

Joint prediction of travel mode choice and purpose from travel surveys: A multitask deep learning approach

Honghan Bei^{a,b,c}, Huanfa Chen^{d,*}, Lin Li^{a,b}, Xiaowei Gao^e, Yutong Xia^f, Yutong Sun^g

^a School of Maritime Economics and Management, Dalian Maritime University, Dalian, People's Republic of China

^b Collaborative Innovation Centre for Transport Study, Dalian Maritime University, Dalian, People's Republic of China

^c School of Management, Shanghai University, Shanghai, People's Republic of China

^d Centre for Advanced Spatial Analysis, University College London, United Kingdom

^e SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London, United Kingdom

^f Institute of Data Science, National University of Singapore, Singapore

^g Business School, Liaoning University, People's Republic of China

ARTICLE INFO

Keywords:

Multitask learning
Machine learning
Travel mode choice
Trip purpose
Logit models
Predictive accuracy

ABSTRACT

The prediction and behavioural analysis of travel mode choice and purpose are critical for transport planning and have attracted increasing interest in research. Traditionally, the prediction of travel mode choice and trip purpose has been tackled separately, which fail to fully leverage the shared information between travel mode and purpose. This study addresses this gap by proposing a multitask learning deep neural network framework (MTLDNN) to jointly predict mode choice and purpose. We empirically evaluate and validate this framework using the household travel survey data in Greater London, UK. The results show that this framework has significantly lower cross-entropy loss than multinomial logit models (MNL) and single-task-learning deep neural network models (STLDNN). On the other hand, the predictive accuracy of MTLDNN is similar to STLDNN and is significantly higher than MNL. Moreover, in terms of behaviour analysis, the substitution pattern and choice probability of MTLDNN regarding input variables largely agree with MNL and STLDNN. This work demonstrates that MTLDNN is efficient in utilising the information shared by travel mode choice and purpose, and is capable of producing behaviourally reasonable substitution patterns across travel modes. Future research would develop more advanced MTLDNN frameworks for travel behaviour analysis and generalise MTLDNN to other travel behaviour topics.

1. Introduction

The analysis and prediction of travel mode choice are of pronounced importance for transport planning and travel demand forecast. Traditionally, the predominant approach used for modelling travel mode choice is random utility models (DCM), which rely on predefined utility specifications for each alternative in the choice set (Domencich and McFadden, 1975). Recently, researchers have been increasingly interested in applying machine learning (ML) techniques such as deep neural networks (DNN) to analyse travel mode choices. Rather than relying on predefined utility functions and pre-selected input variables, ML methods are capable of automatically identifying the non-linear relationship between input features and mode choice. Therefore, ML methods have demonstrated considerable predictive power and

interpretability in mode choice analysis.

On the other hand, trip purpose prediction has also received increasing attention in transport planning and research. Trip purpose refers to the purpose of a trip and why people travel, such as education, work, recreation, and business. The methods for predicting trip purpose are classified into three categories, namely rule-based models (using predefined heuristic rules), statistical methods (using logistic models), and ML methods (Gong et al., 2014). Similar to mode choice analysis, ML methods are proved to outperform rule-based and statistical methods in terms of predictive accuracy for trip purposes.

Until now, the prediction of travel mode choice and trip purpose are two separate research fields and no previous attempts have been introduced to simultaneously predict these two trip characteristics. The joint predicting of travel mode and trip purpose is made possible by two

* Corresponding author.

E-mail addresses: beihh@dlnu.edu.cn (H. Bei), huanfa.chen@ucl.ac.uk (H. Chen), lilin@dlnu.edu.cn (L. Li), xiaowei.gao.20@ucl.ac.uk (X. Gao), yutong.xia@u.nus.edu (Y. Xia).

<https://doi.org/10.1016/j.tbs.2023.100625>

Received 2 July 2022; Received in revised form 28 April 2023; Accepted 24 June 2023

Available online 5 July 2023

2214-367X/© 2023 The Authors. Published by Elsevier Ltd on behalf of Hong Kong Society for Transportation Studies. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

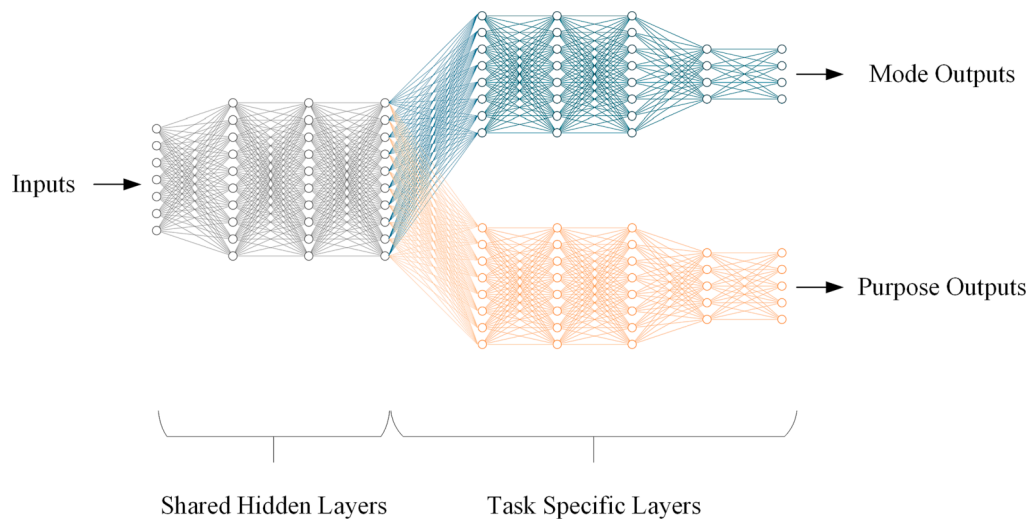


Fig. 1. The MTLDNN architecture for jointly predicting mode choice and trip purpose.

aspects. First, empirical studies show a considerable dependence between travel mode (“how people travel”) and trip purpose (“why people travel”). Second, the recent development in multitask learning (MTL) and DNNs provides an opportunity to address the problem of jointly predicting travel mode and trip purpose, as an alternative to separate prediction of mode or trip purpose.

This study proposes a framework of using multitask learning deep neural networks (MTLDNNs) to analyse and predict travel mode and trip purpose based on travel surveys. This framework starts with several shared hidden layers that capture the shared information between two tasks and ends with some task-specific layers that model each task, as shown in Fig. 1. We first describe the structure and components of this framework and then apply this framework to a travel survey dataset collected in London, England, which is a household survey designed to monitor long-term trends in personal travel and to inform the development of transport policy. In the experiments, we demonstrate the predictive performance of this MTLDNN framework in comparison with the classical DNN and multi-nomial logit models. To demonstrate the behavioural interpretability of this framework, we illustrate the relationship between choice probabilities and key input variables and the substitution patterns between mode alternatives. Overall, this study shows that the MTLDNN framework is appropriate for jointly analysing and predicting travel mode and trip purpose due to the theoretical flexibility and predictive performance.

