Reframing Choice Set Formation: An Attention-Enhanced ResLogit Model for Destination Choice

Houman Haghi [1], Zachary Patterson [2], and Bilal Farooq [3]

[1]PhD Student, Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada , Email: houman.haghi@mail.concordia.ca
[2]Professor, Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada , Email: zachary.patterson@concordia.ca
[3]Associate Professor, Laboratory of Innovations in Transportation, Toronto Metropolitan University, Toronto, ON, Canada , Email: bilal.farooq@torontomu.ca

Abstract
Destination choice modelling is a fundamental step in transportation planning as it informs demand forecasting, infrastructure investment, and policy design. A reliable model should not only achieve high predictive accuracy, but also provide interpretable insights to support decision-making. Due to typically very large numbers of choices available in universal choice sets and lack of information on the choice availability at the individual level, the formulation of consideration choice sets is a crucial component in modelling individual preferences in destination choice modelling.

Traditional discrete choice models, such as the multinomial logit model, provide interpretable parameter estimates but often struggle to capture complex decision-making processes, limiting their predictive power. In recent years, deep learning models have been explored as an alternative due to their ability to capture non-linear relationships between choices and explanatory variables [1]. However, a significant drawback of deep neural networks is their lack of interpretability, making it difficult to extract behavioural insights and understand how individuals make decisions. To bridge this gap, this study applies the ResLogit framework, an interpretable deep learning model for discrete choice analysis [2]. ResLogit integrates residual connections into a logit-based structure, allowing it to capture non-linear effects while preserving the behavioural foundations of discrete choice models. The first contribution of this study is the application of ResLogit to destination choice modelling, providing a framework that balances predictive accuracy and interpretability.

Another major challenge in destination choice modelling is determining which alternatives a decision-maker considers before making a choice. The exact consideration set is often unobservable, leading to potential biases in estimated parameters [3]. To address this, researchers have developed various deterministic and stochastic methods for choice set generation [4, 5, 6]. Deterministic approaches rely on predefined rules, such as travel time thresholds or geographic constraints, to define feasible alternatives. However, these methods often impose rigid structures that fail to capture individual variability, leading to overly restrictive or unrealistic choice sets. Stochastic methods, such as random sampling or importance sampling, introduce variation in the alternatives considered by different individuals. However, a key limitation of these approaches is their reliance on assumptions about the probability of inclusion, which may not accurately reflect true consideration behaviour. While previous studies have primarily focused on generating choice sets through external models before integrating them into discrete choice models, little attention has been given to incorporating choice set formation directly within a deep learning-based model. To address this,

we propose incorporating a self-attention mechanism within ResLogit to dynamically determine relevant alternatives for each decision-maker. Unlike conventional choice set generation methods, which impose rigid rules or rely on heuristics, attention mechanisms allow the model to learn which alternatives are most relevant in a data-driven manner. This integration provides a behaviorally consistent representation of choice sets within the ResLogit structure, offering a novel solution to the challenge of handling large alternative sets.

This study uses data from the 2018 Montreal Origin-Destination (OD) survey, conducted by the Autorité régionale de transport métropolitain (ARTM). The OD survey, collected every five years, provides detailed information on household demographics, individual characteristics, and travel behaviour for residents of the Greater Montreal area. For this study, we focus specifically on home-based work trips, where individuals select a work destination from a large set of possible locations. The dataset includes records of individual trips alongside attributes such as age, employment status, household size, and income level. Additionally, zone-level characteristics, such as population and employment density, are derived from census tract data. Travel impedance is measured using a distance matrix, capturing the spatial separation between origins and potential destinations. To ensure data consistency and reliability, several preprocessing steps were applied. First, only work trips were retained, and trips with unreasonable travel times, such as those exceeding three hours or recorded as zero travel time, were excluded from the dataset. Next, the study area was restricted to the Island of Montreal, which includes 533 census tracts, ensuring a focused analysis on the central urban region. To improve model performance, continuous variables such as population density, employment rates, and distances were standardized, ensuring numerical stability during training. The final dataset consists of 19,389 observations, each representing an individual's work trip decision among 533 possible destinations. This dataset serves as the foundation for training the ResLogit model with attention-based choice set formation.

