INCORPORATING HETEROGENEITY IN DECISION RULE REPRESENTATIONS OF TRAVEL CHOICE DECISIONS

Soora Rasouli, Eindhoven University of Technology, s.rasouli@tue.nl

Harry J. P. Timmermans, Eindhoven University of Technology

ABSTRACT

The inclusion of heterogeneity in different choice modeling approaches has been limited to applying mixing distributions to utility weights and the identification of latent classes. The nature of the assumed utility functions has however assumed to the same across individuals/classes. In this paper, we explore the usefulness of boosting to examine heterogeneity in different decisions rules. The approach is applied the Dutch National Survey data with special attention to transport mode choice decisions. Results indicate that the inclusion of different decision trees improves the goodness-of-fit of the model.

Keywords: Heterogeneity, Decision rules, Travel survey, Boosting

INTRODUCTION

Although it has been widely recognised that individuals may differ widely in terms of their preference and decision strategies, heterogeneity has been treated in rather limiting ways in travel behaviour research. Unobserved heterogeneity has been typically addressed by estimating mixed logit models in which point utility estimates are replaced by some mixing distribution or by latent classes, which identify different classes of individuals who share the same utility function (e.g., Brownstone and Train, 1998, 2008; Greene and Hensher, 2003, 2007; Greene et al., 2006; Hensher et al., 2008; Hess and Train, 2011; Wen et al., 2012). Both approaches are limited in that, although they allow for variation in attribute weights affecting overall utility and therefore choices, the nature/form of the assumed utility function is exactly the same for all individuals. Because predominantly a linear function is used, researchers have assumed that individuals are all involved in a compensatory decision making process. In case decision trees have been used, again researchers have typically assumed that the decision tree model equally applies to all individuals (e.g., Wets, et al., 2000; Arentze and Timmermans, 2004; Janssens, et al., 2006)

A more flexible approach would allow for a model with different utility functions or decision trees. Different functions, allowing for different variables and possibly different functional forms, would capture different decision making styles. In the context of decision trees, this theoretical notion is close to the technique of boosting, although the fundamental underpinnings are different. Boosting is based on the idea that a simple decision tree may depict a certain part of the variability in de dependent variable. Different trees may then capture different sources of variability. The approach involves extracting a sequence of decision trees, where each next tree is based on the prediction residuals of the preceding tree. Thus, at each step, the data is partitioned in the best possible way, considering the maximum number of child nodes, and the deviations of the observed values from the respective means are computed.

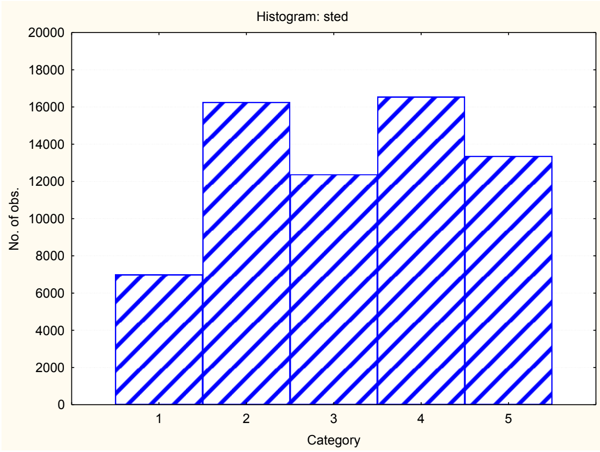
Common to all machine learning algorithms, the approach may result in over-fitting. Therefore, the data are usually split into training and test data, and the quality of the fitted model is evaluated by considering the performance of the model for the test data. The optimal number of decision trees is determined by calculating the smallest average squared error in the test data. Weights are used to combine the predictions from the successive decision trees into a single classification. The concept of risk, defined as the proportion of cases incorrectly classified by the set of decision trees, is used to evaluate the performance of the model.

The aim of this paper is to further explore this approach. The boosting approach is applied to the problem of transport mode choice.

