Travel Behaviour and Society 30 (2023) 118–134

Contents lists available at ScienceDirect

Travel Behaviour and Society

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

A Random Effect Bayesian Neural Network (RE-BNN) for travel mode choice analysis across multiple regions

Yutong Xia [a,b], Huanfa Chen [a,*], Roger Zimmermann [b,c]

[a] Centre for Advanced Spatial Analysis (CASA), University College London, London, UK
[b] Institute of Data Science, National University of Singapore, Singapore
[c] School of Computing, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
Bayesian neural network
Travel mode choice modelling
Regional behaviour comparison
Random effect model

ABSTRACT

Travel mode choice modelling plays a critical role in predicting passengers' travel demand and planning local transportation systems. Researchers commonly adopt classical Random Utility Models to analyse individual decision-making based on the utility theory. Recently, with an increasing interest in applying Machine Learning techniques, a number of studies have used these methods for modelling travel mode preferences for their excellent predictive power. However, none of these studies proposes machine learning models that investigate the regional heterogeneity of travel behaviours. To address this gap, this study develops a Random Effect-Bayesian Neural Network (RE-BNN) framework to predict and explain travel mode choice across multiple regions by combining the Random Effect (RE) model and the Bayesian Neural Networks (BNN). The results show that this model outperforms the plain Deep Neural Network (DNN) regarding prediction accuracy and is more robust across different datasets. In addition, in terms of interpretation, the capability of RE-BNN to learn the travel behaviours across regions has been demonstrated by offset utilities, choice probability functions and local travel mode shares.

1. Introduction

Modelling travel mode choice plays a critical role in predicting passengers' preferences and travel demand, providing references for policymakers to plan transportation systems and understand the underlying factors (de Dios Ortúzar and Willumsen, 2011). Traditionally, researchers commonly understand the passengers' travel mode choice based on the theory of utility (Domencich and McFadden, 1975; McFadden, 1986). Utility-based models are predominantly Random Utility Models (RUMs), most notably the logit family. The Multinomial Logit (MNL) model, the simplest logit model, is widely adopted for analysing individual decision-making due to its high level of interpretability (McFadden, 1981). However, this model is based on an oversimplified assumption that the alternative's utility is a linear specification, and it is hard to capture the potential regulations in a dataset. Another limitation of the MNL model is that it assumes that the probabilities of each choice are independent, leading to biased predictions. Although other logit models, such as the mixed logit model, the Nested Logit (NL) model and the Cross-Nested Logit (CNL) model, have better capability to yield the probabilities for alternatives when the correlations among the choices exist (Hensher and Greene, 2003), the parameters in these models are much tricky to estimate.

Recently, with a growing interest in applying Machine Learning (ML) technology in numerous research fields, including transportation (Ren et al., 2020; Li et al., 2018; Paredes et al., 2017), this data-oriented approach has also been considered an alternative to RUMs for modelling travel mode preferences. By comparing many classifiers' predictive performances, such as Deep Neural Networks (DNNs), Support Vector Machines (SVMs), Decision Trees (DTs), Random Forest (RF) and MNL model, researchers have verified that the predictive power of ML methods, especially DNNs and RF, is better than traditional Discrete Choice Models (DCMs) (Hagenauer and Helbich, 2017). Besides predictive performance, the explanation power of models for deriving behavioural insights is equally vital in travel behaviour studies. Nevertheless, these ML approaches have always been regarded as low interpretability (Lipton, 2018) since they compute the probability relying on the structure of the dataset rather than a theory of the underlying data structure. Zhao et al. (2020) pointed out that there is a tradeoff between prediction performance and interpretation power when selecting between ML methods and conventional DCMs. Recently, Wang et al.

* Corresponding author.
E-mail address: huanfa.chen@ucl.ac.uk (H. Chen).

https://doi.org/10.1016/j.tbs.2022.08.011
Received 6 December 2021; Received in revised form 16 June 2022; Accepted 24 August 2022
Available online 18 September 2022

(2020) demonstrated that DNNs not only have high predictive accuracy but also provide economic knowledge as complete as conventional DCMs, which makes an outstanding contribution to the interpretation of DNNs for choice analysis.

Nevertheless, none of the existing studies on travel mode choice using ML methods focuses on the heterogeneity of travel behaviour among different population groups, such as people from different regions. In the behaviour contexts, researchers have suggested that travel behaviour differs from area to area. Globally, people's travel behaviours in different countries are diverse (Delbosc and Ralph, 2017). Locally, the various culture and public policies lead to passengers' travel attitudes varying across cities. For example, the introduction of the Congestion Charge in London and the one-day-a-week driving restriction policy in Beijing greatly affect individuals' decisions about driving in these cities (Ambühl et al., 2018; Wang et al., 2014). None of the studies using ML methods for travel mode choice modelling has considered these regional differences, which is crucial to understanding the impact of local socio-economic factors on residents' travel behaviour.

To fill the above-mentioned gap, this paper develops a Random Effect-Bayesian Neural Network (RE-BNN) framework to predict and explain travel mode choice across multiple regions. This paper first proposes a region-specific RUM by introducing the RE model into the traditional RUM and then discusses how it can be realised in a BNN framework. The paper then introduces the dataset used to test the capability of RE-BNN. After that, it empirically investigates the predictive power of RE-BNN in comparison with three other models. In addition, in terms of interpretability, based on the behaviour insights derived from RE-BNN, the paper presents the travel behaviour analysis of passengers from different regions from three aspects: offset utilities, choice probability functions and local travel mode shares.

The contributions of this study are threefold as follows. (1) It proposes a new architecture of Neural Network (NN) called RE-BNN by combining the Random Effect (RE) model and Bayesian Neural Network (BNN) that achieves a single model to predict the travel mode choice in multiple regions. (2) The RE-BNN has better predictive performance and generalisation than DNN due to the introduction of Bayesian's uncertainty, which is demonstrated by a case study of predicting mode choice across nine regions in the UK. (3) The experimental results reveal the interpretability of RE-BNN in travel behaviour across multiple regions and the remarkable predictive power of local travel mode market shares. Codes used in this study are available on Github: https://github.com/yutong-xia/RE-BNN.

The rest of the paper is organised as follows. Section 2 reviews the related studies on the determinants of travel mode choice, conventional logit models and ML for travel mode choice. Section 3 describes related theories and models used in this study. Section 4 introduces the dataset used for empirical study, the process of hyperparameter searching in NN models, and the specifications for utility functions in the MNL model. Section 5 describes the results of the experiment, including an exploratory analysis of the travel mode choice in the UK, the predictive performances of five models and the behaviour insights derived from RE-BNN. Lastly, Section 6 summarises the findings and suggests future research directions.

2. Literature review

2.1. Determinants of travel mode choice

The choice of travel mode are influenced by a range of factors. Based on previous studies, these factors can be divided into three groups: (1) demographic characteristics, (2) trip-related characteristics, and (3) environmental characteristics.

First, demographic factors, such as gender, age, car ownership, education, and income, are essential in determining travel modes. Adopting multinomial logistic regression, Ashalatha et al. (2013) investigated the mode choice behaviour of commuters in Thiruvananthapuram, and they found that with the increase in age, the preference for driving significantly increases, as it is less likely for senior citizens to ride on short trips than their younger counterparts (Johansson et al., 2006). In addition, due to the higher demand for comfort and convenience, commuters with higher income prefer private vehicles to other travel approaches (Hensher and Rose, 2007).

Second, the importance of trip-related characteristics in determining residents' travel mode choices has been confirmed, including trip time, distance and purpose. Trip distance plays a crucial role in travel modes decisions and is a significant variable in predicting travel modes (Hagenauer and Helbich, 2017). In addition, as the trip distance, time or cost increases, commuters prefer to choose cars or bicycles instead of public transportation (Ashalatha et al., 2013). Passengers are usually more susceptible to the trip time in the morning and evening peak than other times of the day; therefore, in peak hours, people prefer public transportation to avoid traffic congestion (Habibian and Kermanshah, 2013). In addition, the comfort, flexibility, convenience as well as safety of the travel mode can also impact people's decisions of travel mode alternatives (Heinen et al., 2011).

Third, for later discussion about the research gap, this study classifies environmental characteristics into two sub-characteristics: the built and social environments. An increasing number of studies investigating the environmental impact on determining travel modes focus on built environment characteristics (Ding et al., 2017; Munshi, 2016). Ewing and Cervero (2010) defined built environment as 6Ds: Density (population and jobs per unit area), Diversity (land use mix and balance index), Design (neighbourhood design), access to Destination (accessibility to jobs by transportation), Distance to transit stops, and Demand management (measures to encourage the usage of public transportation). Density is regarded as a critical factor in travel mode choice decisions, especially in host-based travel behaviours (Chen et al., 2008). Moreover, density at the workplace plays a more significant role in tour-based travel behaviours than that at the residential location (Ding et al., 2014). The intersection design, street network and distance to jobs are also essential factors in passengers' travel choice decisions, especially in bicycle and walking modes (Ewing and Cervero, 2010). When living near their workplace or the street connectivity is good, the choice of walking and bicycle modes increases significantly.

In addition to the built environment, the social environment factors also play a significant role in determining the travel mode. A person's social environment is defined by his or her living and working environment as well as the characteristics of the community, which can be experienced on multiple scales, including family, neighbourhood, city and region (Casper, 2001). The social environment can impact passengers' decisions by affecting their subjective norms[1] and attitudes to different means of transportation. Willis et al. (2015) highlighted the importance of social environmental factors' influences on decisions on bicycle commuting and believed that thinking beyond built environment factors is necessary for understanding or predicting bicycle use. Based on a survey of 1991 inhabitants in three German cities, Hunecke et al. (2010) found that the attitude-based approach has better predictive performance than that based on demographic and geographic factors. In addition to the impact on the moral norm, the differences in public policies related to transport can also influence the attitude of commuters to choose their travel modes, such as the one-day-a-week driving restriction policy in Beijing (Wang et al., 2014). However, although the importance of these social factors for determining travel mode has been demonstrated, few researchers have paid attention to this regional heterogeneity of travel behaviours when predicting travel mode choice.

[1] Subjective norm is 'the perceived social pressure to perform or not to perform a behaviour' (Ajzen, 1991).

2.2. Logit models for modelling travel mode choice

Previous researches typically adopt the utility theory to understand the passengers' travel mode choice (Domencich and McFadden, 1975; McFadden, 1986). Utility-based models have been the predominant approach for modelling travel mode alternatives for several decades, most of which are Random Utility Models (RUMs), developed by McFadden (1981). The logit family is a class of RUMs based on random utility maximisation, including the MNL model, the Nested Logit (NL) model, the Cross-Nested Logit (CNL) model and the mixed logit model (Ben-Akiva and Lerman, 2018). These models gained popularity and are widely used in travel behaviour research for they have simple mathematical formula structures and can represent individual choice decisions realistically.

In the MNL model, which is the simplest logit model, the probabilities for each travel choice are generated by a logistic function of the utilities. The simple mathematical structure in the MNL model mitigates the challenge of parameter estimation (Koppelman and Wen, 1998). However, the model's basis on the assumption of independence of irrelevant alternatives (IIA) limits its practical application, especially when dealing with panel data[2]. If the IIA assumption is violated, parameter estimation and prediction results will be biased (McFadden, 1986).

Some flexible logit models were developed to relieve the limitations of the MNL model, including the mixed logit model. Compared with the MNL model, the mixed logit model is not based on the IIA assumption and can better deal with preference heterogeneity (Hensher and Greene, 2003). Other extensions, such as the NL model and the CNL model, can also generate the probabilities for mode choice when the correlations among the choices exist (Ben-Akiva and Lerman, 2018). However, the estimation of parameters of these improved models is more complex and challenging than the MNL model. Moreover, although the mixed logit model has been substantially improved based on the MNL model to refine the assumption and enhance the interpretation of the model, its predictive accuracy becomes worse (Cherchi and Cirillo, 2010).

