Knowledge-Based Systems 218 (2021) 106882

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Extrapolation-enhanced model for travel decision making: An ensemble machine learning approach considering behavioral theory

Kun Gao [a], Ying Yang [b,*], Tianshu Zhang [a], Aoyong Li [c], Xiaobo Qu [a]

[a] Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg SE-412 96, Sweden
[b] School of Behavioral and Health Sciences, Australian Catholic University, Australia
[c] Institute for Transport Planning and Systems (IVT), ETH Zürich, Zürich CH-8093, Switzerland

ARTICLE INFO

Article history:
Received 24 October 2020
Received in revised form 4 January 2021
Accepted 16 February 2021
Available online 18 February 2021

Keywords:
Travel behavior
Machine learning
Intelligent transport system
Behavioral theory

ABSTRACT

Modeling individuals' travel decision making in terms of choosing transport modes, route and departure time for daily activities is an indispensable component for transport system optimization and management. Conventional approaches of modeling travel decision making suffer from presumed model structures and parametric specifications. Emerging machine learning algorithms offer data-driven and non-parametric solutions for modeling travel decision making but encounter extrapolation issues (i.e., disability to predict scenarios beyond training samples) due to neglecting behavioral mechanisms in the framework. This study proposes an extrapolation-enhanced approach for modeling travel decision making, leveraging the complementary merits of ensemble machine learning algorithms (Random Forest in our study) and knowledge-based decision-making theory to enhance both predictive accuracy and model extrapolation. The proposed approach is examined using three datasets about travel decision making, including one estimation dataset (for cross-validation) and two test datasets (for model extrapolation tests). Especially, we use two test datasets containing extrapolated choice scenarios with features that exceed the ranges of training samples, to examine the predictive ability of proposed models in extrapolated choice scenarios, which have hardly been investigated by relevant literature. The results show that both proposed models and the direct application of Random Forest (RF) can give quite good predictive accuracy (around 80%) in the estimation dataset. However, RF has a deficient predictive ability in two test datasets with extrapolated choice scenarios. In contrast, the proposed models provide substantially superior predictive performances in the two test datasets, indicating much stronger extrapolation capacity. The model based on the proposed framework could improve the precision score by 274.93% than the direct application of RF in the first test dataset and by 21.9% in the second test dataset. The results indicate the merits of the proposed approach in terms of prediction power and extrapolation ability as compared to existing methods.

1. Introduction

Modeling and predicting spatiotemporal travel demands in a transport system is one of the vital cores and indispensable inputs for the optimization and operation management of transportation systems. These require explicit and precise understandings regarding how travelers make their decisions in terms of transport modes, routes and departure time for daily activities in the transport systems with considerations of both objective (e.g., level-of-service variables of transport choices such as travel time, cost, and comfort) and subjective factors (e.g., socio-economic attributes and subjective attitudes). For instance, if 10 000 people commute from area A to area B every day, the transport system managers need to predict how many people will choose different transport modes (e.g., bus, metro, car, and bike) and which routes they will select, for planning, managing and optimizing the transport systems. Thus, accurate modeling concerning various individuals' travel decision making is crucial for efficient transportation management and optimization.

The conventional methods for modeling travel decision making such as Multinomial Logit Model [1] and Mixed Logit Model [2] have dominated travel demand analysis for several decades. The traditional methods are founded on recognized behavioral theories such as Utility Maximization Theory and Regret Minimization Theory. Such behavioral theories define the criteria of measuring the individuals' objectives or the optimal outcomes in decision making. For instance, the foundation of Utility Maximization Theory assumes that the decision-makers are utility maximizer, while the Regret Minimization Theory deems the

* Corresponding author.
E-mail address: ying.yang@acu.edu.au (Y. Yang).

https://doi.org/10.1016/j.knosys.2021.106882
0950-7051/

decision-makers aims to minimize the regrets in decision making rather than to obtain the highest utility. Based on the behavioral theories, specific mathematical frameworks and model structures (e.g., logit models) have been developed for modeling travel decision making regarding departure time, route, and mode choices [2]. However, conventional models have essential limitations such as wooden model structures (e.g., generally using linear functions), parametric model specifications, assumptions about behavioral stochastics and parameter distributions, and difficulty to consider complex interactions among explanatory variables and nonlinear effects of explanatory variables [3–6].

Recently, Machine Learning (ML) algorithms such as Support Vector Machines and Decision Trees, provide potential solutions for the above-mentioned limitations of conventional methods [4,6,7]. ML algorithms do not impose strict assumptions into the prior model formulations and learn the complex relationships between influencing factors (i.e., explanatory variables) and travelers' choices (i.e., dependent variable) in a data-driven way [7,8]. The essential characteristics and data-driven mechanisms of ML algorithms enable them to solve the above-mentioned deficiencies of conventional methods such as addressing the nonlinear relationships among the dependent variable and explanatory variables, and automatically learning complex interactions among explanatory variables. The advantages of ML have attracted many researchers to utilize them in empirical studies concerning travel decision making (e.g., [4,6,8–11]). However, the merits of ML algorithms are accompanied with prices as well. The data-driven mechanisms of ML algorithms also mean ignoring the behavioral mechanisms underlying travelers' decision making in the modeling framework. More specifically, the fitted relationships between explanatory variables and dependent variables based on ML algorithms are entirely data-driven and not established on explicit behavioral theories or causality. These will result in low generality and extrapolation of the trained models. The extrapolation in the contexts of travel decision makings refers to the ability of trained models to predict decision making in choice scenarios that do not exist in the training samples. ML algorithms such as Decision Trees have a vital issue: the extrapolation issue. They cannot predict data with features outside the range of the training samples due to their data-driven mechanisms and ignorance of causality in model structures [12]. Nonetheless, one of the top aims of modeling travel decision making is to serve travel demand forecasting and appraisals in the future with currently nonexisting choice scenarios. For instance, transport managers calibrate a travel choice model based on empirical datasets, obtain model parameters and use it to predict travelers' choices if a new policy (e.g., congestion pricing and intelligent management strategies) or a new transport service (e.g., a new rapid bus line) is introduced. Therefore, the extrapolation ability to predict the travel decision making choices in the currently nonexistent situations is indispensable and crucial.

This study stands in the wake of relevant literature to propose an extrapolation-enhanced approach for modeling travel decision making by integrating ensemble machine learning with knowledge-based decision-making theory. The proposed approach takes advantage of tree-based ensemble ML algorithms to address the complex and nonlinear relationships between the dependent variable and explanatory variables, as well as the complicated interactions among explanatory variables. Differing from the literature that overlooked the behavioral mechanisms in the modeling, we, simultaneously, embed the recognized knowledge-based decision-making theories from behavioral science in the modeling framework to fully consider the behavioral mechanisms underlying travel decision making. The integration of knowledge-based decision making theories and data-driven ensemble ML algorithms in this study makes the best of their mutual advantages to enhance both the modeling power and extrapolation ability in prediction. As far as we are concerned, this study is the first work to fuse the knowledge-based behavioral theories into data-driven ML algorithms for modeling travel decision making. Moreover, the proposed approach is examined using three different datasets of travel decision making, including one estimation dataset (for cross-validation) and two test datasets (for model extrapolation tests). Importantly, we use two test datasets containing extrapolated choice scenarios with features that exceed the ranges of training samples, to particularly examine the predictive ability of proposed models in extrapolated choice scenarios, which have never been investigated by relevant literature. This study provides a novel method for modeling and predicting travel decision making, and thus supports the efficient and scientific management and optimizations of transport systems.

The rest of the paper is structured as follows. The next section provides a literature review of relevant studies and identifies the research gaps with discussions. Section 3 presents details about the proposed approach and modeling framework. Section 4 describes the used datasets and provides detailed processes of model estimations and tests. The results and analysis are presented in Section 5, followed by concluding remarks in the last section.

2. Literature review

ML algorithms have been utilized for modeling travel decision making by treating travel decision making as a classification problem. More specifically, the task of modeling travel decision making is to predict the traveler's chosen option when available alternatives for a trip and their attributes are given. For instance, there are four possible transport modes (e.g., bus, bicycle, private car, and metro) from place A to place B, and their attributes (e.g., travel time, cost, and comfort) are known by the travelers (e.g., from Google map); the task is to model which transport mode a traveler will choose for the trip. This task is analogous to making classifications (e.g., choosing car or not for a trip) based on features of possible options. Therefore, prevalent ML classification algorithms such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Decision Trees (DTs), have been adopted in the relevant literature for modeling travel decision making. Shmueli et al. [13] conducted one of the earliest studies that applied ML methods for modeling travel decision making. They used neural networks including a simple Multi-layer Perceptron (MLP) and Classification and Regression Trees (CART) for modeling travel mode choices. They found that both MLP and CART performed similarly and theoretically advantageous than conventional models such as Multinomial Logit Model. Several other empirical studies (e.g., [14–18]) reported that the ANNs presented better predictive capability as compared to traditional Multinomial Logit Model and Nest Logit models as well. Besides, some studies used SVMs for modeling travel decision making. Zhang and Xie [17] and Omrani [19] demonstrated that the SVMs had the highest accuracy for predicting travel choice in contrast to ANNs and Multinomial Logit Model. As for DTs, Sekhar and Madhu [20] compared the prediction accuracy of Random Forest and traditional Multinomial Logit Model, and Cheng et al. [4] developed a robust Random Forest for modeling travel mode choices by incorporating built environment factors. Hagenauer and Helbich [9] conducted a comparative study on the predictive accuracy of seven different machine learning classifiers and reported that Random Forest performed the highest accuracy. Recently, a comparative study was executed by Zhao et al. [6] to investigate both the predictive accuracy and behavioral interpretations of traditional models and seven ML algorithms for modeling travel decision making. They also implied that the Random Forest showed the best performances. In summary, the tree-based ML algorithms such as Random Forest are declared to perform much better than other categories of ML classifiers as per the literature (e.g., [4,6,8,10]).

