A deep learning model for predicting route choice in public transport

Alessio Daniele Marra, IVT, ETH Zurich
Francesco Corman, IVT, ETH Zurich

Conference Paper STRC 2021

STRC | **21th Swiss Transport Research Conference**
Monte Verità / Ascona, September 12 – 14, 2021

A deep learning model for predicting route choice in public transport

Alessio Daniele Marra
Institute for transport planning and systems
ETH Zurich, Switzerland
E: alessio.marra@ivt.baug.ethz.ch

Francesco Corman
Institute for transport planning and systems
ETH Zurich, Switzerland
E: francesco.corman@ivt.baug.ethz.ch

September 2021

Abstract

Deep learning has been recently applied as an alternative method for several choice problems, such as mode choice. Nevertheless, this method has not been particularly explored for route choice, despite its possible advantages.

This work proposes a novel model for predicting route choice in public transport based on a convolutional neural network. The model has several advantages compared to the state of the art (e.g., Path Size Logit model). First, the model can infer a nonlinear utility function for the available routes. Second, it can also easily include any non-alternative-specific variable, such as socioeconomic characteristics or weather conditions, allowing complex interactions with all other variables. Third, the model generalizes the Path Size Logit, and thus can obtain the same or better performance.

The model is tested on a large-scale study based on GPS tracking, observing more than 2700 public transport trips of Zurich residents. The model is tested also on a synthetic dataset, to study its properties, performance, and ability to describe different utility functions. Finally, the performances of the model are compared with the state of the art.

Keywords

Deep Learning, Public Transport, Route Choice, Choice Models, Machine Learning

1. Introduction

Route choice in public transport is typically analyzed with discrete choice models, estimated from observed passengers' trips and a choice set of available alternatives for each observation. These models are used both to predict a choice outcome among a set of available alternatives, and to understand the choice process leading to a certain choice. The most common model used for route choice in literature is the Path Size Logit (Marra and Corman, 2020; Nielsen et al., 2021; Yap and Cats, 2021), which is a variation of the Multinomial Logit, including a correction term for correlated alternatives. Despite this and similar models are widely used, they also have limitations, such as the definition of a linear utility function.

Machine learning, and in particular deep learning, gained significant attention in recent years, as an alternative method to discrete choice models for choice problems (Wang et al., 2020b; Zhao et al., 2020). However, most of the recent works applying deep learning in transportation focuses on mode choice (Buijs et al., 2021; Han et al., 2020; Lee et al., 2018; Sifringer et al., 2020; Wang et al., 2020a), while lower attention is given to route choice. The two problems have important differences, such as in the definition of the alternatives, which can make a neural network designed for mode choice not suitable for route choice, and vice versa. In fact, the alternatives in mode choice represent different entities (e.g., car and bus), with their own variables, which can have assigned a predefined order in the input layer of the network (e.g., the input vector contains always first the variables of the car and then the variables of the bus). In contrast, for route choice, the alternatives do not have any order in general, and therefore the neural network should work regardless of the order of the alternatives in input. To the best of our knowledge, there is no work in literature designing a deep learning model specifically to analyze route choice in public transport. Such a model can allow more complex interactions among the input variables than traditional choice models and possibly obtain better performance in route prediction or model estimation. In this regard, the literature applying machine learning to route choice focuses mainly on car driving (Lai et al., 2019; Sun and Park, 2017). For instance, Lai et al. (2019) compare different machine learning models with standard choice models for drivers' route choice, showing also the former ones can be adopted for behavioral analysis.

In this work, we propose a novel approach to predict the route choice in public transport, based on deep learning. This work aims to compare the proposed model with a traditional model (the Path Size Logit), highlighting the advantages and disadvantages of the former one. In particular, the proposed model:

- Is an alternative method for predicting route choice in public transport based on deep learning. The model extends the Path Size Logit, allowing a more complex interaction among the variables.

- Allows estimating a nonlinear utility function.
- Can include non-alternative specific variables (e.g. socio-demographic or weather information), allowing complex interactions with all other variables.