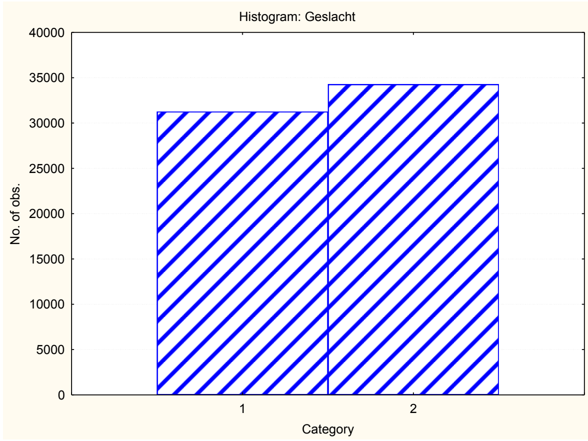
DATA

The current analysis is based on the 2009 Dutch National Travel Survey. In 2009, a total of 23679 respondents participated in the data collection. Respondents were asked to report all their trips for a designated day and a set of characteristics for each trip, including time, purpose and transport model. In total, 65535 trips were recorded.

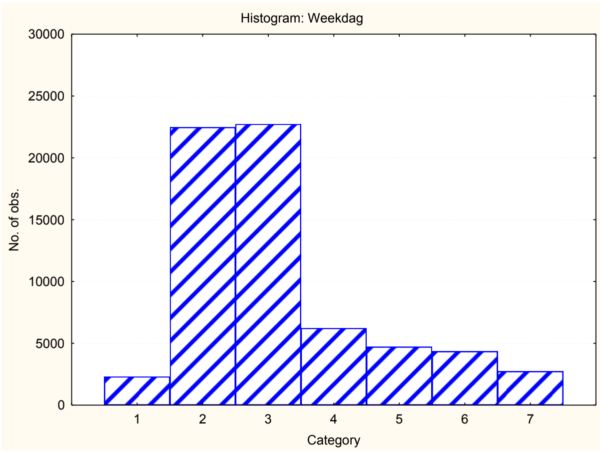
The following explanatory variables were used in the analysis: (i) type of municipality, (ii) day of the week, (iii) gender, (iv) age category, (v) education, (vi) possession of driver's license, (vii) activity duration, (viii) trip purpose and categorized trip distance. The survey classifies municipalities into the following classes: (i) very strongly urban = 2500 addresses per km2; (ii) strongly urban = 1500 to 2500 addresses per km2; (iii) moderately urban = 1000 to 1500 addresses per km2; (iv) weakly urban = 500 to 1000 addresses per km2; and (v) non-urban = no less than 500 addresses per km2. Age is categorized into 18 categories: 0 - 5; 6 - 11; 12 – 14; 15 – 17; 18 – 19; 20 – 24; 25 – 29; 30 – 34; 35 – 39; 40 – 44; 45 – 49; 50 – 54; 55 – 59; 60 – 64; 65 – 69; 70 – 74; 75 – 79, and 80 years of age and older. Education has 7 categories, from elementary school to university. A total of 13 categories are used to classify trip distance: international trip; 0,1 - 0,5 km ; 0,5 - 1,0 km ; 1,0 - 2,5 km; 2,5 - 3,7 km; 3,7 - 5,0 km; 5,0 - 7,5 km; 7,5 - 10 km; 10 - 15 km; 15 - 20 km; 20 - 30 km; 30 - 40 km; 40 - 50 km and more than 50 km. Travel purposes was divided into work, personal care, shopping, education, social visits, recreation, touring and other. Driver's license is a binary variable. The dependent variable, transport mode, was categorized into Car driver, Car passenger, Train, Bus/tram/metro, Motorbike, Bike, Walking and Other.