2.3. Machine Learning for Modelling Travel Mode Choice

Recently, there has been a tremendous interest in the application of Machine Learning (ML) methods in numerical transportation fields, such as traffic flow prediction (Ren et al., 2020), license plate recognition (Li et al., 2018), and car ownership prediction (Paredes et al., 2017). Since the modelling of travel mode alternatives can also be regarded as a general classification problem, the ML classification algorithm has become an alternative to the traditional logit models. A large body of literature has used these algorithmic non-parametric methods to model the decisions of travel mode. Hillel et al. (2019) systematically reviewed 60 articles where ML methods were applied for modelling mode choice. The commonly used ML classifiers include Support Vector Machines (SVMs; Zhang and Xie, 2008), Classification Trees (CTs; Tang et al., 2015), Random Forest (RF; Lhéritier et al., 2019) and Deep Neural Networks (DNNs; Wang et al., 2020).

ML classifiers can automatically identify the relationships between the input features and the mode choice instead of predefining the utility functions in the logit models. The effectiveness of ML methods has been demonstrated in the field of behavioural research. For example, based on the San Francisco Travel Survey datasets, Xie et al. (1854) demonstrated that DT and NN offer better performances than the MNL model in terms of the modelling results. Zhang and Xie (2008) suggested that due to the SVM model's promising performance and easy implementation, it can be used as an alternative procedure for travel mode choice modelling. More recently, Hagenauer and Helbich (2017) compared the predictive accuracy of seven ML classifiers for travel mode choice modelling and found that RF outperforms other modelling approaches, including MNL, which is in line with the findings in the research of Lhéritier et al. (2019) and Zhao et al. (2020). Meanwhile, Neural Networks have been demonstrated by Salas et al. (2022) to have outstanding performances in terms of accuracy and interpretation.

Compared with the predetermined model structures in the traditional logit models, ML can form more flexible modelling structures, contributing to their better predictive performance and high compatibility with data (Xie et al., 1854). The other reason for the higher predictive accuracy of ML is that the development of ML focuses on accurately predicting, while the extension of logit models focuses on modifying model assumptions and enhancing the interpretability ability for modelling individual travel behaviour (Brownstone and Train, 1998; Hensher and Greene, 2003).

However, despite the remarkable predictive performance, a commonly acknowledged disadvantage of ML models is their weak interpretation power due to their complex structures, and they are often viewed as black boxes. In addition, the lack of capability to use previously acquired knowledge is often associated with these methods. Researchers have been focusing on the interpretation of ML models in recent years (Doshi-Velez and Kim, 2017). For example, Hinton et al. (2015) distilled the knowledge by training a smaller model for interpretation from an ensemble or a large, highly regularised model. Based on variable importance, a commonly applied ML explanation tool, Cheng et al. (2019) assessed the importance of variables in the RF model. Zhao et al. (2020) conducted a comparative study between logit models and ML models. They interpreted RF based on variable importance and partial dependence plots and explained Neural Network (NN) based on a sensitivity analysis. Wang et al. (2020) interpreted the DNN model and derived the economic information from DNN, including market shares, elasticities, and marginal rates of substitution. They demonstrated that the reliability of this extracted economic information could be improved by using a larger sample, ensemble model and hyperparameter searching.

Nevertheless, in the study by Wang et al. (2020), they also pointed out that NN has a high propensity to overfit noise, leading to incorrect classification results (also see Guo et al., 2017). Several approaches have been proposed to avoid the over-fitting issue and improve the generalisation capability of NN models. One way is to add a weight-decay or a regularisation term in the process of parameters estimation (Liang and Wong, 2001). However, Marzban and Witt (2001) argued that this approach of improving NN models' generalisation has a negative effect on their capability to approximate non-linear. On the other hand, they proposed that via adopting the Bayesian method to infer, it is possible to improve the generalisation capability of NN models and at the same time keep the strong non-linear approximating capability of NNs. Bayesian Neural Networks (BNNs), first developed by Mackay (1992), have proven to have the capability to effectively reduce the overfitting phenomenon by introducing uncertainty on the weights (Blundell et al., 2015). Liang (2005) improved the BNN model via incorporating a prior on the connections between networks and the weights, which makes it more flexible for NNs to select their hidden neurons as well as the input features. Due to the above-mentioned advantages of BNN models over the other NN models, this method has been applied in many sub-fields of transportation, such as traffic crash prediction models (Xie et al., 2007). However, it has rarely been used for travel mode choice modelling.

Based on the review of previous literature, it is found that nearly all of them are focusing on improving models' predictive power or explanatory power applied to a single city or country. However, due to the different social environment characteristics (e.g., culture and public policy) in different cities or regions, passengers' attitudes to transportation vary, leading to the regional heterogeneity of travel mode utility and preferences. In other words, when other features of a trip and the built environment are the same, one person in a different city may not make the same decision regarding travel mode. Therefore, this paper

[2] Panel data refers to the data containing multiple observations (i.e., travel records) for the same subjects (i.e., travellers).

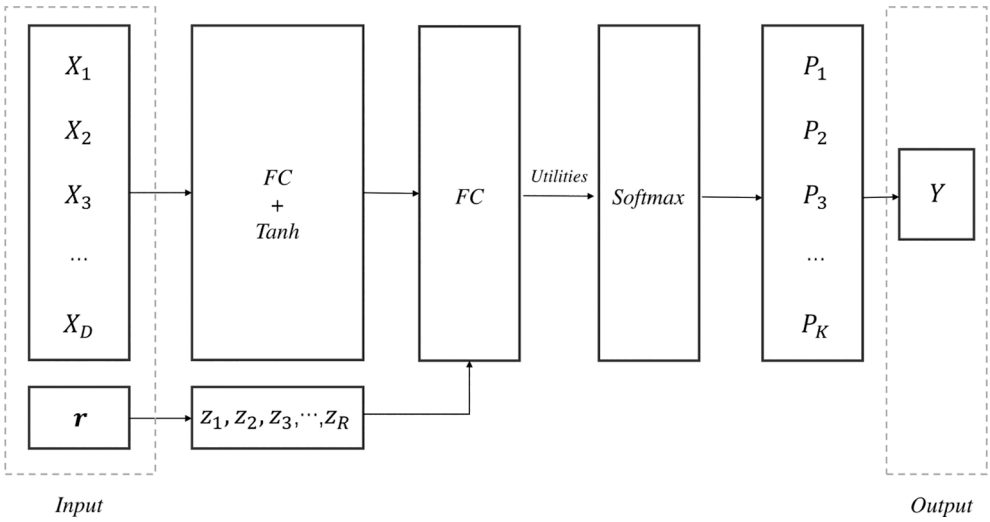

Fig. 1. RE-BNN architecture.

Table 1
Hyperparameters space of RE-BNN and DNN.

Hyperparameters	Values
Panel 1. Invariant Hyperparameters	
Activation functions	Tanh and Softmax
Loss	ELBO (RE-BNN) and Cross Entropy (DNN)
Panel 2. Varying Hyperparameters of RE-BNN and DNN	
Learning rate	[0.1,0.01, 1e-3, 1e-4]
Width *n*	[5,10,15, 25, 50, 100]
Standardisation	[True,False]
Panel 3. Varying Hyperparameters of DNN	
Batch Size	[10,50,100]

introduces the RE model into travel mode alternatives modelling to model travel mode choice across multiple regions. In addition, considering the over-fit issue of NNs and that BNNs can avoid over-fitting by introducing uncertainty and improving the predictive power, this paper incorporates Bayesian inference to neural networks. In general, to fill the research gap, this study combines the RE model and BNNs to develop a new model RE-BNNs for travel mode choice modelling across multiple regions in the context of behaviour analysis.

Table 2
Selected variables of the alternative-specific utility function of MNL and MXL models.

Mode	Selected variables
Walk	Household_car, Trip_distance, Household_licence, Trip_time
Bicycle	Household_bike, Trip_purpose(1), Trip_time
Driving	Household_car, Population_density, Household_licence, Trip_time, Individual_age, Household_settlement(1)
Bus	Household_car, Household_licence, Population_density, Trip_time
Rail	Population_density, Trip_time, Trip_distance, Trip_purpose_1, Individual_education(1)

Table 3
Trip frequency by travel mode choice.

Mode	Frequency	Percentage
Driving	131621	78.48%
Walk	16649	9.93%
Bus	8867	5.29%
Rail	7415	4.42%
Bicycle	3165	1.89%

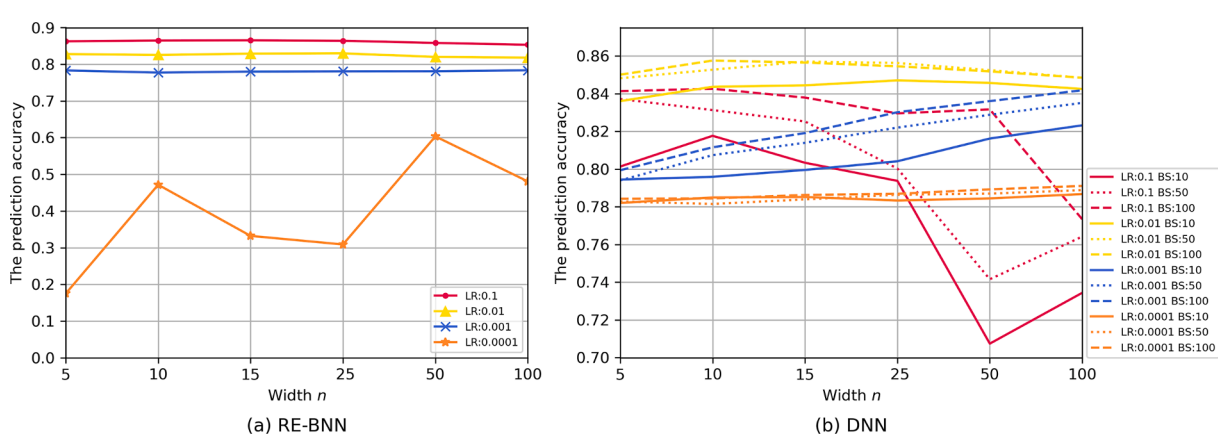

Fig. 2. Hyperparameter searching results for RE-BNN and DNN.

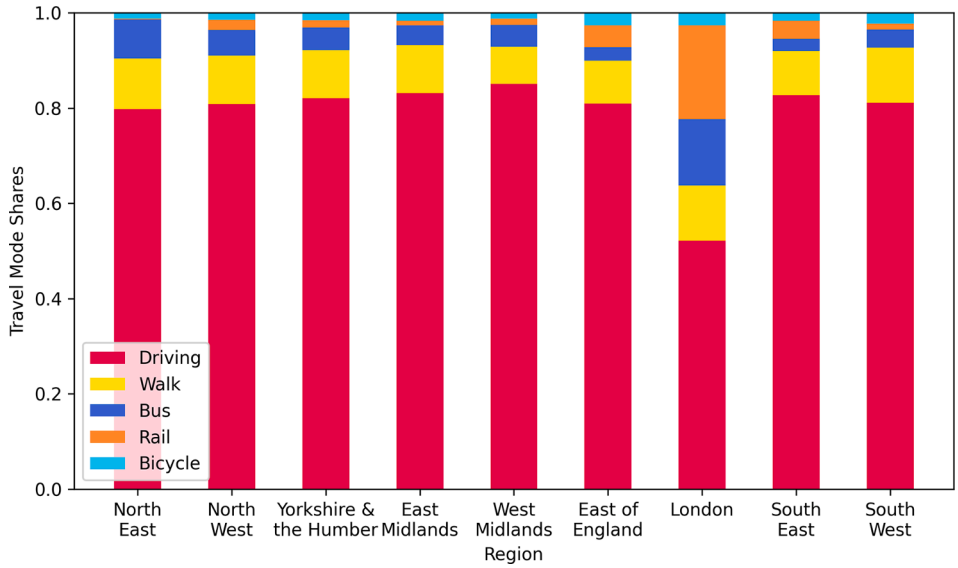

Fig. 3. Travel mode share by region.