In this work, we focus on the prediction ability of the proposed model, while we leave to a future work the analysis of the interpretability of the model and its ability to explain the choice process of passengers.

2. Methods

In this Section, we first show how a neural network can represent the Path Size Logit model. Hence, we present the proposed network, which extends the Path Size Logit.

2.1 Path Size Logit with a Neural Network

The state of the art in literature for route choice in public transport is the Path Size Logit model (Marra and Corman, 2020; Nielsen et al., 2021; Yap and Cats, 2021). This model is an extension of the Multinomial Logit, including an additional variable in the utility function, the Path Size factor. The utility function considered in this work is the following:

$$U_i = \beta_{tram} * tram\ time + \beta_{bus} * bus\ time + \beta_{train} * train\ time + \beta_{walk} * walk\ time$$
$$+\beta_{tt} * transfer\ time + \beta_{transfer} * transfers + \beta_{PS} * PathSize_i + \varepsilon_i \quad (1)$$

$$PathSize_{trip} = - \sum_{stage\ s \in trip} \frac{duration(s)}{duration(trip)} \ln(times\ s\ occurs\ in\ choice\ set) \quad (2)$$

The utility function includes the travel time in tram, bus and train, the walking time, the transfer time, a transfer penalty and the Path Size factor (Equation 2), defined as in Marra and Corman (2020) and based on the formulation in (Bovy et al., 2008). This last variable penalizes overlapping alternatives in the choice set (i.e. trips with one or more stages in common, using the same lines). As in the Multinomial Logit, the probability of choosing an alternative in the choice set is the following:

$$P(trip|choice\ set) = \frac{e^{U_{trip}}}{\sum_{j \in choice\ set} e^{U_j}} \quad (3)$$

Figure 1 shows how the Multinomial Logit (and therefore the Path Size Logit) can be represented by a neural network (as also shown in previous works, such as Sifringer et al., 2020). The input layer is formed by the variables of each alternative (3 variables and 3 alternatives in Figure). Afterwards, a convolutional layer with one filter performs a dot product between each set of variables and the filter, containing the βs to estimate (one β for each variable). The output of this product is the utility function. Finally, a Softmax layer computes the choice probability of each alternative, as in Equation 3. Estimating this network is equivalent to estimating the Multinomial Logit (excluding approximations of the estimation algorithm). The loss function minimized in the neural network is the categorical cross-entropy, described as follows:

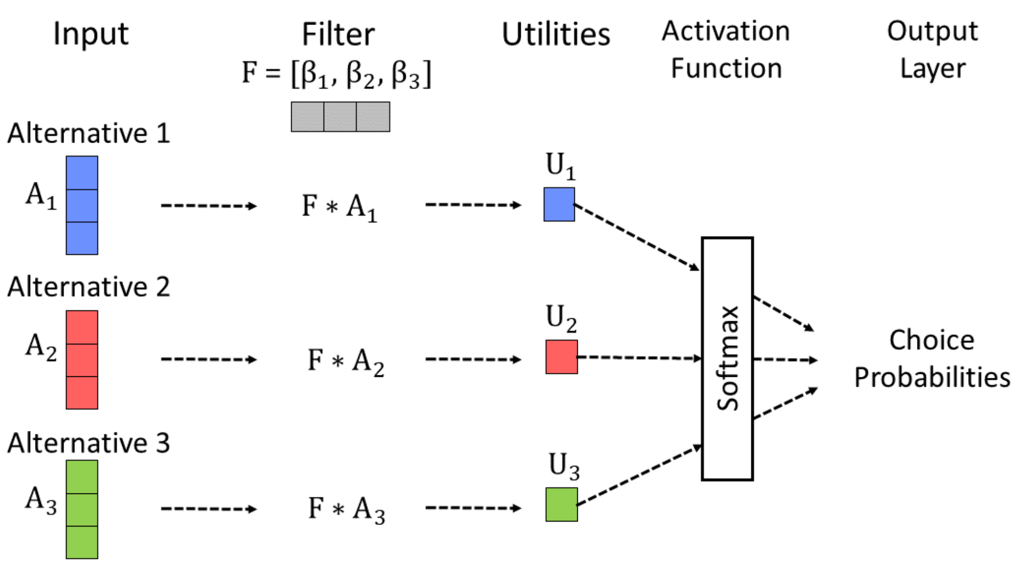
$$H(p,y) = -\sum_{j=1}^{K} y_j \log p_j \tag{4}$$