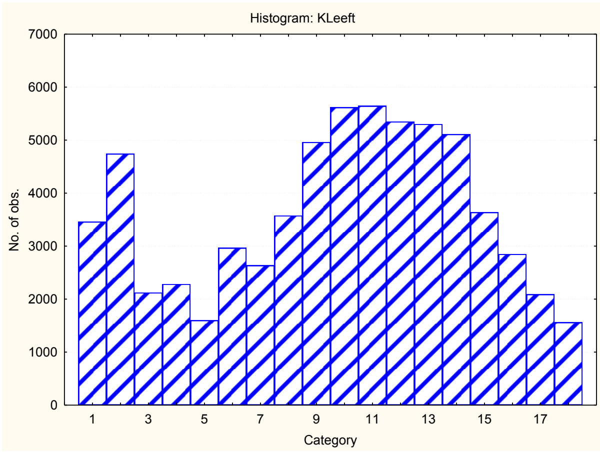
Urban density



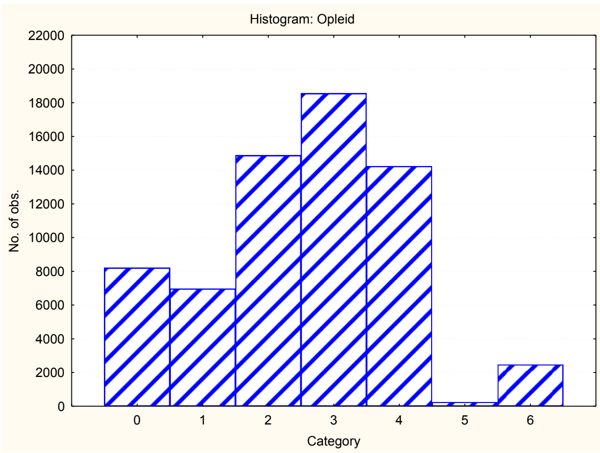
Gender



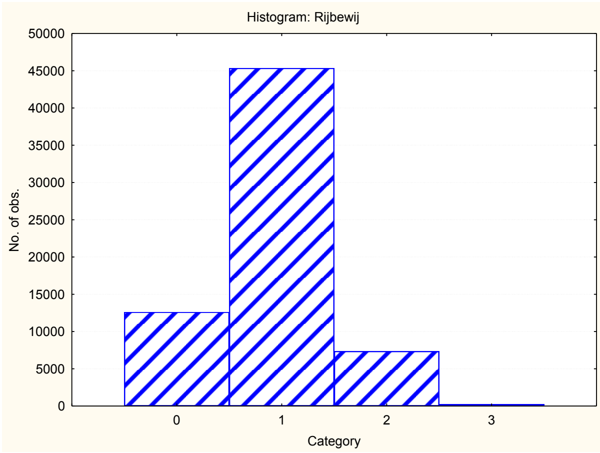
Days of the week



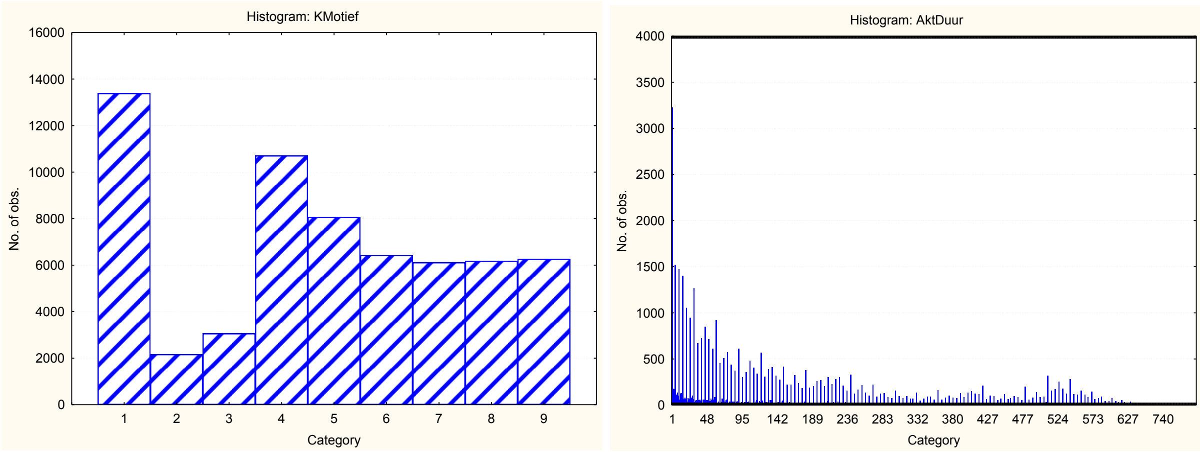
Age categories



Education



Driver’s license



Trip purpose Activity duration

Figure 1 – Overview of distributions of variables

Figure 1 shows the frequency distributions of the selected variables. It shows that the number of observations in very strongly urbanized areas is smaller than in the other categories of urban density. The number of observations for men is almost equal to the number of observations for women in the sample. As for days of the week, Figure 1 shows that the number of observations is significantly higher for Tuesdays and Wednesdays. It also shows that the number of observations in the 12-18 years of age categories is smaller than for the older age categories. The number of observations for the higher education categories is higher. Note that the last two categories denote "unknown" and "other". In the vast majority of observed trips, respondents do have a driver's license. The number of observations is similar for several trip purposes, but lower for