However, every coin has two sides. The data-driven mechanism endows ML algorithms the ability to depict the complicated relationships between influencing factors and travelers' choices without imposing predefined model structures and specifications. However, the data-driven paradigm of ML algorithms has its shortcomings due to lacking causality, especially in the contexts of modeling travel decision making. As we all know, ML classification algorithms determine the class of a data point merely based on the features of the data point. For instance, classifying a fruit is apple or pear as per its features such as shape, color, and tuber. In the contexts of modeling travel decision making, plenty of literature (e.g., [4,16,18,21]) directly transplanted the paradigm of ML classification algorithms into modeling travel decision making and make classifications (whether a transport mode would be chosen or not by a traveler for a trip) based on the attributes of using the transport mode for the trip. For instance, Cheng et al. [4] collected the travel attributes of every trip made by the respondents within the 24 h of the previous day, including used transport mode for each trip, travel time of the trip, trip purpose, built environment around the origin of each trip and personal attributes. They established a Random Forest to predict the travel mode choices, in which the dependent variable was the selected transport mode of each trip and input explanatory variables were previously mentioned attributes of the trip. In other words, the inputs of the Random Forest in [4] for predicting the traveler's decision making regarding transport mode are merely the features (e.g., travel time) of using a transport mode for a trip. The underlying assumption of modeling travel decision making in such a way is that the travelers make the travel decision merely based on the attributes of using a transport mode for a trip, which is implausible. Let us use an example (shown in Table 1) for illustration. An individual could use the bus or metro from home to shopping mall 1 and 2 for shopping. The attributes of using bus and metro to shopping mall 1 and 2 are demonstrated in Table 1. In choice scenario A, using the bus is superior due to shorter travel time and selected by the traveler, whilst using the metro is better in choice scenario B and chosen by the same traveler. If an ML algorithm like Cheng et al. [4] makes predictions based on the attributes of an alternative, it will predict the bus is either chosen or not chosen in both scenarios A and B as the attributes of the bus are the same in the two scenarios. Such predictions are obviously not in line with reality. The theoretical reason is that the real behavioral mechanisms behind travel decision makings are that the decision-makers make trade-offs among available alternatives (comparing their attributes) for a specific trip and select the option with the highest subjective benefits (e.g., maximum utility or minimum regret) after elaborate comparisons [22,23]. Direct transplantations of ML classification algorithms into travel decision making cannot reflect such behavioral mechanisms and causality underlying travelers' choices. More importantly, the ignorance of behavioral mechanisms will result in noticeable biases in predictions, as illustrated in Table 1. Unfortunately, most relevant literature utilizing ML algorithms for modeling travel decision making (e.g., [16,18,21]) used the same modeling methods as in [4] and did not contemplate the compatibility of ML classifiers in the contexts of modeling travel decision making.

The behavioral mechanisms underlying travelers' decision making are that they compare the differences in features of different available alternatives for a trip to select the perceived best one, rather than make a choice merely relying on the attributes of an alternative. It is not plausible to determine whether a transport mode will be chosen for a specific trip merely by the attributes of the transport mode. What makes differences in the travelers' decision-making process is the comparative superiority of a transport mode for a specific trip as compared to other competing options. From the technical perspective, the ML classification algorithms such as SVMs, ANNs and DTs are initially customized to execute classifications based on the features of a data point, rather than designed to make classifications based on the discrepancies in the features of different data points. These may be the reason that many studies using ML classification algorithms determine whether a transport mode is chosen merely relying on the attributes of the transport mode. Nonetheless, these are not in accord with travelers' behavioral mechanisms. Lhéritier et al. [10] and Zhao et al. [6] were aware of the biases of not considering competitive alternatives in the ML algorithms in modeling travel decision makings and put the non-chosen alternatives in their data. Nevertheless, they just included both the chosen and the foregone alternatives for a trip in the training dataset and used a similar modeling structure as previous studies. More especially, in their models, whether a transport mode will be chosen or not, still merely depended on the attributes of a transport mode, rather than the discrepancies among different available transport modes for finishing a trip. Therefore, their methods still do not construct the structure of ML algorithms in a proper way that could reflect the behavioral mechanisms behind travel decision making

Table 1
An illustration example.

Factors	Choice scenario A: from home to shopping mall 1		Choice scenario B: from home to shopping mall 2	
	Bus	Metro	Bus	Metro
Cost	5 CNY	5 CNY	5 CNY	5 CNY
Travel time	45 min	65 min	45 min	35 min
Real choice	Chosen	Not chosen	Not chosen	Chosen

Note: 1 CNY = 0.14 dollar.

The consequences of neglecting behavioral mechanisms in modeling travel decision making are the insufficiency in model generality and extrapolation, namely the ability to predict the decision making in choice scenarios that are not existing in the training samples. Relevant literature generally reported that ML algorithms had better prediction accuracy as compared to traditional models such as MNL based on their datasets. However, it should always be noted that the so-called predictive accuracy in the relevant literature was the predicting outcomes in cross-validations. The used datasets were generally from questionnaire-based SP surveys and contained limited possible choice scenarios and sample size, meaning the variances in the features of data points were not substantial. The typical cross-validation process randomly splits a dataset into several (e.g., 5 or 10) subsets with the same size, and selects one subset for validation and the rest for training. The features (e.g., travel cost and time) of data points in the validation samples are generally in similar ranges with those in the training samples and do not contain features that are out of the scope of the training samples. It is thus not surprising and even expected that the so-called predictive accuracy of ML algorithms is better than traditional models such as Multinomial Logit Model. Since ML algorithms do not restrict the model structures, prioritize the predictive power in model formulation and train the model in a data-driven manner. These are all beneficial for increasing the model fitness in used datasets [6,8,24]. ML algorithms such as Random Forest with large tree depths and tree amount could even enumerate all possible choice scenarios and reach a very high predictive accuracy in cross-validations, on account that there are only a few choice scenarios in the used datasets in the literature. As far as we are concerned, no empirical study concerning travel decision making has examined the predictive accuracy of the ML algorithms in forecasting the decision makings in the choice scenarios that are out of the range of used training samples, namely the model extrapolation. Data-driven ML algorithms essentially have vital extrapolation issues, as they are not capable of predicting the classes of samples with features beyond the range of the training samples [12]. From the perspectives of behavioral modeling, the issue of extrapolation in ML algorithms for modeling travel decision making originates from the absence of considering behavioral mechanisms behind travel decision making in the models and just modeling travel decision making in a purely data-driven way. More specifically, the trained ML algorithms in the relevant literature do not reflect the behavioral mechanisms that travelers make the final decision depending on the comparisons among available alternatives for a trip. Nonetheless, the extrapolation ability to predict the travel choices in the currently non-existent choice situations is crucial in the contexts of modeling travel decision making. Simultaneously, It is rather common that available datasets for training a travel choice model only cover limited choice scenarios from sampled respondents, and it is practically implausible to obtain datasets covering all possible choice situations in reality as every individual has different travel contexts and thus different choice scenarios. Traditional models such as Multinomial Logit Model have acceptable extrapolation abilities since they are based on explicit behavioral theories and reflect the behavioral mechanisms behind travel decision making in the models. Although ML algorithms show superiority in some aspects (e.g., flexible model structures), they are still insufficient in model extrapolations due to ignoring behavioral mechanisms in the model.

To fill the above-mentioned existing gaps, this study proposes an innovative extrapolation-enhanced approach for modeling travel decision making by leveraging the mutually complementary merits of ensemble machine learning and knowledge-based decision-making theory. More precisely, the proposed approach integrates the recognized knowledge-based decision-making theories from behavioral science into the modeling framework of ensemble ML models to take the behavioral mechanisms into accounts. This is, as far as we are concerned, the first study to embed the behavior theories into ML algorithms for modeling travel decision making to enhance predictive ability and extrapolation. Moreover, three different empirical datasets are used to examine the modeling capacity and prediction power of proposed approaches as compared to the methods in the literature. Particularly, we empirically examine the extrapolation ability of proposed approaches and models in existing studies in predicting travelers' decision making in extrapolated choice scenarios that are out of the range of training samples, which, to our best knowledge, have been seriously overlooked in the relevant literature.

3. Methodology

3.1. A new method of reformulating travel decision making as a classification problem

This subsection introduces how we reformulate travel decision making to be a classification problem using a new paradigm. Modeling travel decision making aims to depict and predict how decision-maker i makes his/her choices facing n alternatives for finishing a trip (e.g., commuting from home to workplace) under a specific choice scenario j. We take travel decision making concerning transport mode choices for example in this study. Let us use a cell with tuples in Table 2 to demonstrate the choice scenario. x_{ijqm} represents the value of mth feature of qth alternative (i.e., transport mode) in the choice scenario j. We use $\mathbf{X}_{ij} = \{x_{ijqm} | q = 1, 2, \dots, n; m = 1, 2, \dots, k\}$ to represent the matrix containing the features of all available alternatives in the choice scenario j for decision-maker i. The features of alternatives in $\mathbf{X}_{ij}$ can be different types of influencing factors related to travel decision making, including the level-of-service attributes of alternatives (e.g., travel cost and time) in the choice scenario, travelers' personal characteristics (e.g., age and gender), and contextual factors (e.g., built environments). The level-of-service attributes vary with different alternatives, but features such as travelers' characteristics and contextual factors are the same for all alternatives in a choice scenario (because the same person evaluates all the alternatives in a choice scenario for the same trip). Available alternatives (e.g., the number of available alternatives in a choice scenario) and the features of alternatives may be divergent across different choice scenarios. $C_{ijq} \in \{0, 1\}$ denotes a binary variable representing whether the qth alternative is chosen (i.e., 1) or not chosen (i.e., 0) by decision-maker i in the choice scenario j. The ultimate aim is to model the relationships between C_{ijq} and $\mathbf{X}_{ij}$ based on training data and use the trained model to predict travel decision making in other choice scenarios.

Treating travel decision making as a classification problem, most existing work does not include the features of non-chosen alternatives in a choice scenario and merely considers the features of chosen alternatives in training models (e.g., [4,16,21]). These apparently cannot reflect the behavioral mechanisms that travelers make the travel decisions based on comparisons among alternatives in a choice scenario. We believe that it is indispensable to include information about available but non-chosen alternatives in a choice scenario to develop a general model with enhanced extrapolations. Some studies included the features of all alternatives $\mathbf{X}_{ij}$ as explanatory variables in the models and treated the travel decision making as a multi-class classification regression. The dependent variable is the class that a data point belongs to, which is an analogy to the chosen transport mode. For instance, Zhao et al. [6] considered the attributes of chosen and non-chosen alternatives in a choice scenario as the input features and the chosen transport mode as the output for training ML classification algorithms. In [6], each attribute of a possible alternative in a choice scenario is a feature in the model. However, this method merely considered the features of all possible alternatives in a choice scenario but are still not able to reflect underlying behavioral mechanisms. Moreover, another vital drawback of this approach is that it is incompatible for modeling datasets with various numbers of alternatives in different choice scenarios, which is a common phenomenon in travel decision making analysis. For instance, there are 3 and 4 available commuting alternatives for individual i and z due to their divergent commuting contexts, respectively. If an ML algorithm is trained based on datasets of 3 alternatives in a choice scenario and 4 features for each alternative (thus $3 \times 4 = 12$ features for the model), the algorithm cannot be used to predict the choices in the scenarios with 4 alternatives and the same 4 feature for each alternative (in total $4 \times 4 = 16$ features). The reason is that ML algorithms cannot predict data points with nonexistent features in the training samples.

To solve the above shortcomings, we propose an alternative modeling method. We treat a choice scenario with a set of possible alternatives (one of which is chosen by the decision-maker) as a data point. For each data point shown in Table 2, the explanatory variables are the matrix of features of available alternatives in the choice scenario (namely $\mathbf{X}_{ij}$) and the dependent variable is the matrix $\mathbf{C}_{ij} = (C_{ij1}, C_{ij2}, \dots, C_{ijq}, \dots, C_{ijn})$ denoting whether an available transport mode is chosen in the choice scenario. The probability of C_{ijq} being 1 (i.e., chosen) or 0 (i.e., not chosen) in a choice scenario is predicted based on $\mathbf{X}_{ij}$ using a soft binary classification algorithm. The soft classification refers to the process of estimating the conditional probability $P(C_{ijq}|\mathbf{X}_{ij})$

Table 2
An example of choice scenarios and data formatting.

