Distributed heterogeneous data learning approach for choice modelling using residual neural networks

Melvin Wong[a], Bilal Farooq[a]

[a]*Laboratory of Innovations in Transportation (LiTrans), Civil Engineering Department, Ryerson University, Toronto, Canada*

Introduction

The trend towards machine learning in discrete choice modelling has been driven by the straightforward observation that deep neural networks (DNN) increase model prediction accuracy on large datasets compared to discrete choice models (DCM), by applying non-linear layers to capture data heterogeneity. Current discrete choice modelling large scale choice datasets have faced the issue of complexity and representing error correlation in unobserved attributes. Thus, implementing DNN have been the go-to solution for recent choice modelling applications (Golshani et al., 2018). The motivation behind this trend is that non-linear layers can be used extract high-level correlations (Mullainathan and Spiess, 2017). However, because of the difficulty in performing model identification tests, parameter analysis and lack of model interpretability, the applicability of machine learning has not been clearly justified in behavioural demand modelling applications such as shared mobility or data-driven connected vehicles.

In recent years, a new development in DNN models known as residual neural networks (ResNet) found that simplifying the deep neural network architecture – i.e. by incorporating shortcut connections and bypassing the non-linearities – improves the model interpretability and accuracy. Our preliminary mathematical analysis have shown that the ResNet architecture closely resembles a Generalized Extreme Value (GEV) model structure with multiple correction terms (residuals) in the deterministic utility function. In light of this, we hypothesize that ResNets can be easily adapted to model choice behaviour without the general loss of model interpretability while capturing richer correlation structure in large behavioural datasets.

In this paper, we develop and examine the theoretical and experimental properties of ResNet in the context of capturing higher order correlations specific to human behaviour in making choices. We also develop the associated model estimation procedure using machine learning algorithms e.g. stochastic gradient descent and provide econometric analysis on parameter flexibility, utility correction scale changes and variable elasticities. We also compare the model with mixed-logit specification, which is considered as one of the most flexible form of GEV.

Background

Srivastava et al. (2015) showed that in conventional deep neural network, with the increase in the number of layers (i.e. increasing non-linearity) it may lead to worse performance compared to a simpler model, contradicting the logic that a N-layer network should produce a higher model accuracy than a $(N-1)$-layer network by capturing higher level detail in the model. ResNets exploit the use of identity shortcuts to enable flow of information across layers without causing model degradation from repeated non-linear transformations (He et al., 2015). It has been developed to learn deep feature representations in visual recognition tasks, which enables significant improvements in model accuracy (Veit et al., 2016; Wu et al., 2019). This recent discovery in machine learning literature provided researchers a way to train very deep neural networks.

Methodology

In discrete choice analysis, correlation structure and data heterogeneity can be represented by GEV models with specification of an error corrected utility function. ResNet retains the same utility formulation while improving model optimization by capturing correlation and heterogeneity in the residual components. This concept of residual modelling is analogous to the way utility functions behave in discrete choice models. While residual layers map the unobserved correlation factors between utility of each alternatives, the identity shortcut maps the inputs directly onto the utility by a linear parameter vector.

We interpret the ResNet as a model that captures the harmonic components in the data: we use the identity shortcut to represent the fundamental harmonic, i.e. systematic observed utility. Each residual layer would then add higher level harmonics to the basic utility function, representing the unobserved components. The residual layers are represented as a sum of parameterized error functions in the utility. Let $V_{0,j}$ be the input to the ResNet for a choice model with j alternatives. $V_j = \sum_m \beta_{jm} x_m$ is the observed portion of the utility. Each residual layer performs a nonlinear function on the input $F(V_j; \lambda_j) = \ln(1 + e^{\lambda_j V_j})$ and each layer is a recursive addition of the previous layer:

$$V_{1,j} = V_{0,j} + F(V_{0,j}) \tag{1}$$

$$V_{2,j} = V_{1,j} + F(V_{1,j}) = V_{0,j} + F(V_{0,j}) + F(V_{1,j}) \tag{2}$$

where λ_j is a $(j \times 1)$ matrix that represents the scaling factor of each residual layer. The utility function after N residual layers is the last output of the ResNet:

$$U_j = V_{0,j} + \sum_{n=0}^{N} F(V_{n,j}) + \varepsilon_j \tag{3}$$

$$= V_{0,j} + \sum_{n=0}^{N} \ln(1 + e^{\lambda_{n,j} V_{n,j}}) + \varepsilon_j \tag{4}$$

Data and Analysis

For experiments, we apply the ResNet model on a large scale trip dataset with 293,330 observations, collected from the Greater Montreal Region. We train a set of ResNet mode choice models between 1 and 200 layers to compare the effects and significance of higher level correlations. We use standard econometric attributes e.g. distance, time, purpose, etc., in the model. Based on the theoretical observations and modelling results, we address the question of whether ResNets are a compatible formulation for discrete choice models to capture high-level correlation in large datasets.

Conclusion

In the context of choice behaviour modelling, we propose a distributed heterogeneous data learning approach based on the recently developed residual neural network architecture. Our model specification uses multiple residual layers to learn high level correlations in the data. This work provides a behaviourally plausible case of integrating interpretable machine learning into GEV models. It allows for greater model flexibility on large datasets and heterogenous spatial-temporal data.

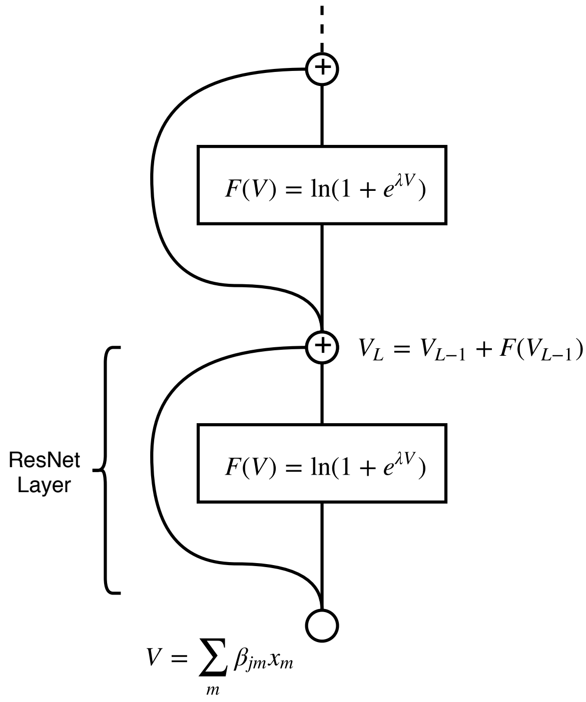

Figure 1: A schematic of a ResNet model

References

Golshani, N., Shabanpour, R., Mahmoudifard, S.M., Derrible, S. and Mohammadian, A., 2018. Modeling travel mode and timing decisions: Comparison of artificial neural networks and copula-based joint model. Travel Behaviour and Society, 10, pp.21-32.

Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. Journal of Economic Perspectives, 31(2), pp.87-106.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.

Veit, A., Wilber, M.J. and Belongie, S., 2016. Residual networks behave like ensembles of relatively shallow networks. In Advances in Neural Information Processing Systems, pp. 550-558.

Wu, Z., Shen, C., van den Hengel, A., 2019. Wider or Deeper: Revisiting the ResNet Model for Visual Recognition, Pattern Recognition, in press.

Srivastava, R.K., Greff, K. and Schmidhuber, J., 2015. Training very deep networks. In Advances in neural information processing systems, pp. 2377-2385.