Applying Machine Learning to Model Freight Vehicle Type Choice

Usman Ahmed
Matthew Roorda

9th METRANS International Urban Freight Conference, Long Beach
May 25, 2022
Freight Business and Logistics Decisions Simulation Framework
Freight Mode v/s Vehicle Type

➢ Mode Choice:
  • Road, rail, air, water
  • Most relevant for inter-city, statewide, and national level studies

➢ Vehicle Type Choice:
  • Road-based mode: Passenger car, trucks, vans, etc.
  • Most relevant for city or metropolitan area level studies
Background and Motivation

- Freight flows have been increasing in Canada.
  - 16.7% increase in freight shipments from 2011 to 2017 (Statistics Canada 2020)
- Economic development of regions
- Global competitiveness of industries
- Changing trends in supply chain and logistics
- Major contribution to greenhouse gas emissions!
Background and Motivation
Background and Motivation

- Implications on quality of life of urban residents
  - Noise pollution
  - Traffic congestion
  - Safety impacts
  - Parking problems
  - Pavement damage
Study Objectives

- Study the factors behind freight vehicle type choice

- Comparison of discrete choice with machine learning methods
  - Discrete choice: Multinomial and mixed logit model
  - Machine learning: Random Forest
Study Area
Data Source

- Commercial Travel Survey
  - Region of Peel (2006/07), Region of Durham (2010), Toronto Area (2012)

- Outbound Shipments
  - 1,439 shipments
  - 385 firms

- Explanatory Variables
  - Industry type, commodity type
  - Shipment origin and destination (cities)
  - Employment and shipment weight
Data Source

Frequency

Employment (Logarithmic Scale)
Data Source

![Histogram](image-url)
Data Source – Vehicle Types

- Passenger Car: 10%
- Pickup/Cube Van: 35%
- Single Unit Truck: 29%
- Tractor Trailer: 26%
Random Forest

- Training Data
  - Bootstrap Sample 1
    - Prediction 1
  - Bootstrap Sample 2
    - Prediction 2
  - Bootstrap Sample K
    - Prediction K

Majority Vote

Trees
mtry
Min_n
Random Forest – Variable Importance

- Shapley Additive Explanation (SHAP)
  - To assess the impact of explanatory variables on the model output
  - Comparison of model prediction with and without the variable
  - SHAP value is calculated for every observation
  - Variables are sorted based on the impact
  - Color of the point shows its value
    - Red: high value
    - Blue: low value
Training v/s Testing Data

- Models are developed on training data
  - RF: 10-fold cross validation
- Model prediction accuracy is calculated on testing data

- Training and testing data are divided based on firms
  - Training data: 269 firms with 1114 shipments (70%)
  - Testing data: 116 firms with 325 shipments (30%)
Results
Variable Importance

- Log Weight (kg)
- Log Employment
- Destination Outside Toronto Area
- C_Food and Food Products
- Intracity Shipment
- C_Metal and Metal Products
- I_Wholesale Trade and Transportation Handling
- C_Manufactured Products

SHAP value (Pickup/Cube Van)
Variable Importance

MNL: 1.64
Mix-MNL: 1.79

- Log Employment
- Log Weight (kg)
- Destination Outside Toronto Area
- C_Food and Food Products
- I_Wholesale Trade and Transportation Handling
- Intracity Shipment
- C_Metal and Metal Products
- C_Manufactured Products

SHAP value (Single Unit Truck)
Variable Importance

- Log Employment
- Log Weight (kg) \textit{MNL: 0.88}
- Destination Outside Toronto Area
- C_Food and Food Products
- Intracity Shipment
- I_Wholesale Trade and Transportation Handling
- C_Manufactured Products
- C_Metal and Metal Products

SHAP value (Tractor Trailer)
Variable Importance

- Log Employment
- Log Weight (kg)
- Destination Outside Toronto Area
- Intracity Shipment
- C_Metal and Metal Products
- C_Food and Food Products
- I_Wholesale Trade and Transportation Handling
- C_Manufactured Products

SHAP value (Passenger Car)
Discrete Choice Methods

➢ Larger firms are more likely to use larger vehicles
➢ Larger vehicles are more likely to be used for heavier shipments
➢ Intracity shipments are more likely to be transported using smaller vehicles
➢ Larger vehicles are more likely to be used for shipments destined outside of Toronto Area
Model Predictions

Observed VS Predicted

- Pickup or Cube Van: Observed 31%, RF 45%, MNL 37%, Mixed MNL 37%
- Single Unit Truck: Observed 33%, RF 26%, MNL 31%, Mixed MNL 31%
- Tractor Trailer: Observed 21%, RF 21%, MNL 21%, Mixed MNL 21%
- Passenger Car: Observed 15%, RF 8%, MNL 11%, Mixed MNL 11%
Prediction Accuracy

Observed VS Correct Predictions

- Pickup or Cube Van: Observed 65%, RF 51%, MNL 50%
- Single Unit Truck: Observed 32%, RF 35%, MNL 34%
- Tractor Trailer: Observed 67%, RF 55%, MNL 47%
- Passenger Car: Observed 31%, RF 19%, MNL 20%

MNL: Mixed Multi-Attribute Logit Model
Results Summary

- Overall prediction accuracy
  - Random Forest: 50%
  - MNL: 42%
  - Mix-MNL: 40%
Conclusion

➢ Applications in policy analysis
  • Demand for parking facilities
  • Greenhouse gas emissions
  • E-commerce, same-day deliveries
Conclusion

- Freight vehicle type choice is studied using discrete choice and RF methods
- Commercial travel survey data are used to develop models for the Toronto Area
- RF results are interpreted using SHAP based variable importance
- RF model has higher prediction accuracy than DCM
More about this work!

• More details about models and results can be found in: