Modeling Drivers' Route Choices and Route Compliance when Interacting with an Eco-Routing Navigation System

Bo Yu [b], Shan Bao [a, b, *], Zeyang Li [b], Eric M Rask[c], Henry Liu [b], Jim Sayer [b]

(a) Industrial and Manufacturing Systems Engineering Department, University of Michigan-Dearborn, 4901 Evergreen Rd, Dearborn, MI 48128
(b) University of Michigan Transportation Research Institute, 2901 Baxter Rd, Ann Arbor, MI, USA, 48109-2150
(c) Argonne National Laboratory, 9700 S Cass Ave, Lemont, IL 60439

(*) Corresponding author: **Shan Bao,** E-mail: shanbao@umich.edu

Abstract

Eco-routing navigation systems have become a promising application to reduce fuel consumption by optimizing driving routes through energy efficiency prioritization instead of solely travel time or distance minimization. Current studies have put limited efforts to investigating whether and why drivers will choose and comply with the eco route recommended by eco-routing navigation systems. Thus, to fill this research gap, this study developed a smartphone-based eco-routing navigation application (app) and collected naturalistic driving data to examine and model drivers' route choices and compliance behavior when interacting with the recommended route. It was observed that drivers chose the eco-routing option with the most energy-saving feature in approximately 78.6% of all the selected routes in this study. To further explore the impacting factors on the eco-routing choice, mixed model analyses were conducted. The results showed that drivers were more likely to select the eco-route when this trip had shorter distance and higher per mile gas consumption. It was also found that giving priority to recommend the eco-route could guide drivers to choose the eco-route. Since drivers' route choices belonged to the multi-label problem, this study applied a Multi-label random forests (MLRF) model to predict route choice behavior. In this model, independent variables were derived from three aspects, including driver characteristics, subjective data, and route information. The overall accuracy and AUC (the area under the receiver operating characteristic curve) of this MLRF model were 88.3% and 0.86 respectively. Overall, the average proportion that participants complied with the recommended route while driving was 56.7%. Mixed model analyses found that when drivers chose the eco or fast routes, they were more likely to fully follow the recommended route. Compared with driving alone or with only one household passenger, drivers preferred to comply with the recommended route when there were three or more household passengers. The findings of this study can help to understand drivers' decision making in route planning and therefore to improve eco-routing navigation system designs, which will be beneficial to the eco-friendly transportation system.

Keywords: Eco-routing navigation system; Route choice; Route compliance; Multi-label random forests; Mixed model analysis

1. Introduction

The transportation sector, as one of the most significant contributors to greenhouse gas emissions, is greatly responsible for the growing environmental problems, such as the fossil fuel shortage crisis, global climate change, air pollution, and so on (Aziz and Ukkusuri, 2014; Salvi and Subramanian, 2015). It has been reported that the transportation sector made up the largest proportion (approximately 29%) of the total U.S. greenhouse gas emissions in 2017, of which 41% was represented by passenger cars (EPA, 2019). With the rapid development of intelligent transportation systems and data, the eco-routing navigation system becomes a potential application to reduce fuel consumption, which optimizes the route based on the highest energy efficiency instead of minimizing the travel time or distance (Ahn and Rakha, 2013; Wang et al., 2019; Zhao et al., 2019). The most energy-economic route (hereinafter called "the eco route") is not always consistent with the fastest or the shortest route, due to higher travel speed or traffic congestion (Zeng et al., 2017; Boriboonsomsin and Barth, 2014). If the eco route is adopted by drivers to take the place of traditional route choices (i.e., the fastest or shortest route), a 4% ~ 20% decrease in energy consumption can be achieved (Ahn and Rakha, 2008).

Currently, eco-routing models designed for advanced navigation systems have been proposed and improved, as more influencing factors on vehicle fuel consumption have been investigated and optimized. For example, hilly routes consumed 15% to 20% more fuel than flat routes (Boriboonsomsin and Barth, 2009), while arterial routes yielded approximately 22% better fuel economy compared to highway routes (Fiori et al., 2018). Traffic congestion showed significantly negative effects on fuel economy, and an increase in fuel consumption ranged from 20% ~ 40% based on different levels of congestion (Sivak and Schoettle, 2012; Lois et al., 2019). An embedded data fusion method was employed to involve both historical and real-time traffic information into an eco-routing navigation system, where consumption-related factors like vehicle type, roadway characteristics, and traffic conditions were taken into consideration (Boriboonsomsin et al., 2012). A calculation model of fuel consumption and greenhouse gas emissions was established by Nie and Li (2013), and their study found that consumption rates were associated with vehicles' physical parameters (e.g. weight and engine displacement) and operational properties (e.g. acceleration, turning movements, and idling at intersections). To achieve the real-time and vehicle-specific route recommendation, a dynamic eco-routing model was presented and tested in a micro-simulation framework, in which several energy cost functions were combined to capture microscopic transient behavior, including acceleration, driving speed, and road grade variations within each road segment (Wang et al., 2019).

Besides traditional route recommendations such as the shortest or fastest routes, eco-routing navigation systems offer drivers an extra option, the eco route, but a new question arises: when faced with different routing options, which one driver would prefer to choose? Route choice models can provide a better understanding of drivers' route choice preferences and their influencing factors; however, current route models mostly focus on traditional routes without considering the eco one. Drivers' route choice was affected by their experience and habits, and experienced drivers appeared to be more likely to choose the route with less travel time (Prato and Bekhor, 2007). Personalities also had effects on route choices, for example, some drivers would like to select the longer but more scenic route to enjoy the driving time, while other drivers who were interested in driving also preferred a longer route (Handy et al., 2005). A reinforced learning-based model was applied to examine the combined effects of real-time information and experience on drivers' route choice behavior (Ben-Elia and Shiftan, 2010). The results showed that compared with non-informed drivers, informed drivers could learn faster from their personal experience, exhibited more risk-seeking behavior, and were more sensitive to the change of travel time. A mixed logistic regression model was used to investigate heterogeneity in route choice behavior (Li et al., 2016). This study found that male and younger drivers cared more about the number of intersections in their selected routes, and if drivers were familiar with the origin-destination, they would be more likely to choose the fastest route. According to the results of a survey study, travel time was the most important factor when drivers chose the route (Papinski et al., 2009), and the significant effect of travel time on route choice behavior was also confirmed in another study (Shakeel et al., 2016).

With the development of advanced navigation systems, the trade-off between extra travel time and lower fuel consumption has begun to be discussed in several studies by questionnaires or simulation experiments. A survey study showed that participants rated energy savings as the third important feature when they decided which route to travel, while travel time ranked first with its importance score 16% higher than that of energy savings (Wang et al., 2020). Since the eco route recommendation was not available for the most current navigation systems, in this survey study, some participants needed to imagine that they were using eco-routing navigation systems. A web-based experiment was conducted by Aziz and Ukkusuri (2014) to mimic real-world travel scenarios. This experiment found that female drivers would trade more additional travel time to choose the eco route than male drivers during both work and non-work trips, while drivers from higher-income families also showed greater willingness to exchange extra travel time for lower emissions due to their generally higher levels of education and awareness of environmental problems (Aziz and Ukkusuri, 2014). Another study through a simulation experiment observed that when the trip distance was less than 10 miles, with the increase of the trip distance, the eco-routing would result in more fuel savings but would take longer travel time (Boriboonsomsin et al., 2014).