The paper is organised as follows. Section 2 reviews the dependence between travel mode and trip purpose, as well as the existing studies that use multitask learning for travel behaviour and choice models. Section 3 introduces the test of independence between mode and purpose and the MTLDNN for analysing mode and purpose. Section 4 presents data and experiment settings, and then Section 5 discusses model performance and the behavioural information in MTLDNN. Finally, Section 6 summarises the key findings and proposes future research directions.

2. Literature review

2.1. Dependence between travel mode and trip purpose

Travel mode and trip purpose are among the most crucial characteristics for travel surveys and transport behaviour research. Empirical studies have shown that these two attributes are correlated and dependent on each other and that one attribute plays an important role in modelling and predicting the other.

Multiple studies have adopted travel mode as one of the input variables for predicting trip purpose and reported that travel mode has a

significant effect on the prediction of trip purpose. In the work by Ermagun et al. (2017), travel modes are statistically significant variables in the nest logit model for predicting trip purpose, and the mode of “Bus” is within the top ten most important variables in the random forest model for predicting trip purpose. Likewise, travel mode was considered as the second most important variable in a trip-purpose random forest model that is developed for large-scale GPS-based travel surveys (Yazdizadeh et al., 2019). On the other hand, a body of research has confirmed the feasibility and importance of using trip purpose for accurate prediction of travel mode in travel surveys. Cheng et al. (2019) designed a random forest approach to predicting travel choice, in which trip purpose ranked fifth regarding variable importance in the 20 input variables.

Although using trip purpose to predict travel mode (or the other way round) leads to improved predictive accuracy, this approach requires the presence of one attribute and hence does not apply to datasets where both attributes are unavailable. Unlike other trip information (e.g., distance, travel time) that can be passively collected by smartphones, either travel mode or purpose needs user input and is expensive to collect, especially at a large scale. For these reasons, it is of crucial importance to jointly predict mode and purpose based on the shared information between them.

2.2. Discrete choice and machine learning methods for travel behaviour analysis

For decades, DCMs have been used to model and examine individual decision making in transportation, including travel modes, trip purposes, travel scheduling, travel route, among others (Annaswamy et al., 2018; Ben-Akiva et al., 1996; Cantarella and de Luca, 2005; De Dioszar Ortu and Willumsen, 2011). DCMs consist of a wide range of models, including Multinomial Logit (MNL), Nested Logit (NL), Cross-Nested Logit (CNL), and Mixed Logit (MXL) models (Ben-Akiva and Lerman, 2018). These models have been widely used in travel behaviour research as they have clear mathematical structures and can provide economic information of travel behaviours. In turn, the economic information extracted from DCMs can provide insights to guide transport policies. For instance, DCMs can compute and derive the market shares of different modes that reveal the popularity of mode alternatives. Moreover, the substitution pattern of mode alternatives uncovers how the choice probability and market share vary with input variables (e.g. travel distance). However, each DCM is based on model assumptions, such as the independence of irrelevant alternatives for MNL model. If these assumptions are violated, the parameter inference and model

Table 1
A comparison of three types of methods for travel behaviour analysis.

Method	Advantages	Disadvantages
DCM	Clear mathematical structure; good model interpretability	Subject to model assumptions; not applicable to panel data
Single-task ML	Having few model assumptions and good applicability; high predictive accuracy; fair model interpretability	Some model interpretation is unreliable
MTLDNN	Having few model assumptions and good applicability; better predictive performance than single-task ML; fair model interpretability; low risk of overfitting on a single task	Some model interpretation is unreliable

prediction will be biased and inaccurate. These assumptions limit the applicability of DCMs, especially when dealing with panel data.

ML algorithms have been adopted and applied in different fields of quantitative travel behaviour studies, including car ownership prediction (Paredes et al., 2017), license plate recognition (Li et al., 2019), and traffic flow prediction (Ren et al., 2020). Notably, ML models have been used as an alternative to the traditional DCMs for modelling and predicting travel mode (Wang et al., 2020a) and trip purposes (Ermagun et al., 2017). The theoretical foundation of this line of research is that the modelling of travel mode or purpose can be considered as a general classification problem, which can then be addressed by ML classification algorithms (or ML classifiers). The main advantage of ML classifiers is that it is more flexible than DCMs due to fewer model assumptions, which leads to a higher predictive accuracy. Moreover, compared with DCMs, machine learning models have more complicated model structures, which makes it possible to model non-linear relationship between variables. The ML classifiers that are commonly used for travel mode or purpose prediction include support vector machine, classification trees, random forest, and DNNs. Notably, comparative studies consistently report that DNNs achieved a higher predictive accuracy for travel mode and purposes than DCMs (Xia et al., 2023; Zhao et al., 2020).

A major limitation of ML models for travel behaviour research (and also other fields) is that these models are not readily interpretable and it is challenging to extract reliable economic information from these models, when compared to DCMs. (Wang et al., 2020a) demonstrate the feasibility of generating a wide range of economic information regarding travel mode choice from DNN models. While the economic information extracted from DNNs is mostly reasonable and consistent with DCMs, some of the information is unreliable, which is caused by three inherent issues of DNNs: high sensitivity to model hyperparameters, model non-identification, and local irregularity. Therefore, challenges still exist in terms of extracting reliable economic information from DNNs for travel behaviours.

Table 2
Contingency table of mode and purpose and the Chi-square test.

Panel 1. Contingency table of mode and purpose								
	Commuting	Business	Education	Shopping	Personal business	Leisure	Other	mode share
Walk	259	24	247	471	285	492	537	2315 (11.68%)
Bicycle	193	23	33	78	46	155	4	532 (2.69%)
Car or van	2049	590	771	2293	1204	3377	1262	11,546 (58.29%)
Bus	722	81	221	591	280	601	62	2558 (12.91%)
Rail	1537	232	177	151	120	607	32	2856 (14.42%)
Purpose share	4760 (24.03%)	950 (4.80%)	1449 (7.32%)	3584 (18.09%)	1935 (9.77%)	5232 (26.41%)	1897 (9.58%)	

Panel 2. Relationship between mode and purpose (Chi-square test)	
Degree of freedom	24
Chi-square statistic	3059.947
p-value	0.00

2.3. Multitask learning for travel behaviour research

As discussed above, most ML classifiers for travel behaviour analysis are designed for a single task, such as estimating travel mode or trip purpose. In contrast, MTL is a mechanism that improves the generalisation of ML models on a task by sharing the information and representations between related tasks (Caruana, 1997). To achieve this, MTLDNN trains tasks in parallel whilst using a shared representation. In comparison with traditional single-task ML, MTLDNN would achieve a better prediction performance on different tasks without the loss of model interpretability, due to the shared information across different tasks. In addition, MTLDNN would greatly reduce the risk of overfitting on one task, as the model has to seek a representation that simultaneously captures all of the given tasks (Ruder, 2017). A summary of the advantages and disadvantages of three types of methods (including DCM, single-task ML, and MTLDNN) is provided in Table 1.

