Transportation Choice Analysis with a Deep Learning Perspective

Madeleine Gates
Department of Statistics
Stanford University
mgates@stanford.edu

Abstract

While deep learning methods, including LSTM model variations, are fairly common in the transportation modeling literature, applications in the discrete choice modeling for transportation mode choice is more sparse. This work aims to compare methods used commonly in the literature (namely, single-layer and 3-4 layer neural networks) with a model that incorporates temporal embeddings (single time step history and weather).

1 Introduction

Understanding the transportation habits and decisions of individuals is an important focus of transportation policy and organization. Travel mode choice is a traditional survey mainstay in the discrete choice modeling literature, where methods have primarily focused on logit and probit models. Mixed logit and multinomial logit models in particular have been heavily used in the discrete choice modeling space.

Use of machine learning, and deep learning in particular, has become much more common in recent years when conducting prediction tasks such as transportation demand modeling, accident prediction, congestion prediction, and driver behavior studies. However, use of deep learning in travel mode choice is less represented, with a greater focus placed on stochastic gradient boosting trees, random forests, and support vector machines.

Additionally, many analyses of transportation mode choice focuses primarily on demographic factors, rather than the trips taken themselves. This data set is novel in its connection of choice data with attributes of alternative choices in terms of duration of trip and trip cost in real-life situations. My addition of weather data allows for a further analysis of the factors at play when choosing transportation modes. However, basic demographic indicators, such as gender, age, and rough socioeconomic grouping, are still incorporated. These data also include repeated observations of individuals, in some cases. Individuals have recorded between one and sixteen trips during their travel diaries. I use embeddings from this previous trip mode information.

2 Related work

The data used in this paper uses data from a study focused on recreating alternative choice situation features, and, from this data, determining the most salient features (Hillel et al., 2018). This application then focused on use of gradient boosting trees for mode choice prediction.

A meta-analysis of transportation studies and their relative predictive accuracies (Varghese et al., 2020) includes mention of choice modeling. However, they count only 3 studies incorporating deep learning for travel mode choice and which use more "vanilla" neural networks. The authors cite the need for behavior interpretability as a factor hindering the more widespread use of neural networks in

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

this space. In areas such as traffic modeling, there is a more diverse use of models such as LSTM and other RNN models.

Some studies in the transportation modeling choice space have focused on fine-tuning the architecture of artifial neural networks (single layer networks) or shallow neural networks with up to 3 layers (Buijs et al, 2021). These researchers have conducted other work focused on applying new methods of extracting features from specific users in choice data sets, before then using these data in shallow neural networks.Other studies have used various methods, including single layer networks along with methods such as stochastic gradient boosting trees, random forests, and support vector machines (Hagenauer Helbich, 2017). Other work has built up neural networks. In general, LSTM or time dependent methods are not found in the transportation choice modeling literature.

3 Dataset

This data set combines data from the London Travel Demand Survey (LTDS) with London temperature data. Survey data was augmented by researchers at University of Cambridge with estimated trip duration data using a directions API. The final data set includes 81,086 trips taken during the period of April 2012- April 2015 from 17,616 distinct households (31,954 unique individuals overall).

The data includes details of the trip as recorded by participants (date, time, purpose, and mode of travel). Trip purposes are categorized as either employer's business, home-based work, home-based education, home-based other, or non-home based other. Mode of travel includes driving, walking, cycling, or public transportation. Additionally, some demographic variables (age, gender, car ownership, proposed public transportation fare type based on socioeconomic factors, and driver's license status) are collected. Journey duration, based on different transportation methods, were also included using a directions API. In other words, for a proposed trip by a given individual at a given time, the time required to take the trip given the different transportation alternatives are provided.

Based on the date and time of recorded trips, I then added weather data from the UK's Center for Environmental Data Analysis (C.E.D.A.) Archive's MIDAS Open UK hourly weather observation data and UK hourly precipitation data, version 201901, for years 2012-2015. Specifically, the weather observations were measured in the greater London region at the Heathrow observation station. Multiple observations were recorded for precipitation data, so the maximum recorded value was utilized. Additionally, missing values were back-imputed (i.e. if hour 2 for a given day was missing, the value recorded for hour 1 was imputed for hour 2) for wind speed, visibility, air temperature, relative humidity, and sun duration. Missing values for snow depth were set to 0.

Data were then split into training, development, and test groupings. Households, rather than individuals or observation trips themselves, were sampled randomly. A total of 15,416 households were randomly selected for the training set, with 1,100 households in the dev set and 1,100 households in the test set. This corresponds to 70,913 trips in the training set (87.45%), 5,104 trips in the dev set (6.29%), and 5,069 trips in the test set (6.25%). Finally, all data were normalized using the mean and standard deviation calculated from the training data set.