With K = number of alternatives; p_j = probability to choose alternative j (i.e., the output of the j-th neuron); y_j = 1 if j is the target class, 0 otherwise. Minimizing this function is equivalent to maximizing the log-likelihood function in the Multinomial Logit (as also explained in Sifringer et al. 2020).

$$L_{log} = \sum_{i=1}^{N}\sum_{j=1}^{K} y_{i,j} \ln p_{i,j} \tag{5}$$

With N = number of samples; K = number of alternatives; $y_{i,j}$ = 1 if sample i has label j, 0 otherwise; $p_{i,j}$ = probability sample i has chosen j.

Figure 1: Multinomial Logit with a neural network



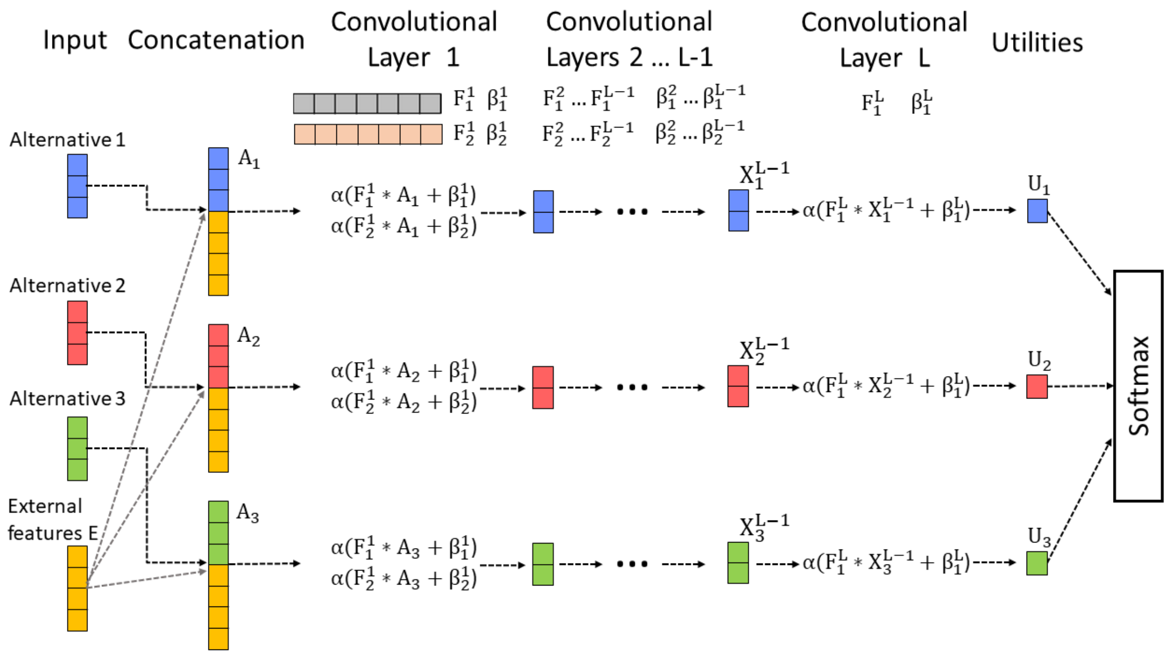
We remark that including more than 3 alternatives does not increase the number of parameters to estimate, i.e. the βs, as for the Multinomial Logit. Unavailable alternatives can be represented by an additional variable in the utility function, equal to 0 if the alternative is available, while equal to infinity if unavailable (practically, a big number of a higher order than the utility). This variable acts as a penalty. The relative β can be fixed to a negative value, e.g. -1. The other variables of an unavailable alternative can be set to 0.