business trips, and higher for commuter trip. As for the distribution of activity duration, most activities take a relatively short time, with some exceptions as indicated by the peak at the right end of the distribution.

ANALYSES AND RESULTS

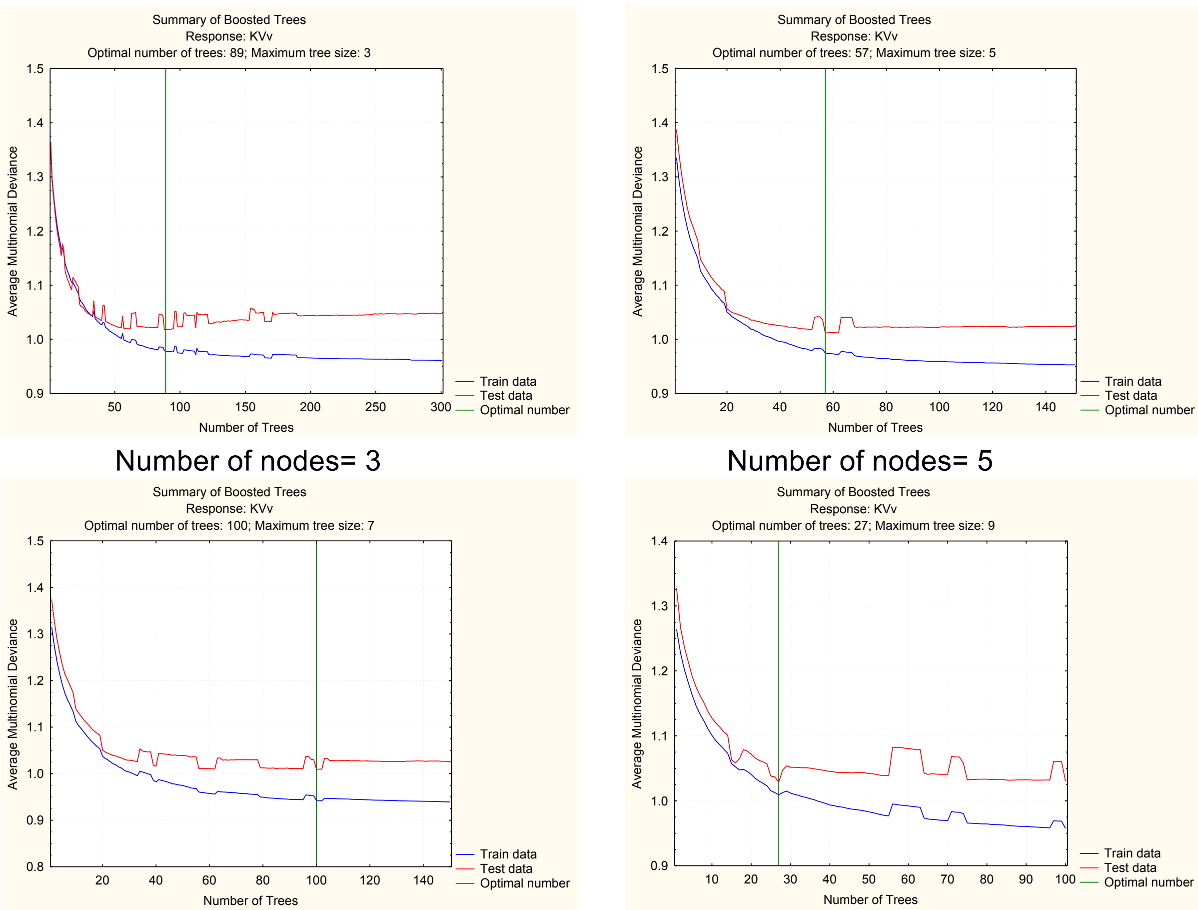
Creating boosted decision trees requires a number of operational decisions. In context of this study, the following decisions were made. First, the total sample was randomly split into training and a test sub-sample. The test subsample consists of 30% of the total sample. The construction of decision trees involves deciding when to stop splitting. Four different stopping parameters were used: (i) the minimum number of cases should at least be 100; (ii) the minimum number of cases in the child node should be at least 10; (iii) the maximum number of levels should be equal to 10. The fourth stopping criterion that was used is the maximum number of nodes. This criterion influences tree complexity, i.e. the degree of heterogeneity in decision rules. To examine the impact of this operational decision on final results, the decision tree induction was repeated for 3, 5, 7 and 9 nodes and results were compared.