MTLDNN has been widely and successfully applied to related tasks of different domains. In natural language processing, MTLDNN has been used to simultaneously predict semantic components of different levels, including part-of-speech tags, chunks, named entity tags (Collobert and Weston, 2008; Hashimoto et al., 2016). In image recognition, MTLDNN has displayed remarkable performance in two pairs of tasks, semantic segmentation and surface normal prediction, and object detection and attribute prediction (Misra et al., 2016). In urban analytics, MTLDNN has been utilised to learn individual geodemographic attributes (including age, gender, income level, and car ownership) from public transport travel patterns (Zhang et al., 2020).

MTLDNN shares similar ideas as the simultaneous estimation of choice models in travel behaviour research. Specifically, MTLDNN has the potential to jointly model and estimate the household car ownership and vehicle kilometre travelled (Zegras, 2010), auto ownership and mode choice (Train, 1980), mode choice and psychological or attitudinal factors (Lyon, 1984; Morikawa et al., 2002). However, the use of MTLDNN for choice models is very limited. The only exception is that researchers used MTLDNN to jointly model revealed and stated preferences in travel surveys (Wang et al., 2020b).

In recent years, various MTLDNN architectures have been developed. The first MTLDNN was proposed by Caruana (1997), which contains shared hidden layers between all tasks and task-specific layers. The MTLDNN architecture was then improved via designing varying regularisation mechanisms and adding network components that control the differences and similarities of tasks (Argyriou et al., 2007; Evgeniou, 2005; Long et al., 2015; Misra et al., 2016; Ruder et al., 2019; Yang and Hospedales, 2019). A comprehensive overview of the history and development of MTL and MTLDNN is provided by Ruder (2017). Despite the advanced MTLDNN architectures, this study uses the classical MTLDNN architecture, as it is straightforward and efficient for the tasks of modelling and predictive mode and purpose.

Table 3
The selected and fitted models for comparison.

Model	Description
MTLDNN-M	the selected MTLDNN that has the lowest cross-entropy of predicting mode choice
MTLDNN-P	the selected MTLDNN that has the lowest cross-entropy of predicting trip purpose
STLDNN-M	the selected STLDNN for predicting mode choice and having the lowest cross-entropy
STLDNN-P	the selected STLDNN for predicting trip purpose and having the lowest cross-entropy
MNL-M	the fitted MNL for predicting mode choice
MNL-P	the fitted MNL for predicting trip purpose

3. Methods

3.1. Test of independence between mode and purpose

This study uses Pearson’s chi-square test to test the independence between mode and purpose choices in travel survey data. Pearson’s chi-square test is a hypothesis testing method that tests the independence between multiple categorical data. This test is based on a contingency table between these data, which displays the multivariate frequency distribution of the variables (see Table 4 for an example). Assuming that the dimensions of mode and purpose choices are K_m and K_p , and O_{ij} is the co-occurrence frequency of mode i and purpose j , then the chi-square test statistic is calculated as:

Table 4
Performance comparison of six models.

	MTLDNN-M	MTLDNN-P	STLDNN-M	STLDNN-P	MNL-M	MNL-P
Panel 1. Cross-entropy loss						
Mode (Testing)	0.506	0.627	0.720	0.795	0.882	–
Purpose (Testing)	1.388	1.249	1.500	1.460	–	1.638
Mode (Training)	0.240	0.115	0.097	0.091	0.871	–
Purpose (Training)	0.529	0.713	0.514	0.546	–	1.626
Panel 2. Predictive accuracy						
Mode (Testing)	0.842	0.867	0.860	0.862	0.680	–
Purpose (Testing)	0.622	0.581	0.605	0.611	–	0.328
Mode (Training)	0.906	0.956	0.959	0.961	0.688	–
Purpose (Training)	0.782	0.733	0.786	0.777	–	0.337

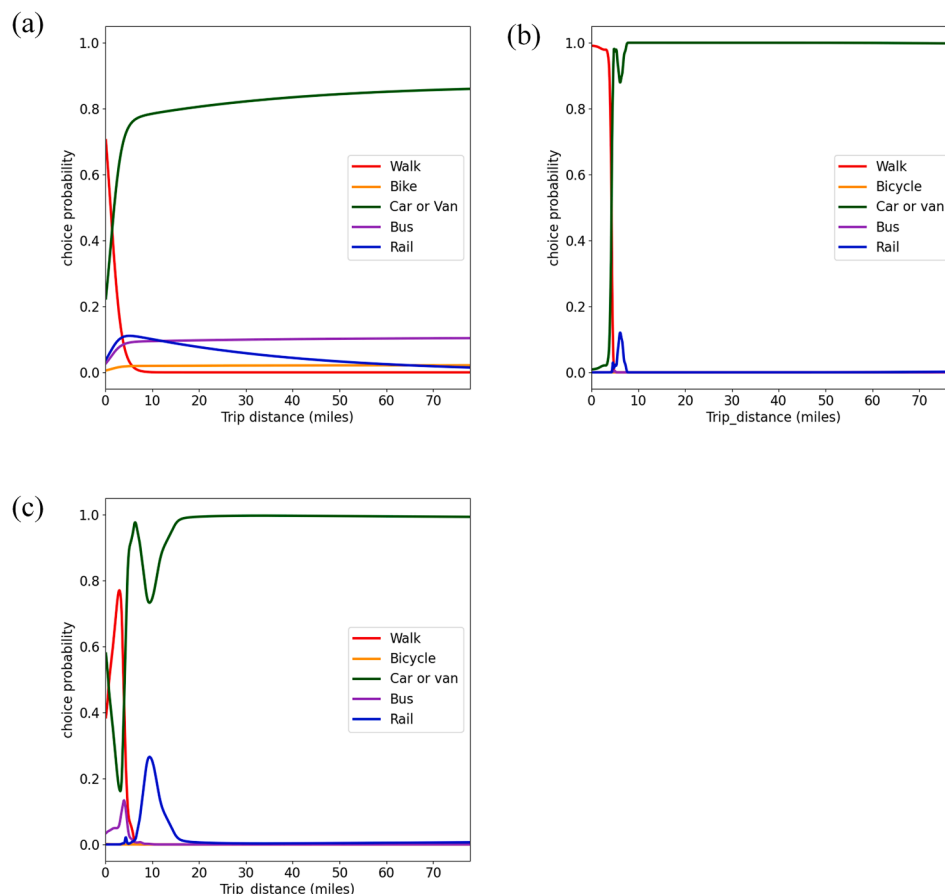


Fig. 2. Choice probability and substitution patterns of all alternatives regarding the trip distance. (a) MNL; (b) STLDNN; (c) MTLDNN.

Table A1
Descriptive statistics of the survey data.