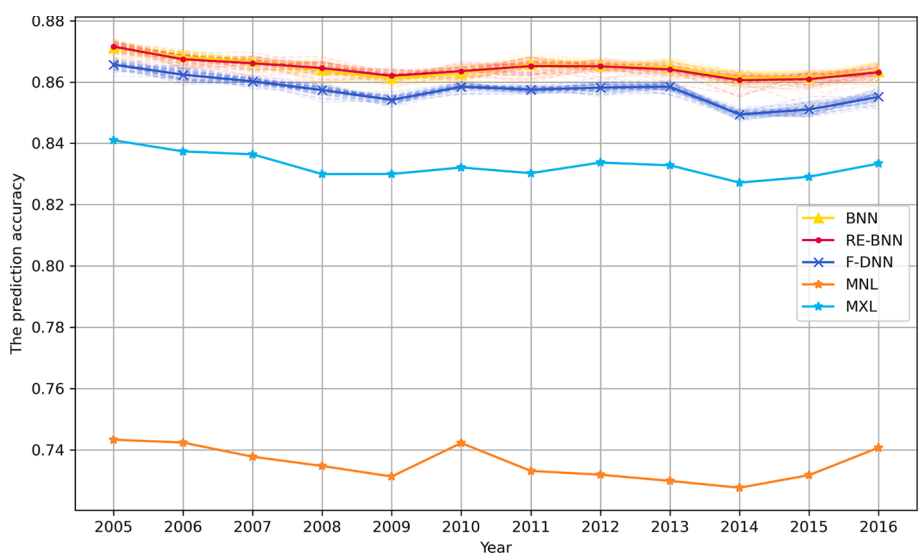

Fig. 4. Prediction accuracy of RE-BNNs, BNNs, DNNs, MNL and MXL models in different year data.

3. Methodologies

This section describes how the Random Effect (RE) model is combined with RUM and how to build the NN framework based on region-specific RUM. As mentioned in Section 2.3, in order to mitigate the sensitivity issue of NNs to parameters, Bayesian's uncertainty is introduced to the construction of the NN model. Since the family of logit models have been widely adopted for exploring different kinds of behaviour choice in the field of travel behaviour analysis, we will fit some logit models as a benchmark for comparison in the latter experiment. Therefore, some basic formulas of the logit model are described in the last subsection.

3.1. Region-specific random utility model

RUM has been a predominant method for passenger mode choice modelling since the seminal paper by McFadden (1981). Based on the assumption that each alternative provides a certain level of utility to an individual, RUM relies on utility specifications for each alternative. In reality, it is impossible for researchers to observe all the utility of the individual. Therefore, for each alternative, the utility $U_{ik}$of selecting choice k out of $[1, 2, \ldots, K]$ choices for individual i is the sum of the deterministic utility V_{ik} and the random utility ϵ_{ik}, which represent the effects of observed variables and unobserved factors respectively (Ben-Akiva and Lerman, 2018):

$$U_{ik} = V_{ik} + \epsilon_{ik} \tag{1}$$

The observed utility V_{ik} is a function of x_i, where x_i is the ith observations, including individual-related and trip-related attributes. $i \in \{1, 2, \ldots, N\}$ and $k \in \{1, 2, \ldots, K\}$. K denotes the total number of travel mode alternatives, and N denotes the total number of trip records. The random utility ϵ_{ik}often results from the specification of the deterministic utility V_{ik}.

The Random Effects (RE) model is a statistical method for analysing multilevel data, including longitudinal repeated-measures data (Gardiner et al., 2009). A well-specified RE model can provide more information than a Fixed Effects (FE) model (Bell et al., 2019; Shor et al., 2007). Unlike FE models where all observations share a common effect size, RE models allow the effect size to vary (Borenstein et al., 2010). Thus they can describe different behaviour to different 'clusters' of

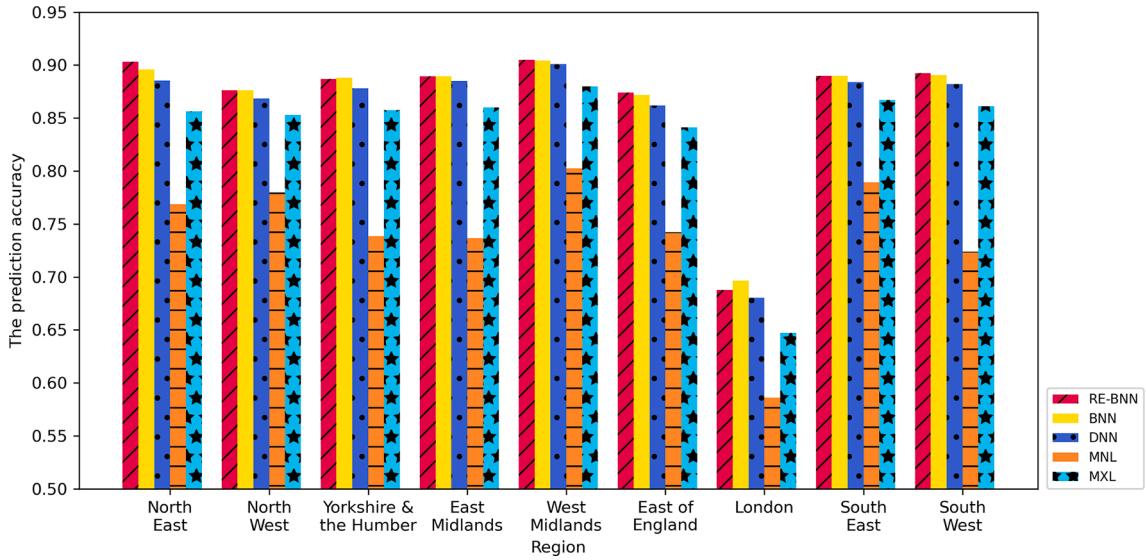

Fig. 5. Prediction accuracy of RE-BNNs, BNNs, DNNs, MNL and MXL models trained by the 2016 year training set for different regions.

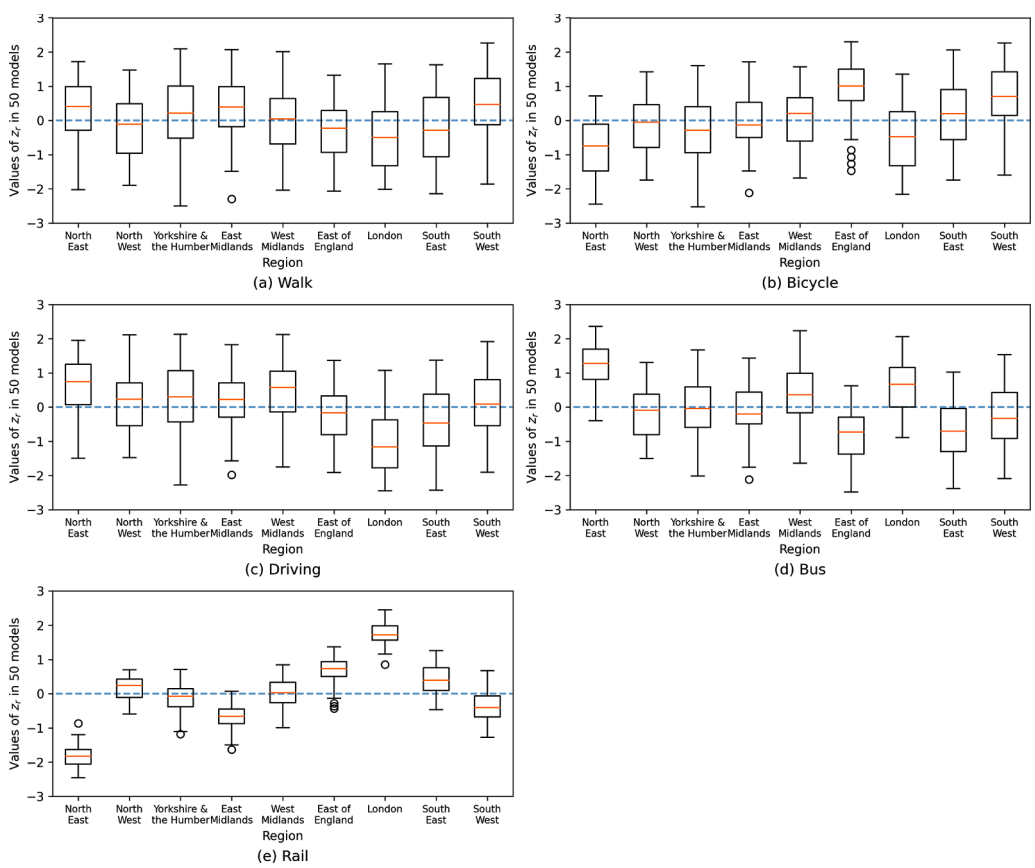

Fig. 6. The offset utility of five alternatives for passengers living in different regions in 2016. The values of z_{rk} parameter are standardised.

observations.

In this study, as the preference for travel mode choice may differ from region to region, the RE model is used to describe the different travel behaviour of individuals from different regions. The RE model is combined with RUM by introducing a random effect factor z_{rk} into Eq. (1), which allows us to accommodate preference heterogeneity from different regions and achieve a region-specific RUM:

$$U_{irk} = V_{irk} + \epsilon_{irk} \tag{2}$$

The observed utility V_{irk} is a function of x_i and z_{rk}, where z_{rk} is a region-varying independent variable. In the context of utility theory, it can be regarded as the offset utility of alternative k for people living in region $r \in \{1, 2, \ldots, R\}$, where R refers to the total number of regions. Note that this random effect parameter differs from the dummy variable, which is the fixed effect. A RE model treats random effects as random draws from a normal distribution, while dummy variables in a FE model are regarded as unconnected entities and estimated separately (Bell et al., 2019).

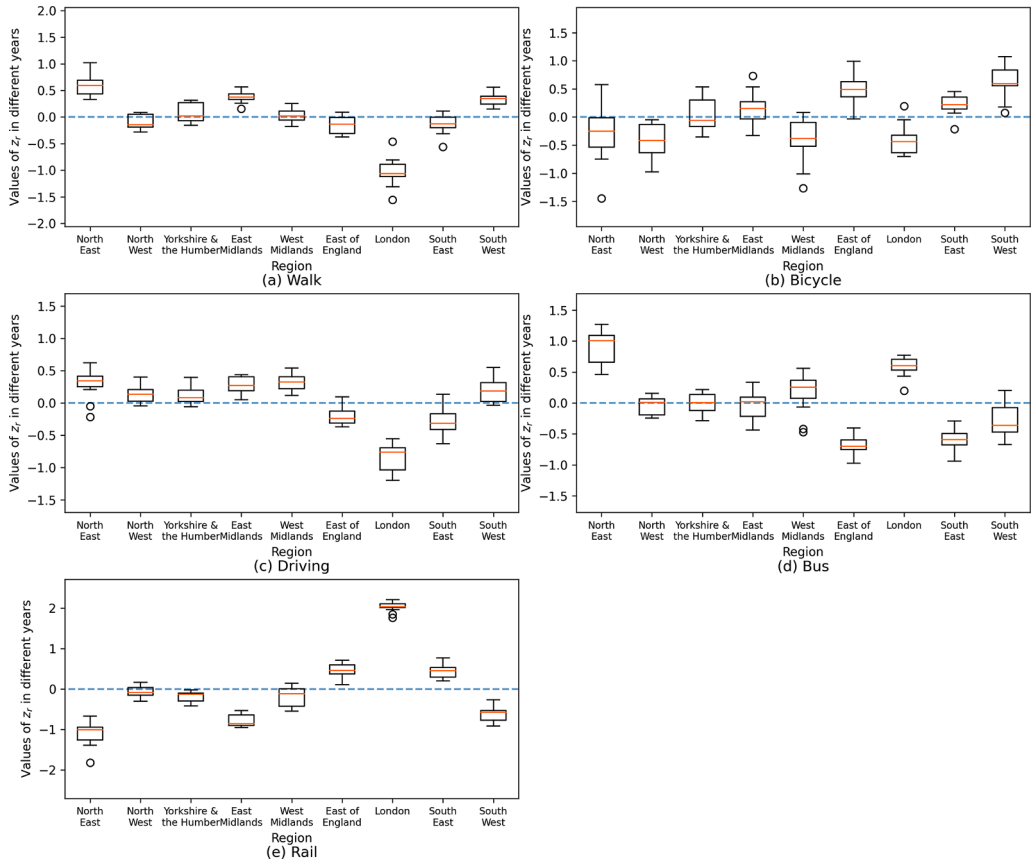

Fig. 7. The variation of the offset utility of five alternatives for passengers living in different regions from 2005 to 2016.