The learning rate was set equal to 0.1 as experience indicates this value tends to produce good results. In the case of 3 nodes, the number of derived decision trees was set equal to 300. When it turned out that the optimum number was considerably less, tit was reduced to 150 for the 5-7 number of nodes options, and to 100 for 9 nodes.

The analyses involved the following steps. First, the optimum number of decision trees for the different settings related to the maximum number of child nodes was compared. Second, the goodness-of-fit of these alternative settings and the nature of mis-predictions was evaluated. The latter analysis was based on an inspection of confusion matrices. Third, the importance of the selected explanatory variables in classifying transport mode used was compared for the different settings, reflecting a difference in the complexity of decision trees.

The optimal number of decision trees

A first analysis concerned a comparison of the optimal number of decision trees as a function of the maximum number of child nodes. Results visualised in Figure 2 shows that the optimal number of decision trees fluctuates with an increasing number of maximum child nodes. This optimal number of decision trees is equal to respectively 89, 57, 100 and 27 for a maximum of respectively 3, 5, 7 and 9 nodes. These numbers are based on the best performance of the combination of decision trees in predicting the test data. A closer inspection of the behaviour of the various graphs shows that both a small number and a high number of child nodes also lead to more and longer fluctuations in the predictive performance of the ensemble of decision trees for the training data.



Number of nodes= 3 | Number of nodes= 5

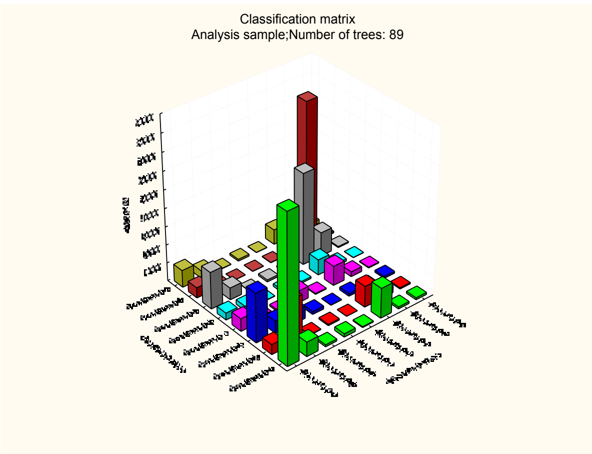
Number of nodes= 7 | Number of nodes= 9

Figure 2 - Optimal number of trees by maximum number of nodes

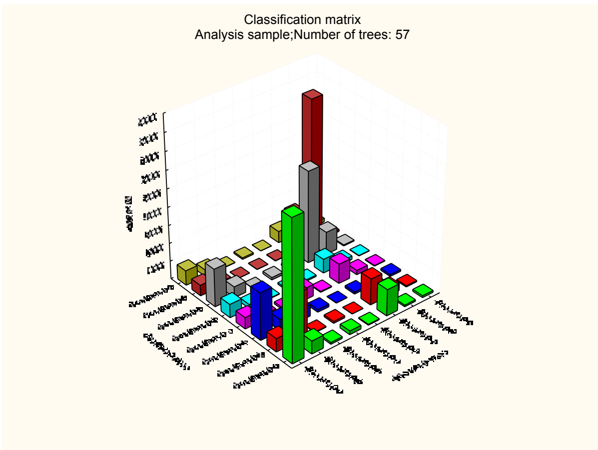
Table 1 - Risk estimate of training and test data by maximum number of nodes