Panel 1. Numerical Variables							
Variable	mean	std	min	25%	50%	75%	max
Trip_distance (miles)	7.67	18.17	0.1	1.5	3.5	7.2	600.5
Trip_time (minutes)	29.16	28.18	1	12	20	35	614
Household_children	0.73	1.04	0	0	0	1	5
Household_bike	1.07	1.30	0	0	1	2	5
Household_car	1.17	0.89	0	1	1	2	4
Household_licence	1.65	0.88	0	1	2	2	5
Panel 2. Categorical Variables							
Variable	Category	Percent (%)					
Purpose	Purpose (Commuting)	24.03					
	Purpose (Business)	4.80					
	Purpose (Education / escort education)	7.32					
	Purpose (Shopping)	18.09					
	Purpose (Personal business)	9.77					
	Purpose (Leisure)	26.41					
	Purpose (Other)	9.58					
Mode	Mode (Walk)	11.69					
	Mode (Bicycle)	2.69					
	Mode (Car or van)	58.29					
	Mode (Bus)	12.91					
	Mode (Rail)	14.42					
Household employed	None	70.85					
	1 part time or full time	0.00					
	2 part time or full time	25.08					
	3 or more part time or full time	4.07					
Individual employment	Full-time	80.48					
	Part-time	19.52					
Individual age	0 - 16 years	0.63					
	17 - 20 years	4.10					
	21 - 29 years	17.31					
	30 - 39 years	24.91					
	40 - 49 years	23.70					
	50 - 59 years	14.99					
	60 years +	14.36					
Individual education	Degree level or above	44.08					
	Other types of qualification	55.92					
Individual income	Less than £25,000	66.45					
	£25,000 to £49,999	24.10					
	£50,000 and over	9.45					
Individual gender	Male	45.80					
	Female	54.20					
Household settlement	Urban	99.94					
	Rural	0.06					

Table A2
Performance of the MNL models.

	MNL-M	MNL-P
Null Loglikelihood	-22314.9	-26980.0
Final Loglikelihood	-12077.0	-22545.5
Number of parameters	36	39
Rho Square	0.459	0.164

$$X^2 = \sum_{i=1}^{K_m} \sum_{j=1}^{K_p} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

where E_{ij} is the expected frequency of mode i and purpose j under the null hypothesis of independence.

The p-value is then calculated by comparing the above test statistic and the X^2 distribution. By comparing the p-value and the predefined significance level, one can decide whether the null hypothesis of no relationship should be rejected. In this study, the selected significance level is 0.01, and a p-value smaller than 0.01 indicates a statistically significant dependence between travel mode and purpose.

3.2. Multitask learning neural network for mode and purpose

The MTLDNN for analysing mode and purpose is described as follows. We let $x_i \in R^d$ denote the input variables for mode and purpose, where $i \in \{1, 2, \dots, N\}$ are the indices of observations, and d represents the input dimension. The output choices of mode and purpose are denoted by $y_{m,i}$ and $y_{p,i}$, where m and p stand for mode and purpose, respectively, $y_{m,i} \in \{0, 1\}^{K_m}$ and $y_{p,i} \in \{0, 1\}^{K_p}$; K_m and K_p are the dimensions of mode and purpose choices, respectively. Both $y_{m,i}$ and $y_{p,i}$ are binary vectors. Due to the constraint that exactly one alternative of mode (or purpose choice) is true, $\sum_{k=1}^{K_m} y_{m,i}[k] = 1$ and $\sum_{k=1}^{K_p} y_{p,i}[k] = 1$. As represented by Fig. 1, the feature transformation of mode and purpose can be represented as:

$$V_{m,i} = (g_m^{M_2} \circ g_m^{M_2-1} \circ \dots \circ g_m^1) \circ (g_0^{M_1} \circ g_0^{M_1-1} \circ \dots \circ g_0^1)(x_{m,i}) \quad (2)$$

$$V_{p,i} = (g_p^{M_2} \circ g_p^{M_2-1} \circ \dots \circ g_p^1) \circ (g_0^{M_1} \circ g_0^{M_1-1} \circ \dots \circ g_0^1)(x_{p,i}) \quad (3)$$

where M_1 denote the depth of shared layers; M_2 denotes the depth of task-specific (or non-shared) layers for each task; g_0 represents the transformation of one shared layer; g_m and g_p represent the transformation of one layer in mode and purpose, respectively. Specifically,

except for the output layer, the transformation functions (including g_m , g_p , and g_0) comprise ReLU and linear transformation: $g^l(x) = \max\{W^l x, 0\}$, $\forall l \neq M_2$. Overall, equations (2) and (3) depict the multi-task learning deep neural networks (MTLDNN) architecture as shown in Fig. 1: ($g_0^{M_1} \circ g_0^{M_1-1} \circ \dots \circ g_0^1$) represent the shared layers, while ($g_m^{M_2} \circ g_m^{M_2-1} \circ \dots \circ g_m^1$) and ($g_p^{M_2} \circ g_p^{M_2-1} \circ \dots \circ g_p^1$) represent task-specific layers for mode analysis and purpose analysis, respectively. The choice probability functions in mode and purpose are calculated by a standard softmax activation function, which is commonly used in multi-class classification, as follows:

$$P(y_{m,i}[k]; w_m, w_0) = \frac{e^{V_{m,i}[k]}}{\sum_{j=1}^{K_m} e^{V_{m,i}[j]}} \quad (4)$$

$$P(y_{p,i}[k]; w_p, w_0) = \frac{e^{V_{p,i}[k]}}{\sum_{j=1}^{K_p} e^{V_{p,i}[j]}} \quad (5)$$

in which w_m and w_p represent the task-specific parameters in g_m and g_p ; w_0 represent the shared parameters in g_0 . Equation (4) computes the choice probability for each of the K_m mode alternatives, and Equation (5) computes the probability of each of the K_p trip purposes, given the input data.

The MTLDNN model is trained by empirical risk minimisation (ERM), which contains classification errors and regularisation terms:

$$\begin{aligned} & \min_{w_m, w_p, w_0} R(X, Y; w_m, w_p, w_0; c_H) \\ & = \min_{w_m, w_p, w_0} \{ \theta_m \times \widehat{L}_M(w_m, w_0) + \theta_p \times \widehat{L}_P(w_p, w_0) + \lambda^T g(w_m, w_p, w_0) \} \\ & = \min_{w_m, w_p, w_0} \left\{ -\frac{\theta_m}{N} \sum_{i=1}^{N_x} \sum_{j=1}^{K_m} y_{m,i}[j] \log P(y_{m,i}[j]; w_m, w_0; c_H) - \frac{\theta_p}{N} \sum_{i=1}^{N_x} \sum_{j=1}^{K_p} y_{p,i}[j] \log P(y_{p,i}[j]; w_p, w_0; c_H) + \lambda_1 (\|w_0\| + \|w_m\| + \|w_p\|) + \lambda_2 (\|w_0\|_2^2 + \|w_m\|_2^2 + \|w_p\|_2^2) \right\} \end{aligned}$$