Scenarios	Alternatives (e.g., transport modes)	Features						Choices
	1	x_{ij11}	x_{ij12}	…	x_{ij1m}	…	x_{ij1k}	C_{ij1}
	2	x_{ij21}	x_{ij22}	…	x_{ij2m}	…	x_{ij2k}	C_{ij2}
j	…	…	…	…	…	…	…	…
	q	x_{ijq1}	x_{ijq2}	…	x_{ijqm}	…	x_{ijqk}	C_{ijq}
	…	…	…	…	…	…	…	…
	n_j	x_{ijn1}	x_{ijn2}	…	x_{ijnm}	…	x_{ijnk}	C_{ijn}
	1	$x_{i(j+1)11}$	$x_{i(j+1)12}$	…	$x_{i(j+1)1m}$	…	$x_{i(j+1)1k}$	$C_{i(j+1)1}$
	2	$x_{i(j+1)21}$	$x_{i(j+1)22}$	…	$x_{i(j+1)2m}$	…	$x_{i(j+1)2k}$	$C_{i(j+1)2}$
$j+1$	…	…	…	…	…	…	…	…
	q	$x_{i(j+1)q1}$	$x_{i(j+1)q2}$	…	$x_{i(j+1)qm}$	…	$x_{i(j+1)qk}$	$C_{i(j+1)q}$
	…	…	…	…	…	…	…	…
	$n_{(j+1)}$	$x_{i(j+1)n1}$	$x_{i(j+1)n2}$	…	$x_{i(j+1)nm}$	…	$x_{i(j+1)nk}$	$C_{i(j+1)n}$

of each alternative belonging to a class (e.g., $C_{ijq} = 1$) in a choice scenario when $\mathbf{X}_{ij}$ is given, rather than to merely provide a predicted class label (i.e., so-called hard classification). Therefore, the model predicts the chosen probability of each alternative in a choice scenario based on $\mathbf{X}_{ij}$. We use the estimated chosen probability of each alternative in a choice scenario as scores to rank the alternatives. The alternative with the highest chosen probability is regarded as the preferred and chosen option by the decision-maker under the given $\mathbf{X}_{ij}$ in the choice scenario. All other alternatives in the choice scenario are judged to be not chosen, on account that only one alternative will be selected by the decision-maker in a choice scenario in the contexts of travel behavior. Through the above process, the proposed method can consider all available alternatives in a choice scenario and accommodate modeling choice scenarios with different numbers of alternatives, which are especially crucial for the model extrapolation to predict choice scenarios that are not existent in the training samples.

3.2. The modeling framework and process

As aforementioned, the insufficiency of directly using ML classification algorithms in the current literature for modeling travel decision making is that the model structure deems the probability of choosing an alternative under a specific choice scenario is determined by the features of the alternative, and have no relationships with other possible alternatives in the choice scenario. In such a way, the model is unable to reflect the behavioral mechanisms underlying travelers' observed choices and merely build on a purely data-driven manner. These result in the deficiency of model extrapolation. We propose a novel solution for this issue by integrating knowledge-based decision-making theories into ML algorithms, as demonstrated in Fig. 1. The basic thinking of the proposed solution is to reconstruct the features of each alternative in a choice scenario described in Section 3.1, based on knowledge-based decision-making theories. This aims to reflect the facts that decision-makers make decisions hinging on the differences across alternatives rather than the absolute values of features of an alternative. We herein use a detailed example to demonstrate the process. Table 3 provides an illustrative example of a choice scenario. The original features of possible transport modes in the choice scenario are the level-of-service attributes of the available alternatives such as travel time, cost, and crowding level in the carriage. Tailored feature transformation methods based on behavioral theories are proposed to reformulate each feature (e.g., x_{ijqm}) of an alternative to be a new transformed feature (e.g., $T(x_{ijqm})$) that reflects the superiority or inferiority of the feature of the alternative as compared to the same feature of other alternatives in the choice scenarios, namely $x_{ijqm} \to T(x_{ijqm})$. This embeds the behavioral mechanism that what matters in travel decision making is the comparison with competing options. For instances, the cost of car in the example of Table 3 is 30 and will be transformed to be T_{cost} (30) where the function $T\left(x_{ijqm}\right)$ will be introduced in details later; the travel time of metro of the example in Table 3 is 50 and will be transformed by $50 \to T_{time}$ (50). All the original features concerning the level-of-service variables of alternatives in the choice scenario are processed by the proposed feature transformation approaches based on behavioral theories. The transformed features of each alternative are not the absolute values of the features of each alternative and will contain the superiority or inferiority of the alternative as compared to other alternatives in the choice scenario, due to embedding knowledge-based decision making theories by feature transformations. These will reflect the behavioral mechanisms underlying individuals' choices that decision-makers make the choices depending on comparisons of available alternatives in a choice scenario.

Afterward, an ML algorithm is employed to model the relationships between the chosen option (i.e., dependent variable) and transformed features (i.e., explanatory variables) by leveraging its advantages in terms of modeling nonlinear relations between explanatory variables and the dependent variable, flexible model specifications, and learning complex interactions among explanatory variables in a data-driven manner. Taking the example in Table 3 for illustration, the input features into the ML algorithm is the transformed features of an alternative in the choice scenario, and the output is the probability of the alternative to be chosen by the decision-maker in the choice scenario via soft classification depicted in Section 3.1. Afterward, the process described in the last paragraph of Section 3.1 is used to determine which option is the predicted choice in the choice scenario. The novelty of the proposed framework is integrating behavioral theories into ML algorithms by feature reconstructions. The proposed approach fuses the advantages of behavioral theories in behavioral interpretations and extrapolation, and the superiority of the ML algorithms in modeling the complex relationships. The proposed approach is expected to improve both modeling power and predictive accuracy for depicting travel decision making. Especially, the proposed approach embeds the behavioral theories in the model structure to enhance the model extrapolation, which is a very important requirement in modeling travel decision making. We will particularly test the extrapolation ability of the proposed models in predicting travelers' decision making in choice scenarios that are out the ranges of training samples based on empirical datasets.

3.3. Integrating knowledge-based decision-making theories into tree-based ensemble ML algorithms

To integrate the knowledge-based decision-making theories into ML algorithms in Step 2 of Fig. 1, we utilize and examine three different types of renowned behavior theories, including

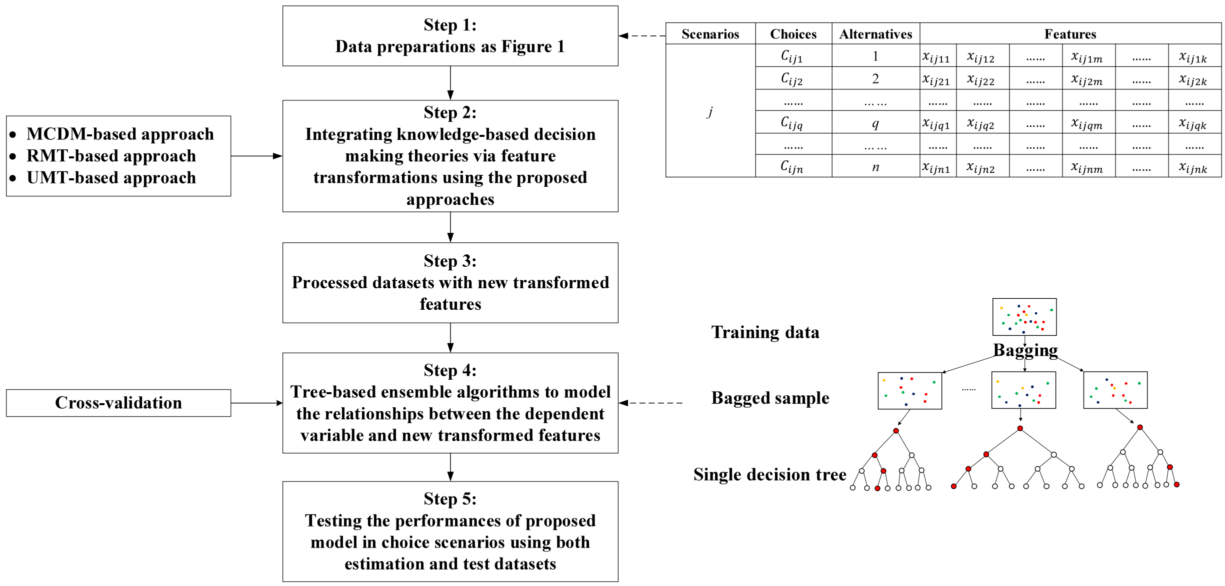

Fig. 1. The proposed framework.

Table 3
An illustrative example of a choice scenario about mode choice.

Scenarios	Alternatives (e.g., transport modes)	Features				Choices
		Cost	Travel time	Travel time reliability	Passenger density in the carriage	
	1 (car)	30 CNY	25 min	8	0	C_{ij1}
j	2 (metro)	4 CNY	50 min	2	0	C_{ij2}
	3 (Park and Ride)	16 CNY	35 min	10	0	…
	4 (bus)	3 CNY	35 min	14	6	C_{ij4}

Multiple-criteria Decision Making, Regret Minimization Theory, and Utility Maximization Theory. The three theories predefine the decision-making criteria from different perspectives. However, the common merits of the three theories are able to reflect the comparative superiority or inferiority in the features of an alternative as compared to others in a choice scenario and reflect the plausible behavioral mechanisms underlying travelers' choices. These behavioral theories have been widely used in various decision-making models. Therefore, they are utilized herein to reconstruct the features of alternatives in each choice scenario for embedding behavioral mechanisms in the modeling framework, as depicted in Section 3.2. We compare the performances of using the three theories based on empirical analysis.

Multiple-criteria Decision Making (MCDM) depicts the process of how decision-makers explicitly evaluate multiple conflicting criteria in decision making and has been widely applied to decision making in various domains [25]. The foundation of MCDM is that decision-makers execute their decisions based on discrepancies among the attributes of alternatives [26]. There are many different model formulations of MCDM [25,26]. Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is one of the most prevalent MCDM methods in both research and practical applications [27,28]. The TOPSIS is built on the concept that the decision maker's preferred alternative should have the shortest geometric distance from the best alternative and the farthest geometric distance from the worst alternative [27,28]. Herein, we adopt the TOPSIS to represent the application of MCDM in our proposed approach on account of its excellent compatibility and prevalence. However, when applying the TOPSIS into multi-attribute decision making requires the prior weights of different features (namely the importance degree of a feature in decision making). Generally, the weights of different features are assumed to be known in TOPSIS. Nonetheless, the weights of features are not known aforehand in our cases (e.g., the weight of travel time in mode choice decision making) and are actually the model parameters to be estimated. Therefore, we cannot apply the TOPSIS for MCDM with known weights of features as most literature did. We, herein, utilize the concept of TOPSIS for comparing the superiority of a single feature of an alternative as compared to the same feature of other alternatives in a choice scenario. Based on TOPSIS, a feature of an alternative in each choice scenario shown in Table 2 could be reconstructed using

$$
\begin{aligned}
&T_{TOPSIS}(x_{ijqm}) = \frac{|x_{ijqm} - x^{*}{}_{ijm}|}{|x^{\#}{}_{ijm} - x^{*}{}_{ijm}|} \\
&x^{*}{}_{ijm} = \max\{x_{ijqm} | q = 1, 2, \dots, n\} \\
&x^{\#}{}_{ijm} = \min\{x_{ijqm} | q = 1, 2, \dots, n\}
\end{aligned} \tag{1}
$$

where $T_{TOPSIS}(x_{ijqm})$ denotes the transformed value for the mth feature of the qth alternative in the choice scenario j for individual i. x^{*}_{ijm} and $x^{\#}_{ijm}$ are the worst and best situations of the mth feature of all alternatives in the choice scenario j, respectively. In the contexts of travel mode choices, the worst and best values are the maximum and minimum values among all alternatives, respectively. Because the considered features such as travel cost and time show negative influences on the utility of an alternative. For instance, the worst and best values of the travel cost in the example of Table 3 are 30 CNY and 3 CNY, respectively. It should be noted that the feature transformation process shown in Eq. (1) is executed in a feature-by-feature manner as we do not know the weights of features aforehand. The transformed value of the cost of Park and Ride in the example of Table 3 is $\frac{|16-30|}{|2-30|} = 0.5$; the transformed value of travel time of Park and Ride in the example of Table 3 is $\frac{|35-50|}{|25-50|} = 0.6$. The weights of transformed features are outputs of the used ML algorithm and estimated based on empirical data. The same process goes for the below feature transformation approaches.