In the proposed framework, self-attention is applied to preprocess the choice set. The attention mechanism computes importance scores for all available destinations and selects the top-K destinations that are most relevant to the individual, resulting in the consideration choice set (C_n). This ensures that the model learns from a refined set of alternatives, preventing issues associated with universal choice sets that may include irrelevant options. The attention mechanism operates as follows: for an individual n considering destinations d, the attention score is computed as:

$$\text{Score}_{n,d} = \frac{Q_n K_d^T}{\sqrt{d_k}} \tag{1}$$

where Q_n represents the query (e.g., an embedding of the traveller's characteristics), K_d^T represents the key (e.g., an embedding of the destination's attributes), and d_k is the dimension of the key vector. The attention weights are then obtained using a softmax function:

$$\alpha_{n,d} = \frac{\exp(\text{Score}_{n,d})}{\sum_j \exp(\text{Score}_{n,j})} \tag{2}$$

These attention weights determine the relative importance of each alternative for the decision-maker. To ensure behavioural realism, only the top-K destinations with the highest attention weights are retained for modelling. The selected destinations are then used as inputs to ResLogit, where the systematic utility function remains:

$$U_{n,d} = V_{n,d} + \varepsilon_{n,d} = \text{ASC}_d + \beta x^*_{n,d} + g_{n,d} + \varepsilon_{n,d} \tag{3}$$

where $V_{n,d}$ represents the deterministic component of utility, which consists of the alternative-specific constant ASC_d, the systematic component $\beta x^*_{n,d}$, and the residual correction term in ResLogit $g_{n,d}$. The term $x^*_{n,d}$ corresponds to the adjusted input vector incorporating attention-based filtering, and $\varepsilon_{n,d}$ is the random error term, assumed to follow a Gumbel distribution, consistent with the logit framework. The residual component is formulated as:

$$g_{n,d} = -\sum_{m=1}^{M} \ln\left(1 + \exp\left(\theta^{(m)} h_{n,d}^{(m-1)}\right)\right) \tag{4}$$

where $\theta^{(m)}$ are the learned parameters of the residual layer m that adjust for heterogeneity and substitution effects, and $h_{n,d}^{(m-1)}$ represents the non-linear transformation of inputs from the previous residual layer. The final choice probability after accounting for the choice selection probability in the consideration set becomes:

$$P(d|C_n) = \frac{\exp(V_{n,d} + \ln(\alpha_{n,d}))}{\sum_j \exp(V_{n,j} + \ln(\alpha_{n,j}))} \tag{5}$$

where $P(d|C_n)$ represents the probability of choosing alternative d given the consideration set C_n. The term $\alpha_{n,d}$ is the attention-derived probability that alternative d is included in the consideration set, ensuring that alternatives with higher relevance receive greater weight in the final choice probability calculation. The logarithmic transformation of the attention weights ensures consistency with the utility scale in the logit model, allowing attention-derived relevance scores to be integrated in an additive form that preserves the probabilistic foundations of the model.

The total number of parameters in the proposed model depends on the dimensionality of the input features, which include zone-level attributes as well as impedance measures. To safeguard against overfitting, standard techniques such as regularization, validation-based tuning, and early stopping will be applied during training. While the attention mechanism adds parameters, its structure remains compact, particularly in the case of single-head implementations. The complete parameter count and diagnostics will be provided in follow-up analyses to contextualize model complexity in relation to the feature space.

The ResLogit-attention framework is trained using maximum likelihood estimation (MLE).The process begins with the initialization of attention weights and ResLogit parameters. Attention scores are then computed for each alternative in the choice set, and a refined subset of top-K destinations is selected. The filtered alternatives are passed through ResLogit, where systematic utility components are estimated. Model parameters are updated iteratively using a loss function based on the negative log-likelihood of the observed choices. To maintain behavioural consistency, a constraint is imposed to always include the chosen alternative in the attention-filtered set, ensuring that no relevant destinations are excluded due to attention misalignment. This constraint preserves model interpretability and ensures realistic choice behaviour in the prediction process. The process of integrating self-attention with ResLogit is illustrated in Figure 1.