	Training data		Test data	
Max number of nodes	Risk estimate	Standard error	Risk estimate	Standard error
3	0.4868	0.00226	0.5110	0.00226
5	0.4860	0.00226	0.4936	0.00433
7	0.4839	0.00226	0.5153	0.00433
9	0.5131	0.00256	0.5489	0.00432

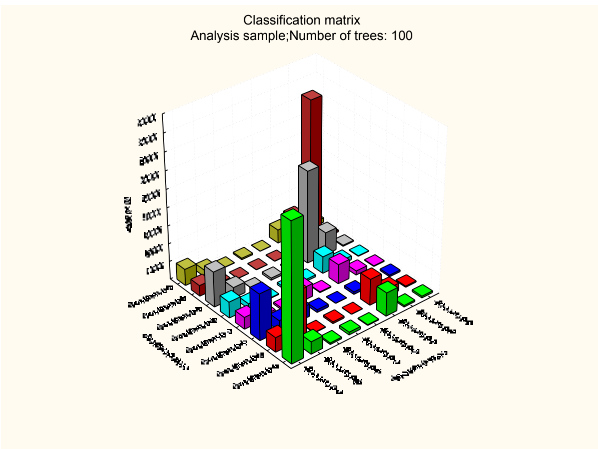
These results indicate that a limited representation of heterogeneity of decision styles (3 nodes) may not sufficiently capture the variation in the choice data, while too much heterogeneity (9 nodes), depicting many different decision styles, may tend to over-represent the data. Thus, given the other non-varied parameters settings, in this study, a moderate complexity, reflected in a maximum of 5 nodes, seems to provide the best performance in terms of robustly predicting transport mode choices in the test data. The optimal number of decision trees in this case is 57. In this context, it is important to also examine the goodness-of-fit of the ensemble of decision trees.



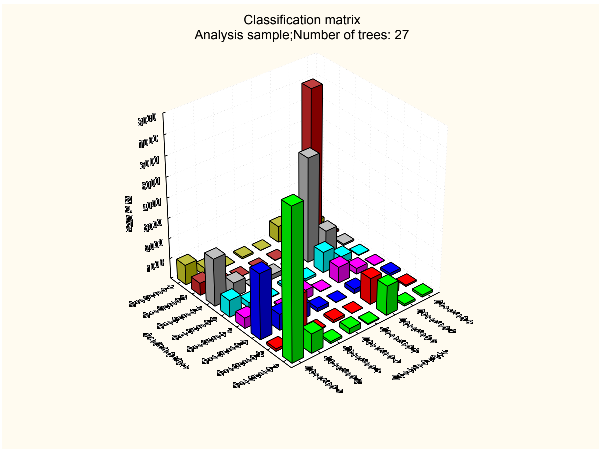
Number of nodes= 3



Number of nodes= 5



Number of nodes= 7



Number of nodes= 9

Figure 3 - Confusion matrices

Goodness-of-fit

Table 1 gives on overview of the goodness-of-fit of the ensembles of decision trees, measured in terms of the concept of risk estimate as a function of the maximum number of nodes. Several conclusions may be drawn from the findings summarized in Table 1. First, risk estimates vary around 48 per cent for the training data, except for the setting when the number of child nodes is equal to 9 (for test data). Differences, however, are small and non-significant at conventional levels.