Even if drivers initially choose a route recommended by the navigation system, they may still not follow the route while driving. A study reported that 20% of participants did not comply with their planned routes, and these route changes took up approximately 44% of the total trip distance (Papinski et al., 2009). Participants who had shorter trip lengths were more likely to deviate from their planned routes (Papinski et al., 2009). Drivers' route compliance was modeled by Radial basis function networks, where independent variables were socio-economic features, expected savings of travel time, and familiarity with road conditions (Dia and Panwai, 2007). Due to this black-box model used in this study, the detailed correlation between these input factors and route compliance behavior could not be further explained. Real-time information and experience-based knowledge may also make drivers divert to an alternative route while driving (Tawfik et al., 2010; Abdel-Aty and Abdalla, 2004; Ben-Elia and Shiftan, 2010). However, it is still not clear about drivers' route compliance under the eco-routing condition.

Given the above, eco-routing navigation systems can greatly contribute to the reduction of fuel consumption and greenhouse gas emissions, if drivers are willing to choose and comply with the eco routes provided by these systems. However, current studies put limited efforts into drivers' route choices and compliance behavior when interacting with eco-routing navigation systems. Therefore, there are two main purposes in this study based on naturalistic driving data: one is to investigate and predict what kind of route drivers will choose from recommendations offered by eco-routing navigation systems; the other is to explore whether and why drivers will follow the selected route while driving. Influencing factors on drivers' route choices and compliance are examined from three aspects: driver characteristics, subjective data, and route information. This study hypothesizes that fuel-saving benefits will attract drivers to choose and follow the eco route.

We expect that this study can help to provide design recommendations to new advanced navigation systems so that drivers are willing to use and comply with the eco route choices, which will be beneficial to the overall eco-friendliness of the transportation system.

2. Methodology

The methodology section contains four parts: (1) a smartphone-based eco-routing application (app) was developed to collect driving and survey data; (2) naturalistic driving experiments were designed and conducted; (3) data reduction and mixed model analyses were applied to interpret drivers' route choices and route compliance; (4) the Multi-label random forests algorithm was used to predict drivers' route choice behavior.

2.1 Eco-routing application development

2.1.1 Eco-routing app functionality

The main goal of this turn-by-turn eco-routing navigation app was to collect all the answers to the questions of a survey presented at the beginning of the route, the number of detours with all the newly created directions, and all the GPS coordinates followed by the driver. This app was developed jointly by the Argonne National Laboratory (ANL) and the University of Michigan Transportation Research Institute (UMTRI). As shown in Figure 1, this app was able to recommend three different routes to drivers with varying the time and the fuel consumption on the routes: (1) the eco route: the least energy consumption; (2) the fast route: the shortest travel time; (3) the balanced route: a balance between energy consumption and travel time. The app could also offer turn-by-turn guidance using the selected route while recording the GPS data gathered along the route. The route and navigation services were provided by the MapQuest API (application program interface). This API could calculate the route after inputting the desired destination and display all the data in the turn-by-turn navigation screen. Trip purpose, drivers' vehicle model, passenger information, and other relevant information were also collected in the survey through the app at the beginning of each trip (Figure 2).

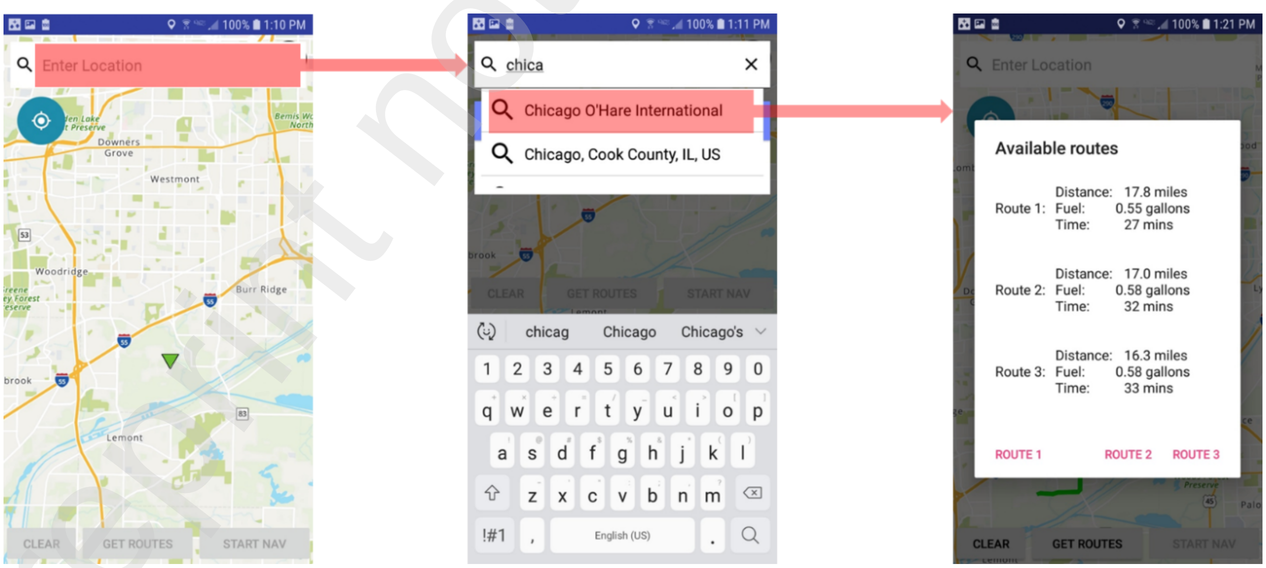


Fig. 1. Three different routes provided by the eco-routing navigation app

Data recording was made locally in a database file using SQLite API. In the code of the app, there were two different databases, one was called "*Trips.db*" and the other was "*cars.db*" (See

Figure 2). The "*Trips.db*" database included all the data for the analysis, such as survey data, trip information, driver route decisions, GPS, etc. The "*cars.db*" file just recorded the different cars used in the experiments to introduce corrections in the required fuel estimation made by the app at the beginning of the route.