Equation (5) consists of four parts. The first two parts

$$\widehat{L}_M(w_m, w_0) = -\frac{1}{N} \sum_{i=1}^{N_x} \sum_{j=1}^{K_m} y_{m,i}[j] \log P(y_{m,i}[j]; w_m, w_0; c_H) \quad (6)$$

$$\widehat{L}_P(w_p, w_0) = -\frac{1}{N} \sum_{i=1}^{N_x} \sum_{j=1}^{K_p} y_{p,i}[j] \log P(y_{p,i}[j]; w_p, w_0; c_H) \quad (7)$$

are the empirical risk of predicting mode and purpose, respectively, which take the form of cross-entropy loss. The third part, $\lambda_1 (\|w_0\| + \|w_m\| + \|w_p\|)$ is the L1-class regularisation term with λ_1 weight, where $\|w_0\| = \sum_j |w_{0j}|$ and so forth. The fourth part $\lambda_2 (\|w_0\|_2^2 + \|w_m\|_2^2 + \|w_p\|_2^2)$ is the L2-class regularisation term with λ_2 weight, where $\|w_0\|_2^2 = \sum_j w_{0j}^2$ and so forth. Equation (5) incorporates four hyperparameters ($\theta_m, \theta_p, \lambda_1, \lambda_2$) with a constraint that $\theta_m + \theta_p = 1$. Specifically, θ_m and θ_p adjust the relative importance of mode and purpose prediction, while λ_1 and λ_2 adjust the absolute magnitudes of layer weights. The larger λ_1 or λ_2 , the larger weight decay in the training of DNN models.

3.3. Multinomial logit models for mode and purpose

This study compares the DNNs for analysing mode and purpose against two baseline MNL models that analyses mode and purpose, respectively. For convenience, the MNL models that predict mode and purpose are called MNL-M and MNL-P, respectively. The utility function of MNL-M follows a linear structure:

$$U_{m,i} = V_{m,i} + \varepsilon_{m,i} = w_m^T x_{m,i} + \varepsilon_{m,i} \quad (8)$$

where w_m is the parameters for travel mode analysis; $\varepsilon_{m,i}$ is the random utility term; $x_{m,i}$ is the independent variables for mode choice.

Then, the choice probability function of each of the K_m mode alternatives in DCMs is computed as follows:

$$P(y_{m,i}[k]; w_m) = \frac{e^{[w_m^T x_{m,i}] [k]}}{\sum_{k=1}^{K_m} e^{[w_m^T x_{m,i}] [k]}} \quad (9)$$

Similar to MNL-M, the utility function of MNL-P follows a linear structure, as follows:

$$U_{p,i} = V_{p,i} + \varepsilon_{p,i} = w_p^T x_{p,i} + \varepsilon_{p,i} \quad (10)$$

where w_p is the parameters for trip purpose; $\varepsilon_{p,i}$ is the random utility term; $x_{p,i}$ is the independent variables for analysing trip purpose.

The probability function of each of the K_p trip purposes in DCMs is as follows:

$$P(y_{p,i}[k]; w_p) = \frac{e^{[w_p^T x_{p,i}] [k]}}{\sum_{k=1}^{K_p} e^{[w_p^T x_{p,i}] [k]}} \quad (11)$$

In the specification of the utility functions (9) and (11), we included all theoretically relevant independent variables. We used the maximum likelihood estimation method to estimate the parameters.

4. Data

4.1. Data summary

We utilised the annual National Travel Survey (NTS) data of the UK from 2005 to 2016, which are publicly provided by the Department for Transport (Department for Transport, 2020). Respondents are divided into nine regions according to the household address they provided, namely North East, North West, Yorkshire and the Humber, East Midlands, West Midlands, East of England, London, South East, and South West. The NTS data is collected through a combination of face-to-face interviews and 7-day self-reported written travel diaries. The NTS data also provides individual socio-demographic information about respondents and their households, including gender, income, and car ownership. This dataset enables researchers to connect travel patterns with individual characteristics. Each year, the survey encompasses individuals from all age groups and involves around 16,000 participants from 7,000 households in England. From 2005 to 2016, a total of 121,765 respondents from 69,208 households participated in this survey. After simple data cleaning, the database contains a total of 2,100,492 observations with detailed travel information, including participant ID, mode, purpose, date and time of start and end point, location of the start and end point (government office region level, coarse-grained), travel duration, and travel distance. As the specific start and end location of each trip is not available and the home/work location of volunteers are not available, constructing the trip chain of a participant is challenging. Therefore, we consider each trip as an instance in the NTS data without constructing trip chains. In the future,

Table A3
Estimated coefficients of the MNL-M model.

Mode	Variable name	Value	Std Err	t-test	p-value
Bike	A_bike	-1.680	0.100	-16.700	0.000
Bike	B_bike_household_bike	0.732	0.051	14.500	0.000
Bike	B_bike_individual_gender	1.260	0.122	10.400	0.000
Bike	B_bike_trip_time	-0.910	0.074	-12.300	0.000
Bus	A_bus	-0.031	0.070	-0.450	0.653
Bus	B_bus_age_individual_age_17_20	1.470	0.108	13.600	0.000
Bus	B_bus_household_bike	-0.175	0.041	-4.310	0.000
Bus	B_bus_household_car	-0.028	0.086	-0.327	0.744
Bus	B_bus_individual_income_less_than_25000	0.832	0.077	10.800	0.000
Bus	B_bus_household_licence	-0.468	0.044	-10.700	0.000
Bus	B_bus_purpose_other	-1.280	0.172	-7.470	0.000
Bus	B_bus_trip_time	-0.321	0.034	-9.540	0.000
Car or Van	A_car	2.610	0.050	52.600	0.000
Car or Van	B_car_individual_age_21_29	-0.649	0.063	-10.300	0.000
Car or Van	B_car_household_bike	0.035	0.024	1.470	0.141
Car or Van	B_car_household_car	1.230	0.077	16.000	0.000
Car or Van	B_car_household_licence	-0.120	0.034	-3.550	0.000
Car or Van	B_car_purpose_commuting	-0.640	0.068	-9.480	0.000
Car or Van	B_car_purpose_leisure	0.355	0.053	6.710	0.000
Car or Van	B_car_trip_time	-0.717	0.029	-24.500	0.000
Rail	A_rail	0.508	0.072	7.080	0.000
Rail	B_rail_individual_age_21_29	0.307	0.075	4.070	0.000
Rail	B_rail_household_car	0.422	0.079	5.320	0.000
Rail	B_rail_trip_distance	-0.499	0.038	-13.200	0.000
Rail	B_rail_individual_education	0.775	0.061	12.700	0.000
Rail	B_rail_individual_income_less_than_25000	-0.196	0.064	-3.080	0.002
Rail	B_rail_purpose_other	-1.680	0.212	-7.920	0.000
Rail	B_rail_purpose_commuting	0.555	0.075	7.450	0.000
Rail	B_rail_purpose_shopping	-0.936	0.117	-8.030	0.000
Rail	B_rail_trip_time	0.502	0.039	12.900	0.000
Walk	A_walk	-2.810	0.129	-21.900	0.000
Walk	B_walk_trip_distance	-15.300	0.488	-31.300	0.000
Walk	B_walk_household_car	0.677	0.081	8.380	0.000
Walk	B_walk_purpose_other	0.562	0.087	6.480	0.000
Walk	B_walk_purpose_commuting	-0.602	0.101	-5.950	0.000
Walk	B_walk_trip_time	1.450	0.058	25.000	0.000

if more details of the trips and the home/work locations of the volunteers are available, we can construct the trip chains to achieve more accurate prediction of travel behaviours.

