New England University Transportation Center
77 Massachusetts Avenue, E40-279
Cambridge, MA 02139
utc.mit.edu

Year 25 Final Report

Grant Number: DTRT13-G-UTC31

Project Title:

Advanced Behavioral Analysis of High Resolution Mobility Data

Project Number:	Project End Date:	Submission Date:
MITR25-56	6/30/19	2 September 2019

Principal Investigator:	Christopher Zegras	Co-Principal Investigator:	
Title:	Professor	Title:	
University:	MIT	University:	
Email:	czegras@mit.edu	Email:	
Phone:	617 452 2433	Phone:	

Brief Description of Project

This project is motivated by the possible value of integrating theory-based discrete choice models (DCMs) and data-driven neural networks. How to benefit from the strengths of both is the overarching question. We propose hybrid structures and strategies to flexibly represent taste heterogeneity and improve predictability while keeping model interpretability. Also, we utilize neural networks' training machinery to speed up and scale up the estimation of Latent Class Choice Models (LCCMs).

First, we embed neural networks in DCMs to enable flexible representations of taste heterogeneity and enhance prediction accuracy. We propose two neural-embedded choice models: TasteNet-MNL and nonlinear-LCCM. Both models provide a flexible specification of taste as a function of individual characteristics.

- TasteNet-MNL extends the Multinomial Logit Model (MNL). A feed-forward neural network (TasteNet) is utilized to predict taste parameters as a nonlinear function of individual characteristics. Taste parameters generated by TasteNet are further fed into a parametric logit model to formulate choice probabilities. We demonstrate the effectiveness of this integrated model in capturing nonlinearity in tastes without a priori knowledge. Using synthetic data, TasteNet-MNL is able to recover the underlying utility specification and predict more accurately than some misspecified MNLs and continuous mixed logit models. TasteNet-MNL also provides interpretations close to the ground truth. In an application to a public dataset (Swissmetro), TasteNet-MNL achieves the best out-of-sample prediction accuracy and discovers a broader spectrum of taste variation than the benchmark MNLs with linear utility specifications.

- Nonlinear-LCCM enriches the class membership model of a typical LCCM. We represent an LCCM by a neural network and add hidden layers with nonlinear transformations to its class membership model. The nonlinearity introduced by the neural network provides a flexible approximation of the mixing distribution for both systematic and random taste heterogeneity. We apply this method to model Swissmetro mode choice. The nonlinear-LCCM outperforms an LCCM with a linear class membership model with respect to the out-of-sample prediction accuracy. Nonlinear-LCCM also provides interpretable taste parameters for each latent class.

Second, we embed DCMs in neural networks to (partially) maintain model interpretability. We propose two strategies: imposing special structure and parameter constraints to incorporate expert knowledge and improve interpretability. Both TasteNet-MNL and nonlinear-LCCM can be seen as special neural networks with DCMs embedded. In TasteNet-MNL, the downstream choice model defines the meaning of each output unit of the TasteNet, through the utility definition. As for nonlinear-LCCM, the class-specific choice models define meanings of the latent classes.

In addition to the special structure, we impose parameter constraints according to prior knowledge in two ways: using a transformation layer in the neural network (TasteNet-MNL); and adding a violation penalty to the objective function (nonlinear-LCCM). Using these strategies, we show both TasteNet-MNL and nonlinear-LCCM can achieve realistic behavioral/economic indicators, such as values of time, demand elasticities and choice probabilities at the disaggregate model level and at the individual level.

Thirdly, we demonstrate the benefits of using the neural network training machinery to estimate LCCMs, especially for models with a large number of classes/parameters, estimated on a large dataset, or under certain challenging scenarios such as highly unbalanced classes and relatively small sample size(s) in the minority latent class(es). We contrast neural network estimation by Adam, a stochastic gradient descent algorithm, with the standard maximum likelihood by quasi-Newton routine, and with Expectation-Maximization (EM). Synthetic data results show similar accuracy and estimation time when the dataset is relatively small and the model has a small number of classes. Traditional estimation approaches scale poorly to a large dataset and a model with a large number of classes/parameters. They tend to encounter the small class vanishing problem. Estimation based on stochastic gradient descent is shown to be 70 times faster on a large synthetic dataset, and can achieve more stable and accurate estimates for latent classes with relatively small membership.

This project takes an initial step towards developing a framework to integrate theory-based and data-driven methods for discrete choice modeling. We highlight the strengths and weaknesses of econometric DCMs and neural networks, and explore several ways to take advantage of both: DCMs' rigorous theory and domain knowledge; and neural networks' function approximation capability and many techniques developed to scale up estimation for big data and high-dimensional optimization.

The neural-embedded choice models proposed in this study can be integrated into trip-based model systems (e.g., a four-step model) and activity-based model systems. They can empower flexible model specification, scale up estimation, and at the same time maintain behavioral interpretability to a satisfactory degree, such that the hybrid approaches can support scenario analysis and discover new knowledge from behavior data with increasing granularity and complexity.

Research Outcomes

We present research outcome in two studies each of which is under preparation for submission to peer-reviewed journals. These studies are also described in a PhD Dissertation resulting from the project.

I. A Neural Embedded Choice Model for Capturing Systematic Taste Heterogeneity

This study introduces a neural embedded choice model – TasteNet-MNL to model heterogenous taste in discrete choice.