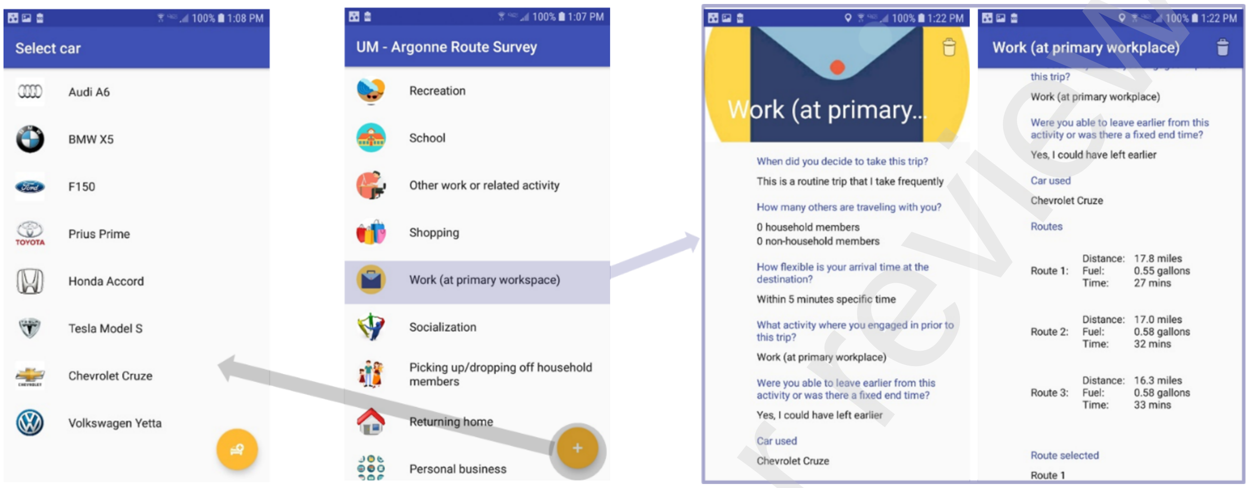


Fig.2. Recording survey data and car types

2.1.2 Fuel consumption estimation

The displayed three recommended routes (eco vs. fast vs. balanced) showed an estimation of the fuel consumption that was calculated considering three main effects: the average velocity, the traffic, and the car used. Specifically, the effects of the velocity and car used were modeled by inputting the MPG (miles per gallon) values in the Urban Dynamometer Driving Schedule (UDDS) cycle (average velocity of 21.2 mph) and in the highway cycle (average velocity of 48.3 mph). These two values could be easily found on the U.S. EPA (Environmental Protection Agency) website for different car models. The U.S. Department of Energy (2020) explained that the maximum MPG occurred around 40-50 mph with a low effect of the speed in this area, and when the speed was 80 mph there was a reduction of around 35% in the maximum value of the MPG. Knowing this and using the two provided values of MPG, the dependency of MPG with the speed observed in Figure 3 was calculated. The effect of traffic was estimated using the data provided by MapQuest API, since this API was able to classify segments of the route according to the different types of traffic, including *"FREE_FLOW"*, "*SLOW*", and "*STOP_AND_GO*".

With these three main effects on the MPG, the calculation was made using the following procedure after splitting the whole route into different segments according to:

(1) The whole route was split into different segments according to the traffic classification made by MapQuest API;

(2) Calculation of the average MPG on each route segment with "*FREE_FLOW*" traffic was made by obtaining the average speed in the segment and using the plot of Figure 3;

(3) The global MPG of the route was estimated by assuming that the vehicle had the value calculated before in the "*FREE_FLOW*" segments of the route, and the city cycle value in the "*STOP_AND_GO*" and "*SLOW*" segments. The values of all these segments were averaged using their distances.

$$MPG_{global} = \frac{\sum Distances_{FREE\_FLOW} * MPG(Average\ speed) + \sum Distances_{NON\_FREE\_FLOW} * MPG\_city}{Total\ distance} \quad (1)$$

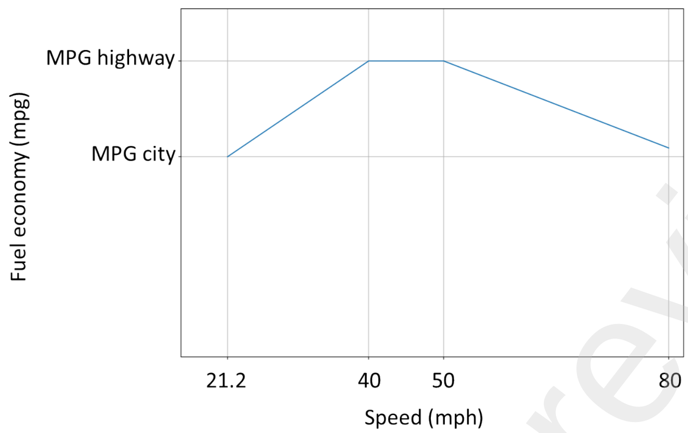


Fig.3. The effect of speed on the MPG during a highway driving

2.2 Experiment design and conduction

Forty-three participants with valid driver's licenses were recruited and participated in the study. Each participant received a cell phone handset with the "Eco-Routing" application installed for a period of two weeks. During the two weeks, participants were expected to interact with the app by completing a route-choice survey and selecting from recommended driving routes prior to the beginning of some of their driving trips. Participants were expected to record data for at least 20 trips and received $100 as their participation compensation.

All participants reviewed and completed the eco-routing informed consent document before they participated in the study. After they signed the consent form, participants were invited to come to UMTRI to pick up the eco-routing device and to learn how to operate it. The expectations for the interaction process with the eco-routing cell phone application were then explained to the participant. Researchers walked through all the procedures with all participants, including setting up the application, inputting a trip, choosing a suggested route, and using the navigation function.

2.3 Data reduction

Valid trips were firstly identified as those with completed surveys and GPS data. In addition, if the participant did not travel to their inputted destination, the trip data was removed from further analysis. The final database contained a total of 737 valid trips from 39 participants, which included the responses to the questionnaire for each trip, information about the selected route, and the GPS data collected on the device for the specific trips. These 39 participants consisted of 22 female and 17 male drivers, aged from 20 to 72 years old (Mean=47.3, S.D.=15.3). The detailed process of experiments was described as follows:

The eco-routing application on the cell phone was set up by inputting vehicle types and the EPA-provided MPG values for both city and highway driving. Before each trip began, the interaction with the cell phone application included answering eight questions about the nature of the trip and the participant's trip planning process via a short survey embedded in the application. Participants then input their desired destination for their planned trip. Next, the application offered them one, two, or three different driving routes to their destination from which to choose. Each suggested route was listed with its expected distance, fuel consumption, and time duration.

Participants were instructed: “The application will provide you with the distance, time, and estimated fuel consumption for up to three routes. Choose the route that is most desirable to you. It is likely that one of these routes is the best route for you to take, but you are not required to follow them.” After choosing a route, the device provided route guidance to the participant via both live turn-by-turn directions on the screen and via verbal directions from the cell phone speaker. Maps and navigation used the MapQuest API.

The number of suggested routes was a function of the complexity of the route. For a very short route, there was likely to be only one reasonable route as this would be the quickest temporally, the shortest distance-wise, and the most economical for fuel consumption. For more complex routes there could be more fundamental differences between route choices, often such as whether to take the highway or surface streets, where to get on or off the highway, or sometimes two opposing directions around a city area.

Google API was used to correct real driving GPS data from cellphones to produce the real route maps, shown in Figure 4 (a). As shown in Figure 4 (b), Ovitalmaps software was used to create the trace maps of the selected route from the recommendations of the eco-routing app, by stepping through the turn-by-turn directions and placing pins at critical points on a map. Then, corrected real driving GPS data were compared with the selected route data to examine drivers’ compliance. Any deviation on surface streets or highways from the recommended route would result in a trip being scored as “not follow the route”. Slight deviations within parking lots, shopping centers, apartment communities, and small, unmarked subdivisions at the beginning and end of a trip were outside of the scope of directions and would not result in a route being scored “not follow the route”. As illustrated in Figure 4 (c), the overlapping distance (the sector of CD) between a real driving trip (the sector of BCD) and a recommended route (the sector of ACD) was calculated based on the longitudinal and lateral coordinates.