To demonstrate the utility of the proposed framework, in this case study, we focus on the subset of travel survey data of London in the year 2005.

4.2. Experiment setup

This study compares the performance of MTLDNNs to STLDNNs and two baseline MNLs. The utility specification and estimation results of the MNLs are presented in Table A3 and Table A4. When training DNN models, the model hyperparameters that define the architecture and regularisation should be carefully selected, as the model performance largely depends on the hyperparameters. Therefore, we predefined the hyperparameter space (see Table A5) and used a grid search approach to select the optimal hyperparameters. The model performance is evaluated by the hold-out method, in which the original survey data is divided into training and testing sets with the ratio of 7:3. The training set is used for model training while the testing set is used for model evaluation.

5. Results and discussion

5.1. Data summary and the dependence between mode and purpose

The market share of mode and proportion of travel purpose is shown in Table 2. The most popular travel mode, Car or Van, takes up 58.29% of total trips, as opposed to 2.69% of the least popular mode, Bicycle. Per travel purpose, the two most common purposes are Leisure (26.41%) and Commuting (24.03%), while the least common purpose is Business (4.80%). Moreover, Table 2 cross tabulates the mode and purpose, with

the most popular travel purpose for each travel mode highlighted in bold. It is shown that the distribution of trip purpose varies considerably across travel modes, and vice versa.

In addition, we performed the Chi-square test to test whether mode and purpose are independent of each other. As shown in Panel 2 of Table 2, the Chi-square statistic is 3059.947, with a degree of freedom of 24 and a p-value smaller than 0.01. This indicates that there is a statistically significant dependence between mode and purpose, which is the foundation for the utility of multitask learning that co-predicts mode and purpose.

5.2. Model performance

Here, we compare cross-entropy loss and predictive accuracy of MTLDNN for predicting mode and purpose, in comparison with the baseline STLDNN and MNL. The notation of the selected models is detailed in Table 2. The hyperparameter values of the selected DNNs are in Table A6. The model performance is detailed in Table 3.

Generally, MTLDNN models outperform STLDNN and MNL models regarding cross-entropy loss in the testing set. When measured by cross-entropy loss for predicting mode, MTLDNN-M outperforms STLDNN-M and STLDNN-P by 0.214 and 0.289. On the other hand, per the cross-entropy loss for predicting purposes, MTLDNN-P outperforms STLDNN-M and STLDNN-P by 0.251 and 0.211. In addition, the MTLDNNs is less likely to overfit on the training samples than the STLDNNs, as MTLDNNs have a smaller difference of cross-entropy loss between training and testing set. This demonstrates the advantage of MTLDNN that it considerably reduces the risk of overfitting (Baxter, 1997; Ruder, 2017). On the other hand, the cross-entropy loss of MTLDNN or STLDNN is much lower than MNL, which indicates the superior predictive power of MTLDNN and STLDNN than MNL.

Table A4
Estimated coefficients of the MNL-P model.

Purpose	Variable name	Value	Std Err	t-test	p-value
Business	A_business	-0.261	0.051	-5.110	0.000
Business	B_business_individual_income_less_than_25000	-1.010	0.083	-12.200	0.000
Business	B_business_trip_time	0.472	0.032	14.900	0.000
Commuting	A_commuting	0.838	0.077	11.000	0.000
Commuting	B_commuting_household_employed	0.244	0.027	8.920	0.000
Commuting	B_commuting_individual_age_21_29	0.108	0.055	1.940	0.052
Commuting	B_commuting_individual_age_60_or_more	-1.310	0.097	-13.600	0.000
Commuting	B_commuting_individual_employment	0.824	0.078	10.600	0.000
Commuting	B_commuting_individual_income_less_than_25000	-0.315	0.048	-6.510	0.000
Commuting	B_commuting_mode_car	-0.488	0.062	-7.850	0.000
Commuting	B_commuting_mode_rail	0.367	0.074	4.950	0.000
Commuting	B_commuting_mode_walk	-0.973	0.097	-10.100	0.000
Commuting	B_commuting_trip_time	0.280	0.025	11.200	0.000
Education	A_education	-1.560	0.087	-17.800	0.000
Education	B_education_household_bike	0.121	0.033	3.660	0.000
Education	B_education_household_children	0.470	0.032	14.900	0.000
Education	B_education_individual_age_0_16	2.740	0.224	12.200	0.000
Education	B_education_individual_age_17_20	1.320	0.110	12.000	0.000
Education	B_education_individual_age_60_or_more	-2.170	0.285	-7.610	0.000
Education	B_education_individual_income_less_than_25000	1.200	0.108	11.200	0.000
Education	B_education_trip_time	-0.163	0.047	-3.440	0.001
Leisure	A_leisure	0.739	0.039	19.200	0.000
Leisure	B_leisure_mode_car	0.283	0.050	5.690	0.000
Leisure	B_leisure_trip_distance	0.344	0.036	9.680	0.000
Leisure	B_leisure_trip_time	0.019	0.032	0.606	0.545
Personal	A_personal_business	-0.116	0.038	-3.060	0.002
Personal	B_personal_business_individual_age_60_or_more	0.295	0.075	3.920	0.000
Personal	B_personal_business_mode_rail	-0.694	0.126	-5.500	0.000
Personal	B_personal_business_trip_time	-0.175	0.042	-4.200	0.000
Shopping	A_shopping	0.440	0.034	12.800	0.000
Shopping	B_shopping_individual_age_60_or_more	0.347	0.060	5.780	0.000
Shopping	B_shopping_mode_rail	-1.030	0.116	-8.820	0.000
Shopping	B_shopping_trip_distance	-0.326	0.095	-3.450	0.001
Shopping	B_shopping_trip_time	-0.276	0.046	-6.050	0.000
Other	A_other	-0.084	0.042	-1.980	0.047
Other	B_other_mode_bus	-1.460	0.159	-9.180	0.000
Other	B_other_mode_rail	-2.050	0.225	-9.090	0.000
Other	B_other_mode_walk	0.790	0.074	10.700	0.000
Other	B_other_trip_time	-0.157	0.044	-3.570	0.000

Table A5
Hyperparameter space of MTLDNN and STLDNN.

Hyperparameter	Values
Number of shared layers	[2,3,4]
Number of task-specific layers	[1,2]
λ_1 constant	1^{-20}
λ_2 constant	1^{-10}
θ_m	[0.05,0.5,0.95]
Learning rate	0.001
Dropout rate	0.01
Batch size	200

Table A6
The hyperparameters of the selected DNN models.

