Ensemble-Based Fractional Split Multinomial Logit Model: An Application to Work Commute Mode Shares for Localized Transportation Planning

Srinath Ravulaparthy* (srinath@geog.ucsb.edu), Senior Director, Geospatial Sciences, CoreLogic LLC., Irvine California, USA

Konstadinos G. Goulias (goulias@geog.ucsb.edu), Professor, Department of Geography, University of California, Santa Barbara

ABSTRACT

United States (U.S) transportation policy is shifting from a 20th-century automobile-oriented market towards a 21st-century multimodal transportation-oriented market - choices including walking, bicycling, and public transportation. This focus is particularly amplified by emerging mobility options and technologies (e.g., ride-hail services and alternative fuel vehicles) leading to various decarbonization pathways that encourage efficient energy mobility systems (CARB, 2021). As a result, transportation planning agencies embark on developing advanced transportation models, wherein aspects of budget, time and other constraints act as barriers for agile decision making and policy analysis. We propose to fill this gap by developing an agile sketch-planning neighborhood-level work commute mode share model that leverages nationally available data sources from the U.S. Census Bureau and Environmental Protection Agency (EPA) and OpenStreet Maps. In this study, we make use of the flexible fractional split multinomial logit (FMNL) model formulation as proposed by Sivakumar and Bhat (2002) that predicts shares of commute modes (or proportion of commuters by mode) at the neighborhood-level (at block-group) for the entire U.S. Also, FMNL model application in transportation planning and modeling literature has been scarce with a few exceptions as applied to activity time-use patterns and aggregate land-use planning (Ye and Pendyala 2005; Schnieder et al., 2018).

Although this structure provides a great advantage in predicting homogenous block-group scale commute mode shares. It is evident from past research that commute shares are both spatially correlated and have a heterogeneous distribution due to first and second order effects that impact travel behavior of the commuters. However, there is very little work-to-date that addresses these issues within FMNL models in a holistic framework. As a result, we propose to leverage bagging-based methods, i.e., utilizing an ensemble of FMNL models aimed at reducing model instability, capturing spatial heterogeneity and improving model goodness-of-fit and performance. This work contributes to the field in two ways - First, we demonstrate that bagging-based fractional split multinomial logit model (BAG-FMNL) combines desirable properties of FMNL of being stable and theoretically sound, along with bagging, which is noise resistant and applicable to large feature spaces in model building. Second, we describe a feature selection algorithm for BAG-FMNL based on statistical properties of FMNL that ensures computational time is not wasted on statistically insignificant features.

In order to demonstrate the effectiveness of the proposed algorithm, we make use of the 2019 U.S. Census Journey-to-Work (J2W) flow data as compiled for the New York Metropolitan Transportation Commission (NYMTC) region - consisting of roughly 20 million in population. For this exercise, we specifically build on open-source data to make the sketch-level planning tool and BAG-FMNL algorithm accessible to planners and policymakers. For this purpose, we define

block-group level commute mode shares (e.g., drive alone, carpool, bus, subway, telecommute, walk, bike and other) as the outcome variable, wherein their proportion sums to unity. We also conduct a comprehensive analysis based on literature review to perform feature engineering that represents various influencing aspects of commute behavior such as sociodemographic characteristics and built-environment characteristics (e.g., accessibility, density and diversity). For instance, Figure-1 shows a sample box plot distribution of differences between observed and predicted (on Y-axis ranging from -1 to +1) commute shares by mode (as on X-axis) at the block-group level in the NYMTC study area. Overall, the distribution showcases that a preliminary specification of BAG-FMNL with an ensemble of 10 training models matches the observed mode shares value very closely with significantly lower variation for modes related to bike, walk, telecommute (work from home) and public transportation (bus and subway). Similarly, Table-1 documents the performance results of the BAG-FMNL model with observed data along the prediction accuracy and mean absolute error measurements. The results (although still preliminary) demonstrate the model stability and predictive nature, wherein model stability and goodness of fit is robust. As an application, we also propose to demonstrate in this study the policy-case of COVID-19 restrictions, wherein aspects of shelter-in-place and operational constraints on businesses were imposed early in the pandemic.