To evaluate the model's predictive performance, we will employ several accuracy measures, including the percentage of correct predictions, log-likelihood of observed choices, and comparison of predicted versus observed distance distributions as a form of behavioral validation. This approach helps assess the realism of spatial patterns and the extent to which the model reflects underlying decision tendencies. Evaluation will be conducted on both training and held-out test sets to ensure generalizability, with a minimal gap between training and test accuracy interpreted as evidence of strong out-of-sample performance. To benchmark performance, we will compare the ResLogit-attention model against both traditional interpretable models and more opaque black-box alternatives. Specifically, we will estimate a multinomial logit (MNL) model to represent a fully interpretable baseline. This comparative evaluation will demonstrate how the proposed framework balances predictive power with interpretability, and whether it offers improvements over existing modeling paradigms without compromising behavioral transparency.

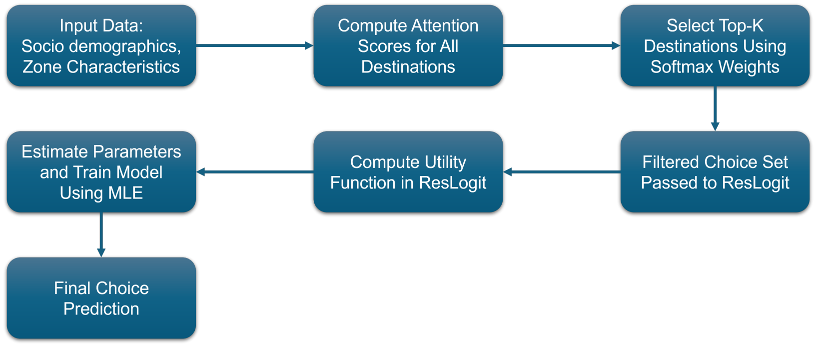

Figure 1: Flowchart representation of the methodology integrating self-attention with ResLogit.

The study presents an innovative approach for destination choice modelling by integrating the ResLogit framework with a self-attention mechanism. The proposed methodology goal is to address three major challenges in destination choice modelling: (1) improving predictive accuracy, (2) ensuring model interpretability, and (3) handling large choice sets. ResLogit provides a behaviorally consistent framework by maintaining a logit-based structure, while self-attention enables dynamic choice set formation by identifying the most relevant alternatives for each individual. The key advantage of this approach is that it allows for data-driven choice set formation without relying on predefined rules or heuristics. Unlike traditional deterministic and stochastic choice set generation methods, the attention mechanism learns from observed behaviour and dynamically assigns relevance scores to alternatives. This ensures that the model captures individual-specific preferences while maintaining a structured utility-based decision-making process. Additionally, by applying attention, we refine the input space while preserving the interpretability of the estimated parameters. The proposed methodology has broader applicability beyond destination choice modelling, as its underlying principles is transferable to other domains, such as route choice models, where decision-makers face a large set of alternatives.

References

[1] Shenhao Wang, Qingyi Wang, and Jinhua Zhao. Deep neural networks for choice analysis: Extracting complete economic information for interpretation. *Transportation Research Part C: Emerging Technologies*, 118:102701, 2020.

[2] Melvin Wong and Bilal Farooq. Reslogit: A residual neural network logit model for data-driven choice modelling. *Transportation Research Part C: Emerging Technologies*, 126:103050, 2021.

[3] Rui Yao and Shlomo Bekhor. A variational autoencoder approach for choice set generation and implicit perception of alternatives in choice modeling. *Transportation Research Part B: Methodological*, 158:273–294, 2022.

[4] Moshe E Ben-Akiva and Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.

[5] Jean-Claude Thill. Choice set formation for destination choice modelling. *Progress in human geography*, 16(3):361–382, 1992.

[6] Panagiotis Tsoleridis, Charisma F Choudhury, and Stephane Hess. Probabilistic choice set formation incorporating activity spaces into the context of mode and destination choice modelling. *Journal of Transport Geography*, 108:103567, 2023.