Model	θ_m	Shared layers	Task-specific layers
MTLDNN-M	0.05	2	1
MTLDNN-P	0.95	2	1
STLDNN-M	-	2	1
STLDNN-P	-	4	2

The predictive accuracy of MTLDNN models is close to STLDNN. Specifically, per predicting mode in testing data, MTLDNN-P slightly outperforms STLDNN, whereas MTLDNN-M performs worse than the STL models. Regarding the predictive accuracy of purpose, MTLDNN-M performs slightly better than the STL models whilst MTLDNN-P underperforms STLDNN. In addition, the predictive accuracy of MTLDNNs and STLDNNs is significantly higher than MNL.

The model comparison demonstrates the inconsistency between cross-entropy loss and predictive accuracy, which can be explained by the differing weighting mechanisms (Wang et al., 2020b). On the one hand, the predictive accuracy assigns equal weights to all instances based on the binary loss. On the other hand, the cross-entropy loss assigns a large weight to the ‘confident but incorrect predictions’ (Wang et al., 2020b), which is shown in Equation (6) and (7).

We argue that cross-entropy loss is a more appropriate measure than predictive accuracy in travel behaviour research, which is also discussed by (Train, 2009). On the one hand, the cross-entropy loss is equivalent to negative log likelihood in the maximum likelihood estimation, therefore it has a clear theoretical interpretation. In fact, log likelihood and the derived measures are often used to measure the discrete choice models that fit the data. On the other hand, predictive accuracy indicates “precent correctly predicted” and “should actually be avoided” in practice. The reason is that this measure presumes that the decision-maker is expected to choose the alternative with the highest probability given by the model, which is inconsistent with the meaning of probabilities and the objective of specifying choice probabilities.

5.3. Analysis of substitution patterns of mode choice

The substitution pattern of the mode alternatives is one of the most important economic information extracted from travel mode analysis. It reveals how the mode choice probability and market share vary with input variables (e.g. travel distance) (Wang et al., 2020a). In practice, it provides researchers and transport planners with an understanding of how market share of each mode changes with trip attributes. Here, we

compute and visualise the substitution patterns of choice alternatives regarding travel distance across these models as an example of extracting economic information and behavioural insights from different mode choice models. In Fig. 2, the x-axis represents the varying travel distance, and each line illustrates how the choice probability changes with travel distance while the other variables are fixed at the sample average. Note that the choice probability of MTLDNN and STLDNN is averaged over 50 trainings due to randomness in the model training.

The substitute patterns of MTLDNN and STLDNN are overall reasonable and intuitive. As is a common trend in Fig. 2a, 2b, and 2c, when the trip distance is smaller than 20 miles, the alternatives are substitute to each other, and the probability of Car or van is generally increasing while the probability of the other alternatives is decreasing. When the trip distance exceeds 20 miles, people are more likely to choose Car or van than the other modes. However, the substitute pattern of MTLDNN is more flexible than MNL and STLDNN, as when the trip distance is less than 10 miles, the choice probability of Car or van decreases before it rapidly increases, and the choice probability of Walk increases before it dramatically declines. This complex pattern is possibly due to the model non-identification problem of the DNN models (Wang et al., 2020c), which indicates that each training of DNNs can lead to very different models, even though conditioned on the fixed training data and hyperparameters. Overall, this analysis demonstrates that our proposed MTLDNN is interpretable and can derive reasonable and intuitive substitution patterns of mode choices.

6. Conclusions

This work proposes a multitask learning deep neural network (MTLDNN) framework that jointly predicts travel mode choice and trip purpose. The main advantage of this framework is its capability of automatically capturing and leveraging the shared information between travel mode and trip purpose. The case study using the annual travel survey data in London reveals that the MTLDNN achieves significantly lower cross-entropy loss than multinomial logit (MNL) models and single-task deep neural network (MTLDNN) models. Interestingly, in terms of predictive accuracy, while MTLDNN significantly outperforms MNL, it achieves similar accuracy with STLDNN. In addition, the MTLDNN produces reasonable choice probability patterns regarding input variables, which largely agrees with MNL and STLDNN. Overall, this framework is capable of providing an accurate joint prediction of travel mode choice and trip purpose, as well as intuitive behaviour analysis of travel behaviours.

Multitask learning is theoretically and empirically appealing for a wide range of applications relating to transport behaviours and planning. Further studies can apply multitask learning to classical topics in transport behaviours, such as auto ownership and mode choice, as well as mode choice and psychological factors. On the other hand, further studies should improve the MTLDNN performance for transport research by using advanced MTLDNN architecture or developing an MTLDNN framework specific for transport-related research.

CRedit authorship contribution statement

Honghan Bei: Conceptualization, Methodology, Writing – original draft. **Huanfa Chen:** Conceptualization, Data curation, Supervision, Writing – original draft. **Lin Li:** Software, Visualization, Investigation, Writing – review & editing. **Xiaowei Gao:** Investigation, Writing – review & editing. **Yutong Xia:** Data curation, Methodology, Writing – review & editing. **Yutong Sun:** Software, 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.

Acknowledgements

This study has been partly supported by the project of “Research and application of numerical method for optimal control problem based on sequential convex programming” (2021 Education Department of Liaoning Province, LJKZ0066).

This study has been partly supported by the project of “Optimal control problem for a class of nonlinear stochastic hybrid systems in microbial fermentation process” (2021 Natural Science Foundation of Liaoning Province, 2021-BS-076).

This study has been partly supported by the project of “The Research on Multi-party Freight Game and Coupling Optimization of Multimodal Transport Logistics Network” (List of Key Science and Technology Projects in Transportation Industry of Ministry of Transport in 2022, 2022-MS3-102).

Appendix

Appendix 1. Data Descriptive Statistics

The key statistics of the survey data are presented in Table A1.

Appendix 2. Performance and Estimation Results of the MNL models

The performance of MNL-M and MNL-P is summarised in Table A2. The estimated coefficients of the MNL-M and MNL-P models are presented in Table A3 and Table A4, respectively. In Table A3, the first column is the mode category, and the second column is the variable, where ‘A’ implies mode-specific constant and ‘B’ implies the mode-specific variable coefficient. In Table A4, the first column is the purpose category, and the second column is similar to Table A3.