Table 1: Percentage of Travel Mode Responses

Travel Mode	Training Set	Dev Set	Test Set
cycle	2.97%	2.84%	3.02%
drive	44.19%	44.85%	43.03%
pt	35.34%	33.99%	35.65%
walk	17.49%	18.32%	18.31%

4 Methods

Baseline Model: Conditional Logit Model/Artificial Neural Network (ANN)

A conditional logit, or a one-layer neural network, was also used as a baseline comparison. This is commonly referred to in the literature as an Artificial Neural Network (ANN) and is heavily relied upon when considering machine learning methods for application to choice situations. I initially implemented this model in R using the "nnet" package, which had been mentioned in various papers in the transportation modeling space. I used a softmax activation function to model the probability of choosing each of the possible transportation modes (car, public transportation, cycling, or walking). Models were trained with 100 epochs and a weight decay parameter of 0.001 (selected after testing a range of decay parameters on dev set performance). Additionally, the number of units in the hidden layer were selected based on the percentage of correct predictions generated by the model on the dev set. As can be seen in Figure 1, dev set performance remained fairly constant once reaching approximately 9 neurons in the hidden layer. In order to prevent over-fitting while still retaining complexity, I chose 13 neurons in the hidden layer, which was the second-highest predictive accuracy behind a hidden layer of 24.

Unfortunately, the functionality of this R package precluded selection of specific loss functions, so I ultimately decided to generate the final ANN model using Keras, which allows for much more tuning in terms of loss function and other parameters. Figure 2 and 3 depicts the loss and accuracy over number of epochs trained. Due to the flattening out at approximately epoch 100, I chose to train the model on 100 epochs. Adam optimization was found to perform best on dev set, and was used for training the ultimate model.

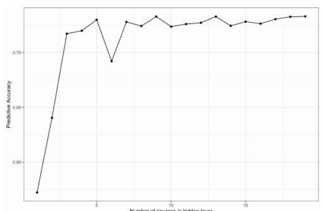


Figure 1: Number of neurons in hidden layer

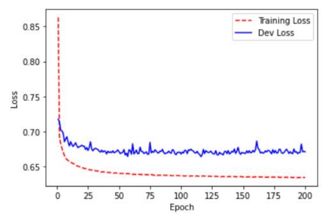


Figure 2: Training and Development Loss

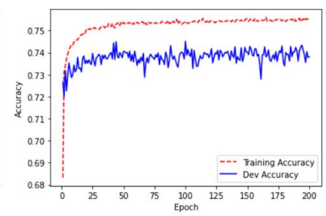
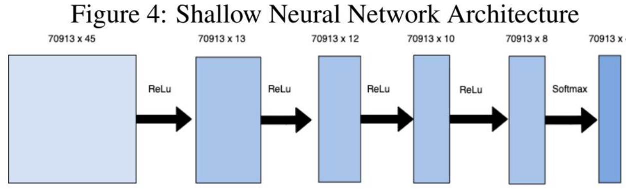


Figure 3: Training and Development Accuracy

Neural Network

To build upon the more common models used in the transportation modeling space, I additionally built a shallow neural network to learn more complex features of the data. I used a 4 layer neural network, utilizing ReLu activation functions in the hidden layers and a softmax output layer to model the probability of choosing each of the four transportation modes. The first layer again uses 13 nodes, while further node numbers decrease as layer index increases (see Figure 4). After testing optimizers including Adam, SGD, and RMSprop, I chose RMSprop due to highest dev set performance.

Figure 4: Shallow Neural Network Architecture

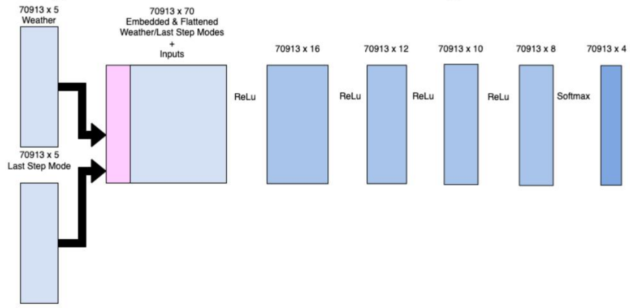


Neural Network with Embedding and Time Lag

These data had an element of time structure, given that most individuals reported multiple trips occurring over a single day or multiple days. However, many individuals reported only 2 trips in total, which made using LSTM techniques rather impractical. Rather than using a recurrent neural network, I created a one-step time lag variable to account for the method of the individual's last trip.

