data advantages provided by the navigation systems. In the rest of this paper, "Individual route choice model" is used to refer the standalone individual model.

It should be noted that traditional route choice analysis includes two major research questions, namely route set generation and route choice decision (C. Prato, 2009). This paper focuses on the study of drivers' route choice decisions, i.e., how to better predict drivers' preferred route when route set is already available. The route set can be generated with methods in the literatures such as efficient paths (Dial, 1971), labeling (M. Ben-Akiva et al., 1984), K-shortest paths (Yen, 1971), link elimination (Barra et al., 1993), or branch and bound algorithm (C. G. Prato & Bekhor, 2006).

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 (M. E. Ben-Akiva & 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 the probability that an individual driver chooses each alternative route, such as multinomial logit models (M. E. Ben-Akiva & Lerman, 1985; C. 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 & Yu, 2005), support vector machine (Sun & Park, 2017), decision tree (Park et al., 2007). 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 & Xie, 2008). Another variation of aggregate route choice modeling is the multi-class route choice model. It first divides drivers into different groups based on certain criteria (e.g., learning and choice evolution pattern) and then build a model for each group (Peeta & Yu, 2004; Tawfik & Rakha, 2013). Drivers within the same group share the same route choice preference. Therefore, aggregate models lack sufficient preference heterogeneity to accurately predict each user's route choice, as drivers who have the same sociodemographic characteristics can still have different 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 such sociodemographic information due to privacy concerns, for example, age, income level, education, number of cars in family, etc., which are usually included in aggregate route choice models (Jan et al., 2000). It is also difficult to pick proper exogenous criteria for making segmentations (Hensher & Greene, 2003; Peeta & Yu, 2004) and the heterogeneity existing within a class is still difficult to determine.

The second route recommendation strategy is the individual route choice model. It is established based on individual driver's own route choice preference data. 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 from his/her own preference data and 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 navigation (Nadi & Delavar, 2011; Pahlavani et al., 2012; Pahlavani & Delavar, 2014). 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 contains 675 accumulated trips (Park et al., 2007) and 232 driver-rated virtual routes (Pahlavani & Delavar, 2014). In real life, it will take a certain length of time to get this amount of data. In practice, drivers may give up using a new navigation system after several times of unsatisfied experiences, so a good navigation system should capture individual drivers' preference even when only several trips' (for example, 3 to 4 trips) preference data is available. In addition, 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, departure time, and trip distances, etc. It is impossible to cover all possible trip characteristics of any particular driver may face with his/her own historical data.

More recent studies of individual route choice model capture personalized preference by allowing users to input preference weights on different criteria. Hayes et al. (2020) asked users to input weights on three criteria of travel time, distance and the number of crash incidents. In a personalized indoor route guidance system for people with visual impairment, Z. Wang et al. (2022) also asked users to put weights on five criteria that represent different obstacles. Ceder and Jiang (2020) developed a lexicographical shortest path method for personalized public transport services. The method combines users' input preference rating and human perception thresholds of the difference between attributes of alternatives in route generation. As pointed by Ceder and Jiang (2020), a good personalized route guidance system should minimize the input efforts required from users. Asking users to input their preference requires additional effort from users and also may not necessarily capture users' actual preferences as stated preference may not be consistent with users' actual preference.

As discussed above, to capture individual drivers' route choice preferences in navigation systems, aggregate route choice models have assumptions that drivers' preferences can be distinguished with sociodemographic characteristics or other criteria which are required modeling inputs but not easy to obtain in practice. 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. Therefore, in this paper, a human centric machine learning based technique, Multi-Task Linear Classification Model Adaptation (MT-LinAdapt) was introduced to help navigation systems understand every individual user's preference and provide better personalized route recommendations. And the method was tested with a large-scale and real-world navigation systems usage data. With the growing interests of understanding individuals' preference, MT-LinAdapt is expected to serve as a machine learning based approach in recommendation systems of the transportation domain.

Methodology

MT-LinAdapt model can be used for personalized route recommendation with its capability of capturing drivers' route choice preferences at the individual level. 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 human'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 members'. 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, MT-LinAdapt model adapts the aggregate route choice preference to individual drivers' level so that individuals' preference can be captured. Meanwhile, the change of individual drivers' preference could lead to drivers' aggregate preferences shifting and evolving which 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's own evolving. A linear classification model typically has a form of $y=sign(\boldsymbol{wX+c})$ in which **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 scenarios. $\boldsymbol{w}$ 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 affect drivers' route choice decisions.