In utility maximisation behaviour, the passenger i tends to choose the alternative k that provides the largest utility out of K choices (Ben-Akiva and Lerman, 2018). The probability of choosing alternative k can be described as:

$$P_{irk} = Prob(V_{irk} + \epsilon_{irk} > V_{irj} + \epsilon_{irj}, \forall j \in \{1, 2, \ldots, N\}, j \neq k) \tag{3}$$

Assuming that the random utility $\epsilon_{irk}$is followed by Gumbel distribution, the probability of choosing alternative k can be obtained as follows (McFadden, 1986):

$$P_{irk} = \frac{exp(V_{irk})}{\sum_{j=1}^{K} exp(V_{irj})} \tag{4}$$

3.2. Neural network for region-specific choice modelling

NNs are proven to be highly adaptable in identifying non-linear interactions and they can be adopted to analyse travel mode choice. When assuming that the random utility term $\epsilon_{irk}$is obeying the Gumbel distribution, the input variables into the Softmax activation function of NNs can be regarded as utilities, and the outputs from the Softmax activation function represent the probabilities of alternatives (Wang et al., 2020).

$$P_{irk} = Softmax(V_{irk}) = \frac{exp(V_{irk})}{\sum_{j=1}^{K} exp(V_{irj})} \tag{5}$$

An NN for region-specific choice is as follows:

$$V_{irk} = (g_L^k \circ g_{L-1} \ldots \circ g_1)(x_i) + z_{rk} \tag{6}$$

where

$$g_l(x) = Tanh(\omega_l x + b_l), l \in \{1, 2, \ldots, L\} \tag{7}$$

Each $g_l(x)$ consists of linear and non-linear unit (i.e., activation function) transformation, where $\omega_l$refers to a parameter matrix containing random weights in the lth layer. g_L^k transforms the last layer into the utility of choice k, L represents the total number of layers in a NN. z_{rk} is an offset term of alternative k for people living in region r. This term has two contributions to the model. First, it represents the region term as a random effect in this NN model. Second, it contributes to the interpretability of NNs, for it represents the offset utility. The variation of this term can be used to analyse the differences in travel behaviours and preferences among regions.

3.3. Random effect-bayesian neural network

Due to the sensitivity to hyperparameters, the economic information extracted from DNN may be unreliable (Wang et al., 2020). To mitigate this problem, the Bayesian Neural Network (BNN) is chosen as the baseline of this study. In contrast to the other plain feedforward NNs, which are prone to over-fitting issues and make decisions only relying on point prediction (Guo et al., 2017; Pereyra et al., 2017), BNNs can avoid the over-fitting issue by introducing uncertainty on the weights and improve the quality of prediction (Blundell et al., 2015).

Based on the theory of utility and the random effect model, a Random Effect-Bayesian Neural Network (RE-BNN) model was designed to predict and analyse the travel mode choice across regions. Fig. 1 illustrates this RE-BNN architecture.

Given a dataset $D = \{x_i, r_i, y_i\}_{i=1}^{N}$, containing N individual trip record (i.e., training input) $x_i \in \mathbb{R}^P$, region of the record $r_i \in \{1, 2, \ldots, R\}$, and label (i.e., corresponding output) $y_i \in \{e_1, e_2, \ldots, e_K\}$, where e_k is the one-hot encoded vector with the only kth element being '1' and the others being '0'. P represents the dimension of the inputs (i.e., the number of independent variables in the model). Let ω denote the collection of all

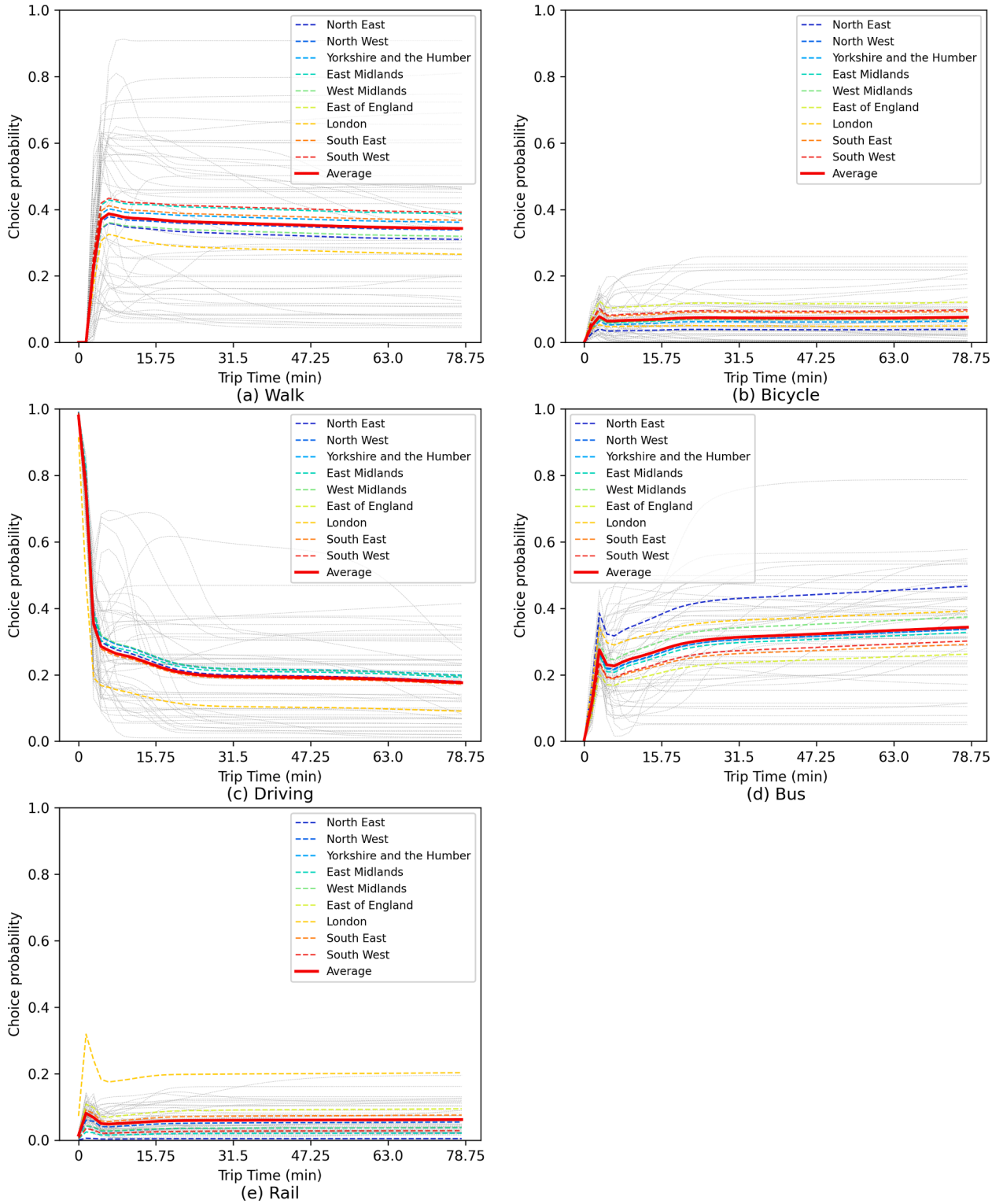

Fig. 8. Choice probability functions with trip time in RE-BNN models (trained by 2016 data) for different regions.

parameter matrices: $\omega = (\omega_1, \omega_2, \ldots, \omega_L)$. Following the Bayesian approach, a prior distribution $p(\omega)$ is placed on the parameter vector ω, which represents the prior belief as to which parameters are likely to have yielded the data before the data points are observed. When some data points are observed, this distribution will be changed to capture different parameters. The distribution $p(\omega)$ can be described as follows by invoking Bayes' theorem[3].

[3] Note that for notational convenience, the bias parameters $b = (b_1, b_2, \ldots, bL)$ and the random effect parameters $z = (z_1, z_2, \ldots, z_R)$ are not represented above.

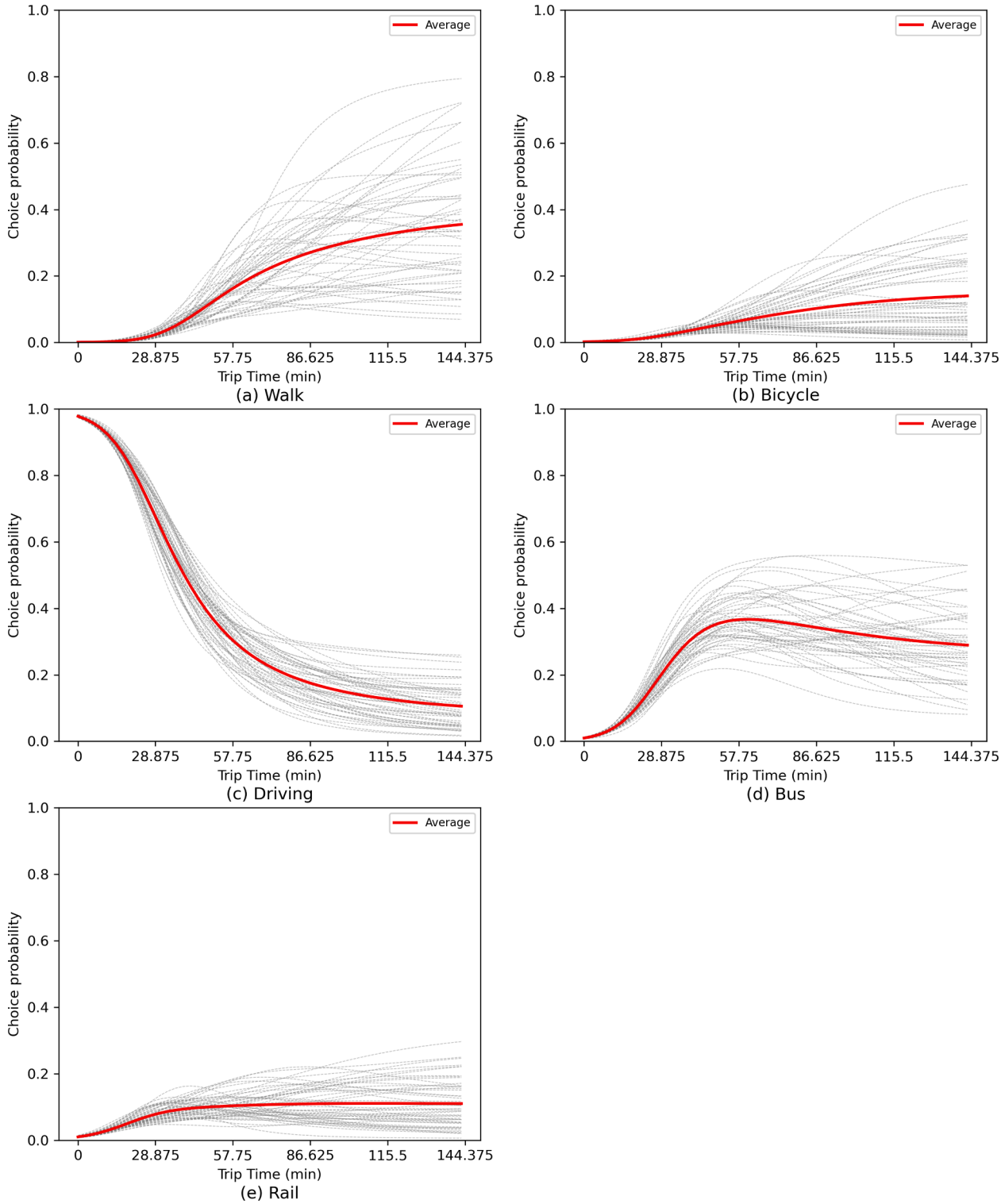

Fig. 9. Choice probability functions with trip time in BNN models (trained by 2016 data) for different regions.