The second utilized theory is Regret Minimization Theory (RMT). RMT proposed by Loomes and Sugden [29] is based on the notion that decision-makers make their choices to avoid the situations where foregone alternatives are better than the chosen one, which results in regrets after choices. In other words, RMT deems that individuals minimize the anticipated regret in making choices by executing comparisons among alternatives. The RMT has been utilized under the framework of discrete choice models for modeling travel decision making such as the Regret Minimization Model [23,30,31]. Based on RMT, we propose two different RMT-based approaches for reconstructing the features of alternatives in a choice scenario to reflect the behavioral mechanisms. The first RMT-based approach assumes that individuals evaluate the associated regret of an alternative by accumulating the regret of each feature of the alternative, referring to Chorus et al. [23] and Quiggin [32]. The regret of a feature of an alternative in a choice scenario is attained by comparing it with the best value of the feature among all alternatives. Based on RMT, the formula for reconstructing features of alternatives is

$$T_{RMT1}(x_{ijqm}) = \max\{\max\{0, x_{ijqm} - x_{ijq'm}\} | q' = 1, 2, \ldots, n\} \tag{2}$$

where $T_{RMT1}(x_{ijqm})$ denotes the transformed value for the mth feature of the qth alternative in the choice scenario j. On account that the considered features in the contexts of travel decision making are negatively related to the utility of an alternative, the larger a feature of an alternative is, the more regret the decision-maker will get if choosing the alternative. The second RMT-based approach also calculates the regret of an alternative on a feature-by-feature basis. However, the second RMT-based approach postulates the r"-egret of an alternative is related to all foregone alternatives, rather than the best one [30]. On the basis of this assumption, the regret of a feature of an alternative is conceived to be the sum of binary regrets by comparing the feature of the alternative with the same features of all other alternatives in a choice scenario, as shown in Eq. (3).

$$T_{RMT2}(x_{ijqm}) = \sum_{q'=1}^{n} (\max\{0, x_{ijqm} - x_{ijq'm}\}) \tag{3}$$

where $T_{RMT2}(x_{ijqm})$ is the transformed value for the mth feature of the qth alternative using the second RMT-based approach.

The third utilized theory is Utility Maximization Theory (UMT), which is the most prevalent behavioral theory used for modeling travel decision making. The well-known Random Utility Models and their derivative models are all founded on UMT [33]. In brief, UMT deems that individuals seek to get the highest satisfaction or positive utilities during the decision process. The decision-makers are assumed to select the alternative with the highest subjective utility after making trade-offs among the features of all available alternatives in a choice scenario. The proposed UMT-based approach for reconstructing features reflects the behavioral mechanism that travelers make decisions to select the option with the highest subjective utility after comparisons. If a feature of the chosen alternative is superior (i.e., smaller) to the same features of other forgone alternatives, it will lead to a positive utility. The average positive utility of a feature of an alternative in a choice scenario is calculated by the following formula

$$T_{UMT}(x_{ijqm}) = \sum_{q'=1}^{n} (\min\{0, x_{ijqm} - x_{ijq'm}\})/(n-1) \tag{4}$$

where $T_{UMT}(x_{ijqm})$ is the transformed feature for the mth feature of the qth alternative using the UMT-based approach.

For the ML algorithm in Step 4 of Fig. 1, this study employs the Random Forest (RF) [34], which is a prevalent tree-based ensemble learning method for classifications. Even though several other ML classification algorithms such as ANNs, SVM, and DTs could be used, RF has been recognized to be superior in modeling travel decision making due to its advantages in model structure, ensemble characteristics, and behavioral interpretations [6,35]. RF was tested by several literature and was reported to superior for modeling travel decision making, as compared to algorithms such as ANNs, SVM, and DTs (e.g., [4,6,8,10]). In contrast to a single decision tree, RF alternatively combines a collection of simple randomized decision trees to nicely enhance the model generality and reduce the variance in predictions [6,35]. In a bootstrapping manner, RF determines class assignments in predictions based on the majority voting of all decision trees in the forest and can output the probability of a data point belonging to a class (namely soft classification). Moreover, differing from algorithms such as ANNs and SVMs, an attractive feature of RF for modeling travel decision making is its ability to extract the feature importance, namely the aforementioned weights of features in the model. The merit of RF can obtain the feature weights of different factors and thus used for behavioral interpretations (e.g., deriving the value of time based on the results).

For each decision tree in RF, a subset with the replacement of the training samples is randomly selected to train the decision tree (namely bagging), and a partial set of randomly selected features are used in each splitting node in the decision tree. Through the two randomnesses (i.e., bagging and randomly selecting features for a splitting node), RF achieves the diversity of different decision trees, robustness to noisy data, and the capacity to handle redundant prediction [34]. RF can also accommodate both numeric and categorical variables in the model [6,35]. In each decision tree, each internal node executes a splitting process to categorize the data into child nodes. By recursively repeating the splitting process, the decision tree will divide the training datasets into sub-regions with fewer observations and stop until the predefined stopping rules are satisfied. By bootstrapping, RF trains many decision trees and acquire the final consequences by averaging over all predictions from majority voting [34]. Finally, RF will output the chosen probability of each alternative in a choice scenario. Training an RF is to select the most discriminative features and thresholds for each node in the decision trees for minimizing the impurity in the partitioned sub-regions. Two impurity criteria are commonly used including GINI and entropy impurity [36]. The GINI impurity is selected in our case for its fast computation and there are no noticeable differences in the results of using the two criteria. The GINI impurity (GIM) of a node is calculated by

$$GIM = \sum_{i=1}^{N} f_i(1 - f_i) \tag{5}$$

where f_i is the frequency of label i (namely transport mode in the contexts of modeling travel mode choices) at a node, and N is the number of unique labels.

4. Datasets, model training and testing

4.1. Experiments and collected datasets

Three datasets including one dataset for model estimation and two other datasets for model tests, are used to fully examine the predictive performances of proposed models in various choice scenarios. The estimation dataset is used to train the proposed models by cross-validations. In the estimation dataset, the data are split into training samples and validation samples through the typical 10-folder cross-validation process. Although the cross-validations randomly select the training and validation samples, the explanatory variables (i.e., features) in the validation samples are generally in similar scales or ranges with those in training samples. Therefore, the predictive performances in cross-validations based on the estimation dataset merely represent the ability of trained models to predict travel decision makings in choice scenarios where features have similar ranges with those in the training samples. In other words, only using the cross-validation in the estimation dataset is not able to test the predictive performances of the trained models in predicting decision makings in the choice scenarios with out-of-range features, namely the extrapolation of trained models. Nonetheless, as aforementioned, it is one of the fundamentally required capacities for modeling travel decision making in choice scenarios that do not exist in the estimation dataset. The limitation encourages us to utilize another two different test datasets to further examine the extrapolation of the proposed methods. The two test datasets consist of choice scenarios that are not existing in the estimation dataset, and features that are out of the ranges of those in the estimation dataset. As far as we are concerned, there is no existing literature in the contexts of modeling travel decision making that examined the model extrapolation ability of ML algorithms. The trained ML algorithms based on the estimation dataset are used to test their prediction power in the two test datasets.

The estimation dataset came from a stated-preference (SP) survey concerning commuting mode choice in 2017, Shanghai of China. In the choice scenarios of the survey, four commonly used commuting modes (car, metro, Park & Ride, metro, and bus) were incorporated and four key level-of-service variables including cost, travel time, travel time reliability and in-vehicle crowding were considered. The respondent is asked to choose their preferred transport mode in the given SP scenarios (see Appendix for more details). Besides the level-of-service attributes of transport modes, respondents' personal characteristics including demographic attributes (e.g., gender, age, education, occupation, income, car ownership) and commuting contexts (e.g., commonly used mode and the distance) were gathered as well on account of their influences on travel decision making.[1] Eighteen different choice scenarios are designed using efficient design methods [37]. A respondent was asked to finish six SP choice scenarios, which were randomly selected from the eighteen scenarios. The features and descriptions about the SP scenarios and surveys are summarized in the Appendix. The dataset has 2316 effective observations from 386 respondents with a wide coverage of socio-economic attributes. A more detailed description of the survey and collected respondents are available in [37].

The considered features of each alternative include both numerical and categorical variables. Their definitions and coding for analysis are demonstrated in Table 4. For the level-of-service variables of alternatives in a choice scenario (e.g., travel time, cost, in-vehicle crowding, and travel time reliability), the proposed approaches described in Section 3.3 are utilized to do feature transformations, which are aimed to integrate knowledge-based behavioral mechanisms. Besides the features of level-of-service variables, we also consider the subjective factors such as demographic attributes (e.g., age and income) and travel contexts as they are also important influencing factors towards the travelers' decision making. For instance, different people have different predilections for different transport modes (e.g., high-income individuals generally prefer private cars more) and different weights towards level-of-service variables (e.g., high-income individuals show fewer weights to cost). Differing from level-of-service variables of alternatives, the subjective factors of a decision-maker are the same for all alternatives in the choice scenarios as the same decision-maker evaluates all alternatives. The subjective factors are treated as categorical variables referring to Gao et al. [37]. The detailed stratification rules for subjective factors are described in Table 4. For categorical variables in Table 4, one-hot encoding is used to transform a categorical variable into a matrix of dummy variables for handling categorical variables in the Random Forest model.

The first test dataset is a synthesis dataset constructed based on the above estimation dataset. The synthesis dataset is customized to test the extrapolation ability of trained models. To achieve the purpose of testing extrapolation, we artificially create a synthesized alternative for each choice scenario in the estimation dataset. Table 5 demonstrates an example of how the synthesis dataset is created based on the above estimation dataset. The contents with shadows are an actual choice scenario in the estimation dataset. We observed that the respondent chose car in the choice scenario as per his/her subjective preferences. This indicated that the car in the choice scenario of Table 5 had the highest subjective benefit for the respondent as compared to all other options. The contents without shadows are the corresponding synthesis choice scenario. The corresponding synthesis scenario adds a new synthesis option, of which level-of-service variables are 75% of the values of the chosen alternative in the actual choice scenario. The variables of other alternatives keep unchanged, as demonstrated in Table 5. Therefore, the added synthesis option is much better than the chosen alternative in the previous choice situation. On account that the respondent chose the car in the actual choice scenario, it can be deduced that the same respondent would choose the synthesis option if he faced with the synthesis choice scenario. By the same process, a synthesis choice scenario could be created for each actual choice scenario in the estimation dataset, which makes up the synthesis dataset. The synthesis scenarios contain features that are out of the ranges of choice scenarios in the estimation dataset (i.e., the synthesis option) and are used to examine the extrapolation of proposed models. The trained models based on the estimation dataset are further tested to predict the travel decision making in the synthesis choice scenarios.