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

$$\begin{aligned}\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_i, \ldots, \boldsymbol{w}_n] &= \begin{bmatrix} w_{11} & w_{21} & \ldots & w_{i1} & \ldots & w_{n1} \\ w_{12} & w_{22} & \ldots & w_{i2} & \ldots & w_{n2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{1k} & w_{2k} & \ldots & w_{ik} & \ldots & w_{nk} \end{bmatrix} \\ &= [\boldsymbol{w}_s, \boldsymbol{w}_s, \ldots, \boldsymbol{w}_s] \circ \boldsymbol{A}_u + \boldsymbol{B}_u \\ &= \begin{bmatrix} w_{s1} & w_{s1} & \cdots & w_{s1} \\ w_{s2} & w_{s2} & \cdots & w_{s2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{sk} & w_{sk} & \cdots & w_{sk} \end{bmatrix} \circ \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{bmatrix} \\ &+ \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{n1} \\ b_{12} & b_{22} & \cdots & b_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1k} & b_{2k} & \cdots & b_{nk} \end{bmatrix}\end{aligned} \tag{1}$$

$\boldsymbol{W}$ 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 $\boldsymbol{A_u}$ and $\boldsymbol{B_u}$ represent the specific adaptation operations based on driver i's preference data. $\circ$ 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 such preference changes. Therefore, the similar adaptation process can be conducted to $\boldsymbol{w_s}$ as well.

$$\boldsymbol{w_s} = \boldsymbol{w}_0 \circ \boldsymbol{A_s} + \boldsymbol{B_s} = \begin{bmatrix} w_{01} \\ w_{02} \\ \vdots \\ w_{0k} \end{bmatrix} \circ \begin{bmatrix} a_{s1} \\ a_{s2} \\ \vdots \\ a_{sk} \end{bmatrix} + \begin{bmatrix} b_{s1} \\ b_{s2} \\ \vdots \\ b_{sk} \end{bmatrix} \tag{2}$$

$\boldsymbol{w}_0$ is a k dimensional vector representing a prior weight vector which can be obtained by building a model based on a dataset that consists of a small portion from every individual's data. Any linear classification model can be incorporated into Eqs. (1) and (2). 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. The problem can be viewed as a joint optimization problem. By selecting different linear classifier models, the objective function can be different.

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 $y=1$ for driver i ($i=1, 2, \ldots, n$) in scenario j is:

$$P_{ij}(y=1|\boldsymbol{x}_j) = \frac{\exp(\boldsymbol{w}_i\boldsymbol{x}_{j1})}{\exp(\boldsymbol{w}_i\boldsymbol{x}_{j1}) + \exp(\boldsymbol{w}_i\boldsymbol{x}_{j0})}$$
$$= \frac{\exp((\boldsymbol{a}_i\circ\boldsymbol{w}_s + \boldsymbol{b}_i)\boldsymbol{x}_{j1})}{\exp((\boldsymbol{a}_i\circ\boldsymbol{w}_s + \boldsymbol{b}_i)\boldsymbol{x}_{j1}) + \exp((\boldsymbol{a}_i\circ\boldsymbol{w}_s + \boldsymbol{b}_i)\boldsymbol{x}_{j0})}$$
$$= \frac{\exp((\boldsymbol{a}_i(w_0\circ\boldsymbol{A}_s + \boldsymbol{B}_s) + \boldsymbol{b}_i)\boldsymbol{x}_{j1})}{\exp((\boldsymbol{a}_i(w_0\circ\boldsymbol{A}_s + \boldsymbol{B}_s) + \boldsymbol{b}_i)\boldsymbol{x}_{j1}) + \exp((\boldsymbol{a}_i(w_0\circ\boldsymbol{A}_s + \boldsymbol{B}_s) + \boldsymbol{b}_i)\boldsymbol{x}_{j0})} \tag{3}$$

$\boldsymbol{X_j}$ 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:

$$L_i(\boldsymbol{a}_i, \boldsymbol{A}_s, \boldsymbol{b}_i, \boldsymbol{B}_s) = \sum_{j=1}^{J}[y_j \log P_{ij}(y_j=1|x_j) + (1-y_j)\log P_{ij}(y_j=0|x_j)] \tag{4}$$

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 (5 b), 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 (i.e., scaling weight vectors by 1 and shifting weight vectors by 0). Therefore, taking n drivers' preference data together, the objective function is:

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

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. Readers who are interested in more details about the model can refer to (Gong et al., 2016).

Data

Real-world route choice behavior data were obtained and processed from a mobile navigation application, Kakao Navigation, in South Korea. Kakao Navigation is widely used in South Korea and has a large user population. When people use the navigation, the App provides two routes for users to choose, as shown in Figure 1(a). If the user does not take actions to choose one, the App automatically goes with the first route (referred as "default route" later in the paper) which is usually the one with shorter travel time. The other route is referred as the "alternative route." App records the routes recommended to users as well as the routes they chose. Such data collected from Kakao navigation users in Daegu, South Korea was used to test the proposed model.

The data included active drivers who made at least 30 trips with the navigation App from January 1st, 2018 to December 31st, 2019, in Daegu. This includes 9.5 million trips collected from approximately a hundred thousand drivers. When the App recorded that a driver went with the alternative route, it means the driver actively chose and preferred the alternative route. When the app recorded that a driver went with the default route, it could be the result of "did not take an action to actively choose" or "actively chose after consideration." Though the former situation can also be interpreted as a kind of driver preference (i.e., preferring what the App recommends), this algorithm targets on the latter case where drivers consciously compare two optional routes and actively chose one route to go. We call these drivers the conscious drivers, who are defined as the proportion of his/her own trips choosing either default or alternative route should be within 10% to 90%. In other words, a conscious driver should have chosen both the default and alternative routes in his/her historical trip records and the less-frequent chosen one should be at least 10%. This criterion excludes drivers who solely chose default or alternative route therefore creates a dataset which can reflect drivers' preference in route evaluation. This results in 20,278 drivers with 2,518,951 trip records for final analysis. This also means that around 30% of navigation users did not follow the system recommended top route in at least 10% of