(a) Real route recorded by GPS data; (b) Selected route from recommendations; (c) Road overlap correction

Fig.4. The real route and selected route Road

2.3 Mixed model analyses

To investigate the impacting factors on the drivers’ eco route choice and route compliance behavior, mixed model analyses were used in this study. Mixed models contain both fixed and random effects (Wu et al., 2016). In this study, fixed effects were variables from driver demographic characteristics, subjective data, and route information, while individual drivers and interactions between individual drivers and any fixed effects were chosen as random effects. Compared with those models that assumed the impacting factors on drivers' route choices and route compliance were the same across different observations, mixed models could provide more accurate estimations of contributing factors on drivers' behavior by considering individual heterogeneity (Wang et al., 2017; Yu et al., 2019a; Jermakian et al., 2017).

2.4 Multi-label random forests

Since each route recommended by navigation systems may have several different features at the same time, for example, one recommended route is the most fuel-efficient as well as the fast one, drivers' route choices belong to the multi-label problem. Thus, this study employed a Multi-label random forests (MLRF) method to model drivers' route choice behavior when interacting with eco-routing navigation systems. Different from classical multi-class classification problems, where each example can only belong to one single label from a given set of n labels ($n > 2$) that were mutually exclusive, the multi-label classification can simultaneously assign each instance with m target labels ($1 \leq m \leq n$) from all the n given labels (Madjarov et al., 2012). Rational multi-label classification models are capable to take underlying correlations among different labels into consideration (Huang and Zhou, 2012). There are several kinds of techniques to deal with multi-label problems, including problem transformation, adapted algorithms, and ensemble approaches, but many studies have examined that ensemble methods always perform better than other state-of-the-art approaches (Zhang et al., 2015; Rokach et al., 2014; Read et al., 2009).

The MLRF method pertains to ensemble learning approaches, which combines a multitude of multi-label decision trees (week learners) to achieve more accurate and stable prediction results (a strong learner) (Agrawal et al., 2013). The types of decision trees used in the traditional random forests model are typically used for multi-class classification but cannot solve the multi-label problems. To extend random forests to be able to perform multi-label classification, multi-label decision trees that can generalize the entropy of the sample set to adapt to multi-label data sets, are employed to develop the MLRF method (Qu et al., 2017; Clare and King, 2001). Specifically, decision trees are constructed top-down, and at each node, data are divided into subsets by finding the attribute that leads to the highest information gain. The information gain is calculated by the decrease in entropy after partitioning data based on an attribute, as shown in Equation (2):

$$IG(A,X) = entropy(A) - \sum_{x \in X} \frac{|A_x|}{|A|} entropy(A_x) \quad (2)$$

Where: $IG(A,X)$ is the information gain; A is a set of training data; X is an independent variable; A_x is a subset of A when the value of X is equal to x.

Entropy is a measure of impurity or disorder. In the traditional decision trees used for multi-class problems, the computational formula of entropy is as follows:

$$entropy(A) =- \sum_{i=1}^{N} p(Y_i) log_2 p(Y_i) \quad (3)$$

Where: N is the number of the labels of the dependent variable Y; $p(Y_i)$ is the probability of label Y_i in this set.

Multi-label decision trees extend the Equation (3) to calculate entropy for multi-label data sets, as shown in the following Equation (4):

$$entropy(A) = -\sum_{i=1}^{N}((p(Y_i)log_2 p(Y_i) + (1 - p(Y_i))log_2(1 - p(Y_i))) \quad (4)$$

One of the most important features of the MLRF model is that it can measure the importance of each variable (Wu et al., 2019). This model can also quickly and accurately process high-dimensional data without overfitting problems (Yu et al., 2019b). In addition, since each multi-label decision tree is built by different bootstrap samples, where approximately one-third of original data are left out as "Out of bag" (OOB) at random and are not used to grow this tree, the MLRF model is able to provide an internal unbiased estimate of the prediction accuracy (called the OOB accuracy) based on these OOB data (Qu et al., 2017). As illustrated in Figure 1, the process of the MLRF model consists of four steps: generating bootstrap samples, growing each multi-label decision tree, voting for the final results, and calculating the OOB accuracy and variable importance.

Tab. 1. The pseudo-code of the MLRF method

Algorithm: Multi-label random forests

Input: Original dataset D with the dimension o |D|=l*m (i.e., l samples and m independent variables).

Required parameters: a) N_{tree}: the number of multi-label decision trees constructed in the model;
b) K: the number of input variables tried at each node to search for the best split;
c) M: the maximum depth of the tree;
d) P_{in}: the minimum number of samples required for splitting at the internal node;
e) P_{ex}: the minimum number of samples required for splitting at the external node.

For i= 1 to N_{tree}

1. Generate a bootstrap sample D_i with | D_i |=|D| by randomly drawing with replacement from D, which includes around 2/3 of the observations from D.
2. Grow a multi-label decision tree T_i based on the bootstrap sample D_i: (a) randomly try K input variables at each node to search for the best split according to the highest IG calculated by Equations (2) and (4); (b) When the internal or external node reaches the threshold P_{in} or P_{ex}, this node will stop splitting; (c) When the depth of each tree reaches M, it will stop growing.

End

3. Vote for the set of labels by combining the results of all multi-label decision trees $\{T_1, T_2, \ldots, T_{Ntree}\}$. Then the final set is the one with the majority votes.
4. Estimate the OOB accuracy by considering the performance of each subset of OOB data in the corresponding multi-label tree, and calculate the variable importance based on the mean decrease in impurity.

3. Results

3.1 Modelling drivers' route choices

As mentioned above, drivers' route choices pertained to the multi-label problem. To mitigate the effects of unbalanced sample sizes for different drivers, the average probability of route choices was calculated based on each driver, as shown in Figure 5. In general, drivers were more likely to choose the fast route, having the highest average probability of 83.4%, followed by the eco route with a selection probability around 78.6%, while the routes with the balanced feature had the least likelihood to be selected, averagely 70.7% for each driver. There was a total of 18 variables coded for each trip, derived from driver characteristics, subjective data, and route information. After eliminating the highly correlated variables, such as gas consumption and driving time of the selected route were excluded due to their high correlations with distance, 14 variables were finally chosen as the input variables for further analysis. Table 2 demonstrates the detailed descriptions and distributions for all these candidate variables.