The second test dataset is from another SP survey regarding mode choice behavior for commuting, which was collected in 2019 and also in Shanghai of China. The choice scenarios in the survey assumed that the respondent's company changed the workplace to a new location, and there were several available transport modes for commuting to the new workplace. The respondents chose which transport mode they would like to use in the given SP scenarios (see Appendix for details). The possible alternatives include car, metro, bus, and taxi. The considered level-of-service variables consist of travel cost, travel time, and in-vehicle crowding. The same personal characteristics as those in the estimation dataset were collected in the survey as well. It should be noted that in contrast to the scenarios in the estimation dataset, the second test dataset does not include the feature "travel time reliability" in the scenarios, so the feature of travel time reliability is set to be default values (i.e., 0) in the second test dataset. Thirty-six different SP choice scenarios were designed for the survey [38]. A respondent finished five scenarios randomly chosen from these scenarios. The features and their ranges in the SP scenarios are described in Table A.2 of the Appendix. More details about the survey design and data collections are available in [38]. The dataset contains 2405 choice scenarios from 481 valid respondents with wide demographic coverage [38]. The choice scenarios in the second test dataset differ from the estimation dataset in terms of attribute values and thus could be used to test the extrapolation of the proposed models. However, the ranges of features of alternatives in the second test dataset are not very distinct as compared to those in the estimation dataset (see more descriptions in Appendix).

[1] Please note a respondent finished several SP choice scenarios in a survey. The personal characteristics of a respondent in the SP choice scenarios that he/she finished are invariant.

Table 4
Used features in the proposed model and their definitions.

Feature	Type	Cardinality and coding
Level-of-service variables		
Travel cost	Numerical	The value of overall travel cost in CNY.
Travel time	Numerical	The value of travel time in minutes.
Travel time reliability	Numerical	The value of travel time reliability measured by standard deviation of the travel time distributions in minutes.
In-vehicle crowding	Numerical	The overcrowding level in transport carriage measured by standing density in persons/m^2.
Other variables		
Alternative-specific parameter (ASP)	Categorical	Four possible values: {ASP_car, ASP_metro, ASP_bus, ASP_pr}. These are analogous to the alternative-specific constants for each transport mode in the conventional discrete choice models to reflect travelers' unobserved preference for a certain transport mode.
Age	Categorical	Three possible values: {young age, middle age, old age}. Young, middle, and old age are defined to be age less than 30, age between 30 and 50, and age over 50, respectively.
Education level	Categorical	Three possible values: {education level 1, education level 2, education level 3}. Level 1, 2, and 3 refer to an education level below undergraduate, undergraduate, and master/doctor, respectively.
Income level	Categorical	Three possible values: {low income, middle income, high income}. Low, middle, and high income are defined as the monthly income less than 6000 CNY, between 6000 ∼10,000 CNY, and over 10,000 CNY, respectively.
Commuting distance	Categorical	Four possible values: {short distance, middle distance, long distance, ultra distance}. Short, middle, long, and ultra distance refer to a commuting distance of fewer than 10 km, 10∼20 km, 20∼30 km, and over 30 km, respectively.
Commonly used commuting mode	Categorical	Four possible values: {CM_car, CM_metro, CM_bus, CM_pr}. Considering the effect of currently used transport mode on travel behavior is to incorporate the state-dependence effects of past travel behavior [39]. For instance, if the respondent commonly uses a car for commuting, the value of the variable is CM_car.
Continue the table in the last page		
Male	Dummy	The variable equals 1 when the respondent is male, otherwise 0.
Departure constraint	Dummy	Departure constraint denotes the time scheduling limits of departure to work (e.g., "cannot depart early due to the duty of making meals or delivering kids to school before work" or "have to take the bus at a timeslot"). The variable equals to 1 when the respondent has departure constraints, otherwise 0.
Feature	Type	Cardinality and coding
Flexible work time	Dummy	Flexible work time equals to 1 when the respondent's work time is flexible and is otherwise 0
Hupai license plate	Dummy	In Shanghai, there are different kinds of license plates. Only hupai license plate can use the motorway systems in peak hours. The variable is 1 if the respondent has a hupai license plate, otherwise 0.

Table 5
An example of creating a synthesis choice scenario based on an actual choice scenario in the surveys.

<table>
<tr><th rowspan="3">Scenarios</th><th rowspan="3">Choices</th><th rowspan="3">Alternatives</th><th colspan="5">Features</th></tr>
<tr><th colspan="4">Level-of-service variables</th><th>Other variables in Table 4 such as demographic attributes</th></tr>
<tr><th>Travel cost</th><th>Travel time</th><th>Travel time reliability</th><th>In-vehicle crowding</th><th>……</th></tr>
<tr><td rowspan="4">j</td><td>1</td><td>Car</td><td>30</td><td>25</td><td>8</td><td>0</td><td>……</td></tr>
<tr><td>0</td><td>Metro</td><td>4</td><td>50</td><td>2</td><td>0</td><td>……</td></tr>
<tr><td>0</td><td>P&R</td><td>16</td><td>35</td><td>10</td><td>0</td><td>……</td></tr>
<tr><td>0</td><td>Bus</td><td>3</td><td>35</td><td>14</td><td>6</td><td>……</td></tr>
<tr><td rowspan="5">j (synthesis)</td><td>0</td><td>Car</td><td>30</td><td>25</td><td>8</td><td>0</td><td>……</td></tr>
<tr><td>0</td><td>Metro</td><td>4</td><td>50</td><td>2</td><td>0</td><td>……</td></tr>
<tr><td>0</td><td>P&R</td><td>16</td><td>35</td><td>10</td><td>0</td><td>……</td></tr>
<tr><td>0</td><td>Bus</td><td>3</td><td>35</td><td>14</td><td>6</td><td>……</td></tr>
<tr><td>1</td><td>Car</td><td>30×0.75 = 22.5</td><td>25×0.75 = 18.75</td><td>8×0.75 = 6</td><td>0×0.75 = 0</td><td>……</td></tr>
</table>

		Predicted class	
		Class = 1	Class = 0
Actual class	Class = 1	True Positive (TP)	False Negative (FN)
	Class = 0	False Positive (FP)	Ture Negative (TN)

Fig. 2. The confusion matrix for model performance measurements.

4.2. Model tuning, training and testing

Three main hyperparameters in RF influence its performances: the number of decision trees, the maximum depth of each tree, and the maximum number of features for each splitting node. The number of trees should be suitably selected to balance the model complexity and computation cost. Deeper tree structure and more features in each splitting node are anticipated to improve the impurity of partitioned sub-regions, but would lead to overfitting issues [4,34]. Therefore, the three hyperparameters should be tuned. We employ several performance metrics to quantitatively measure the predictive performances of models including accuracy, precision, recall, and F1-score, whose equations are demonstrated in Eq. (6). A confusion matrix shown in Fig. 2 demonstrates the definitions of the notations used in Eq. (6). The meaning of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) refer to Stehman [40]. The Area Under the Receiver Operating Characteristic (AUROC) curve score is also used for measuring the performance of a classification model at different classification thresholds.

To tune the models, we firstly test the tree number through 10-folder cross-validations based on the estimation dataset. The result indicates that the number of trees beyond 200 does not lead to additional improvements in the performances but notably increases the computing time. Thus, we set the forest size to be 200. After confirming the forest size, we further tune the tree depth and the maximum number of features in a splitting node. The suggested value for the number of features in each splitting mode is $\log_2 p$ where p is the number of used features. We experimented the maximum number of features in a splitting node from 1 to 33 in increments of 1 and the maximum depth of a tree from 1 to 100 with an increment of 2 by cross-validations based on the estimation dataset. The performance is measured by the AUROC score, and the experimental results are depicted in Fig. 3. It turns out that the combination of 10 maximum features in a splitting node with a maximum depth of 10 for each tree, exhibits the best performance. We also tuned the hyperparameters using other performance metrics and show similar results. Therefore, we settled the above hyperparameters in our implementations. Moreover, it is worth noting that the class of the dependent variables (i.e., $\mathbf{C}_{ij}$) is imbalanced as a choice scenario in the estimation dataset has four alternatives, and merely one of them is chosen by the decision-maker in a choice scenario. We thus use a balanced RF to handle the imbalance by incorporating class weights. The weight of a class is inversely proportional to class frequencies in the input data, which penalizes misclassifying the minority class harder [41]. The average performances of a model in the validation samples are used to measure its predictive performances. It should be noted that the dataset is partitioned in the cross-validation as per the choice scenarios, rather than the alternatives. A choice scenario in Table 2 is regarded as a sample point in the random splitting process of cross-validations.

$$\begin{aligned}
&Accuracy = (TP + TN)/(TP + FP + FN + TN)\\
&Precision = TP/(TP + FP)\\
&Recall = TP/(TP + FN)\\
&F1 - score = 2 \cdot (Recall \cdot Precision)/(Recall + Precision)
\end{aligned} \tag{6}$$

Based on the proposed framework, six different models are developed and examined to fully compare their performances in predicting travel decision making in different choice scenarios with a focus on the extrapolation of the models. We utilize another two test datasets that are different from the estimation dataset for examining the extrapolation of the models in predicting choice scenarios that do not exist in the estimation dataset and have features beyond the feature ranges in the estimation dataset.

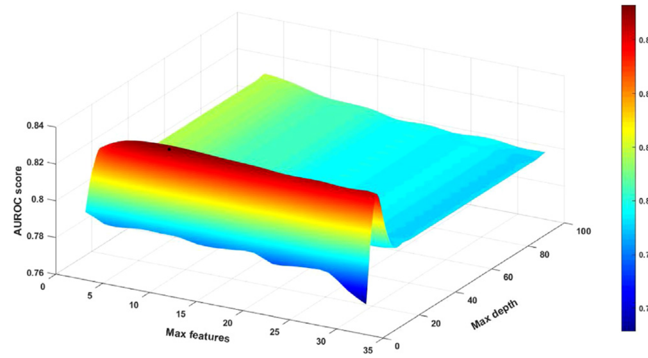

Fig. 3. The AUROC score for different hyperparameters.