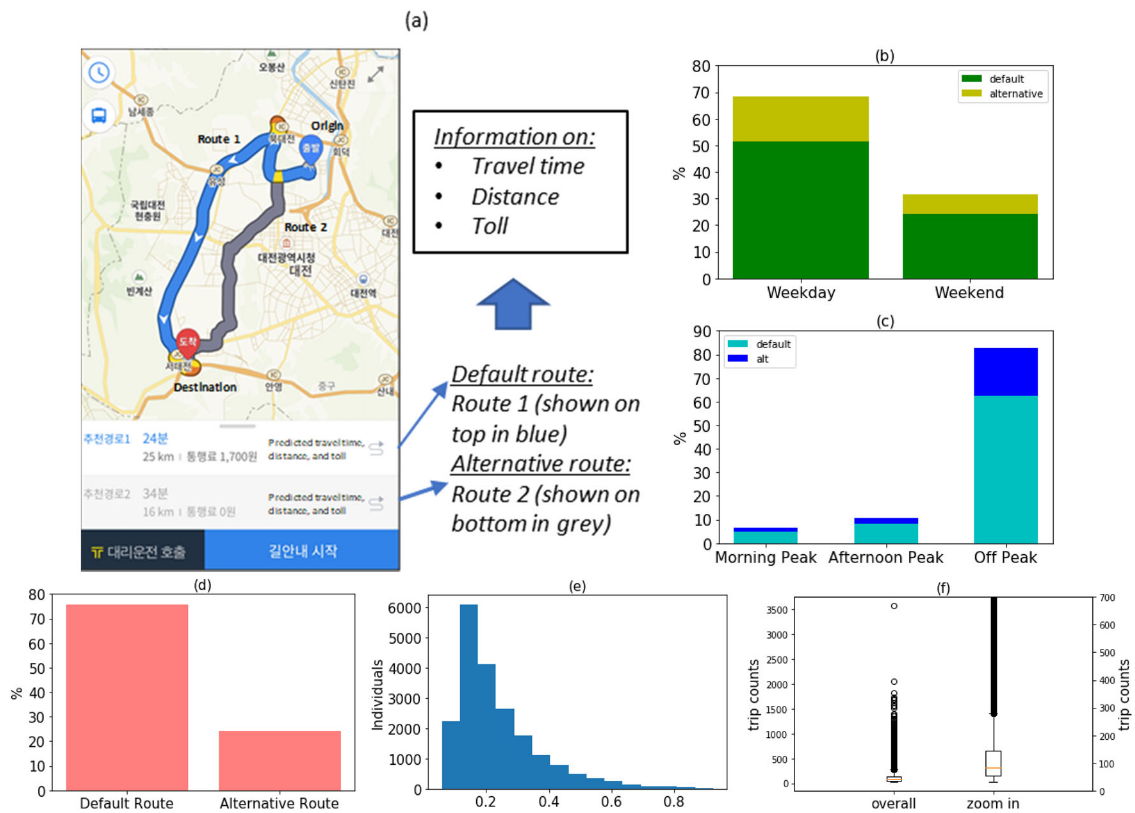


Figure 1. Data Statistics: (a) KaKao Navi interface and provided information; (b) weekday and weekend trips percentages; (c) percentage of trips in different time of day; (d) default and alternative chosen percentages; (e) driver counts histgrame by percetnage of trips that chosen alnterative routes; (f) boxplot of trip counts per user.

their guided trips. This further confirms the necessity of having personalized route recommendation.

To have a better understanding of the dataset, some of the data description is shown in Figure 1.

As shown in Figure 1(b) and (c), 68.3% of the trips were made on weekdays and 31.7% of them were made on weekend. 6.7% and 10.8% of the trips happened during morning peak and evening peak, while 82.5% of them were made during other time of the day.

Among trips made by all conscious drivers, 76% took the default routes and 24% took the alternative routes, as shown in Figure 1(d). A histogram of drivers with different alternative-default chosen ratios is shown in Figure 1(e). Only 5.5% of conscious drivers chose alternative routes more often than the default routes in their historical trip records.

When look at the number of trips made by each individual, the boxplot in Figure 1(f) shows that a portion of drivers used the navigation App extensively. The maximum number of trips that a conscious driver made in the dataset is 3,573 trips in two years, which is roughly 4.89 times a day on average. As shown on the right-hand side of Figure 1(f), a closer look at the boxplot shows that conscious drivers in our dataset used navigation 41 times (median value) a year (roughly one time every 9 days).