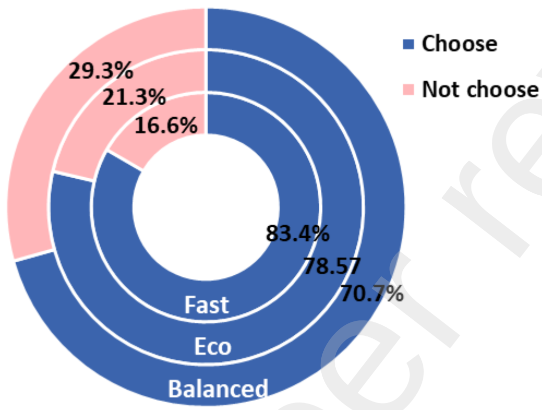


Fig.5. The average probability of route choices

Tab. 2. Definitions and distributions of input variables

Variables	Description (units)	Min	Max	Mean	S.D.
Route information					
Distance	Distance for the selected route (mile)	0.30	132.60	9.34	11.87
Distance saving	Distance differences between the longest and the shortest recommended routes (mile)	0	13.70	0.82	1.39
Average gas consumption	Gas consumption per mile for the selected route (gallon per mile)	0.02	0.09	0.04	0.01
Sequence	The recommendation sequence for different routes	1st (75.88%), 2nd (19.78%), 3rd (4.34%).			
Number of routes	The number of recommended routes	1 (36.45%), 2 (47.94%), 3 (15.58%).			
Driver characteristics					
Age	Younger: 20~50, Older:50~75	Younger (48.72%), Older (51.28%).			
Gender	Gender of drivers	Male (43.59%), Female (56.41%).			
Subjective data					
Purpose	Purpose of this trip	Household errands (5.01%), Personal business (9.08%), Picking up/dropping off (3.52%), Recreation (9.62%),			

		Returning home (29.27%), Shopping (11.11%), Socialization (6.10%), School /work (19.24%), Other (7.04%).			

Tab. 2. Definitions and distributions of input variables (continued)

Variables	Description (units)	Min	Max	Mean	S.D.
Decision time	When did the driver decide to take this trip?	Earlier today (10.43%), Several days or longer (8.54%), Just now (18.56%), Not sure (0.41%), Routine (56.37%), Yesterday (5.59%).			
Household passenger	How many household passengers were traveling with the driver?	0 (85.09%), 1 (9.49%), 2 (3.52%), 3 or more (1.90%).			
Non-household passenger	How many non-household passengers were traveling with the driver?	0 (95.93%), 1 (3.79%), 2 (0), 3 or more (0.27%).			
Flexibility	How flexible was the driver's arrival time at the destination?	Whenever (18.70%), Within 15 - 30 mins (4.20%), Within 5 - 15 mins (9.08%), Within 5 mins (68.02%).			
Prior activity	What activity where the driver engaged in prior to this trip?	Household errands (9.89%), Personal business (13.41%), Picking up/dropping off (2.30%), Recreation (9.21%), Returning home (6.37%), Shopping (10.43%), Socialization (6.64%), School/work (26.70%), Other (15.04%).			
Leave earlier	Was the driver able to leave earlier from the prior activity?	Maybe (9.89%), No (23.31%), Yes (66.80%).			

Figure 6 illustrates the trip purposes of the National household travel survey in 2017 (U.S. Department of Transportation Federal Highway Administration, 2020) and our experiment. The results showed that our study data were consistent with national household survey data on trip purposes, indicating our study data were representative. In both national survey data and this study, driving back to home accounted for the largest proportion of trip purposes, and the following was driving to work/school. Shopping/errands and social/recreational purposes ranked third and fourth, respectively.

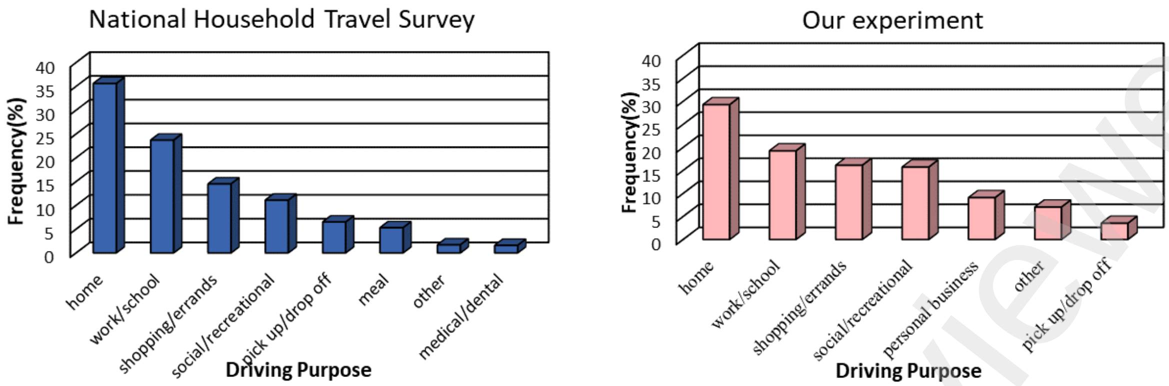


Fig.6. Trip purposes of national household travel survey data and our experiment

To further explore the impacting factors on the eco-routing choice, mixed model analyses were conducted in the statistical software package SAS 9.2 by using the PROC GLIMMIX procedure. All the input variables and their interactions were chosen as the fixed effects, while individual drivers and interactions between individual drivers and these fixed effects were treated as random effects. The dependent variable was whether to choose the eco route or not. After excluding the insignificant factors, the final model was shown in Table 3. Distance had a negative impact on the eco route choice ($t(643)=-5.56$, $p<0.001$), while average gas consumption positively affected the eco route choice ($t(643)=2.35$, $p=0.019$), indicating that drivers were more likely to select the eco route when this trip had short distance and higher gas consumption per mile. In addition, the route recommendation sequence also had a significant effect on choosing the eco route (all $p<0.001$), and giving priority to recommend the eco route could guide drivers to choose the eco way.

Tab. 1 Mixed model results for the eco-routing choice

Effect	Estimate	Standard error	DF	t Value	Pr>|t|
Intercept	1.139	0.715	38	1.59	0.119
Distance	-0.056	0.010	643	-5.56	0.019
Average gas consumption	40.868	17.395	643	2.35	<0.001
Sequence					
1st*	0				
2nd	-1.925	0.316	52	-6.10	<0.001
3rd	-2.518	0.491	52	-5.13	<0.001

*Note: * denotes reference group for categorical variables; only significant factors were demonstrated in this Table*

To predict drivers' route choice behavior, a Multi-label Random forests (MLRF) classification model was established by using the "scikit-learn" package in Python software (version 3.6). Those 14 variables mentioned above were selected as the independent variables, while the independent variable was what kind of route drivers would choose, which was a multi-label variable with three candidate features, i.e., eco, fast, and balanced. All 737 samples were partitioned randomly into 70% for training and 30% for testing. After 5- fold cross-validation, the parameters in MLRF were determined to make the model reach the stable and maximum accuracy: the number of trees was 550; the number of variables considered in each split was 4; the maximum depth of the tree was none; and the minimum number of samples required for splitting at the internal and external nodes were 30 and 10, respectively. The final prediction result was shown in Table 4. In the training group, the out-of-bag (OOB) accuracy was 88.3%, and the overall testing accuracy was 86.8%. As for the prediction results in each label, their precisions were greater than 80.0%. For comparison, several other machine learning methods that are commonly used for multi-label classification were also tried in this study, including Multi-label k-neighbors classifier (MLKNC), Multi-label support vector classification (MLSVC); Neural network multi-layer perceptron classifier (NNMLPC). For brevity, the introductions of these methods were not provided in this paper, and please see "scikit learn" (2020) for more information. The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the performance of different algorithms, and the ROC curve in the multi-label classification was measured by the average value of all labels. As shown in Figure 7, the AUC of the MLRF classification was 0.86 which was greater than others, indicating that the MLRF classification had a better performance.

