A Decision Tree Approach to Analyze Freight Mode Choice Decisions

Amir Samimi *

Department of Civil Engineering,

Sharif University of Technology, Tehran, Iran

Hesamoddin Razi-Ardakani

Department of Civil Engineering,

Sharif University of Technology, Tehran, Iran

Kouros (Abolfazl) Mohammadian

Department of Civil and Materials Engineering

University of Illinois at Chicago, Chicago, IL, USA

***** Email: asamimi@sharif.edu

1 Introduction

Freight mode choice modeling is an integral part of the commodity demand forecasting procedure. Complexity of agent behaviors and diversity of supply chains, along with the unavailability of an appropriate dataset, has hindered realistic freight planning frameworks. Although optimization approaches [1] and discrete choice models are the state of practice methods [2], there are still some practical limitations. Optimization approaches, on one hand, are data intensive and not easy to formulate and solve in practice. On the other hand, discrete choice models have some unavoidable statistical assumptions such as linear property of utility function and pre-defined structures (e.g. probability distributions) that prevent, to some extents, realistic estimation of mode choice behavior. Decision tree models, however, are very easy to build and interpret, and are used in this study to explain freight mode choice behaviors in the U.S. Five different decision tree models, namely CART, CHAID, E.CHAID, QUEST and C5.0 are developed and compared with a binary logit mode choice model.

2 Method and Data

Decision trees are easy to understand, and require little data for calibration. Although a few studies have adopted artificial intelligence methods [3] for the analysis of freight modal selection behaviors, tree-based methods are not utilized yet. Data mining techniques have been implemented in other transportation planning arenas; when the primary concern is basically the prediction power not policy assessment. Decision trees, however, are easy to develop and provide insightful information about factors that influence mode choice decisions.

The analysis in this paper is based on a survey conducted in April and May 2009, providing information of 881 domestic shipments in the United States. Basic information about each establishment along with data on five recent shipments, namely origin, destination, mode of transportation, type, value, weight, and volume of the commodity, were obtained. Two modes are considered in this study: truck only and rail or rail intermodal. In this study, however, choosing rail over truck has only 9% chance of occurrence.

4 Results

Five decision trees are developed and C5.0 classification tree is illustrated in Figure 1, as an example. Prediction power and descriptive ability of the trees are compared to a logit model that is discussed in details elsewhere [4]. Only a few explanatory variables that are used in the logit model are employed for developing the trees. A six fold cross validation technique is also utilized to measure the performance of models. The mean value of six classification accuracies is then considered as the model performance (Table 1). Geometric mean (G-Mean) of correctly predicted cases of rail and truck is used instead of total accuracy to compare the performance of models, since an imbalanced dataset is under study. According to the G-Mean value, C5.0, CART and logit proved to have the best prediction power.

The C5.0 and CART trees convey that long distance, heavy, and containerized shipments have a higher chance for rail haul. In general, the mode choice determinants in the CART tree are very similar to the significant variables in the logit model. Furthermore, based on a variable importance analysis in logit, C5.0 and CART models, shipment weight turned out to be the most influential on freight mode choice decisions. Generally, decision trees proved to be quite powerful in predicting the freight mode choice decisions, although modeling complexities are much less than econometric models. Econometric models, on the other hand, seem to be more appropriate for more complex policy assessments and for taking into account the interactions among explanatory variables.

References

[1] Cunningham, W. H. J., "Freight Modal Choice and Competition in Transportation: A Critique and Categorization of Analysis Techniques", Transportation Journal, 21(4), pp 66-75. , 1982.

[2] de Jong, G., Gunn, H., Walker, W., "National and international freight transport models: An overview and ideas for future development", Transport Reviews, 24(1), pp 103-1240., 2004.

[3] Abdelwahab, W., Sayed, T., "Freight mode choice models using artificial neural networks", Civil Engineering and Environmental Systems, 16(4), pp 267-286., 1999.

[4] Samimi, A., Pourabdollahi, Z., Mohammadian, A., and Kawamura, K., "An Activity-based Freight Mode Choice Microsimulation Model", Proceedings of the 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.