As summarized in Jan et al. (2000), many route attributes could influence travelers' route choice decisions, such as travel time, cost, speed limits, waiting time, congestion, stop signs, etc. From Kakao Navi usage data, five relevant variables were developed regarding travel time, toll, day of the week and time of day:

- Travel time ratio: the ratio of the travel time of the alternative route over that of the default route.
- Cost ratio: the monetary expenditure ratio between taking two routes. The monetary expenditure comprises possible toll and fuel cost. The assumptions of 1550 Korean won per liter (i.e., $4.725 per gallon) gas and fuel efficiency of 10.5 km/L were used to calculate the fuel cost.
- Weekend: a binary variable indicating whether it is a trip on weekend.
- Moring peak: a binary variable indicating whether it is a trip happening in morning peaks, namely 7:00 to 9:00.
- Afternoon peak: a binary variable indicating whether it is a trip happening in afternoon peaks, namely 17:00 to 19:00.

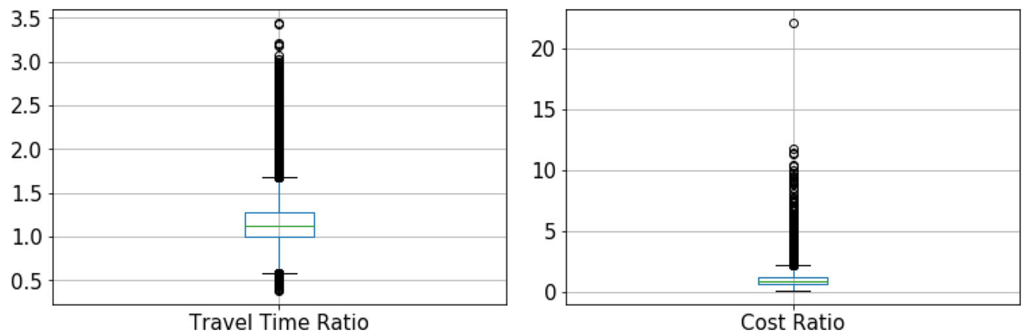


Figure 2. Value ranges of two variables.

It should be noted that route distance information is also available, but it is highly correlated with travel time, so it is embedded in the cost calculation instead of being included as another independent variable. Meanwhile, there are different ways to develop the route features from the dataset. For example, including the characteristics of both routes as model features and using the difference between the two routes' characteristics as features. We explored different options and adopted the one with the best performance in this paper as it can capture the relative level of service of two routes.

Figure 2 shows the range of two model variables. The default route usually has shorter travel time. The travel time of the alternative route is on average 1.15 times of the default route. The ratio can reach up to 3.43 and low to 0.38. For the variable of cost ratio, using the alternative route can be at most 22 times more expensive than using the default route.

For each trip, the chosen route paired with the unchosen route was considered as one route choice scenario. Values of five variables were generated for each trip. Such dataset was developed for evaluating the introduced MT-LinAdapt model.

Performance evaluation and results analysis

Test scenario

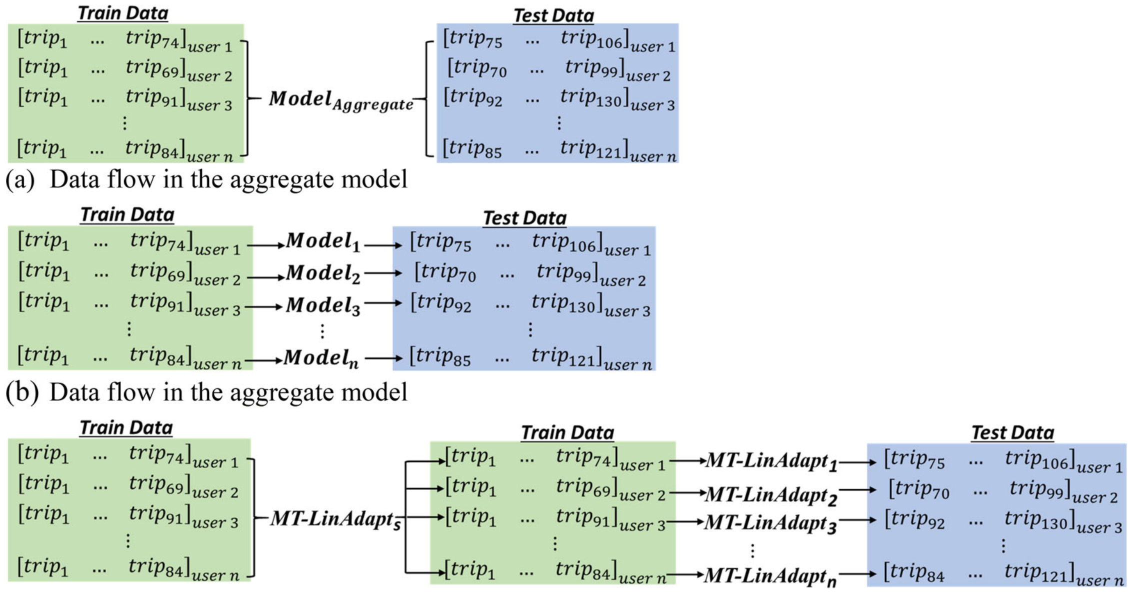
We first identified a test scenario which reflects the difficulty that navigation systems could face in practice. In reality, some of route guidance system users could accumulate adequate amount of preference data either because they are frequent users or have signed up the service for long time, but there are situations where users do not have adequate historical preference data either because they are new users or they do not like the service and do not use it often. The situation that has inadequate preference data is more difficult to capture users' preference but is more important for route guidance systems. as it is important to make good route recommendations so that users' satisfaction can be guaranteed, and the system could keep these users. Therefore, we tested whether the model can capture drivers' route choice preference at different levels of data availability.