Appendix 3. Hyperparameter space of MTLDNN and STLDNN

References

- Annaswamy, A.M., Guan, Y., Tseng, H.E., Zhou, H., Phan, T., Yanakiev, D., 2018. Transactive Control in Smart Cities. *Proc. IEEE* 106, 518–537. <https://doi.org/10.1109/JPROC.2018.2790841>.
- Argyriou, A., Evgeniou, T., Pontil, M., 2007. Multi-task feature learning. *Adv. Neural Inf. Process. Syst.* 41–48. <https://doi.org/10.7551/mitpress/7503.003.0010>.
- Baxter, J., 1997. A Bayesian/Information Theoretic Model of Learning to Learn via Multiple Task Sampling. *Mach. Learn.* 28, 7–39. <https://doi.org/10.1023/A:1007327622663>.
- Ben-Akiva, M.E., Lerman, S.R., 2018. *Discrete choice analysis: theory and application to travel demand*. MIT press.
- Ben-Akiva, M., Bowman, JohnL., Gopinath, D., 1996. *Travel demand model system for the information era*. Transportation 23 (3).
- Cantarella, G.E., de Luca, S., 2005. Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models. *Transp. Res. Part C Emerg. Technol.* 13, 121–155. <https://doi.org/10.1016/j.trc.2005.04.002>.
- Caruana, R., 1997. Multitask Learning. *Mach. Learn.* 28, 41–75. <https://doi.org/10.1023/A:1007379606734>.
- 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. <https://doi.org/10.1016/j.tbs.2018.09.002>.
- Collobert, R., Weston, J., 2008. A unified architecture for natural language processing, in: *Proceedings of the 25th International Conference on Machine Learning - ICML '08*. ACM Press, New York, New York, USA, pp. 160–167. 10.1145/1390156.1390177.
- De Dioszar Ortu, J., Willumsen, L.G., 2011. *Modelling transport*. John Wiley & sons. Department for Transport, 2020. National Travel Survey, 2002–2019. [data collection], 14th ed. 10.5255/UKDA-SN-5340-11.
- Domencich, T.A., McFadden, D., 1975. *Urban Travel Demand - A Behavioral Analysis*. Ermagun, A., Fan, Y., Wolfson, J., Adomavicius, G., Das, K., 2017. Real-time trip purpose prediction using online location-based search and discovery services. *Transp. Res. Part C Emerg. Technol.* 77, 96–112. <https://doi.org/10.1016/J.TRC.2017.01.020>.

- Evgeniou, T., 2005. Learning Multiple Tasks with Kernel Methods Charles A. Micchelli Massimiliano Pontil. *J. Mach. Learn. Res.* 6, 615–637.
- Gong, L., Morikawa, T., Yamamoto, T., Sato, H., 2014. Deriving Personal Trip Data from GPS Data : A Literature Review on the Existing Methodologies. *Procedia - Soc. Behav. Sci.* 138, 557–565. <https://doi.org/10.1016/j.sbspro.2014.07.239>.
- Hashimoto, K., Xiong, C., Tsuruoka, Y., Socher, R., 2016. A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks. *EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc.* 1923–1933.
- Li, H., Wang, P., Shen, C., 2019. Toward End-to-End Car License Plate Detection and Recognition with Deep Neural Networks. *IEEE Trans. Intell. Transp. Syst.* 20, 1126–1136. <https://doi.org/10.1109/TITS.2018.2847291>.
- Long, M., Cao, Z., Wang, J., Yu, P.S., 2015. Learning Multiple Tasks with Multilinear Relationship Networks. *Adv. Neural Inf. Process. Syst.* 2017-December, 1595–1604.
- Lyon, P.K., 1984. Time-Dependent Structural Equations Modeling: A Methodology for Analyzing the Dynamic Attitude-Behavior Relationship. *Transp. Sci.* 18, 395–414. <https://doi.org/10.1287/trsc.18.4.395>.
- Misra, I., Shrivastava, A., Gupta, A., Hebert, M., 2016. Cross-Stitch Networks for Multi-task Learning. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE*, pp. 3994–4003. <https://doi.org/10.1109/CVPR.2016.433>.
- Morikawa, T., Ben-Akiva, M., McFadden, D., 2002. Discrete choice models incorporating revealed preferences and psychometric data. *Adv. Econ* 16, 29–55. [https://doi.org/10.1016/S0731-9053\(02\)16003-8/FULL/XML](https://doi.org/10.1016/S0731-9053(02)16003-8/FULL/XML).
- Paredes, M., Hemberg, E., O'Reilly, U.M., Zegras, C., 2017. Machine learning or discrete choice models for car ownership demand estimation and prediction? 5th IEEE Int. Conf. Model. Technol. Intell. Transp. Syst. MT-ITS 2017 - Proc. 780–785. 10.1109/MTITS.2017.8005618.
- 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. Geogr. Inf. Sci.* 34, 802–823. <https://doi.org/10.1080/13658816.2019.1652303>.
- Ruder, S., Bingel, J., Augenstein, I., Søgaard, A., 2019. Latent Multi-Task Architecture Learning, in: *Proceedings of the AAAI Conference on Artificial Intelligence. AAAI Press*, pp. 4822–4829. 10.1609/aaai.v33i01.33014822.
- Ruder, S., 2017. An Overview of Multi-Task Learning in Deep Neural Networks [WWW Document]. URL <http://arxiv.org/abs/1706.05098>.
- Train, K., 1980. A Structured Logit Model of Auto Ownership and Mode Choice. *Rev. Econ. Stud.* 47, 357. <https://doi.org/10.2307/2296997>.
- Train, K.E., 2009. *Discrete choice methods with simulation, 2nd ed, Discrete Choice Methods with Simulation.* Cambridge University Press. 10.1017/CBO9780511753930.
- Wang, S., Wang, Q., Zhao, J., 2020. Deep neural networks for choice analysis: Extracting complete economic information for interpretation. *Transp. Res. Part C: Emerg. Technol.* 118, 102701.
- Wang, S., Wang, Q., Zhao, J., 2020b. Multitask learning deep neural networks to combine revealed and stated preference data. *J. Choice Model.* 37, 100236 <https://doi.org/10.1016/j.jocm.2020.100236>.
- Xia, Y., Chen, H., Zimmermann, R., 2023. A Random Effect Bayesian Neural Network (RE-BNN) for travel mode choice analysis across multiple regions. *Travel Behav. Soc.* 30, 118–134. <https://doi.org/10.1016/J.TBS.2022.08.011>.
- Yang, Y., Hospedales, T.M., 2019. Trace norm regularised deep multi-task learning. 5th Int. Conf. Learn. Represent. ICLR 2017 - Work. Track Proc.
- Yazdizadeh, A., Patterson, Z., Farooq, B., 2019. An automated approach from GPS traces to complete trip information. *Int. J. Transp. Sci. Technol.* 8, 82–100. <https://doi.org/10.1016/j.ijst.2018.08.003>.
- Zegras, C., 2010. The Built Environment and Motor Vehicle Ownership and Use: Evidence from Santiago de Chile. *Urban Stud.* 47, 1793–1817. <https://doi.org/10.1177/0042098009356125>.
- Zhang, Y., Sari Aslam, N., Lai, J., Cheng, T., 2020. You are how you travel: A multi-task learning framework for Geodemographic inference using transit smart card data. *Comput. Environ. Urban Syst.* 83, 101517 <https://doi.org/10.1016/j.compenvurbsys.2020.101517>.
- 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. <https://doi.org/10.1016/j.tbs.2020.02.003>.