Implementation

- Model structure
- Monte-Carlo experiments

- Application: Swissmetro mode choice

Key Findings

On synthetic data:

- TasteNet-MNL is able to capture nonlinearity in taste and uncover the true utility form without a priori knowledge. Misspecified MNLs or random coefficient logit (RCL) models result in large bias in parameter estimates.
- TasteNet-MNL's prediction accuracy matches the true model (77% to 79%), higher than MNLs and RCLs with misspecified utility functions (70% to 72%).
- TasteNet-MNL provides interpretable economic indicators, like value of time and demand elasticities, close to the ground truth; while MNLs and RCLs with misspecified utility can produce unreliable interpretations.

On the Swissmetro dataset:

- TasteNet-MNL discovers a wider spectrum of taste variations in the population than the benchmarking MNLs.
- TasteNet-MNL predicts more accurately on hold-out datasets. Its superior predictability is a result of its flexibility in capturing nonlinear taste functions.
- TasteNet-MNL provides interpretable indicators for policy analysis. Values of time and elasticities derived from TasteNet-MNL are reasonable compared to the results from the MNLs. However, the average VOTs estimated by TasteNet-MNL are higher than the MNLs, due to the longer tails on the high end of willingness-to-pay.

II. Latent Class Choice Model Estimation and Extension with Neural Networks

Implementation

- A neural network representation of a latent class choice model
- Monte-Carlo experiments for comparing alternative estimation methods
- LCCM with flexible class membership model: a case study of Swissmetro mode choice

Key findings

On synthetic data: Stochastic gradient descent algorithm Adam has significant advantages for LCCM estimation

- First, Adam scales to large data better than BFGS or EM. For example, EM takes 70 times as long as Adam (69 vs 2 minutes) on a 50k data with 10 classes; and 56 times as long as Adam on a 100k data with 10 classes (140 vs 2.5 minutes).
- Second, Adam is more time-efficient and stable under complex model scenarios, such as with a large number of classes and parameters and unbalanced class membership. For example, when the number of latent classes increases from 3 to 10, with the total number of parameters increasing from 112 to 509, BFGS fails to converge after two days. EM takes about 60 to 70 times as long as Adam, with large variations across runs. Adam's estimation time is shorter with low variability.
- Thirdly, Adam obtains more accurate parameter estimates under certain difficult scenarios. When class membership is highly unbalanced, and the smaller classes do not have sufficient observations, EM frequently fails to identify small classes, generates large parameter

bias, and predicts class membership poorly. Adam is more robust. It is able to estimate small classes and predict class membership more accurately.

On the Swissmetro dataset:
- Nonlinear-LCCM learns a better mixing distribution than LCCM with linear utilities in class membership model; and improves choice prediction accuracy.

Research Benefits

Benefits of methods

1. Flexible specification, reduced bias, and more accurate prediction

Both TasteNet-MNL and nonlinear-LCCM provide a flexible specification of taste as a function of individual characteristics.
TasteNet-MNL is able to recover the underlying utility specification, while exemplary MNLs and continuous mixed logit models with misspecified systematic utility can produce large parameter bias and result in lower prediction accuracy.
The nonlinear-LCCM allows for a more flexible mixing distribution, which captures both systematic and random taste heterogeneity. The nonlinear-LCCM can improve out-of-sample prediction accuracy compared to the LCCM with linear utilities in its class membership model.

2. Interpretability partially maintained

Model interpretability is partially maintained compared to a direct employment of a neural network. Both TasteNet-MNL and nonlinear-LCCM are able to obtain realistic behavioral/economic indicators, such as values of time, choice elasticities and probability at the disaggregate model level and at the individual level.

3. Faster estimation

Estimating LCCM with a neural network can significantly speed up estimation (e.g. by 70 times), especially for large data and complex problems. SGD algorithm also alleviates small class vanishing problem, common for traditional estimation methods.

Practical Implications

Neural-embedded discrete choice models (NEDCMs) can be beneficial to transport modeling and planning, by reducing potential biases and improving model forecast accuracy.

At the model development stage, NEDCMs can complement manually specified DCMs to detect misspecification. For example, modelers can compare a NEDCM and a manually specified DCM's log-likelihood and prediction accuracy on hold-out datasets. If an NEDCM predicts better, the DCM specification can be potentially improved. In practice, we suggest trying both approaches and comparing the economic indicators derived from each, to see where disagreements occur, as a clue to detect data outliers or discover systematic patterns or signals that are ignored.

For model estimation, SGD algorithms can be implemented in software packages to speed up LCCM estimation; and improve estimation accuracy and stability for many-class problems. The improved speed can allow modelers to try many alternative specifications quickly.

For model application, e.g., aggregate demand forecast and policy scenario analysis, NEDCMs can be integrated into trip-based (e.g. a four-step model) or activity-based model systems. For example, certain choice modules, such as the vehicle ownership model or the mode choice model, can have a NEDCM as a second option. The NEDCMs in this study are written in Python using the deep learning platform PyTorch. They can be integrated into a four-step model or a micro-simulation model in the future. By comparing the forecasting results from alternative models, we can better understand the uncertainty due to model specification.

Final Products

The research conducted with support of this project culminated in a PhD Dissertation:

> Yafei Han (2019). Neural-Embedded Discrete Choice Models. Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Transportation, MIT, Dept. of Civil and Environmental Engineering, Sept.

In addition, we are currently preparing two peer-reviewed journal articles based on the research:

> Han, Y., Pereira, F., Ben-Akiva, M., Zegras, C. "Latent Class Choice Model Estimation and Extension with Neural Networks."

> Han, Y., Pereira, F., Ben-Akiva, M., Zegras, C. "A Neural Embedded Choice Model for Capturing Systematic Taste Heterogeneity."