$$p(\omega|D) = \frac{p(D|\omega)p(\omega)}{p(D)} = \frac{\prod_{i=1}^{N} p(y_i|x_i,\omega)p(\omega)}{p(D)} \tag{8}$$

With the distribution $p(\omega)$, we can do the inference process that an output y^* can be predicted given a new input x^* (Kwon et al., 2020):

$$p(y^*|x^*,D) = \int p(y^*|x^*,\omega)p(\omega|D)d\omega \tag{9}$$

Nevertheless, it is analytically intractable to calculate the posterior distribution of weights $p(\omega|D)$ for the reason that it requires taking exceptions $p(\hat{y}|\hat{x}) = \mathbb{E}_{p(\omega|D)}$, which means to adopt an ensemble of infinite NNs (Blundell et al., 2015). Variational inference (Hinton and Van Camp, 1993) is one of commonly used methods for training BNNs. This method approximates the posterior distribution by finding the parameters $\theta$of a tractable variational distribution $q_\theta(\omega)$ that minimises the divergence with the true posterior distribution $p(\omega|D)$. Kullback-Leibler (KL) divergence (Kullback, 1959) is a widely adopted to measure the

Table 4
Travel mode shares in different regions of the UK (testing set of 2016 data).

Mode	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West
Panel 1. True Travel Mode Share									
Walk	10.07%	10.09%	10.65%	10.85%	7.99%	9.08%	12.30%	8.96%	12.35%
Bicycle	1.66%	1.47%	1.80%	1.76%	1.35%	2.97%	2.34%	1.64%	2.64%
Driving	79.99%	80.89%	81.32%	81.25%	84.66%	80.71%	52.08%	83.05%	80.29%
Bus	8.22%	5.63%	4.84%	4.87%	4.77%	2.76%	13.70%	2.51%	3.68%
Rail	0.07%	1.92%	1.40%	1.27%	1.23%	4.48%	19.57%	3.83%	1.04%
Panel 2. RE-BNN									
Walk	10.57%	10.13%	10.1%(0.84%)	10.58%	7.97%	9.37%(0.76%)	12.34%	9.43%	12.48%
	(0.93%)	(0.92%)		(0.82%)	(0.67%)		(1.32%)	(0.71%)	(0.91%)
Bicycle	1.17%	1.54%	1.45%(0.26%)	1.85%	1.17%	2.74%(0.54%)	2.64%	1.64%	2.22%(0.4%)
	(0.23%)	(0.29%)		(0.32%)	(0.23%)		(0.47%)	(0.32%)	
Driving	80.01%	80.77%	82.25%(0.93%)	82.62%	84.82%	80.79%(1.1%)	51.76%	82.57%	80.54%
	(0.94%)	(1.02%)		(0.92%)	(0.76%)		(1.81%)	(1.03%)	(0.9%)
Bus	8.08%	5.36%	4.63%(0.48%)	4.01%	4.62%	2.71%(0.29%)	13.16%	2.55%	3.54%
	(0.74%)	(0.53%)		(0.38%)	(0.44%)		(1.05%)	(0.25%)	(0.36%)
Rail	0.17%	2.2%(0.3%)	1.56%(0.25%)	0.94%	1.43%	4.4%(0.61%)	20.1%	3.81%	1.21%
	(0.03%)			(0.15%)	(0.21%)		(1.58%)	(0.57%)	(0.18%)
Panel 3. BNN									
Walk	11.2%	10.5%(0.8%)	10.03%(0.68%)	9.97%	8.58%	9.32%(0.68%)	12%(1.14%)	8.94%	11.55%
	(0.76%)			(0.64%)	(0.63%)			(0.63%)	(0.76%)
Bicycle	2.15%	1.73%(0.2%)	1.82%(0.2%)	2.07%	1.37%	1.84%(0.23%)	2.61%	1.37%	1.81%
	(0.27%)			(0.25%)	(0.18%)		(0.27%)	(0.17%)	(0.21%)
Driving	79.49%	80.09%	81.54%(0.9%)	81.94%	83.89%	82.75%	51.86%	84%(0.88%)	80.91%
	(0.83%)	(0.99%)		(0.83%)	(0.82%)	(0.87%)	(2.21%)		(0.82%)
Bus	5.78%	5.55%	4.66%(0.43%)	4.04%	4.28%	3.34%(0.3%)	13.8%	2.81%	3.83%
	(0.48%)	(0.47%)		(0.35%)	(0.35%)		(1.16%)	(0.23%)	(0.34%)
Rail	1.39%	2.14%	1.95%(0.35%)	1.99%	1.88%	2.75%(0.47%)	19.73%	2.88%	1.9%(0.33%)
	(0.24%)	(0.36%)		(0.35%)	(0.34%)		(2.27%)	(0.53%)	
Panel 4. DNN									
Walk	12.31%	11.95%	11.21%(0.17%)	11.54%	9.78%	10.59%	10.46%	9.98%	12.74%
	(0.2%)	(0.2%)		(0.17%)	(0.19%)	(0.17%)	(0.77%)	(0.17%)	(0.18%)
Bicycle	2.26%	1.86%	1.88%(0.11%)	2.2%(0.12%)	1.72%	1.91%(0.11%)	3.01%	1.63%	2.04%
	(0.16%)	(0.09%)			(0.07%)		(0.29%)	(0.08%)	(0.09%)
Driving	77.79%	78.27%	79.77%(0.3%)	79.54%	81.81%	80.78%	51.05%	82.15%	78.81%
	(0.3%)	(0.3%)		(0.28%)	(0.28%)	(0.32%)	(0.96%)	(0.29%)	(0.27%)
Bus	5.89%	5.43%	4.76%(0.14%)	4.34%	4.43%	3.54%(0.11%)	13.87%	2.93%	4.09%
	(0.19%)	(0.15%)		(0.14%)	(0.12%)		(0.66%)	(0.08%)	(0.13%)
Rail	1.74%	2.49%(0.1%)	2.37%(0.11%)	2.38%	2.26%	3.18%(0.14%)	21.6%(0.7%)	3.31%	2.32%
	(0.09%)			(0.12%)	(0.09%)			(0.16%)	(0.11%)
Panel 5. MNL									
Walk	7.90%	8.34%	9.45%	9.81%	8.35%	9.80%	11.21%	8.65%	9.93%
Bicycle	7.94%	8.37%	9.55%	10.03%	8.48%	10.00%	10.94%	8.83%	10.06%
Driving	68.02%	65.63%	60.99%	59.65%	65.70%	59.68%	33.96%	64.06%	59.41%
Bus	8.06%	8.39%	9.48%	9.74%	8.23%	9.60%	15.82%	8.43%	9.80%
Rail	8.08%	9.26%	10.52%	10.77%	9.25%	10.93%	28.08%	10.03%	10.80%

closeness of two distributions:

$$KL\{q_\theta(\omega)||p(\omega|D)\} = \int q_\theta(\omega) log \frac{q_\theta(\omega)}{p(\omega|D)} d\omega \tag{10}$$

Minimising this KL divergence is also intractable because it directly depends on the true posterior distribution $p(\omega|D)$. However, minimising the KL divergence is equivalent to maximising the Evidence Lower Bound (ELBO) with respect to the defined variational parameters $q_\theta(\omega)$,

$$-\int q_\theta(\omega) log p(y|x,\omega) d\omega + KL\{q_\theta(\omega)||p(\omega)\} \tag{11}$$

This procedure of variational inference is a standard approach in Bayesian modelling (Jordan et al., 1999). Variational inference casts posterior inference as an optimisation problem, where optimisation can be done by gradient-based methods. In this study, ELBO was adopted as the loss function to find the best distribution of parameter ω, b and z.

3.4. Logit model

In the field of travel behaviour analysis, logit models have been adopted for exploring different kinds of behaviour choices, such as departure time choice (Afandizadeh and Safari, 2020), since the seminal paper by McFadden (1986).

In a logit model, the utilities for choosing a travel mode are generated by a logistic function of the input variables. Taking the MNL model as an example, the utility of travel alternative k is defined as follows:

$$U_k = \beta_k X_k + \epsilon_k \tag{12}$$

where β_k is a parameter vector for alternative k. The unobserved utility ϵ_k is always treated as a stochastic element. It is assumed by researchers to be independent and identically distributed extreme values in the logit model. The probability of choosing mode k for passenger i is defined as follows:

Table 5
Absolute value of errors between the predicted market shares and the true travel mode shares.

Mode	North East	North West	Yorkshire and the Humber	East Midlands	West Midlands	East of England	London	South East	South West	AVE
Panel 1. RE-BNN										
Walk	0.50%	0.04%	0.55%	0.26%	0.02%	0.29%	0.04%	0.47%	0.13%	0.26%
Bicycle	0.49%	0.07%	0.35%	0.09%	0.18%	0.24%	0.30%	0.00%	0.42%	0.23%
Driving	0.03%	0.11%	0.94%	1.37%	0.15%	0.07%	0.31%	0.48%	0.25%	0.41%
Bus	0.14%	0.27%	0.21%	0.86%	0.15%	0.05%	0.55%	0.03%	0.14%	0.27%
Rail	0.10%	0.28%	0.17%	0.34%	0.20%	0.08%	0.52%	0.03%	0.17%	0.21%
SUM	1.25%	0.77%	2.20%	2.92%	0.69%	0.72%	1.72%	1.02%	1.11%	**1.38% (0.72%)**
Panel 2. BNN										
Walk	1.13%	0.41%	0.62%	0.88%	0.59%	0.24%	0.30%	0.02%	0.81%	0.56%
Bicycle	0.49%	0.26%	0.02%	0.30%	0.02%	1.13%	0.27%	0.27%	0.83%	0.40%
Driving	0.50%	0.80%	0.23%	0.69%	0.78%	2.03%	0.22%	0.94%	0.62%	0.76%
Bus	2.44%	0.08%	0.19%	0.83%	0.49%	0.58%	0.10%	0.30%	0.16%	0.57%
Rail	1.32%	0.21%	0.55%	0.72%	0.65%	1.73%	0.15%	0.95%	0.86%	0.79%
SUM	5.88%	1.76%	1.61%	3.42%	2.53%	5.72%	1.04%	2.48%	3.26%	**3.08% (1.62%)**
Panel 3. DNN										
Walk	2.24%	1.87%	0.57%	0.70%	1.79%	1.51%	1.84%	1.02%	0.39%	1.32%
Bicycle	0.61%	0.38%	0.08%	0.44%	0.37%	1.06%	0.66%	0.01%	0.60%	0.47%
Driving	2.19%	2.61%	1.54%	1.71%	2.85%	0.07%	1.02%	0.90%	1.49%	1.60%
Bus	2.33%	0.20%	0.08%	0.53%	0.34%	0.78%	0.17%	0.42%	0.42%	0.59%
Rail	1.67%	0.57%	0.97%	1.11%	1.03%	1.30%	2.03%	0.52%	1.28%	1.16%
SUM	9.04%	5.64%	3.25%	4.49%	6.39%	4.72%	5.73%	2.87%	4.17%	**5.14% (1.75%)**
Panel 4. MNL										
Walk	2.18%	1.75%	1.19%	1.03%	0.36%	0.72%	1.09%	0.31%	2.42%	1.23%
Bicycle	6.29%	6.90%	7.75%	8.26%	7.13%	7.03%	8.59%	7.19%	7.42%	7.40%
Driving	11.96%	15.25%	20.33%	21.60%	18.97%	21.04%	18.12%	18.99%	20.88%	18.57%
Bus	0.16%	2.76%	4.64%	4.87%	3.46%	6.84%	2.12%	5.92%	6.12%	4.10%
Rail	8.01%	7.34%	9.13%	9.50%	8.02%	6.45%	8.50%	6.20%	9.76%	8.10%
SUM	28.60%	34.00%	43.04%	45.27%	37.94%	42.07%	38.42%	38.60%	46.61%	**39.39% (5.34%)**