TABLE 1. Comparing accuracy of decision trees and the logit model

Model		Train				Test			
		Truck	Rail	Total	G-Mean	Truck	Rail	Total	G-Mean
LOGIT	Mean	98.5%	55.6%	95.2%	73.8%	98.0%	58.3%	95.0%	74.9%
	S.D.	0.005	0.075	0.010	0.052	0.025	0.175	0.030	0.119
CART	Mean	98.9%	69.4%	96.7%	82.8%	97.5%	58.3%	94.6%	74.7%
	S.D.	0.006	0.053	0.009	0.034	0.023	0.175	0.029	0.120
CHAID	Mean	98.1%	44.4%	94.1%	64.6%	96.4%	33.3%	91.7%	50.3%
	S.D.	0.011	0.171	0.008	0.148	0.024	0.236	0.015	0.280
E.CHAID	Mean	97.8%	42.2%	93.6%	62.7%	96.2%	33.3%	91.4%	50.2%
	S.D.	0.013	0.171	0.005	0.145	0.026	0.236	0.018	0.280
QUEST	Mean	99.4%	39.4%	94.9%	61.3%	98.0%	42.8%	93.9%	64.5%
	S.D.	0.005	0.173	0.009	0.136	0.028	0.083	0.027	0.064
C5.0	Mean	99.5%	75.6%	97.7%	86.7%	97.3%	66.7%	95.0%	80.1%
	S.D.	0.004	0.040	0.004	0.023	0.024	0.149	0.025	0.091

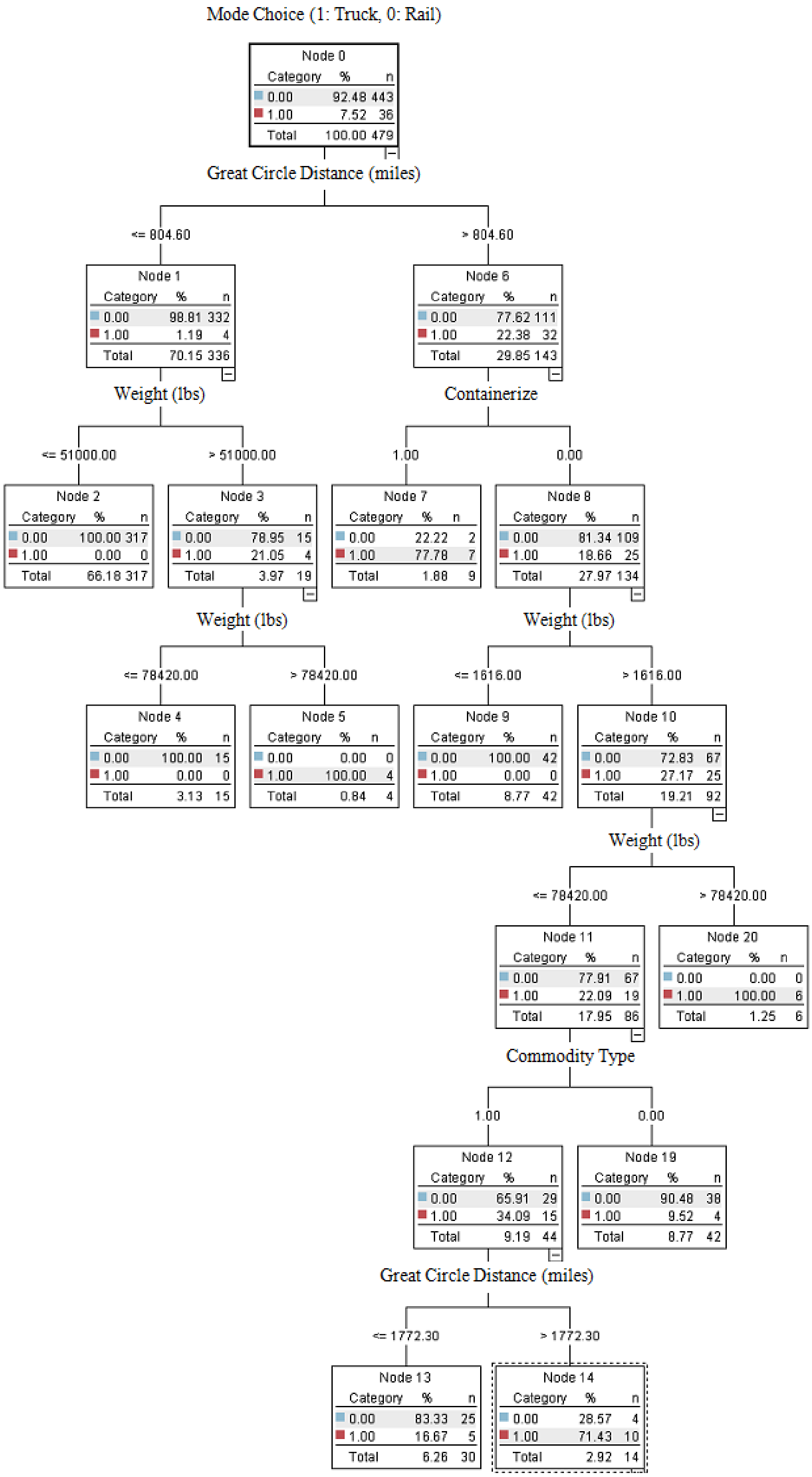

FIGURE 1. C5.0 classification tree for freight mode choice decisions