Tab.2. Prediction results of the MLRF.

Label	Precision	Recall	f1-score
Eco	0.91	0.91	0.91
Fast	0.87	0.94	0.90
Balanced	0.83	0.89	0.86
Overall accuracy	Training (OOB): 0.883; Testing: 0.868		

Note: Precision=TP/(TR+FP); Recall=TP/(TP+FN); f1-score=2(Precision*Recall)/(Precision+Recall)*

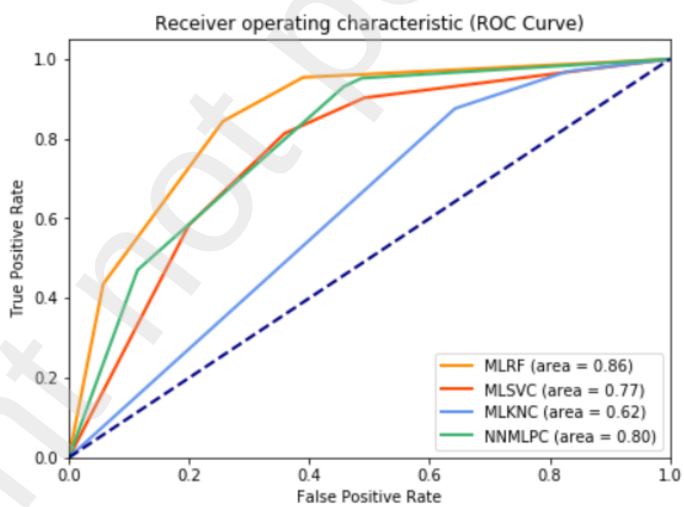


Fig. 7. ROC curves for multi-label classifiers

Figure 8 illustrates the variable importance that represented the statistical prioritization of independent variables regarding their contribution to the prediction model. Variables from route information showed the largest impacts on drivers' route choices, i.e., distance saving, recommendation sequence, distance, average gas consumption, and the number of recommended routes, ranking the top five of the feature importance. The following were subjective data, such as prior activities, the purposes of this trip, decision time, etc. However, no obvious relationships were found in demographic data, indicating that drivers' route choices were less likely to be affected by age and gender differences.

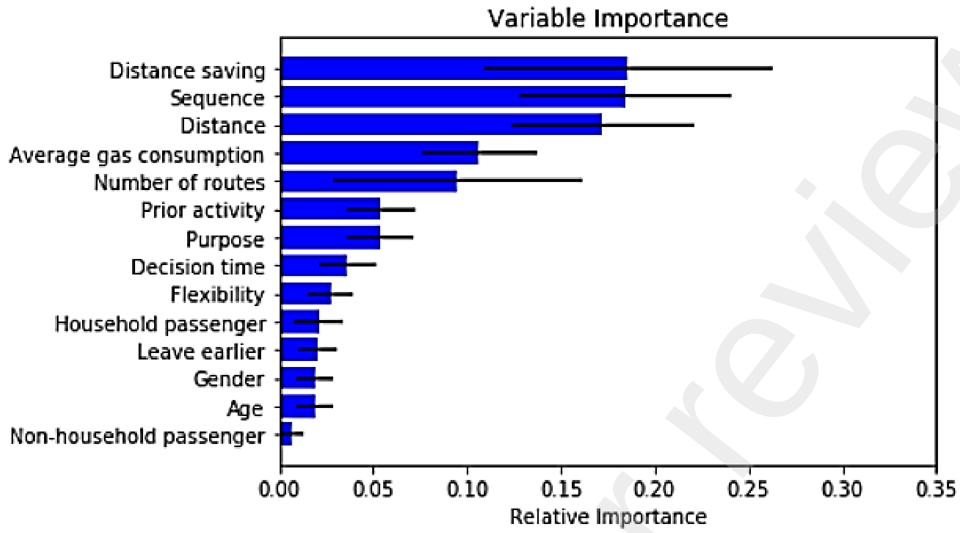
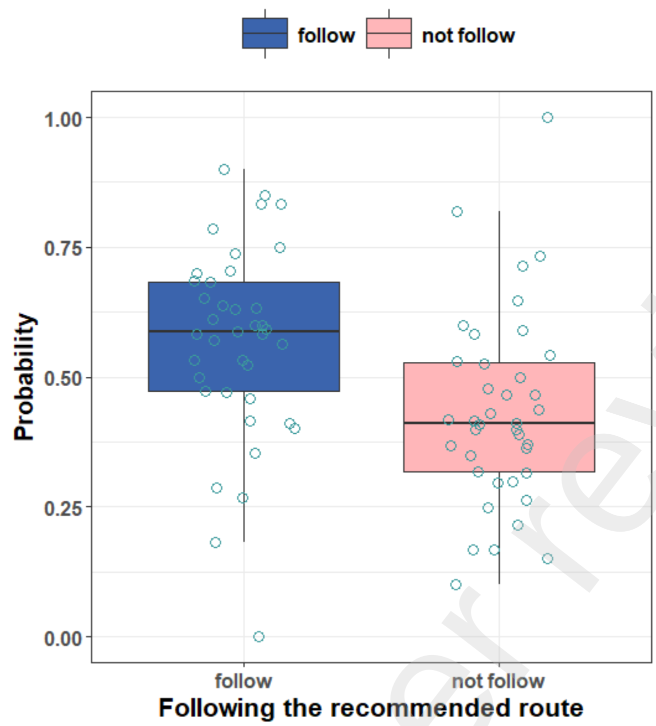


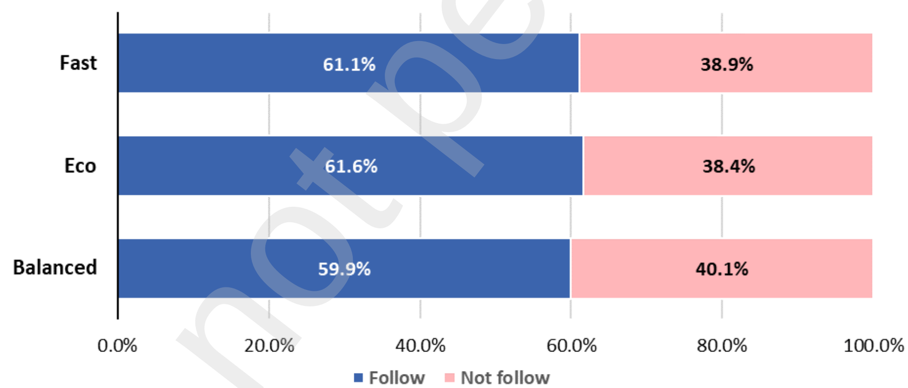
Fig. 8. Variable importance of the MLRF

3.2 Impacting factors on following the recommended route

Generally, the average probability that drivers would actually follow the route after they chose from the recommended options was 56.7%, as shown in Figure 9 (a). The detailed results of the following probability when they selected different categories of recommended routes were illustrated in Figure 9 (b). When drivers chose the eco, they had the largest likelihood (61.6%) to follow the route, followed by selecting the fast one with an averaged probability of 61.1%. When drivers selected the route with the balanced feature, they were the least likely to comply with their option (59.9%).