$$P_{ik} = \frac{exp(\beta_k X_{ik})}{\sum_{j=1}^{K} exp(\beta_j X_{ik})} \tag{13}$$

where K denotes the total number of alternatives. Given the parameter β, the likelihood function of the MNL model can be formulated as follows:

$$L(\beta) = \prod_{i=1}^{N} \prod_{k=1}^{K} \left(\frac{exp(\beta_k X_{ik})}{\sum_{j=1}^{K} exp(\beta_j X_{ik})} \right) \tag{14}$$

The Maximum Likelihood Estimation approach can be used to search for the best parameter $\widehat{\beta} = argmax_{\beta} L(\beta)$.

4. Setup of experiments

This section briefly introduces the dataset used to test the models in this study. Since five models will be compared, including three NN models and one conventional logit model, the process of tuning hyper-parameters for NN models and the specification of the utility functions for the logit model are presented in this section.

4.1. Datasets

The experiments rely on the National Travel Survey (NTS) dataset of Great Britain from 2005 to 2016, which is publicly provided by Department for Transport (2020). The study area is divided into nine regions according to the household address of the respondents: North East, North West, Yorkshire and the Humber, East Midlands, West Midlands, East of England, London, South East, and South West. From 2005 to 2016, there are 121765 respondents from 69208 households being annually interviewed. After simple data cleaning, the database contains a total of 2100492 observations on the details of all their travel activities, including mode, purpose, original, destination, travel time and travel distance. The database also provides information about respondents' households and themselves, including gender, education, income, car-ownership, and employment, et al.

The original database consists of several interconnected sub-tables containing data on the individual, household, attitudes, trip, vehicle, stage and day level. Based on the determinants of travel mode choice, which has been discussed in Section 2.1, fifteen variables are selected from the individual, household, and trip tables for this study. The data on population density in these regions is provided by Greater London Authority (2018).

Data pre-processing minimises the noise of the data while retaining the key information, leading to an effective classification performance (Bishop et al., 1995). Encoding the categorical variables to numerical values plays an essential role in modelling the data and simplifies the learning process, which is commonly used in ML algorithms (Potdar et al., 2017). The dataset used in this study contains both numerical and categorical variables. Some categorical variables are ordinal, such as age and income, which can be transformed into continuous variables by assigning them the interval mid-point values to increase comparability. Other categorical variables are transformed into dummy variables by the one-hot encoding technique. In the dataset, the output y_i denotes the mode alternative. In order to maintain choice's regional consistency, several alternatives have been merged, such as 'bus in London' and 'other local bus'. After merged, there are five travel mode alternatives:

walk, bicycle, driving, bus, and rail (including London Underground). The input x_i is the attributes of each trip (i.e., variables in Table C.9 excluding the region), and r_i represents the household region of each trip's respondent. The descriptions and summary statistics of the dataset used in this paper can be seen in A.

4.2. Hyperparameter searching for ML models

For comparison purposes, in addition to our RE-BNN model, this study also trained BNNs and DNNs using the same datasets as RE-BNN, except for the regional feature.[4] Since NNs are sensitive to the selection of hyperparameters, one of the challenges in the architecture of NN-based models is hyperparameters searching, on which the predictive performance of NN largely depends (Wang et al., 2020). Table 1 shows the crucial hyperparameters in the architecture of RE-BNN and DNN and the range of their values. Width n denotes the number of neurons in each layer of RE-BNN or DNN. Since the data contain various sets of variables with different scales and ranges, standardising them before training the model can speed up the learning process for some ML algorithms using the gradient descent method as an optimisation algorithm (Raschka, 2014). In order to select the best hyperparameters for the ML models, we use the grid search method to find the best hyperparameter combination. To be specific, this study randomly samples 100,000 observations from the original dataset as a sub-dataset. Then the dataset is dividing it into training and testing sets at a ratio of 4:1 to evaluate the model under different combinations of hyperparameters. Fig. 2 shows the prediction accuracies of the RE-BNN and DNN models with different hyperparameters. According to the results, when learning rate = 0.1 and Width n=15, the prediction accuracy of RE-BNN performs best. The same hyperparameters were applied to the architecture of BNN. For DNN, this study used learning rate = 0.01, Width n=10 and Batch size = 100.

4.3. Specification for logit models

Considering that the MNL model are frequently used in travel mode choice modelling and the MiXed Logit (MXL) model can take the random nature of parameters into account (Ben-Akiva and Lerman, 2018), this study has also fitted MNL and MXL models as the benchmark for comparison. The same training dataset is used to calibrate the MNL model.

Unlike ML models that consider a RUM as a supervised probabilistic classifier, classical logit models rely on alternative-specific functions of utilities. The specification of MNL and MXL models is based on the correlation coefficients between alternatives and attributes (see Appendix C.8), and variables selected for specifying the alternative-specific utility functions are listed in Table 2. The number in the parenthesis is the value of the categorical variable before being transformed into dummy variables. Descriptions of these variables are shown in Appendix C.9. Note that it is ensured that all alternative-specific functions contain the same parameter (trip time) for the drawing of choice probability functions in Section 5.3.2.

5. Experiment results

This section first presents a brief exploratory analysis of the travel mode choice in the UK in 2016. Then it shows the predictive performance of the RE-BNN model in comparison with four models (i.e. BNN, DNN, MNL and MXL). After that, it describes the interpretation of the RE-BNN model regarding offset utility, choice probability function, and local travel mode shares.

[4] To avoid errors caused by ignoring regional factors when comparing the training results, for models other than RE-BNN, a population density attribute was added to distinguish different regions.

5.1. Exploratory analysis of travel mode choice in the UK

Table 3 describes the frequency and percentage of trips by travel mode choice in the UK in 2016. Driving has the largest proportion of 78.48%, followed by walking (9.93%) and taking a bus (5.29%). Rail and bicycle are less popular than the others, which occupied 4.42% and 1.89% of the total trips, respectively. Travel mode shares of different regions are presented in Fig. 3. According to the figure, London greatly differs from the other regions with a significantly lower share of driving (about 50%) and a higher share of rail.

5.2. Prediction accuracy

Each NN-based model is trained and evaluated 50 times for each one-year data. Each one-year dataset is divided into training and testing sets in the ratio of 4:1. Fig. 4 shows the prediction accuracy of RE-BNN, BNN, DNN, MNL and MXL for each year. For three NN-based models, curves with relatively light colours represent the predictive accuracy of each single training result, and dark curves are the average accuracy of all training results. A table containing the specific prediction accuracy information of models is attached in Appendix B.7. The result reveals that ML models (i.e., RE-BNN, BNN and DNN) on average outperform the MNL model and the MXL model by around 10 % and 2 % predictive accuracy, respectively, which is in line with the previous researches by Hagenauer and Helbich (2017); Lhéritier et al. (2019) and Zhao et al. (2020). The average prediction performances of RE-BNNs and BNNs are similar, outperforming DNNs by around 1 to 2 percentage points. This result reveals that the improvement of the predictive power of RE-BNN is mainly attributed to the introduction of Bayesian's uncertainty into NNs. In addition to the high predictive accuracy, compared with DNN, the curve of RE-BNN is more stable across different year datasets, indicating that it is more robust and can fit better to different datasets.

Since the main objective of introducing the random effect parameter into our model is to enhance the ability of the model to copy with the regional heterogeneity, this study then investigates the local-level predictive accuracy of RE-BNN and the other three models. Fig. 5 illustrates the predictive accuracy of these five models trained by the 2016 dataset in different regions. Note that the accuracy for three ML models is represented by the average value of 50 training results. The figure shows that, compared with other regions, the accuracy for London's travel mode choice are relatively low for all these five models, including the RE-BNN model. This result indicates that the RE-BNN model does not have obvious advantages in predicting individual-level travel mode choice in different regions. Nevertheless, for all regions, the prediction accuracy of RE-BNN and BNN is higher than that of DNN, proving the advantage of introducing Bayesian's uncertainty in the process of prediction.

5.3. Interpretation of RE-BNN

In this section, we interpreted the RE-BNN model via the behaviour information extracted from the model. Specifically, we analysed the regional heterogeneity of choice preferences, the choice probability functions of five different modes and the market shares across regions.

5.3.1. Offset utility of travel mode across regions

Based on the theory of utility and Eq. (6), the region-specific random effect parameter z_{rk} can be regarded as the offset utility of mode k for region r in the context of behaviour analysis. Therefore, by comparing the values of this parameter, it is possible for us to analyse the various utility of the same travel mode and the heterogeneity in passengers' travel preferences across different regions. Fig. 6 shows the variation of the region-specific random effect parameter z_{rk} in 50 RE-BNN models trained by the 2016 dataset. Because in RE-BNNs, it is the distribution of parameters that are estimated, including the random effect parameter z_{rk}, the mean values of these distributions are extracted for behaviour

analysis. The length of each box in Fig. 6 represents the variation degree of the random effect parameter obtained from 50 RE-BNN training results, and the position of each box represents the average level of the offset utility of choosing a travel mode for people living in different regions.

Among these five alternatives, the regional difference in the behavioural preferences of choosing rail as a mode of transportation is the most significant. When other variables are the same, rail brings significantly more utility to people living in London than those living in other regions, which is consistent with the fact that the Underground as a popular and convenient means of transportation only serves London and some parts of the adjacent counties (Office of Rail and Road, 2014). The utility of choosing rail living in North East is the least among all regions.

In contrast to the highest utility brought by rail in London, the utility of choosing driving as the travel mode is the lowest for London residents, compared to people living in other regions. The potential reason for this phenomenon is that London's social policy environment is not friendly to drive. The London Congestion Charge was introduced in 2003 as a measurement to alleviate traffic congestion in downtown London. Under this policy, cars entering the central area of about 16 km^2 are charged (£5 in 2003, rising to £11.50 in 2016) (Ambühl et al., 2018), which reduces the utility of driving for London residents.

In East of London and South West, passengers prefer to select riding as their travel mode, while people in North East obtain more utility when travelling by bus. Regarding the preference for walking, there is no significant regional difference.

It is also possible for us to investigate the temporal variation of the utility of travel modes for people living in different regions. The average values of the random effect parameter in each year are used to represent the overall level of offset utilities. Fig. 7 shows this parameter of different regions and alternatives varies during the period from 2005 to 2016. The length of each box represents the degree of the variation in the offset utility of five travel modes in this period. The position of each box represents the overall level of the offset utility for residents from different regions during this period, which is similar to the boxes' positions in 2016. This similarity indicates that the regional differences in the behavioural preferences of travel mode choice in 2016 are similar to the average level during the past twelve years. The short length of boxes in Figs. 7a, 7c, 7d, and 7e indicates that the variations of the utilities of choosing walking, driving, bus and rail in these regions are not very significant. In contrast, the utility of riding as a mode of travel have more considerable changes in the same period.