Existing recommendation strategies (baseline models)

Based on the types of route recommendation strategies reviewed in the Literature Review section, an aggregate model and an individual model were selected for performance comparison. One of commonly used machine learning methods, Support Vector Machine (SVM), was chosen to build the aggregate and the individual route choice models, as SVM has demonstrated good performance in many traveler behavior analysis such as route choice and mode choice modeling (Sun & Park, 2017; Yuksel & Atmaca, 2021; Zhang & Xie, 2008). 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 & Christmann, 2008; Zhang & Xie, 2008) for detailed objective functions and constraints. It is noted that multinomial logit model was also explored to establish the individual and aggregate route choice models. While the multinomial logit model demonstrated similar performance as SVM, it sometimes cannot generate a statistically significant model when data amount is too small. Therefore, SVM was adopted in this paper to build the individual and aggregate models for later comparison.

Data preparation

The dataset was divided into training data and test data. The training data was used for establishing



(a) Data flow in the aggregate model

(b) Data flow in the aggregate model

(c) Data flow in the MT-LinAdapt model

Figure 3. Data flow in three models.

models and the test data was used for evaluating the performance of the established models. 30% of each individual driver's data was randomly selected as test data.

Figure 3 shows the data flow in each of the three models. For the aggregate model, all participant's training data 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 were included, as the data that can be collected from navigation systems do not contain such kind of information. 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 the proposed MT-LinAdapt, each user's data contributes to the estimation of the global-level model, *MT-LinAdapt$_s$*. Then, *MT-LinAdapt$_s$* is adapted to each user's own model with his/her own data.

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

To test the performance at different levels of data availability, each individual participant's test data was 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 models used for comparison, 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. For some drivers who has less data, that is equivalent to around 2 data points in the dataset. The data availability of 10% can represent the difficult scenario that users are new to the navigation systems and have limited historical data accumulated.

Implementation of three models

Support Vector Machine (SVM)

The scikit-learn package with Python was used for training SVM. Since SVM models have decent performance in travelers' behavior study (Sun & Park, 2017; Zhang & Xie, 2008) and different kernel functions show similar performance with the data, 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

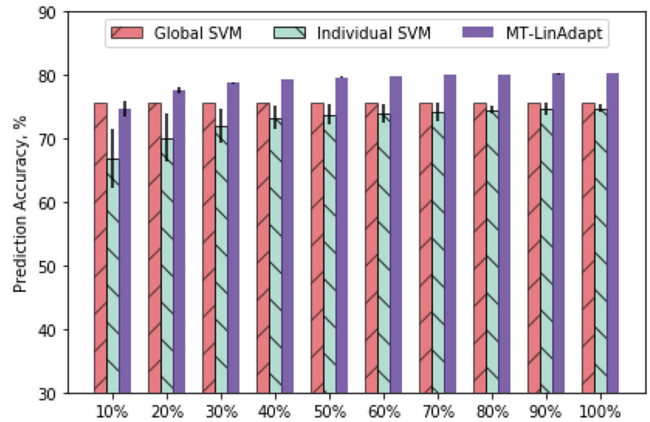


Figure 4. Prediction accuracies at different levels of data.

a commonly used range for penalty parameter in SVM (Ben-Hur & 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. The input data of SVM models includes all the route attributes that were contained in the dataset. The final output of aggregate SVM model is a separation plane represented by a linear equation for the whole group of participants. The final output of the individual SVM models have a linear equation for every single user. Two types of models were tested on the test data to obtain the prediction accuracies of the aggregate models and the individual models.

MT-LinAdapt model

The MT-LinAdapt model described in the Methodology section was coded with Java. In the training process, four parameters 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 tuned 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 the test data. The input data of MT-LinAdapt includes all the route attributes. The output of the MT-LinAdapt model includes a set of linear functions, each of which represents the preference of a particular user. The individual-level preferences (i.e., $\boldsymbol{w}_i$ as shown in Eq. (1)) were used to evaluate MT-LinAdapt model's performance.

Results analysis across three models

Figure 4 shows the prediction accuracies of three models at different levels of data availability.

For all three models, the overall trend of prediction accuracy gradually increases as more data is added for training. The performance of MT-LinAdapt is at least the same and mostly better for all levels of data availability, when compared to the other two models.

Among three models, the individual SVM model has the lowest prediction accuracies in all tested scenarios. As explained in the Methodology section, travelers' route choice behavior can be understood as a social norm. Travelers share certain common preference and still have some differences by individuals. Individual SVM model only uses each driver's own preference data which contains limited data variability. If test data contains new scenarios that are not covered by training data, individual SVM model may fail to predict the route choices correctly.

The performance of aggregate SVM is in between of the other two models. At the 10% of data availability, the aggregate SVM has almost the same prediction accuracy as MT-LinAdapt. That is because both the aggregate SVM and MT-LinAdapt capture the common preference when each user only has limited amount of data available. As data amount increases, the performance improvement of the aggregate SVM is limited because more data does not provide additional information on the common preference. At higher data availability levels, it is hard for the aggregated SVM to have further performance improvements after the common preference is captured, because it lacks the capability to adapt the captured preference to individual driver's level.