2.2 Deep Learning Model for Route Choice

Figure 2 shows the proposed convolutional neural network for route choice in public transport. In input, there are the alternative-specific features (as in Figure 1) and non-alternative specific features, named external features, such as socio-demographic information or weather information. The set of external features is concatenated to each group of alternative-specific features, forming a new input for each alternative. Afterwards, a convolutional layer is applied, based on the activation function **α**. The layer includes **M** filters (2 in Figure). **L** convolutional layers are applied consecutively, to determine the final utility (the last layer has 1 filter). Finally, as in Section 2.1, a Softmax layer computes the probabilities.

To model an unavailable alternative, as explained in Section 2.1, a penalty can be added to the utility in the last layer. This is similar to Nam and Cho (2020), applying an "Availability of Alternatives Function" in the last layer.

Figure 2: Graphical representation of the proposed neural network



The proposed model has a similar structure to the Path Size Logit in Figure 1: a separate input for each alternative; the same function to compute the utility of each alternative (described by one linear filter in the Path Size Logit, while by multiple convolutional layers in the proposed network); a Softmax layer to compute the probabilities.

The advantages of the proposed model are the following:

1) It allows estimating a nonlinear utility function. In fact, if the activation function $\boldsymbol{\alpha}$ is nonlinear, the model can estimate complex interactions among the variables. Given that the network is convolutional, the same utility function is estimated for each alternative and including more or less alternatives does not change the number of parameters to estimate. Moreover, the order of the alternatives in input is not relevant. This last aspect is particularly important to analyze route choices, since in general the alternatives in a choice set do not have a specific order (in contrast, for mode choice the alternatives in input can be ordered, e.g., first car then bus).
2) The model extends the Multinomial Logit (and therefore the Path Size Logit). In fact, considering $\boldsymbol{\alpha}$(x) = x, **L** = 1, **M** = 1, and no external variables, the model is identical to the one in Figure 1. Therefore, with a proper tuning of the parameters, a model similar or more complex than the Logit can be represented.
3) The model can include non-alternative specific variables. We remark there are already models in literature considering non-alternative specific variables, such as models including interaction terms (Menghini et al., 2010; Raveau et al., 2014). Nevertheless, the proposed network allows complex interactions among all variables, without specifying a priori the type of interaction.

We estimated the proposed model using keras (Chollet and others, 2015), with Adam as estimation algorithm. We divided the dataset (described in Section 3) in 60% Training Set, 20% Validation Set and 20% Test Set. During the validation, we estimated the following hyper-parameters: $\boldsymbol{\alpha}$ (relu or tanh), **L**, **M**, **learning-rate, l2-regularization**. The batch size was set to 32. The number of epochs was determined with “Early Stopping” (Chollet and others, 2015). After validation, we evaluated the best model on the Test Set. As external features, we used socio-demographic information (gender, age and income), weather information (humidity and temperature) and the hour of the day. The gender is modelled with a categorical feature (0 or 1), while the income with 4 categorical features (representing low, medium, high, and not declared income). Age, humidity and temperature are modelled as numerical features, normalized to have 0 mean and unit standard deviation ($x' = (x - \mu)/\sigma$).

We compared the proposed model with the Path Size Logit, both in terms of prediction accuracy (percentage of correct detections) and log-likelihood (divided by the number of samples, $L'_{log} = L_{log}/N$).

3. Dataset

3.1 Tracking Data

In this work, we analyze route choices from realized data of public transport passengers, collected via GPS tracking. The dataset consists of 2901 public transport trips of 172 users in the city of Zürich (Switzerland). The data were collected with a smartphone application, the *ETH-IVT Travel Diary*, which collects GPS data continuously throughout the day, without affecting the battery consumption (Marra et al., 2019). After that, a mode detection algorithm determines the travel diaries of the tracked users from the GPS data, identifying activities, trips and modes used. The algorithm can also identify the public transport line and vehicle used, exploiting the Automatic Vehicle Location data (AVL) of the Zurich public transport network. Knowing the lines and vehicles used allows studying route choice of public transport passengers. Further information on the smartphone application and the mode detection algorithm are in Marra et al. (2019).