5.3.2. Choice probability functions across regions

Under the utility maximisation theory, the regional differences in utility brought by various alternatives directly lead to regional heterogeneity in choice probabilities for travel mode choices. The choice probability functions for each alternative can be visualised via numerical simulation (Wang et al., 2020). Fig. 8 and 9 visualise how the choice probability functions in 50 RE-BNN and 50 BNN models (trained by the 2016 training set) in different regions vary as the trip time increases when remaining all other features' values are consistent with their average values. The grey curves represent each training result in each sub-figure, and the colourful curves are the mean value of all the training results. Note that the significant variation of the grey curves in Fig. 8a, 8c and 8d indicates that although trained by the same dataset and showing similar predictive powers, the choice probability functions of these alternatives in the RE-BNN models are quite different. Nevertheless, it is worth noting that these irregular behavioural patterns are not necessarily negative since they may exist in other literature and can be regarded as successfully identifying flexible behavioural patterns (Wang et al., 2020).

According to Fig. 8, overall, the average lines (represented by red colour in the figure) represent the overall trend of the change in the probabilities. When the value of trip time is very close to zero, the probability of driving is relatively high, in contrast to the low probabilities of choosing the other four alternatives. As the trip time increases, the driving probability has a decreasing trend; the probabilities for taking the bus and walking increase and then level off; yet the probabilities for riding and rail still maintain low values. Similar trends can be found in the probabilities function learned by BNNs (Fig. 9), except less flexible. Most curves are intuitive and reasonable, notably driving and taking the bus. People may prefer to drive by themselves on a short trip, but when trips may cost too much time, they may be more likely to take public transportation, such as the bus. Concurrently, the choice probabilities of rail and bicycle are less sensitive to the trip time. However, the walking probability function may suffer the interpretability problem, especially for that learned by BNNs. This issue may attribute to the trip time variable is not alternative-specific, subjected to the data provided by the dataset used in this study.

In terms of the regional differences, the colourful curves describe the choice probability functions for passengers in different regions, demonstrating the power of RE-BNNs to automatically learn the choice probability functions across different regions. For the same travel mode, the more scattered the curves, the more significant the regional differences in the travel mode behaviour. A curve above the regional average curve (represented by the red curve in each sub-figure) indicates that choosing this travel mode can bring more utility to people in this region than those from other regions and vice versa. In Fig. 8e, the orange curve (representing the choice probability for passengers in London) is higher than all the other curves by around 10% to 20%. This value means that when other variables are the same (such as age, income, trip distance, etc.), it is possible that among 100 passengers, 10 to 20 more people in London would choose rail than in other regions. The opposite situation can be observed in the behaviours of choosing walking and driving as the travel mode (shown in Fig. 8a and 8c). While the difference in riding behaviours is not significant across regions the colourful curves in Fig. 8b are very concentrated. In comparison to the choice probability functions' learning ability of RE-BNNs, BNNs are not able to capture the probability functions in different regions.

5.3.3. Travel mode shares prediction across regions

This subsection shows the results of predicting local market shares for each travel mode. Note that if the predictive capability investigated in Section 5.2 is at the individual level, this predictive power of travel mode shares can be regarded as the aggregate-level predictive power.

Table 4 describes the prediction results of travel mode's market shares predicated by the RE-BNN models and the other three models in the 2016 dataset in different regions. Each value refers to the mean value of the travel mode shares predicted by 50 models. The value in the parenthesis refers to the standard deviation. According to the table, the values of RE-BNN's predictive results are very close to the real travel mode shares, demonstrating its considerable capability in predicting travel modes' market shares in different regions. The standard deviations in the RE-BNN mode and the BNN model are on average higher than that in the DNN model by about 0.4%, indicating that the predictive power of the local travel mode share of DNN is stable than that of RE-BNN and BNN.

Then the errors between the prediction results and the true values are calculated to quantify the sum of differences, shown in Table 5. The bold values and numbers in the parenthesis in Table 5 represent the mean values and the standard deviation of the predicted aggregated market shares across regions, respectively, which reveals models' prediction power of market shares. The sum of five modes' average errors across regions of the RE-BNN models (1.38%) is the least among the four models, showing that the prediction result of aggregated market shares of RE-BNN is the most accurate, outperforming the other models. The standard deviation of the RE-BNN models (0.72%) is also less than the other three models, revealing its stable prediction power of aggregated market shares across regions.

6. Conclusions

While many studies have used ML methods for predicting travel mode choice, none of them reveal the regional heterogeneity of travel mode choice. To fill this research gap, this study develops a Random Effect Bayesian Neural Network (RE-BNN) to predict mode choice across regions and reveal the regional heterogeneity. Compared with the DNN model, the RE-BNN model has two improvements in the structure: first, it includes a random effect term representing regional-specific offset utilities; second, the Bayesian's uncertainty is introduced to enhance the stability of the model's predictive accuracy. The merits of the RE-BNN model are demonstrated in a case study that predicts the UK's national travel mode alternatives from 2005 to 2016. Most importantly, the RE-BNN model is demonstrated to have the capability to automatically learn and reveal the regional heterogeneity of mode preferences, which is impossible for other four models. For instance, from the offset utility terms and the choice probability functions derived from the RE-BNN model, it is revealed that compared with other regions, residents in London have a higher preference for rail and a lower preference for driving. In addition to its outstanding interpretability power of regional travel preferences, the RE-BNN model also achieves better predictive performances than DNN and MNL models in terms of individual-level (i.e., individual trips) and aggregated-level (i.e., local travel mode shares) perspectives. Moreover, the RE-BNN model is more robust for datasets of different years, outperforming DNN and MNL models.

This study sheds light on future research directions. First, the RE-BNN model has the potential of exploring travel preferences across other social groups, including age groups. The travel behaviour of different generations, especially millennials, has gained increasing attention from researchers (Rive et al., 2015; Hjorthol, 2016). Second, adding more regional-specific variables related to human travel behaviour, such as regional investment in cycling infrastructure or public transport, is another direction for the improvement of the RE-BNN model in better predictive performance across multiple regions.

CRediT authorship contribution statement

Yutong Xia: Conceptualization, Methodology, Software, Writing - original draft. **Huanfa Chen:** Conceptualization, Data curation, Writing - original draft, Writing - review & editing. **Roger Zimmermann:** 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.

Appendix A. Summary statistics of the dataset

Table A.6.

Table A.6
Summary Statistics of selected variables.

Panel 1. Numerical Variables

Variables	mean	std	min	25%	50%	75%	max
Trip_distance (miles)	8.92	20.03	0.1	1.9	3.6	8	613
Trip_time (minutes)	24.22	27.88	0	10	15	30	1155
Household_children	0.64	0.97	0	0	0	1	8
Household_bike	1.40	1.36	0	0	1	2	8
Household_car	1.40	0.85	0	1	1	2	8
Household_licence	1.79	0.83	0	1	2	2	7
Population_density (Pop per km_2)	1015.70	1700.75	231	307	446	510	5590

Panel 2. Categorical Variables

Variables	Variable values		Count
Trip_purpose	1	Commuting	425122
	2	Business	102443
	3	Education / escort education	131691
	4	Shopping	417029
	5	Personal business	208363
	6	Leisure	569549
	7	Other	246295
Household_employeed	1	None	646871
	2	1 part time or full time	395729
	3	2 part time or full time	814977
	4	3 or more part time or full time	242915
Individual_age	1	0 - 16 years	19301
	2	17 - 20 years	103996
	3	21 - 29 years	258607
	4	30 - 39 years	408984
	5	40 - 49 years	469555
	6	50 - 59 years	371614
	7	60 years +	468435
Individual_education	1	Degree level or above	626540
	2	Other type of qualification	1473952

(continued on next page)

Table A.6 (*continued*)

Panel 1. Numerical Variables

Variables	mean	std	min	25%	50%	75%	max
Individual_income	1	Less than £25,000					1458396
	2	£25,000 to £49,999					487843
	3	£50,000 and over					154253
Individual_gender	1	Male					996838
	2	Female					1103654
Household_settlement	1	Urban					1663277
	2	Rural					437215
Household_region	1	North East					110344
	2	North West					290214
	3	Yorkshire and the Humber					204610
	4	East Midlands					187078
	5	West Midlands					216772
	6	East of England					246156
	7	London					252901
	8	South East					362747
	9	South West					229670

Appendix B. Prediction accuracy

Table B.7.

Table B.7
Prediction accuracy of RE-BNNs, BNNs, DNNs, MNL and MXL models trained by different year data.

	2005	2006	2007	2008	2009	2010
RE-BNN(Average)	0.872	0.867	0.866	0.865	0.862	0.864
BNN(Average)	0.871	0.868	0.867	0.864	0.861	0.863
DNN(Average)	0.866	0.862	0.86	0.857	0.854	0.858
MNL	0.743	0.742	0.738	0.735	0.731	0.742
MXL	0.841	0.837	0.836	0.830	0.830	0.832

	2011	2012	2013	2014	2015	2016
RE-BNN(Average)	0.865	0.865	0.864	0.861	0.861	0.863
BNN(Average)	0.866	0.865	0.865	0.861	0.861	0.863
DNN(Average)	0.857	0.858	0.858	0.849	0.851	0.855
MNL	0.733	0.732	0.73	0.728	0.732	0.741
MXL	0.830	0.834	0.833	0.827	0.829	0.833

Appendix C. Correlation between alternatives and attributes

Tables C.8 and C.9.

Table C.8
Correlation coefficients between alternatives and attributes.

	Walk	Bicycle	Car or van	Bus	Rail
Trip_distance	**−0.1197**	−0.0396	0.0467	−0.0411	**0.1631**
Trip_time	0.0202	**−0.0106**	**−0.1686**	0.0788	**0.2381**
Household_employeed	−0.0494	0.0138	0.0487	−0.0567	0.0331
Household_children	−0.006	−0.0008	0.0355	−0.0362	−0.0207
Household_bike	−0.0324	**0.0779**	0.0726	−0.0991	−0.0358
Household_car	**−0.1281**	−0.0598	**0.2842**	**−0.2291**	−0.0822
Household_licence	**−0.1046**	−0.0287	0.2227	**−0.2158**	−0.0252
Individual_age	−0.0416	−0.0294	**0.1207**	−0.0748	−0.08
Individual_income	−0.0487	0.0093	0.038	−0.09	0.0997
Population_density	0.0153	0.0208	**−0.2319**	**0.1538**	**0.2651**
Trip_purpose_1	−0.073	**0.0697**	−0.0596	0.0369	**0.1446**
Trip_purpose_2	−0.0546	−0.0089	0.0385	−0.0296	0.0449
Trip_purpose_3	0.044	−0.0046	−0.0613	0.0604	−0.0088
Trip_purpose_4	−0.0104	−0.0273	0.0252	0.0382	−0.0666
Trip_purpose_5	−0.0157	−0.0156	0.0267	0.0069	−0.0309
Trip_purpose_6	−0.0607	0.0112	0.066	−0.0308	−0.0173
Household_settlement_1	0.0438	0.0203	**−0.1147**	0.0826	0.0604
Individual_education_1	−0.0063	0.0236	−0.0277	−0.0448	**0.1085**
Individual_gender_1	−0.0178	0.0638	−0.0049	−0.0342	0.0359

Table C.9
Descriptions of selected variables.