- **RF-TOPSIS**: This model applies the proposed framework and uses the TOPSIS-based approach for integrating the knowledge-based decision-making theories.
- **RF-RMT1** and **RF-RMT2**: The two models use the proposed framework and take advantage of RMT for considering behavioral mechanisms. However, they have differences in the feature transformations as shown in Eqs. (2) and (3)
- **RF-UMT**: This model is founded on the proposed framework and utilizes the UMT for integrating behavioral mechanisms into the ML classification algorithms.
- **RF-UMT-RMT2**: This model is a combination of **RF-UMT** and **RF-RMT2**. In this model, a feature of an alternative is processed by Eqs. (3) and (4) to obtain two derivative features, which are both used in the ML algorithms. This idea comes from the notion that decision rules underlying individuals' behavior may not be pure utility maximization or regret minimization, but a mixture of utility maximizations and regret minimizations [42]. Therefore, this model specification is used to test performances as compared to purely applying UMT or RMT.
- **RF**: This model follows the method in the existing literature (e.g., [4,6]) and directly uses the features of alternatives in the model without considering behavioral mechanisms. This model is regarded as the benchmark for comparison.

5. Results and discussions

This section presents the results of the predictive accuracy of the proposed models. Particularly, we compare the models founded on the proposed approach from this study and the existing method in the literature that directly applies RF for modeling travel decision making. In the analysis, the prediction accuracy in the estimation dataset and in the two test datasets are discussed separately. The results from the estimation dataset represent the

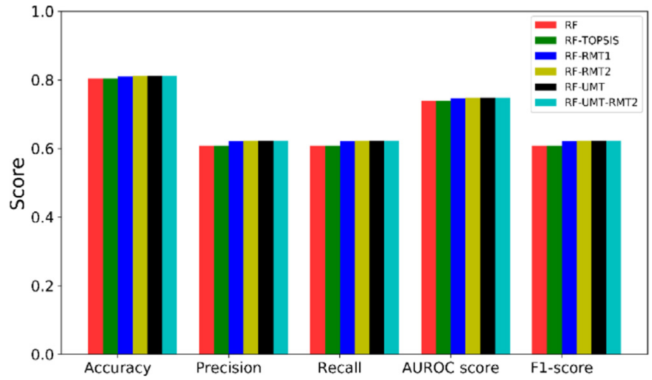

Fig. 4. Predictive performances in the estimation dataset.

ability of models to predict travel decision making in choice scenarios where alternatives have similar features with those in the training samples. The results of two test datasets are customized for testing the extrapolation of the proposed models in choice scenarios that do not exist in the estimation dataset and have extrapolated features.

5.1. Predictive performances in the estimation dataset

The results of the predictive performances in the estimation dataset are illustrated in Fig. 4. The mean predictive accuracy, precision, and recall scores of proposed models including **RF-RMT1**, **RF-RMT2**, **RF-UMT**, and **RF-UMT-RMT2** in 10-folder cross-validation are around 0.8, 0.6, and 0.6, respectively. The accuracy score indicates that the models could correctly predict an alternative to be chosen or not chosen in a specific choice scenario with a probability of around 80%. However, there are more than one foregone alternatives (i.e., not chosen options) in a choice scenario. Generally, travel behavior modelers care more about the probability of correctly predicting the chosen alternative in a choice scenario rather than not chosen options. The probability of correctly predicting the chosen alternative in a choice scenario is reflected by the precision and recall scores. The precision and recall scores of a model are identical since the proposed framework only allows one alternative in a choice scenario to be chosen by the decision-maker. If the chosen alternative in a choice scenario is not correctly predicted by the model, it will result in one-unit increase of values of both False Positives and False Negatives simultaneously. Therefore, the values of False Positives and False Negatives across all choice scenarios for a model are the same, which would lead to the same precision and recall scores as per Eq. (6). The recall scores of the models demonstrate that they could correctly predict the chosen alternative in a choice scenario by a probability of around 60% on average, which is quite good in the contexts of modeling travel decision making on account of large behavioral variances. The AUROC scores of the proposed models are over 0.7, which indicates the models perform quite well [43].

When comparing the models based on our proposed framework with the method in the literature (i.e., **RF**), the predictive performances of proposed approaches in the estimation dataset are similar to the **RF** in terms of accuracy, precision, recall, and AUROC scores. The **RF-RMT1**, **RF-RMT2**, **RF-UMT**, and **RF-UMT-RMT2** only provide slightly more superior predictive performances as shown in Fig. 4. It seems that the proposed framework of integrating knowledge-based decision making theories into RF does not provide substantially better predictive power in the estimation dataset, in contrast to **RF**. However, these are logical and even expected to some extent. The underlying reasons are the choice scenarios in the estimation dataset and the structure of RF. There are eighteen different choice scenarios in the estimation data. In the estimation dataset, the choice scenarios in the training samples selected by the cross-validation process would cover most choice scenarios in the validation sample. Thus, an RF with a size of 200 trees, maximum 10 features in a splitting node, and a maximum depth of 10, can fully fit the travel decision making in all possible choice scenarios in a near enumeration manner, and thus is anticipated to perform quite well in predicting similar choice scenarios in the validation sample. Therefore, the direct application of RF to model the travel decision making in a purely data-driven way without considering behavioral mechanisms, is still capable of predicting the choices in the validation samples. In such cases, the proposed framework in this study does not present a notable advantage in predictive power as compared to **RF**. However, as aforementioned, ignoring the behavioral mechanisms underlying observed choice behavior would result in the deficiency of the model to predict the choice scenarios that are out of the range in training samples, namely the model extrapolation issues. The proposed framework is expected to show its superiority in extrapolation derived from integrating the knowledge-based decision-making principles. These encourage us to further examine the predictive power of the six models in the test datasets.

5.2. Predictive performances in predicting extrapolated choice scenarios

5.2.1. Performances in the first test dataset

The predictive performances of the six models in the first test dataset are summarized in Table 6 and displayed in Fig. 5. In the first test dataset, each choice scenario has a new synthesis alternative that has out-of-range features in contrast to the training samples. The direct application of RF merely has a precision score of 0.2174, which means that the **RF** could only correctly predict the chosen option in a choice scenario by a probability of 21.74%. Given that there are five alternatives in each choice scenario of the first test dataset, a random prediction could give a correct predicting probability of around 20%. This implies that the direct application of RF presents terrible performances in predicting choice scenarios that are out of the ranges in training samples. The AUROC score of **RF** is 0.5108, which is close to 0.5 and indicates that **RF** has nearly no discrimination capacity to predict the chosen alternative in the choice scenarios of the synthesis test dataset.

In contrast, the models based on the proposed framework from this study all provide much better predictive performances in the first test dataset as compared to **RF**, as shown in Fig. 5. The **RF-TOPSIS** improves the predictive accuracy and precision of **RF** from 0.6869 to 0.8483 and from 0.2174 to 0.6208, respectively. A precision score of 0.6208 means that **RF-TOPSIS** could correctly predict the chosen alternative in choice scenarios with a probability of 62%. As for the RMT-based approaches, **RF-RMT1** and **RF-RMT2** both provide better prediction ability than **RF-TOPSIS**. The **RF-RMT2** presents the best prediction ability among the six models. The **RF-RMT2** has an accuracy score of 0.9260 and a precision score of 0.8151, demonstrating that the **RF-RMT2** could make correct predictions of the selected option in all choice scenarios of the first test dataset with a probability of 81.51%. The predictive precision of **RF-RMT2** improves the precision of **RF** from 21.74% to 81.51% by 274.79%, which is very substantial. The UMT-based model **RF-UMT** also gives better predictions as compared to **RF** as well, but are inferior to RMT-based models and **RF-TOPSIS**. The accuracy score of **RF-UMT** is 0.8257 and the precision score is 0.5643, which promotes the precision score of **RF** by 159.56%. The **RF-UMT-RMT2** model has the second-best

Table 6
Predictive performances in two test datasets.

Models	Datasets	Accuracy	Precision	Recall	AUROC score	F1-score
RF	First test dataset	0.6869	0.2174	0.2174	0.5108	0.2174
RF-TOPSIS		0.8483	0.6208	0.6208	0.7630	0.6208
RF-RMT1		0.9014	0.7537	0.7537	0.8460	0.7537
RF-RMT2		0.9260	0.8151	0.8151	0.8844	0.8151
RF-UMT-RMT2		0.9110	0.7776	0.7776	0.8610	0.7776
RF-UMT		0.8257	0.5643	0.5643	0.7277	0.5643

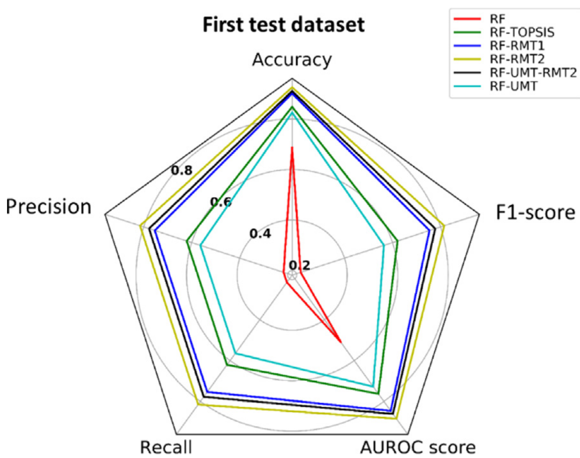

Fig. 5. The predictive performances in the first test dataset.

predictive performances with an accuracy score of 0.911 and a precision score of 0.7776.

The above results indicate that the direct application of RF without considering hidden behavioral mechanisms of decision-makers' choices has a very weak extrapolation ability to predict the choice scenarios with features that are out of the ranges in training samples. The results provide an important implication for the relevant literature that only tested the predictive performances based on one estimation dataset and cross-validation. It is generally reported in the relevant literature that the tree-based ensemble algorithms such as RF provide much better predictive accuracy as compared to conventional models such as the Multinomial Logit Model (e.g., [4,6,8,10]). However, it should always be kept in mind that the so-called predictive accuracy refers to the predictive performances in the validation samples from the process of cross-validation, which have similar choice scenarios as those in training samples in terms of feature values. Through theoretical derivation, the RF that neglects the behavioral mechanisms would not be able to predict travel decision making in the choice scenarios with out-of-range features correctly. Our empirical results confirm the deduction. The direct application of RF is indeed deficient in predicting the out-of-range choice scenarios. The underlying reasons are the essential deficiency of extrapolation of ML algorithms such as tree-based ensemble models and the ignorance of behavioral mechanisms in the model establishments. The arbitrary application of data-driven ML algorithms may perform acceptably in choice scenarios that have similar features with training samples, but are essentially deficient in extrapolation ability. On the contrary, the models founded on our proposed framework provide much more superior predictive ability as compared to **RF**. The reason is straightforward. By integrating the knowledge-based decision-making theories, the proposed framework reflects the behavior mechanisms behind travel decision making, leading to stronger extrapolation to predict the choice scenarios beyond the training samples.