Table 1 provides general information on the dataset. The same dataset was already used in Marra and Corman (2020), to study route choice with traditional methods (Path Size Logit). Therefore, we refer to that work for more details on the dataset, its limitations, and the estimation of a standard route choice model.

Table 1: Information on the tracking data

Period	03/04/2019 – 02/06/2019
Participants	172
Average age	32.6
% female	43%
Average days per person	22
Public transport trips	2901
# transfers per trip	{0: 60%, 1: 29%, 2: 9%, 3+: 2%}
% modes used	{Tram: 52%, Bus: 38%, Train: 10%}
Avg. duration per trip	21.7 min
Avg. air distance per trip	2.88 km

3.2 Analyzed choices

For each observed public transport trip, we applied the choice set generation algorithm described in Marra and Corman (2020), to identify the available alternatives (according to the timetable). In Marra and Corman (2020), the algorithm obtained high coverage (94% of times the choice set contains the same alternative of the user, in terms of public transport lines) and it was already used to estimate successfully the Path Size Logit. In this work, we consider only trips covered by the respective choice set (2719 trips ≈ 94% of 2901), since the calculation of the alternative-specific variables (e.g. walking time and travel time) for non-covered trips is not obvious. In fact, the observations are based on actual trips, during actual operations, which may differ from planned operations, on which the choice set is based (e.g., the user may have used a vehicle not available according to the timetable). This may affect the model estimation, despite in a limited way, given that only 6% of the trips are discarded. We refer to Marra and Corman (2020) for more details on assumptions and limitations of the choice set generation algorithm.

To evaluate our model and test its properties, we analyzed both real and synthetic choices. In particular, given the identified choice sets, we analyzed the following choice scenarios:

1) Real choice: the chosen alternative is the one observed from the tracking.
2) Nonlinear choice: the chosen alternative is determined by the following nonlinear utility function (travel time in seconds): $U = -(transfers + 1)^2 * (busTime^2 + trainTime^3 + 10tramTime) * \sqrt{walkTime + 2 * transferTime}$
3) Group-based choice: the chosen alternative is determined by a linear utility function (Equation 1, time in seconds), different for four groups of participants, based on gender (A or B in this text) and age. The values of the coefficients (β_{tram} , β_{bus} , β_{train} , β_{walk} , β_{tt} , $\beta_{transfer}$, β_{PS}) are the following:
    - Gender = A, age >= 33: -1, -1, -10, -0.1, -0.1, -300, 0
    - Gender = A, age < 33: -1, -1, -0.1, -10, -10, -900, 0
    - Gender = B, age >= 33: -1, -10, -0.1, -1, -1, -300, 0
    - Gender = B, age < 33: -0.1, -1, -0.1, -0.1, -10, -1800, 0

The “nonlinear choice” and the “group-based choice” are synthetic scenarios, defined only to verify the properties of the model, namely, describing a nonlinear utility and including non-alternative specific variables. Therefore, no particular meaning should be associated with the utility functions chosen.

4. Results

Table 1 compares the performance of the Deep Learning model with the Path Size Logit for the two synthetic scenarios. The models are estimated considering 5 alternatives, including the chosen one and the 4 shortest available alternatives (in terms of travel time, considering a transfer penalty of 5 minutes). Considering more alternatives (e.g. 10) does not change significantly the comparison between the two models shown in Table 1. However, the choice set size may affect the tuning of the neural network. We remark 87% of times the observed trip is among the first 5 shortest alternatives and 90% of times among the first 10 (further information on choice set coverage in Marra and Corman, 2020).