Finally, we propose to demonstrate the capability and efficiency of the BAG-FMNL model with other model specifications and we plan to compare this model against a generalized FMNL model and a series of ensemble-based machine learning models (e.g., gradient boosting and random forests). For this purpose, we will document and report the model performance results against standardized benchmarking (e.g., RMSE and MAPE).

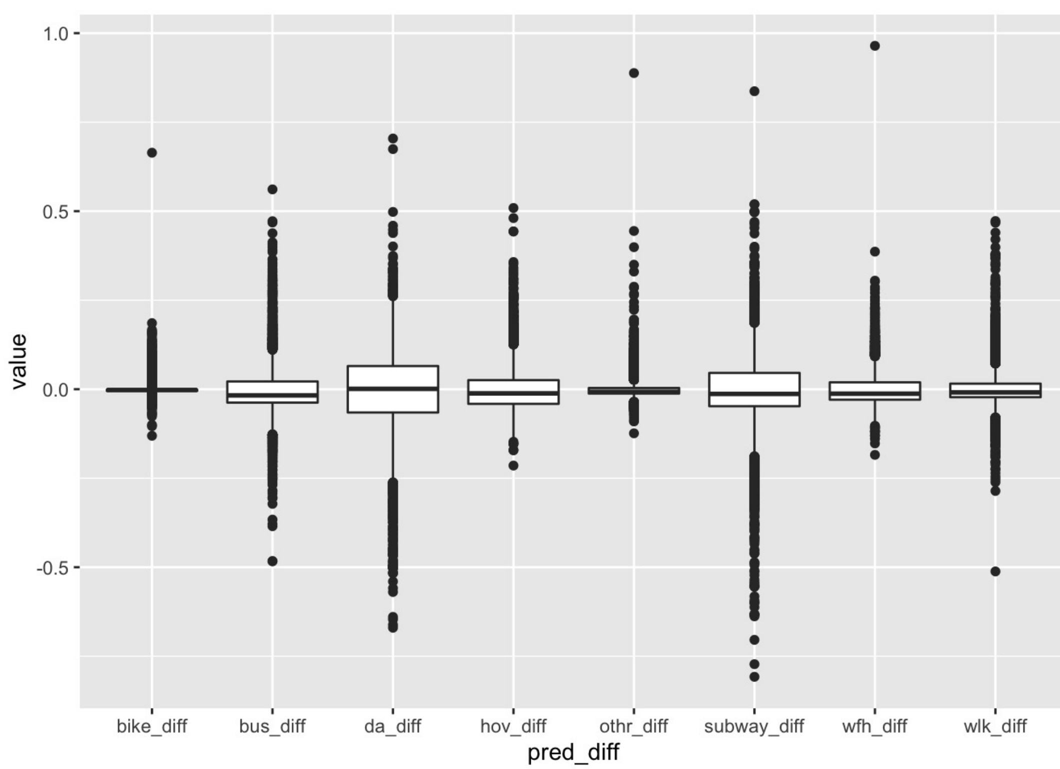

Figure 1: Boxplot of observed and predicted commute shares by mode using the 2019 J2W flow data

Table 1: Goodness of fit measures by Mode

Commute Mode	Prediction Accuracy	Mean Absolute Error
Drive Alone (DA)	87.98%	0.086
Carpool (HOV)	93.74%	0.045
Bus	92.41%	0.052
Subway (SUB)	89.36%	0.072
Bike	98.08%	0.008
Walk (WLK)	94.56%	0.034
Telecommute (WFH)	95.19%	0.034
Other (OTHR)	96.91%	0.017

References

1. CARB, (2021). LCFS Pathway Certified Carbc
2. Sivakumar, A. and Bhat, C.R. (2002). A fractional split distribution for statewide commodity flow analysis. Transportation Research Record, vol.1790, pp. 80-88.
3. Ye, X. and Pendyala, R. (2005). A model of daily time-use allocation using the fractional logit methodology. Proceedings from the 16th International Symposium on Transportation and Traffic Theory.
4. Schnieder, R.J., Hu, L. and Stefanich, J. (2018). Exploring the importance of detailed environment variables in neighborhood commute mode share models. Journal of Transport and Land Use, vol.11, no.1, pp.921-938.