5.2.2. Performances in the second test dataset

We further use another test dataset collected in real experiments to examine the performances of different models. The predictive performances of the six models in the second test dataset are demonstrated in Table 7 and Fig. 6. The **RF** shows a precision score of 0.4706, which is better than its performance in the first test dataset but worse than its performance in the estimation dataset. The **RF-TOPSIS** and **RF-RMT1** give slightly better predictive performances as compared to **RF**, but the differences are not substantial. However, the **RF-RMT2**, **RF-UMT**, and **RF-UMT-RMT2** present more advantages in predictive performances comparing to the **RF**, as displayed in Fig. 6. The **RF-UMT-RMT2** has the best performances among the three models and has a precision score of 0.5738, which is 21.9% larger than the precision score of **RF** (0.4706). The **RF-RMT2** and **RF-UMT** improve the precision score by 19.3% (from 0.4706 to 0.5617) and 17.9% (from 0.4706 to 0.5555) in comparisons to **RF**, respectively. The AUROC scores of **RF-RMT2**, **RF-UMT**, and **RF-UMT-RMT2** show noticeably larger values in contrast to that of **RF**. These all indicate that the **RMT2**, **RF-UMT**, and **RF-UMT-RMT2** have much better predictive abilities in the choice scenarios of the second test dataset as compared to **RF**. The best model **RF-UMT-RMT2** has a probability of 57.38% to correctly predict the chosen alternative in the choice scenarios of the second test dataset and has an AUROC score of 0.6713, indicating it is an acceptable classifier to determine the chosen alternative in various choice scenarios.

The models based on the proposed framework provide better predictive performances in the second test dataset as compared to **RF**, but are not as superior as they are in the first test dataset. The potential reason is that the ranges of features of alternatives in the second test dataset are not very distinct to those in the training samples (see details in the Appendix). Hence, the **RF** that fits the travel choices in a purely data-driven way, does not show pronounced inferiorities to the proposed framework in this study. More specifically, there are not many extrapolated choice scenarios in the second test dataset as compared to the choice scenarios in the training samples. Nonetheless, **RMT2**, **RF-UMT**, and **RF-UMT-RMT2** still show substantial advantages in predicting the choice behavior in the second test dataset, demonstrating the positive effects of integrating the knowledge-based decision-making theories on improving the predictive ability and extrapolation.

6. Concluding remarks

The prevalence of machine learning has attracted researchers to utilize them for modeling travel decision making on account of their advantages in considering complex relations such as the nonlinear effects and interactions. However, machine learning classification algorithms that fit the travel decision making in a purely data-driven manner ignore the behavioral mechanisms underlying travelers' choices. Such models are not capable of predicting decision making in the choice scenarios that have features out of the ranges in training samples, namely the extrapolation issue. However, extrapolation ability is very crucial in the contexts of modeling travel decision making.

Table 7
Predictive performances in the second test dataset.

Models	Datasets	Accuracy	Precision	Recall	AUROC score	F1-score
RF		0.6471	0.4706	0.4706	0.6030	0.4706
RF-TOPSIS		0.6598	0.4898	0.4898	0.6173	0.4898
RF-RMT1	Second test dataset	0.6693	0.5039	0.5039	0.6279	0.5039
RF-RMT2		0.7036	0.5555	0.5555	0.6666	0.5555
RF-UMT-RMT2		0.7158	0.5738	0.5738	0.6803	0.5738
RF-UMT		0.7078	0.5617	0.5617	0.6713	0.5617

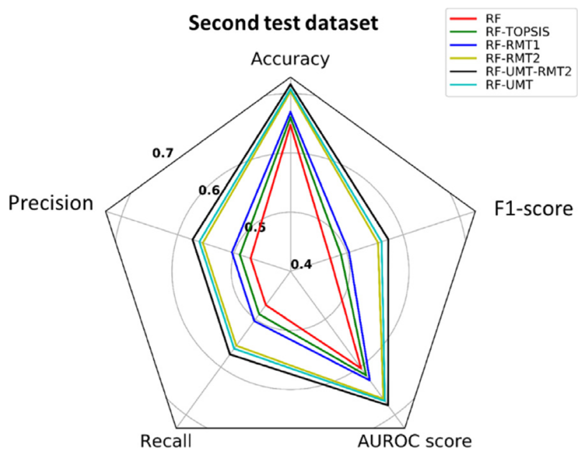

Fig. 6. Comparisons among the predictive performances in the synthesis test dataset.

In this study, we proposed an extrapolation-enhanced approach for modeling travel decision making by integrating knowledge-based decision-making theories into ensemble ML algorithms, which, to our best knowledge, is the first work of proposing such a method for modeling travel decision making. The proposed approach utilizes tree-based ensemble ML algorithms to address the complex and nonlinear relationships between the dependent variable and explanatory variables, as well as the complicated interactions among explanatory variables. Concurrently, the approach integrates the recognized knowledge-based decision making theories from behavioral science to fully consider the behavioral mechanisms underlying travel decision making for enhancing model extrapolation ability. In this manner, the approach is able to reflect the behavioral mechanisms that decision-makers select the subjectively optimal option by making trade-offs among alternatives rather than depending on the attributes of an alternative in a specific choice scenario.

The predictive performances of the proposed approach are compared to the existing method through three different datasets including one estimation dataset and two test datasets. The estimation dataset is used to train the models by cross-validations and to test the predictive performances of models in predicting decision making in choice scenarios within similar ranges of training samples. More importantly, we put a special focus on the prediction ability of the models in the choice scenarios with features that are out of the ranges in training samples, namely the ability of extrapolation. These have hardly been investigated in the relevant literature. Two test datasets are particularly utilized to test the extrapolation ability of different models.

The results show that both proposed models and the method in the literature (i.e., direct application of RF) can both give quite good predictive performances in the choice scenarios that have similar feature ranges with training samples. However, RF has very bad predictive performances, which is close to a random prediction, in the first test dataset with choice scenarios that have out-of-range features as compared to the training samples. The empirical results reveal the deficient extrapolation ability of directly using tree-based ensemble algorithms in modeling travel decision making.

In contrast, the proposed approach provides substantially superior predictive performances as compared to the direct application of RF in the choice scenarios that have out-of-range features, namely much stronger extrapolation ability. In the first test dataset, the model based on our proposed approach (i.e., **RF-RMT2**) could improve the precision score to 81.51%, which is 274.93% larger than that of RF. In another test dataset, the model based on our proposed approach (i.e., **RF-UMT-RMT2**) could improve the precision score up to 21.9% as compared to the direct application of RF as well. These confirm the superiority of the proposed approach in terms of model extrapolation for modeling travel decision making.

Even though the present study proposes a novel approach for modeling travel decision making to support optimization of the transportation system, there are some limitations and future work that could be addressed in future work. Efforts in both theoretical and practical aspects are required to improve the accurate modeling of travel decision making based on machine learning techniques. The machine learning methods are essentially designed to prioritize the prediction rather than behavioral analysis. Therefore, it is necessary to develop methods for extracting more behavioral insights from the results of machine learning algorithms, such as willingness-to-pay and demand elasticities for practical analysis. Actually, one advantage of RF is obtaining feature importance after training. Due to text limitations, the results are not presented and discussed in the current paper. However, it is interesting to get behavioral interpretations based on the results. Moreover, it is interesting to apply the proposed framework to empirical modeling of choice behavior in other contexts such as route choice, departure time choices, and other decision situations beyond travel choices. Re-examining the performances of the proposed approach based on more adequate and various datasets is necessary as well.

CRediT authorship contribution statement

Kun Gao: Conceptualization, Data collection, Methodology, Formal analysis, Writing - original draft. **Ying Yang:** Conceptualization, Methodology, Writing - review & editing. **Tianshu Zhang:** Data collection, Methodology, Formal analysis. **Aoyong Li:** Manuscript revision, Writing, Suggestions. **Xiaobo Qu:** Conceptualization, Methodology, Validation, Resources, Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This project is financially sponsored by EU JPI project SMUrTS.

Table A.2
The features and levels used for SP scenarios in the second test dataset [38].

Transport modes	Features	Possible levels (values)
Short-time commuting		
Car	Travel time Cost (oil, parking fare, and tolls) Crowding inside car	{10, 20, 30} min {5, 15, 25} CNY None
Metro	Travel time Cost (ticket) Crowding inside metro	{15, 25, 35} min {3, 4, 5} CNY {Level 1, Level 2, Level 3}
Bus	Travel time Cost (ticket) Crowding inside bus	{15, 25, 35} min 2 CNY {Level 1, Level 2, Level 3}
Taxi	Travel time Cost Crowding inside taxi	{10, 20, 30} min {10, 20, 30} CNY None
Medium-time commuting		
Car	Travel time Cost (oil, parking fare, and tolls) Crowding inside car	{20, 30, 40} min {10, 20, 30} CNY None
Metro	Travel time Cost (ticket) Crowding inside metro	{25, 35, 45} min {3, 4, 5} CNY {Level 1, Level 2, Level 3}
Bus	Travel time Cost (ticket) Crowding inside bus	{25, 35, 45} min 2 CNY {Level 1, Level 2, Level 3}
Taxi	Travel time Cost Crowding inside taxi	{20, 30, 40} min {15, 25, 35} CNY None
Long-time commuting		
Car	Travel time Cost (oil, parking fare, and tolls) Crowding inside car	{30, 40, 50} min {15, 25, 35} CNY None
Metro	Travel time Cost (ticket) Crowding inside metro	{35, 45, 55} min {4, 5, 6} CNY {Level 1, Level 2, Level 3}
Bus	Travel time Cost (ticket) Crowding inside bus	{40, 50, 60} min 2 CNY {Level 1, Level 2, Level 3}
Taxi	Travel time Cost Crowding inside taxi	{30, 40, 50} min {25, 35, 45} CNY None

Table A.1
Features and levels used in SP scenario design of estimation dataset [37].

Transport mode	Features	Possible levels (values)
Car	Mean travel time Travel time unreliability Cost (oil, parking fare, and tolls) Crowding inside car	{15, 25, 35, 40} min {4, 8, 12, 18} min {10, 25, 35, 45} CNY None
Metro	Mean travel time Travel time unreliability Cost (ticket) Crowding inside metro	{30, 40, 50, 60} min {2, 4, 6, 8} min {3, 4, 5} CNY {Level 1, 2, 3}
Park&Ride	Mean travel time Travel time unreliability Cost (oil, parking fare, and ticket) Crowding inside transit	{25, 35, 45, 55} min {2, 6, 8, 10} min {12, 16, 18, 22} CNY 3 (Level 1, 2, 3}
Bus	Mean travel time Travel time unreliability Cost (ticket) Crowding inside bus	{30, 40, 50, 60} min {4, 8, 14, 20} min {1, 2, 3} CNY {Level 1, 2, 3}

Note: The travel time reliability is measured by the standard deviation of travel time distribution. For the in-vehicle crowding, three levels reflecting three typical situations during commuting in Shanghai were set in the survey. Crowding Level 1, 2, 3 represent the standing passenger density is 0, 3 and 6 persons/m^2, respectively.

Appendix

The used features and their levels for stated preference (SP) scenario design in the estimation dataset are listed in Table A.1. An efficient design method was used to generate the statistical contents of scenarios [37]. The statistical content design refers to determining the features of alternatives in an SP scenario through specific methods such as orthogonal design or efficient approaches [44]. The efficient design method balances the utilities of different alternatives to avoid dominating options in the scenarios referring to Rose et al. [44]. For instance, it should be omitted that the features of an alternative are better than others in a choice scenario. Dominating option in a scenario should be avoided as it cannot reflect trade-offs among features and provide valuable information about how individuals make decisions. Fig. A.1 displays an SP scenario used in the survey. The SP scenario provides the respondent with an assumptive scenario in which the attributes of all alternatives are given. The respondent is asked to choose their preferred option in the given SP scenario. Eighteen scenarios were generated. Each respondent was presented with randomly selected six scenarios in one questionnaire to avoid nimiety. More details about the survey design are available in [37].