(a) The average probability of following the recommended route



(b) The average following probability of each recommended route

Fig. 9. The average probability of drivers' route compliance

To explore the impacting factors on following the recommended route, mixed model analyses were carried out by using the PROC GLIMMIX procedure. In total, 12 candidate variables were obtained from three sources:

(1) **Route information**: eco, fast, and balanced routes;

(2) **Driver characteristics**: age and gender;

(3) **Subjective data**: purpose, decision time, household passenger, non-household passenger, flexibility, prior activity, and leave earlier.

These variables together with interactions among them were treated as the fixed effects, while individual drivers and interactions between individual drivers and fixed effects were regarded as random effects. Whether drivers followed the selected route was the independent variable. The results were demonstrated in Table 5. When drive chose the eco (t(32)=3.61, *p*=0.001) or fast (t(31)=4.68, p<0.001) routes, they were more likely to fully drive along the recommended route. Additionally, compared with driving alone (t(22)=-2.81, p=0.010) or with only one household passenger (t(22)=-2.95, p=0.007), drivers were more willing to comply with the recommended route when there were three or more household passengers.

Tab. 3. Mixed model results for the recommended route following

Effect	Estimate	Standard error	DF	t Value	Pr>|t|
Intercept	-1.199	0.257	22	-4.67	<0.001
Eco route	0.776	0.215	32	3.61	0.001
Fast route	1.071	0.229	31	4.68	<0.001
Household member					
0	-2.394	0.852	22	-2.81	0.010
1	-2.635	0.893	22	-2.95	0.007
2	-2.003	0.989	22	-2.02	0.055
3 or more*	0				

*Note: * denotes reference group for categorical variables; only significant factors were demonstrated in this Table.*

4. Discussion and conclusion

This study aims to understand what factors are influencing drivers' decision-making on route choices and route compliance when interacting with eco-routing navigation systems. Although drivers' route choice and compliance behavior have been discussed in many previous studies, the impacts of eco-routing navigation systems have not been considered and examined in these studies. Therefore, this study developed an eco-routing navigation app and collected naturalistic driving data to explore and model what kind of recommended route drivers would choose and whether they would comply with their selected route.

In general, this study found that drivers would change their route choices under certain conditions when they were provided with information related to different routes by the eco-routing navigation app. With the help of this app, approximately 78.6% of all the selected routes had the most energy-saving feature. In contrast, only 54% of the trips based on drivers' spontaneous choices were the eco way without using eco-routing navigation systems (Ericsson et al., 2006). Results of the mixed model analyses showed that drivers were more willing to choose the eco route when this trip had shorter distances and higher gas consumption per mile. It was also found that prioritized recommendations for the eco route could make drivers prefer to choose the eco way. A study also reported that the relationship between the fuel saving and trip distance was significantly positive when the distance was not greater than 10 miles, while this kind of relationship became non-significant with a trip distance longer than 10 miles (Boriboonsomsin et al., 2014). A route with a high fuel consumption rate did not necessarily have high totals (Frey et al., 2008) so that these two features may have different effects on drivers' route choice. Age and gender differences were found in a route choice study without considering eco-routing systems (Li et al., 2016), and a web-based mimic experiment also observed gender differences of the trade-off between energy saving and travel time (Aziz and Ukkusuri, 2014). However, using the naturalistic driving data, this study did not find any significant effects of age and gender on drivers' route choices when drivers were provided with the eco-routing navigation app.

Since drivers' route choices belonged to the multi-label problem, this study applied a Multi-label random forests (MLRF) model to predict which kind of route drivers would choose from the recommendations offered by the eco-routing navigation app. A study has presented that due to many different label combinations, the accuracy is a harsher metric in the multi-label classification than in the multi-class one, where any possible labels for each sample need to be predicted and exactly match with the corresponding set of true labels (Read et al., 2008). In this study, the overall OOB accuracy and AUC of this MLRF model were 88.3% and 0.86 respectively, indicating that MLRF had a sterling performance in predicting drivers' route choice behavior. In addition, the results showed MLRF also performed better than other commonly-used multi-label classification methods.

Overall, the average proportion that participants followed the recommended route while driving was 56.7%. Mixed model analyses showed that when drivers chose the eco or fast routes, they had a higher probability of compliance with the recommended route. Relative to driving alone or with only one household passenger, drivers were more likely to fully drive along the recommended route when there were three or more household passengers. This might be because driving with three or more passengers makes drivers drive more carefully. It has been reported that as for drivers aged 25 to 64 years old, the crash risk when driving alone was about 12 times higher than that of having three or more passengers (Engström et al., 2008).

To reduce energy consumption and benefit the eco transportation system, findings in this study can help guide drivers to choose and comply with the eco route provided by navigation systems, after understanding factors associated with drivers' decision making. This study also contributes to the design of incentive-based programs to motivate eco-routing choices in drivers to (1) increase awareness about eco-routing options and consequences, (2) develop sustained driving styles in saving energy, and (3) enhance long term effect on public awareness. Fuel consumption feedback of eco-routing navigation systems can be designed and improved based on this study to reward drivers who follow the eco route in a long run. One limitation of this study is that the real-time traffic condition after selecting the route was not included in the analysis of route compliance. In the future, real-time information about energy consumption and traffic condition changes while driving will be provided in the calculation of the eco-routing option, so that corresponding driver behavior can be further analyzed.

Acknowledgments

This project was partially sponsored by the United States Department of Energy and the University of Michigan Center for Connected and Automated Transportation. The authors would like to thank Dr. Huei Peng for his great support. The authors would also like to thank Dillon Funkhouser, Willian Wang, Huajian Zhou, Zeyang Cheng, and Haibo Wang for the data collection and reduction.

References

Abdel-Aty, M., & Abdalla, M. F. (2004). Modeling drivers' diversion from normal routes under ATIS using generalized estimating equations and binomial probit link function. Transportation, 31(3), 327-348.

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

Ahn, K., & Rakha, H. A. (2013). Network-wide impacts of eco-routing strategies: A large-scale case study. Transportation Research Part D: Transport and Environment, 25, 119-130.

Agrawal, R., Gupta, A., Prabhu, Y., & Varma, M. (2013, May). Multi-label learning with millions of labels: Recommending advertiser bid phrases for web pages. In Proceedings of the 22nd international conference on World Wide Web (pp. 13-24). ACM.

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

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(4), 249-264.

Boriboonsomsin, K., & Barth, M. (2009). Impacts of road grade on fuel consumption and carbon dioxide emissions evidenced by use of advanced navigation systems. Transportation Research Record, 2139(1), 21-30.