As to the MT-LinAdapt model, because of its capability to capture the social norm (common part of all drivers' preferences) and then adapt to each individual's preference, it consistently has better performance than other two models. Compare to the aggregate SVM model, it has the same prediction accuracy at the 10% of data availability but starts to develop advantages ranging from 3% to 5% higher prediction accuracy starting from 20% of data availability. 5% of higher prediction accuracy leads to 31,864 more correctly predicted trips. When compare to the individual SVM, the MT-LinAdapt maintains 5% to 7.5% of higher prediction accuracy. As each individual driver accumulates more route choice preference data, the estimated preference that is adapted to individual level can become more accurate. Therefore, the performance can be further improved.

As shown in Figure 4, the performance variation of three model types also differs. The performance variation of the global SVM is minimum across data divisions and its error bars are not noticeable in the figure. It is as expected because the global SVM model captures the preference at the aggregate level, which does not change as data availability increases from 10% to 100%. It also means the captured preference does not change much across data divisions because it is the preference of the same group of drivers. On the other hand, individual SVM models' performance varies more at low data availability levels than high data availability levels. It is also as expected, because individual SVM models only use part of individual's data to build a personalized model for each user. Due to limited observations and variation in individual's historical data, additional data points can have large impact on the estimated individual SVM models. This is especially true at low data availability levels where each driver does not have many historical data points. Thus, the performance of individual SVM models varies more at low data availability levels. Compared to global and individual SVM models, the error bar of MT-LinAdapt starts to become unnoticeable at 30% of data availability and has stable prediction accuracy across data divisions at high data availability levels.

When further assessing model's prediction capability in each type of choice, confusion matrices of three models at 100% data availability level were shown in Table 1 Each column shows the percentage of the test data falling into each category. The column "1-0" represents the percentage of trips in which the predicted choices are 1 (i.e., alternative route) and the observed choices are 0 (i.e., default route). The prediction accuracy calculated in figure is the summation of the columns "1-1" and "0-0" divided by the total number of trips. The global SVM basically predicts all cases to be 1 which is the dominant choice in the data set (as shown in Figure 1(d)). This is the current practice, namely recommending what the group likes to each individual. It should be noted that its performance is likely to be worse when the percentage of the dominant route decreases. For example, the prediction accuracy of the global SVM could become 55% when 55% of trips took the dominant route in the data. For individual SVM, since it is based on individuals' own data, it can capture the preferences of some individuals whose own dominant route is different from the group dominant route. Therefore, the column "1-1" has a slightly larger number of 0.49% than that of the global SVM. Because of the model nature, MT-LinAdapt better captures users' preferences when look at two types of choice together, as shown in the columns "1-1" and "0-0."

Table 1. Confusion matrices of three models at 100% of data availability.

	Predicted label – Observed label			
Model	0-0	0-1	1-0	1-1
Global SVM	75.67%	24.33%	0%	0%
Individual SVM	74.32%	23.84%	1.36%	0.49%
MT-LinAdapt	71.21%	15.28%	4.47%	9.05%

Table 2 summarized the performance of three model types at driver level as well as the pair-T test between MT-LinAdapt and each of the baseline models. It shows that MT-LinAdapt can not only improve the system overall prediction accuracy but also improve the performance for each individual driver. As shown in Table 2, the minimum accuracy achieved by MT-LinAdapt is around 38% which is much higher than 7% of global SVM models and 9% of individual SVM models. That means the worst service a driver can receive is much improved. The maximum prediction accuracy achieved by MT-LinAdapt can reach 100%, while the maximum prediction accuracy that global and individual SVM models can get is 94.6% and 95%. Therefore, the best service that a driver can receive is also improved. The average prediction accuracy at the driver level shows similar trend as the aggregate performance (i.e., Figure 4).

Additionally, paired-test was conducted between MT-LinAdapt and each of the baseline models. The average prediction accuracy of MT-LinAdapt is significantly higher than individual SVM models at all data availability levels. When compared with global SVM models, the prediction accuracy of MT-

Table 2. Performance at individual driver's level and statistical test between performance of different models.

	Maximum accuracy achieved			Minimum accuracy achieved			Average Accuracy Achieved			Paired T test	
Data levels	Global SVM	Individual SVM	MT-LinAdapt	Global SVM	Individual SVM	MT-LinAdapt	Global SVM	Individual SVM	MT-LinAdapt	MT-LinAdapt vs. Global	MT-LinAdapt vs. Individual SVM
10%	94.7%	90.1%	98.0%	6.9%	14.7%	40.8%	76.3%	69.3%	76.4%	=	>
20%	94.6%	91.2%	98.1%	6.9%	11.9%	41.4%	76.3%	67.9%	76.2%	=	>
30%	94.6%	92.2%	98.8%	6.9%	11.7%	41.1%	76.3%	69.8%	77.6%	>	>
40%	94.6%	93.4%	98.7%	6.9%	11.7%	39.3%	76.3%	72.0%	78.5%	>	>
50%	94.6%	94.0%	99.0%	6.9%	11.7%	41.0%	76.3%	72.7%	78.9%	>	>
60%	94.6%	94.0%	99.6%	6.9%	11.7%	39.5%	76.3%	73.0%	79.3%	>	>
70%	94.6%	93.7%	99.2%	6.9%	11.7%	39.3%	76.3%	73.3%	79.5%	>	>
80%	94.6%	95.0%	100.0%	6.9%	8.3%	38.0%	76.3%	74.1%	79.6%	>	>
90%	94.6%	94.2%	99.6%	6.9%	8.9%	37.7%	76.3%	74.4%	79.8%	>	>
100%	94.6%	94.2%	99.1%	6.9%	9.6%	37.7%	76.3%	74.4%	79.9%	>	>