For the SP scenarios in the second test dataset, the used features and corresponding levels are summarized in Table A.2.

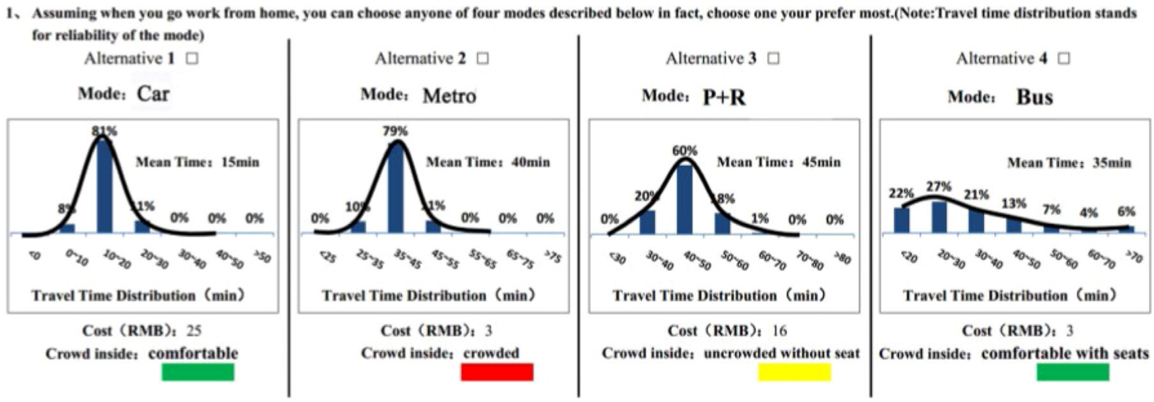

Fig. A.1. The used SP scenario in the estimation dataset [37].

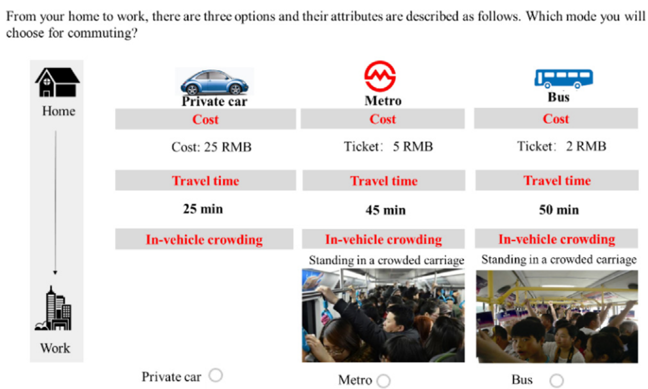

Fig. A.2. The used SP scenario in the second test dataset.

The SP scenario assumes that the respondent's company changed the workplace. The respondents answered which transport mode they would choose in the SP scenarios. An example of the used SP scenario is shown in Fig. A.2. If the respondent had access to a private car, the three alternatives were car, metro, and bus; otherwise, the three options were the taxi, metro, and bus. The SP scenarios were designed as per the actual travel contexts of a respondent. A respondent was categorized into one of three groups based on their common commuting time: short-time commuting (<25 min), medium-time commuting (25∼40 min), long-time commuting (>40 min). For each group, corresponding SP scenarios were generated using the efficient design method. The features and their levels for SP scenarios of each group are demonstrated in Table A.2. Twelve scenarios are generated for each group, and five scenarios are randomly selected for each respondent in the survey. More details about collected respondents are available in [38].

By comparing Tables A.1 and A.2, it can be found that the features of alternatives in the SP scenarios of the estimation dataset are different from those in the SP scenarios of the second test datasets. Therefore, the choice contexts in the SP scenarios in the second test dataset differ from those in the estimation dataset. In other words, the choice scenarios in the second test dataset are plausible and good testing samples for examining model extrapolations. However, the ranges of features of alternatives in the estimation and the second test dataset are not very divergent. For instance, the range of travel time in the SP scenarios of the estimation dataset is from 15 to 60 min; the range of travel time in the scenarios of the second test dataset is from 10 to 55 min.

References

[1] J. Hausman, D. McFadden, Specification tests for the multinomial logit model, Econometrica 121 (1984) 9–1240.
[2] D. McFadden, K. Train, Mixed MNL models for discrete response, J. Appl. Econometrics 15 (5) (2000) 447–470.
[3] Q. Gao, D.-L. Xu, An empirical study on the application of the Evidential Reasoning rule to decision making in financial investment, Knowl.-Based Syst. 164 (2019) 226–234.
[4] L. Cheng, X. Chen, J. De Vos, X. Lai, F. Witlox, Applying a random forest method approach to model travel mode choice behavior, Travel Behav. Soc. 14 (2019) 1–10.
[5] T. Wang, H. Li, X. Zhou, B. Huang, H. Zhu, A prospect theory-based three-way decision model, Knowl.-Based Syst. (2020) 106129.
[6] X. Zhao, X. Yan, A. Yu, P. Van Hentenryck, Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models, Travel Behav. Soc. 20 (2020) 22–35.
[7] Y. Chen, Y. Zhou, Machine learning based decision making for time varying systems: Parameter estimation and performance optimization, Knowl.-Based Syst. 190 (2020) 105479.
[8] J.D. Pineda-Jaramillo, A review of Machine Learning (ML) algorithms used for modeling travel mode choice, Dyna 86 (211) (2019) 32–41.
[9] J. Hagenauer, M. Helbich, A comparative study of machine learning classifiers for modeling travel mode choice, Expert Syst. Appl. 78 (2017) 273–282.
[10] A. Lhéritier, M. Bocamazo, T. Delahaye, R. Acuna-Agost, Airline itinerary choice modeling using machine learning, J. Choice Model. 31 (2019) 198–209.
[11] D.E. Brown, Text mining the contributors to rail accidents, IEEE Trans. Intell. Transp. Syst. 17 (2) (2015) 346–355.
[12] H. Zhang, D. Nettleton, Z. Zhu, Regression-enhanced random forests, 2019, arXiv preprint arXiv:1904.10416.
[13] D. Shmueli, I. Salomon, D. Shefer, Neural network analysis of travel behavior: evaluating tools for prediction, Transp. Res. C 4 (3) (1996) 151–166.
[14] A. Mohammadian, E.J. Miller, Nested logit models and artificial neural networks for predicting household automobile choices: comparison of performance, Transp. Res. Rec. 1807 (1) (2002) 92–100.
[15] D.A. Hensher, T.T. Ton, A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice, Transp. Res. E 36 (3) (2000) 155–172.
[16] C. Xie, J. Lu, E. Parkany, Work travel mode choice modeling with data mining: decision trees and neural networks, Transp. Res. Rec. 1854 (1) (2003) 50–61.
[17] Y. Zhang, Y. Xie, Travel mode choice modeling with support vector machines, Transp. Res. Rec. 2076 (1) (2008) 141–150.
[18] N. Golshani, R. Shabanpour, S.M. Mahmoudifard, S. Derrible, A. Mohammadian, Modeling travel mode and timing decisions: Comparison of artificial neural networks and copula-based joint model, Travel Behav. Soc. 10 (2018) 21–32.
[19] H. Omrani, Predicting travel mode of individuals by machine learning, Transp. Res. Procedia 10 (2015) 840–849.
[20] C.R. Sekhar, E. Madhu, Mode choice analysis using random forest decision trees, Transp. Res. Procedia 17 (2016) 644–652.
[21] F. Wang, C.L. Ross, Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model, Transp. Res. Rec. 2672 (47) (2018) 35–45.
[22] M. Ben-Akiva, S.R. Lerman, Discrete Choice Analysis: Theory and Application to Travel Demand, Transportation Studies, Cambrige, MA, USA, 2018.

[23] C.G. Chorus, T.A. Arentze, H.J. Timmermans, A random regret-minimization model of travel choice, Transp. Res. B 42 (1) (2008) 1–18.
[24] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, USA, 2006.
[25] S. Guo, H. Zhao, Fuzzy best-worst multi-criteria decision-making method and its applications, Knowl.-Based Syst. 121 (2017) 23–31.
[26] M. Kadziński, R. Słowiński, S. Greco, Multiple criteria ranking and choice with all compatible minimal cover sets of decision rules, Knowl.-Based Syst. 89 (2015) 569–583.
[27] K. Yoon, A reconciliation among discrete compromise solutions, J. Oper. Res. Soc. 38 (3) (1987) 277–286.
[28] Z. Xu, Uncertain Multi-Attribute Decision Making: Methods and Applications, Springer, London, 2015.
[29] G. Loomes, R. Sugden, Regret theory: An alternative theory of rational choice under uncertainty, Econ. J. 92 (368) (1982) 805–824.
[30] C.G. Chorus, A new model of random regret minimization, Eur. J. Transp. Infrastruct. Res. 10 (2) (2010).
[31] C. Chorus, Random regret minimization: an overview of model properties and empirical evidence, Transp. Rev. 32 (1) (2012) 75–92.
[32] J. Quiggin, Regret theory with general choice sets, J. Risk Uncertain. 8 (2) (1994) 153–165.
[33] J. Walker, M. Ben-Akiva, Generalized random utility model, Math. Social Sci. 43 (3) (2002) 303–343.
[34] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[35] H.M. Gomes, A. Bifet, J. Read, J.P. Barddal, F. Enembreck, B. Pfharinger, G. Holmes, T. Abdessalem, Adaptive random forests for evolving data stream classification, Mach. Learn. 106 (9–10) (2017) 1469–1495.
[36] L.E. Raileanu, K. Stoffel, Theoretical comparison between the gini index and information gain criteria, Ann. Math. Artif. Intell. 41 (1) (2004) 77–93.
[37] K. Gao, L. Sun, H. Tu, H. Li, Heterogeneity in valuation of travel time reliability and in-vehicle crowding for mode choices in multimodal networks, J. Transp. Eng. A 144 (10) (2018) 04018061.
[38] K. Gao, Y. Yang, L. Sun, X. Qu, Revealing psychological inertia in mode shift behavior and its quantitative influences on commuting trips, Transp. Res. F 71 (2020) 272–287.
[39] V. Cantillo, J.d.D. Ortúzar, H.C. Williams, Modeling discrete choices in the presence of inertia and serial correlation, Transp. Sci. 41 (2) (2007) 195–205.
[40] S.V. Stehman, Selecting and interpreting measures of thematic classification accuracy, Remote Sens. Environ. 62 (1) (1997) 77–89.
[41] C. Chen, A. Liaw, L. Breiman, Using random forest to learn imbalanced data, Univ. Calif. Berkeley 110 (1–12) (2004) 24.
[42] C.G. Chorus, J.M. Rose, D.A. Hensher, Regret minimization or utility maximization: it depends on the attribute, Environ. Plan. B: Plann. Des. 40 (1) (2013) 154–169.
[43] M.E. Rice, G.T. Harris, Comparing effect sizes in follow-up studies: ROC Area, Cohens D, and R, Law Human Behav. 29 (5) (2005) 615–620.
[44] J.M. Rose, M.C. Bliemer, D.A. Hensher, A.T. Collins, Designing efficient stated choice experiments in the presence of reference alternatives, Transp. Res. B 42 (4) (2008) 395–406.