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

Boriboonsomsin, K., Dean, J., & Barth, M. (2014). Examination of attributes and value of ecologically friendly route choices. Transportation Research Record, 2427(1), 13-25.

Clare, A., & King, R. D. (2001, September). Knowledge discovery in multi-label phenotype data. In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 42-53). Springer, Berlin, Heidelberg.

Dia, H., & Panwai, S. (2007). Modelling drivers' compliance and route choice behaviour in response to travel information. Nonlinear Dynamics, 49(4), 493-509.

EPA. (2019). U.S. Transportation Sector Greenhouse Gas Emissions: 1990–2017. Office of Transportation and Air Quality EPA-420-F-19-047.

Engström, I., Gregersen, N. P., Granström, K., & Nyberg, A. (2008). Young drivers—reduced crash risk with passengers in the vehicle. Accident Analysis & Prevention, 40(1), 341-348.

Ericsson, E., Larsson, H., & Brundell-Freij, K. (2006). Optimizing route choice for lowest fuel consumption–potential effects of a new driver support tool. Transportation Research Part C: Emerging Technologies, 14(6), 369-383.

Fiori, C., Ahn, K., & Rakha, H. A. (2018). Optimum routing of battery electric vehicles: Insights using empirical data and microsimulation. Transportation Research Part D: Transport and Environment, 64, 262-272.

Frey, H. C., Zhang, K., & Rouphail, N. M. (2008). Fuel use and emissions comparisons for alternative routes, time of day, road grade, and vehicles based on in-use measurements. Environmental Science & Technology, 42(7), 2483-2489.

Handy, S., Weston, L., & Mokhtarian, P. L. (2005). Driving by choice or necessity?. Transportation Research Part A: Policy and Practice, 39(2-3), 183-203.

Huang, S. J., & Zhou, Z. H. (2012, July). Multi-label learning by exploiting label correlations locally. In Twenty-sixth AAAI conference on artificial intelligence.

Jermakian, J. S., Bao, S., Buonarosa, M. L., Sayer, J. R., & Farmer, C. M. (2017). Effects of an integrated collision warning system on teenage driver behavior. Journal of safety research, 61, 65-75.

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.

Lois, D., Wang, Y., Boggio-Marzet, A., & Monzon, A. (2019). Multivariate analysis of fuel consumption related to eco-driving: Interaction of driving patterns and external factors. Transportation Research Part D: Transport and Environment, 72, 232-242.

Madjarov, G., Kocev, D., Gjorgjevikj, D., & Džeroski, S. (2012). An extensive experimental comparison of methods for multi-label learning. Pattern recognition, 45(9), 3084-3104.

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

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

Prato, C. G., & Bekhor, S. (2007). Modeling route choice behavior: How relevant is the composition of choice set?. Transportation Research Record, 2003(1), 64-73.

Qu, H. & Liu, H. & Li, X. & Huang, J. (2017). Recognition of multiple power quality disturbances using multi-label random forest. Power System Protection and Control, 45(11): 1-7.

Read, J., Pfahringer, B., & Holmes, G. (2008). Multi-label classification using ensembles of pruned sets. In 8th IEEE international conference on data mining (pp. 995-1000). IEEE.

Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2009, September). Classifier chains for multi-label classification. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 254-269). Springer, Berlin, Heidelberg.

Rokach, L., Schclar, A., & Itach, E. (2014). Ensemble methods for multi-label classification. Expert Systems with Applications, 41(16), 7507-7523.

Salvi, B. L., & Subramanian, K. A. (2015). Sustainable development of road transportation sector using hydrogen energy system. Renewable and Sustainable Energy Reviews, 51, 1132-1155.

Scikit learn. (2020). 1.12. Multiclass and multilabel algorithms. Retrieved January 16, 2020, from https://scikit-learn.org/stable/modules/multiclass.html.

Sivak, M., & Schoettle, B. (2012). Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. Transport Policy, 22, 96-99.

Shakeel, K., Rashidi, T. H., & Waller, T. S. (2016). Choice set formation behavior: joint mode and route choice selection model. Transportation Research Record, 2563(1), 96-104.

Tawfik, A. M., Rakha, H. A., & Miller, S. D. (2010, September). An experimental exploration of route choice: Identifying drivers choices and choice patterns, and capturing network evolution. In 13th International IEEE Conference on Intelligent Transportation Systems (pp. 1005-1012). IEEE.

The U.S. Department of Energy. (2020). Driving More Efficiently. Retrieved January 16, 2020, from https://www.fueleconomy.gov/feg/driveHabits.jsp.

U.S. Department of Transportation Federal Highway Administration (2020). National Household Travel Survey in 2017. Retrieved January 16, 2020, from https://nhts.ornl.gov/vehicle-trips.

Wang, Y., Bao, S., Du, W., Ye, Z., & Sayer, J. R. (2017). A spectral power analysis of driving behavior changes during the transition from nondistraction to distraction. Traffic injury prevention, 18(8), 826-831.

Wang, J., Elbery, A., & Rakha, H. A. (2019). A real-time vehicle-specific eco-routing model for on-board navigation applications capturing transient vehicle behavior. Transportation Research Part C: Emerging Technologies, 104, 1-21.

Wang, G., Makino, K., Harmandayan, A., & Wu, X. (2020). Eco-driving behaviors of electric vehicle users: A survey study. Transportation Research Part D: Transport and Environment, 78, 102188.

Wu, Q., Zhang, G., Chen, C., Tarefder, R., Wang, H., & Wei, H. (2016). Heterogeneous impacts of gender-interpreted contributing factors on driver injury severities in single-vehicle rollover crashes. Accident Analysis & Prevention, 94, 28-34.

Wu, X., Gao, Y., & Jiao, D. (2019). Multi-Label Classification Based on Random Forest Algorithm for Non-Intrusive Load Monitoring System. Processes, 7(6), 337.

Yu, B., Bao, S., Feng, F., & Sayer, J. (2019a). Examination and prediction of drivers' reaction when provided with V2I communication-based intersection maneuver strategies. Transportation Research Part C: Emerging Technologies, 106, 17-28.

Yu, B., Chen, Y., & Bao, S. (2019b). Quantifying visual road environment to establish a speeding prediction model: an examination using naturalistic driving data. Accident Analysis & Prevention, 129, 289-298.

Zeng, W., Miwa, T., & Morikawa, T. (2017). Application of the support vector machine and heuristic k-shortest path algorithm to determine the most eco-friendly path with a travel time constraint. Transportation Research Part D: Transport and Environment, 57, 458-473.

Zhang, W., Liu, F., Luo, L., & Zhang, J. (2015). Predicting drug side effects by multi-label learning and ensemble learning. BMC bioinformatics, 16(1), 365.

Zhao, Y., Yao, S., Liu, D., Shao, H., & Liu, S. (2019, June). GreenRoute: A Generalizable Fuel-Saving Vehicular Navigation Service. In 2019 IEEE International Conference on Autonomic Computing (ICAC) (pp. 1-10). IEEE.