LinAdapt is significantly higher at data availability levels of 30% to 100%, and has similar performance at data availability levels of 10% and 20%.

Therefore, MT-LinAdapt has better performance in terms of prediction accuracy not only at the average trend, but also improves the best and worst service scenarios that a driver can experience.

Conclusions

To maximally utilize available data for making personalized route recommendations in navigation systems, we introduced a human-centric machine learning based technique, MT-LinAdapt and demonstrated its capability of accommodating drivers' heterogeneous preferences even when limited preference data is available. The model does not require personal sociodemographic information (e.g., age, gender, income, etc.) or other criteria to differentiate drivers' different route choice preferences. Meanwhile, MT-LinAdapt also works well when a user has limited amount of preference data (e.g., a new user).

This paper compared MT-LinAdapt with two existing recommendation strategies: the aggregate and the individual data-based recommendations. Three types of strategies were tested with route choice preference data that was collected from 30,837 real-world navigation system users' daily usage from South Korea for performance evaluation at different levels of data availability. The results showed, MT-LinAdapt has at least the same and mostly better prediction accuracies than the other two models at all levels of data availability. When each user has only limited amount of data, MT-LinAdapt has 7.5% higher prediction accuracy (i.e., 48,500 more trips were predicted correctly) than the individual model. When more data was added, the MT-LinAdapt can still maintain the advantage around 5% higher than both the aggregate and individual models. From individual driver's perspective, the best and worst service that a user can experience are both improved when compared MT-LinAdapt to baseline models. With more features developed from route choice preference data and including more route characteristics or trip information (e.g., trip purpose developed from land use data), the performance of MT-LinAdapt is expected to be further improved and this will be tested in future research.

These advantages of MT-LinAdapt would help navigation systems considering each individual driver's specific preference when making personalized route recommendations in practice, and consequently improve users' satisfaction, increase users' compliances with the guidance system and potentially achieve better road network performance (Jaber & O'Mahony, 2009). Take the dataset in this paper as an example, among 100,000 randomly selected drivers who used the navigation systems more than 30 times in two years, 30% of these drivers declined to use the system suggested route at least in 10% of their trips. The 5% to 7.5% higher prediction accuracy achieved by MT-LinAdapt could be converted to thirty to nearly fifty thousand more trips in which drivers comply with system suggested routes. Depending on the user population, this potentially can help predict future network conditions and improve network performance when network performance is also considered in route planning (Hu et al., 2017). The network level impacts of considering individual route choice preferences in navigation systems will be evaluated in a future research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the DGIST R&D Program of the Ministry of Science and ICT (18-EE-01). This research was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1A6A3A03037304), and authors are grateful to Kakao Mobility Corp. for their data used in this paper.

ORCID

Bingrong Sun http://orcid.org/0000-0003-2806-0200
B. Brian Park http://orcid.org/0000-0003-4597-6368

References

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*, *6*(4), 240–254. https://doi.org/10.1016/j.ijtst.2017.10.001

Barra, T. D. L., Pérez, B., & Añez, J. (1993). Multidimensional path search and assignment. In PTRC Summer Annual Meeting, 21st, 1993, University of Manchester, United Kingdom.

Ben-Akiva, M., Bergman, M. J., Daly, A. J., & Ramaswamy, R. (1984). *Modelling inter urban route choice behaviour* [Paper presentation]. The Ninth International Symposium on Transportation and Traffic Theory (pp. 299–330). https://trid.trb.org/view/210836

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. MIT Press.

Ben-Hur, A., & Weston, J. (2010). *A user's guide to support vector machines* (pp. 223–239). Humana Press. https://doi.org/10.1007/978-1-60327-241-4_13

Bifulco, G. N., Simonelli, F., & di Pace, R. (2007). *Endogenous driver compliance and network performances under ATIS* [Paper presentation].2007 IEEE Intelligent Transportation Systems Conference (pp. 1028–1033). https://doi.org/10.1109/ITSC.2007.4357722

Ceder, A., & Jiang, Y. (2020). Route guidance ranking procedures with human perception consideration for personalized public transport service. *Transportation Research. Part C, Emerging Technologies*, *118*, 102667. https://doi.org/10.1016/j.trc.2020.102667

Dial, R. B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, *5*(2), 83–111. https://doi.org/10.1016/0041-1647(71)90012-8