In both models, we considered the same alternative-specific variables described in Equation 1. In the group-based scenario, we also included non-alternative specific variables (gender and age) in the Deep Learning model. The dataset has been divided into Training Set (60%), Validation Set (20%) and Test Set (20%). Table 1 reports the performance on the Test Set.

In both scenarios, the Deep Learning model has a very high accuracy (97% and 91%). This shows the model is able to learn a nonlinear utility function and to include non-alternative specific variables. In contrast, the Path Size Logit has lower performance, since it does not have these properties.

Table 2: Performance comparison between the Deep Learning model and the Path Size Logit on the synthetic scenarios

	Nonlinear Choice	**Group-based Choice**
Deep Learning Prediction Accuracy	97%	91%
Path Size Logit Prediction Accuracy	83%	67%
Deep Learning Log likelihood (L'_{log})	0.08	0.27
Path Size Logit Log likelihood (L'_{log})	0.61	0.99
Training Set	1631	1631
Validation Set	544	544
Test Set	544	544
Non-alternative specific variables in the Deep Learning Model	None	Gender, Age

Table 2 compares the performance of the Deep Learning model with the Path Size Logit for the real scenario, corresponding to the observed trips from the tracking data. Also in this scenario, the dataset has been divided into Training Set (60%), Validation Set (20%) and Test Set (20%). We estimated the Deep Learning model with and without non-alternative specific variables. The Deep Learning model with non-alternative specific variables have lower (better) Log Likelihood and a higher prediction accuracy than the Path Size Logit. This shows the proposed model achieves better performance than the Path Size Logit, which represents the state of the art in route choice in public transport.

Table 2 also shows that including non-alternative specific variables in the Deep Learning model does not change significantly the performance (higher precision but similar log-likelihood). This may suggest the chosen variables (listed in Table 2) are not particularly relevant to predict route choice in public transport. Moreover, the additional variables in input increase the number of parameters to estimate in the model. This may lead to overfitting and therefore lower performance on the test set.

In this work, we showed the proposed model is a valid method to predict route choice in public transport, which outperforms traditional methods. However, the moderate improvements in the prediction accuracy in the real scenario leave the following considerations, which we plan to investigate in detail in a future work:

1) The high prediction accuracy of the Path Size Logit suggests that a linear utility function, with only alternative specific variables, is sufficient to explain most of the observed route choices. Moreover, it is possible the wrongly predicted choices (≈ 24%) cannot be explained by the selected variables and they depend on unobserved characteristics of the users or of the alternatives.

Table 2: Performance comparison between the Deep Learning model and the Path Size Logit on the real scenarios

	Prediction Accuracy	**Log Likelihood (L'_{log})**	**Non-alternative specific variables**
Deep Learning (with non-alternative specific variables)	76%	0.73	Gender, Age, Income, Humidity, Temperature, Hour of the day
Deep Learning (only alternative specific variables)	75%	0.73	None
Path Size Logit	73%	0.78	None

2) Weather information and socio-demographic information have marginal effects on route choice (in the real scenario). This is in accordance with Marra and Corman (2020), which estimated a Mixed Path Size Logit in the same dataset, showing low heterogeneity among the users in the costs perception. This may be dependent on the data collection or on the choice set generation algorithm used.

5. Conclusions and Future Directions

In this work, we proposed a novel method for predicting route choice in public transport, based on deep learning. The proposed model generalizes the Path Size Logit, providing also two main advantages. First, it is able to infer a nonlinear utility function. Second, it can include non-alternative specific variables, allowing complex interactions among all variables. We verified these two properties on synthetic data, showing also better performance than the Path Size Logit. We tested the proposed model on a real dataset, collected via GPS tracking. The model obtained slightly better performance than the Path Size Logit, showing its validity for route choice prediction.