In this model, I use an embedding layer for weather categorical variables, and, separately, an embedding layer for the last transportation mode used by individuals. These inputs, along with the remaining inputs used in the previous ANN and NN models, were incorporated into a similarly structured neural network; however, this model used 16 rather than 13 nodes in the first full dense layer (see Figure 5).

Figure 5: Neural Network + Embeddings Architecture



5 Results and Discussion

Model performance was based on sparse categorical cross-entropy loss, as well as predictive accuracy as calculated on both the development set and test set. Referencing other work in the transportation modeling space, accuracy typically ranges from approximately 70% to close to 95% in some applications. Modeling in the transportation mode choice space (when deep learning methods are used) commonly have accuracy levels between 60% and 90%, depending on the application, with most studies resulting in approximately 75% to 80%, a mark which this paper also achieves.

The conditional logit model and 4 layer neural network performed similarly, with the conditional logit actually outperforming the 4 layer model slightly. It appears that the higher complexity of a somewhat deeper network did not aid in predictive power to a high degree, or that overfitting took place.

However, the model that incorporated time and weather embeddings outperformed the other two models. This is expected in that the addition of more data, in particular previous travel mode, is extremely helpful in estimating the next move. The improvement in loss was especially considerable, although accuracy is commonly the metric most cared about when considering transportation mode choice.

Table 2: Summary of Model Performance

Model	Dev Loss	Dev Accuracy	Test Loss	Test Accuracy
Conditional Logit	0.67	74.22%	0.64	75.24%
NN	0.69	72.88%	0.66	74.89%
NN + Time/Weather Embedding	0.57	80.90%	0.56	81.02%

Test set performance is slightly better than dev set performance for all models; however, the data all comes from the same distribution and has similar distributions in terms of travel mode, individual demographics, and last-trip travel mode.

6 Conclusion

Although having in-depth time series of user transportation mode is not likely feasible, it could be worthwhile to incorporate time-series of weather data, incorporating recurrent neural network methods to then be fed into a neural network with other input values. This could allow for

Connecting neural network methodology and more interpretable models (i.e. RUM) is an important area for the transportation choice modeling literature. Deeper and more complex networks, which often have improved predictive power, can hinder the ability of researchers to extrapolate user utility and preferences from specific aspects of their choice situations. Stacking models, incorporating embeddings, and building more complex networks for aspects of the choice landscape that can be abstracted without hindering overall interpretability may useful to consider in future work.

References

[1] Hagenauer, J., amp; Helbich, M. (2017, February 13). A comparative study of machine learning classifiers for Modeling Travel Mode choice. Expert Systems with Applications. `https://www.sciencedirect.com/science/article/pii/S0957417417300738`

[2] Proceedings of the Institution of Civil Engineers. (n.d.). Recreating passenger mode choice-sets for transport simulation: A case study of London, UK. `https://www.icevirtuallibrary.com/doi/epdf/10.1680/jsmic.17.00018`

[3] Varghese, V., Chikaraishi, M. Urata, J. Deep Learning in Transport Studies: A Meta-analysis on the Prediction Accuracy. J. Big Data Anal. Transp. 2, 199–220 (2020). Ret `https://doi.org/10.1007/s42421-020-00030-z`

[4] Buijs, R., Koch, T., Dugundji, E. Applying transfer learning and various ANN architectures to predict transportation mode choice in Amsterdam. Procedia Computer Science, 184, 532-540 (2021). `https://www.sciencedirect.com/science/article/pii/S1877050921006980`

[5] Hagenauer, J., Helbich, M. A comparative study of machine learning classifiers for modeling travel mode choice (2017). Expert Systems with Applications, 78, 273-282. `https://www.sciencedirect.com/science/article/pii/S0957417417300738`

[6] Wang, S., Mo, B., amp; Zhao, J. (n.d.). DNN with Alternative Specific Utility Function. Retrieved May 12, 2022, from `https://mobility.mit.edu/sites/default/files/DNN_with_Alternative_Specific_Util_Function.pdf`

[7] Ripley, B. nnet: Fit Neural Networks in nnet: Feed-forward Neural Networks and multinomial log-linear models. CRAN (2022). `https://rdrr.io/cran/nnet/man/nnet.html`

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

[9] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems (2015). Software available from tensorflow.org.

7 Appendix

Code Repository `https://github.com/madeleinegates/Transportation-Modeling-RNN`