Gong, L., Al Boni, M., & Wang, H. (2016). *Modeling social norms evolution for personalized sentiment classification* [Paper presentation]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 855–865. https://doi.org/10.18653/v1/P16-1081

Hayes, S., Wang, S., & Djahel, S. (2020). *Personalized road networks routing with road safety consideration: A case study in Manchester* [Paper presentation].2020 IEEE International Smart Cities Conference, ISC2 2020. https://doi.org/10.1109/ISC251055.2020.9239085

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: The state of practice. *Transportation*, *30*(2), 133–176. https://doi.org/10.1023/A:1022558715350

Hu, X., Chiu, Y. C., & Shelton, J. (2017). Development of a behaviorally induced system optimal travel demand management system. *Journal of Intelligent Transportation Systems*, *21*(1), 12–25. https://doi.org/10.1080/15472450.2016.1171151

Jaber, X., & O'Mahony, M. (2009). Mixed stochastic user equilibrium behavior under traveler information provision services with heterogeneous multiclass, multicriteria decision making. *Journal of Intelligent Transportation Systems*, *13*(4), 188–198. https://doi.org/10.1080/15472450903299554

Jan, O., Horowitz, A. J., & 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*(1), 37–44. https://doi.org/10.3141/1725-06

Liao, C. H., & Chen, C. W. (2015). Use of advanced traveler information systems for route choice: Interpretation based on a bayesian model. *Journal of Intelligent Transportation Systems*, *19*(3), 316–325. https://doi.org/10.1080/15472450.2014.936289

Ma, J., Zhou, F., & Lee, C. (2016). Providing personalized system optimum traveler information in a congested traffic network with mixed users. *Journal of Intelligent Transportation Systems*, *20*(6), 500–515. https://doi.org/10.1080/15472450.2016.1152549

Mahmassani, H. S., Koppelman, F. S., Frei, C., Frei, A. R., & Haas, R. (2013). *Synthesis of traveler choice research: Improving modeling accuracy for better transportation decisionmaking*. https://www.semanticscholar.org/paper/Synthesis-of-Traveler-Choice-Research%3A-Improving-Mahmassani-Koppelman/08e936ca9037771adad2ac7bdb922a16e315fc3a

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*(3), 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*(2), 159–166. https://doi.org/10.1049/iet-its:20060074

Paz, A., & Peeta, S. (2009). Behavior-consistent real-time traffic routing under information provision.

Transportation Research Part C: Emerging Technologies, *17*(6), 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*(2), 243–256. https://doi.org/10.1109/TSMCA.2003.822272

Peeta, S., & Yu, J. W. (2005). A hybrid model for driver route choice incorporating en-route attributes and real-time information effects. *Networks and Spatial Economics*, *5*(1), 21–40. https://doi.org/10.1007/s11067-005-6660-9

Prato, C. (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

Prato, C. G., & Bekhor, S. (2006). Applying branch-and-bound technique to route choice set generation. *Transportation Research Record: Journal of the Transportation Research Board*, *1985*(1), 19–28. https://doi.org/10.1177/0361198106198500103

Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer.

Sun, B., & Park, B. B. (2017). Route choice modeling with support vector machine. *Transportation Research Procedia*, *25*, 1806–1814. https://doi.org/10.1016/j.trpro.2017.05.151

Tawfik, A. M., & Rakha, H. A. (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*(1), 84–94. https://doi.org/10.3141/2334-09

Wang, J., Lv, J., Wang, C., & Zhang, Z. (2017). Dynamic route choice prediction model based on connected vehicle guidance characteristics. *Journal of Advanced Transportation*, *2017*, 1–8. https://doi.org/10.1155/2017/6905431

Wang, Z., Lin, W. H., & Xu, W. (2021). A data driven approach to assessing the reliability of using taxicab as probes for real-time route selections. *Journal of Intelligent Transportation Systems*, *25*(4), 331–342. https://doi.org/10.1080/15472450.2019.1617142

Wang, Z., Yu, Y., Feeley, C., Herrick, S., Hu, H., & Gong, J. (2022). A route optimization model based on building semantics, human factors, and user constraints to enable personalized travel in complex public facilities. *Automation in Construction*, *133*, 103984. https://doi.org/10.1016/j.autcon.2021.103984

Yamamoto, T., Kitamura, R., & Fujii, J. (2002). Drivers' route choice behavior: Analysis by data mining algorithms. *Transportation Research Record: Journal of the Transportation Research Board*, *1807*(1), 59–66. 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*(2), 199–223. https://doi.org/10.1007/BF01307059

Yen, J. Y. (1971). Finding the K shortest loopless paths in a network. *Management Science*, *17*(11), 712–716. https://doi.org/10.1287/mnsc.17.11.712

Yuksel, A. S., & Atmaca, S. (2021). Driver's black box: A system for driver risk assessment using machine learning and fuzzy logic. *Journal of Intelligent Transportation Systems*, *25*(5), 482–500. https://doi.org/10.1080/15472450.2020.1852083

Zhang, Y., & Xie, Y. (2008). Travel mode choice modeling with support vector machines. *Transportation Research Record: Journal of the Transportation Research Board*, *2076*(1), 141–150. https://doi.org/10.3141/2076-16