Despite the proposed model showed promising performance, we see several future directions, to strengthen and clarify the proposed analysis. First, the prediction accuracy of the proposed model with different choice set sizes should be analyzed in more detail. Second, we plan to analyze the impact of each non-alternative specific variable (and their combinations) on the model performance. Third, in this work we focused only on model performance, in terms of prediction accuracy. Therefore, we plan to investigate the interpretability of the Deep Learning model, also compared to the Path Size Logit. Finally, in the proposed model we considered the taste heterogeneity including users' characteristics as non-alternative specific variables. Therefore, we plan to analyze this heterogeneity, comparing it with traditional methods, such as the Mixed Logit (Nielsen et al., 2021).

6. References

Bovy, P.H.L., Bekhor, S., Prato, C.G., 2008. The Factor of Revisited Path Size: Alternative Derivation. Transportation Research Record 2076, 132–140. https://doi.org/10.3141/2076-15

Buijs, R., Koch, T., Dugundji, E., 2021. Using neural nets to predict transportation mode choice: Amsterdam network change analysis. J Ambient Intell Human Comput 12, 121–135. https://doi.org/10.1007/s12652-020-02855-6

Chollet, F., others, 2015. Keras, https://keras.io.

Han, Y., Zegras, C., Pereira, F.C., Ben-Akiva, M., 2020. A Neural-embedded Choice Model: TasteNet-MNL Modeling Taste Heterogeneity with Flexibility and Interpretability. arXiv:2002.00922 [cs, econ, stat].

Lai, X., Fu, H., Li, J., Sha, Z., 2019. Understanding drivers' route choice behaviours in the urban network with machine learning models. IET Intelligent Transport Systems 13, 427–434. https://doi.org/10.1049/iet-its.2018.5190

Lee, D., Derrible, S., Pereira, F.C., 2018. Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. Transportation Research Record 2672, 101–112. https://doi.org/10.1177/0361198118796971

Marra, A.D., Becker, H., Axhausen, K.W., Corman, F., 2019. Developing a passive GPS tracking system to study long-term travel behavior. Transportation Research Part C: Emerging Technologies 104, 348–368. https://doi.org/10.1016/j.trc.2019.05.006

Marra, A.D., Corman, F., 2020. Determining an efficient and precise choice set for public transport based on tracking data. Transportation Research Part A: Policy and Practice 142, 168–186. https://doi.org/10.1016/j.tra.2020.10.013

Menghini, G., Carrasco, N., Schüssler, N., Axhausen, K.W., 2010. Route choice of cyclists in Zurich. Transportation Research Part A: Policy and Practice 44, 754–765. https://doi.org/10.1016/j.tra.2010.07.008

Nielsen, O.A., Eltved, M., Anderson, M.K., Prato, C.G., 2021. Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. Transportation Research Part A: Policy and Practice 147, 76–92. https://doi.org/10.1016/j.tra.2021.02.010

Raveau, S., Guo, Z., Muñoz, J.C., Wilson, N.H.M., 2014. A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. Transportation Research Part A: Policy and Practice 66, 185–195. https://doi.org/10.1016/j.tra.2014.05.010

Sifringer, B., Lurkin, V., Alahi, A., 2020. Enhancing discrete choice models with representation learning. Transportation Research Part B: Methodological 140, 236–261. https://doi.org/10.1016/j.trb.2020.08.006

Sun, B., Park, B.B., 2017. Route choice modeling with Support Vector Machine. Transportation Research Procedia, World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016 25, 1806–1814. https://doi.org/10.1016/j.trpro.2017.05.151

Wang, S., Mo, B., Zhao, J., 2020a. Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. Transportation Research Part C: Emerging Technologies 112, 234–251. https://doi.org/10.1016/j.trc.2020.01.012

Wang, S., Wang, Q., Zhao, J., 2020b. Multitask learning deep neural networks to combine revealed and stated preference data. Journal of Choice Modelling 37, 100236. https://doi.org/10.1016/j.jocm.2020.100236

Yap, M., Cats, O., 2021. Taking the path less travelled: Valuation of denied boarding in crowded public transport systems. Transportation Research Part A: Policy and Practice 147, 1–13. https://doi.org/10.1016/j.tra.2021.02.007

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 Behaviour and Society 20, 22–35. https://doi.org/10.1016/j.tbs.2020.02.003