Variable		Variables Type	Description		
Trip Variables	Trip_distance	Numrical	Trip distance (miles)		
	Trip_time	Numrical	Total trip travelling time (minutes)		
	Trip_purpose	Categorical	Trip purpose	1	Commuting
				2	Business
				3	Education / escort education
				4	Shopping
				5	Personal business
				6	Leisure
				7	Other
Household Variables	Household_children	Numrical	Number of children in household		
	Household_bike	Numrical	Number of household bicycles		
	Household_car	Numrical	Number of household 3 and 4 wheeled cars		
	Household_licence	Numrical	Number of persons in household with full car licence		
	Household_employeed	Categorical	Number of employed in household	1	None
				2	1 part time or full time
				3	2 part time or full time
				4	3 or more part time or full time
Individual Variables	Individual_age	Categorical	Age	1	0 - 16 years
				2	17 - 20 years
				3	21 - 29 years
				4	30 - 39 years
				5	40 - 49 years
				6	50 - 59 years
				7	60 years +
	Individual_education	Categorical	Level of highest qualification	1	Degree level or above
				2	Other type of qualification
	Individual_income	Categorical	Individual Income	1	Less than £25,000
				2	£25,000 to £49,999
				3	£50,000 and over
	Individual_gender	Categorical	Gender	1	Male
				2	Female
Built Evironment Variables	Population_density	Numrical	Population per square kilometre		
	Household_settlement	Categorical	Settlement	1	Urban
				2	Rural
Social Evironment Variables	Household_region	Categorical	Household Region	1	North East
				2	North West
				3	Yorkshire and the Humber
				4	East Midlands
				5	West Midlands
				6	East of England
				7	London
				8	South East
				9	South West

References

Afandizadeh Zargari, S., Safari, F., 2020, 'Using clustering methods in multinomial logit model for departure time choice', J. Adv. Transp. 2020.

Ajzen, I., 1991. The theory of planned behavior. Organiz. Behav. Human Decision Processes 50 (2), 179–211.

Ambühl, L., Loder, A., Becker, H., Menendez, M., Axhausen, K.W., 2018. Evaluating london's congestion charge: An approach using the macroscopic fundamental diagram, in '7th Transport Research Arena (TRA 2018)'. IVT, ETH Zurich.

Ashalatha, R., Manju, V.S., Zacharia, A.B., 2013. Mode choice behavior of commuters in thiruvananthapuram city. J. Transp. Eng. 139 (5), 494–502.

Bell, A., Fairbrother, M., Jones, K., 2019. Fixed and random effects models: making an informed choice. Qual. Quant. 53 (2), 1051–1074.

Ben-Akiva, M., Lerman, S.R., 2018. Discrete choice analysis: theory and application to travel demand. Transportation Studies.

Bishop, C.M., et al., 1995. Neural networks for pattern recognition. Oxford University Press.

Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D., 2015. Weight uncertainty in neural network. In: Bach, F., Blei, D. (Eds.), 'Proceedings of the 32nd International Conference on Machine Learning', vol. 37 of Proceedings of Machine Learning Research, PMLR, Lille, France, pp. 1613–1622.

Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res. Synth. Methods 1 (2), 97–111.

Brownstone, D., Train, K., 1998. Forecasting new product penetration with flexible substitution patterns. J. Econometr. 89 (1), 109–129.

Casper, M., 2001. A definition of social environment. Am. J. Public Health 91 (3), 465–470.

Chen, C., Gong, H., Paaswell, R., 2008. Role of the built environment on mode choice decisions: additional evidence on the impact of density. Transportation 35 (3), 285–299.

Cheng, L., Chen, X., De Vos, J., Lai, X., Witlox, F., 2019. Applying a random forest method approach to model travel mode choice behavior. Travel Behav. Soc. 14, 1–10.

Cherchi, E., Cirillo, C., 2010. Validation and forecasts in models estimated from multiday travel survey. Transp. Res. Rec. 2175 (1), 57–64.

de Dios Ortúzar, J., Willumsen, L.G., 2011. Modelling transport. John wiley & sons.

Delbosc, A., Ralph, K., 2017. A tale of two millennials. J. Transp. Land Use 10 (1), 903–910.

Department for Transport, 2020, 'National travel survey, 2002–2019', [data collection]. 14th Edition. UK Data Service. SN: 5340, doi: 10.5255/UKDA-SN-5340-10.

Ding, C., Lin, Y., Liu, C., 2014. Exploring the influence of built environment on tour-based commuter mode choice: A cross-classified multilevel modeling approach. Transp. Res. Part D: Transp. Environ. 32, 230–238.

Ding, C., Wang, D., Liu, C., Zhang, Y., Yang, J., 2017. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. Transp. Res. Part A: Policy Practice 100, 65–80.

Domencich, T.A., McFadden, D., 1975. Urban travel demand-a behavioral analysis, Technical report.

Doshi-Velez, F., Kim, B., 2017. Towards a rigorous science of interpretable machine learning.

Ewing, R., Cervero, R., 2010. Travel and the built environment. J. Am. Plann. Assoc. 76 (3), 265–294.
Gardiner, J.C., Luo, Z., Roman, L.A., 2009. Fixed effects, random effects and gee: What are the differences? Stat. Med. 28 (2), 221–239.
Greater London Authority, 2018. 'Land area and population density',https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough.
Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On calibration of modern neural networks. In: Precup, D., Teh, Y.W. (Eds.), 'Proceedings of the 34th International Conference on Machine Learning', Vol. 70 of Proceedings of Machine Learning Research, PMLR, pp. 1321–1330.
Habibian, M., Kermanshah, M., 2013. Coping with congestion: Understanding the role of simultaneous transportation demand management policies on commuters. Transp. Policy 30, 229–237.
Hagenauer, J., Helbich, M., 2017. A comparative study of machine learning classifiers for modeling travel mode choice. Expert Syst. Appl. 78, 273–282.
Heinen, E., Maat, K., Van Wee, B., 2011. The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. Transp. Res. Part D: Transp. Environ. 16 (2), 102–109.
Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation 30 (2), 133–176.
Hensher, D.A., Rose, J.M., 2007. Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study', Transp. Res. Part A: Policy Practice 41(5), 428–443. Bridging Research and Practice: A Synthesis of Best Practices in Travel Demand Modeling.
Hillel, T., Bierlaire, M., Jin, Y., 2019. A systematic review of machine learning methodologies for modelling passenger mode choice.
Hinton, G.E., Van Camp, D., 1993. Keeping the neural networks simple by minimizing the description length of the weights. In: Proceedings of the sixth annual conference on Computational learning theory, pp. 5–13.
Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network.
Hjorthol, R., 2016. Decreasing popularity of the car? changes in driving licence and access to a car among young adults over a 25-year period in norway. J. Transp. Geogr. 51, 140–146.
Hunecke, M., Haustein, S., Böhler, S., Grischkat, S., 2010. Attitude-based target groups to reduce the ecological impact of daily mobility behavior. Environ. Behav. 42 (1), 3–43.
Johansson, M.V., Heldt, T., Johansson, P., 2006. The effects of attitudes and personality traits on mode choice. Transp. Res. Part A: Policy Practice 40 (6), 507–525.
Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K., 1999. An introduction to variational methods for graphical models. Mach. Learn. 37 (2), 183–233.
Koppelman, F.S., Wen, C.-H., 1998. Alternative nested logit models: structure, properties and estimation. Transp. Res. Part B: Methodol. 32 (5), 289–298.
Kullback, S., 1959. Information theory and statistics. Courier Corporation.
Kwon, Y., Won, J.-H., Kim, B.J., Paik, M.C., 2020. Uncertainty quantification using bayesian neural networks in classification: Application to biomedical image segmentation. Comput. Stat. Data Anal. 142, 106816.
Lhéritier, A., Bocamazo, M., Delahaye, T., Acuna-Agost, R., 2019. Airline itinerary choice modeling using machine learning. J. Choice Modell. 31, 198–209.
Li, H., Wang, P., Shen, C., 2018. Toward end-to-end car license plate detection and recognition with deep neural networks. IEEE Trans. Intell. Transp. Syst. 20 (3), 1126–1136.
Liang, F., 2005. Bayesian neural networks for nonlinear time series forecasting. Stat. Comput. 15 (1), 13–29.
Liang, F., Wong, W.H., 2001. Real-parameter evolutionary monte carlo with applications to bayesian mixture models. J. Am. Stat. Assoc. 96 (454), 653–666.
Lipton, Z.C., 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue 16 (3), 31–57.
Mackay, D.J.C., 1992. Bayesian methods for adaptive models, PhD thesis, California Institute of Technology.
Marzban, C., Witt, A., 2001. A bayesian neural network for severe-hail size prediction. Weather Forecasting 16 (5), 600–610.
McFadden, D., 1981. Econometric models of probabilistic choice. Structural analysis of discrete data with econometric applications 198272.
McFadden, D., 1986. The choice theory approach to market research. Market. Sci. 5 (4), 275–297.
Munshi, T., 2016. Built environment and mode choice relationship for commute travel in the city of rajkot, india. Transp. Res. Part D: Transp. Environ. 44, 239–253.
Office of Rail and Road, 2014. An overview of the british rail industry.https://www.orr.gov.uk/media/12376.
Paredes, M., Hemberg, E., O'Reilly, U.-M., Zegras, C., 2017. Machine learning or discrete choice models for car ownership demand estimation and prediction?. In: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), IEEE, pp. 780–785.
Pereyra, G., Tucker, G., Chorowski, J., Kaiser, Ł., Hinton, G., 2017. Regularizing neural networks by penalizing confident output distributions arXiv preprint arXiv: 1701.06548.
Potdar, K., Pardawala, T.S., Pai, C.D., 2017. A comparative study of categorical variable encoding techniques for neural network classifiers. Int. J. Comput. Appl. 175 (4), 7–9.
Raschka, S., 2014. About feature scaling and normalization. Disques, nd Web. Dec, Sebastian Racha.
Ren, Y., Chen, H., Han, Y., Cheng, T., Zhang, Y., Chen, G., 2020. A hybrid integrated deep learning model for the prediction of citywide spatio-temporal flow volumes. Int. J. Geograph. Inform. Sci. 34 (4), 802–823.
Rive, G., Thomas, J., Frith, B., Chang, J., 2015. Public transport and the next generation, number 569.
Salas, P., De la Fuente, R., Astroza, S., Carrasco, J.A., 2022. A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. Exp. Syst. Appl., p. 116253.
Shor, B., Bafumi, J., Keele, L., Park, D., 2007. A bayesian multilevel modeling approach to time-series cross-sectional data. Political Anal. 15 (2), 165–181.
Tang, L., Xiong, C., Zhang, L., 2015. Decision tree method for modeling travel mode switching in a dynamic behavioral process. Transp. Plann. Technol. 38 (8), 833–850.
Wang, S., Wang, Q., Zhao, J., 2020. Deep neural networks for choice analysis: Extracting complete economic information for interpretation. Transp. Res. Part C: Emerging Technol. 118, 102701.
Willis, D.P., Manaugh, K., El-Geneidy, A., 2015. Cycling under influence: Summarizing the influence of perceptions, attitudes, habits, and social environments on cycling for transportation. Int. J. Sustain. Transp. 9 (8), 565–579.
Xie, C., Lu, J., Parkany, E., 2003. Work travel mode choice modeling with data mining: decision trees and neural networks. Transp. Res. Rec. 1854 (1), 50–61.
Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using bayesian neural network models: An empirical analysis. Acc. Anal. Prevent. 39 (5), 922–933.
Zhang, Y., Xie, Y., 2008. Travel mode choice modeling with support vector machines. Transp. Res. Rec. 2076 (1), 141–150.
Zhao, X., Yan, X., Yu, A., Van Hentenryck, P., 2020. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. Travel Behav. Soc. 20, 22–35.
Wang, L., Xu, J., Qin, P., 2014. Will a driving restriction policy reduce car trips?-The case study of Beijing, China. Transp. Res. Part A: Policy Practice 67, 279–290.