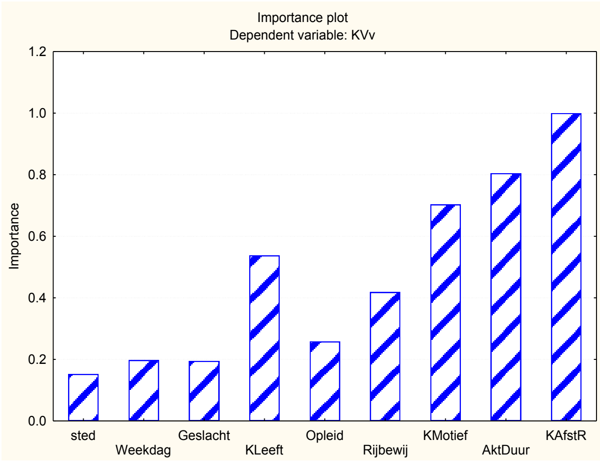
Second, for all different settings for the maximum number of child nodes, Table 1 indicates that the goodness-of-fit is better for the training data than for the test data. One would expect this kind of result as the ensemble of decision trees is derived from the training data. The table, however, also shows that the drop in goodness-of-fit between the training data and the test data is small. Interestingly, the drop in goodness-of-fit is smallest for the setting of a maximum of 5 child nodes and highest if this number is 7 or 9.

In addition to examining overall goodness-of-fit in terms of risk estimates, it is worthwhile to examine the confusion matrices for the ensembles of decision trees of varying complexity. A confusion matrix captures the distribution of misclassified observations across the wrong choice alternatives. Figure 3 then shows that in general the ensembles of decision trees can reasonably well predict observed transport modes. The heights of the columns along the main diagonal are tallest. The general shape of this figure is similar, but some differences do exist as a function of the maximum number of nodes. In particular, the graphs on the top right side of Figure 3 shows that walking is predicted better if the maximum number of nodes is 5.

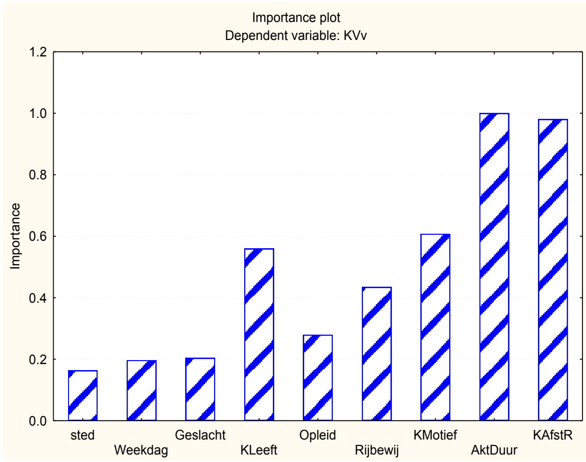
Predictor importance

In addition to the overall goodness-of-fit and misclassification of observations of transport mode choice, it is relevant to compare the relative importance of the explanatory variables in the classification of observations. Predictor importance is calculated as the relative scaled average value of the predictors across all decision tables and nodes.

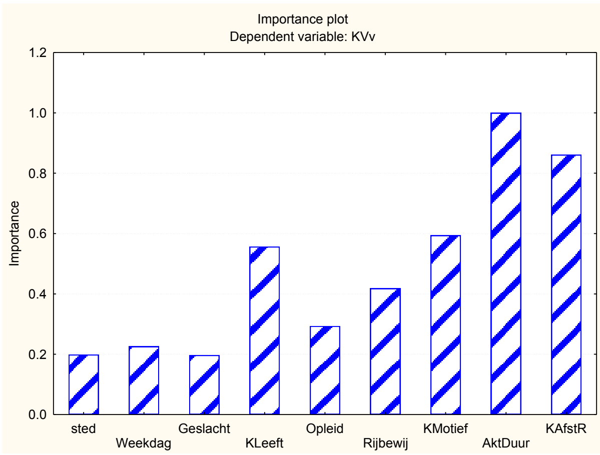
Figure 4, portraying the results for the different settings of the maximum number of nodes, illustrates that the pattern of relative importance is different for the case of a maximum of 3 nodes. In particular, the importance of activity duration is less in this case. Overall, however, transport mode choices in the current sample are mostly influenced by activity duration and travel distance, followed by travel purpose. The selected socio-demographic variables play a lesser role. Least influence is exerted by the urban density variable. Thus, it seems that urban density does not play a critical role in explaining transport mode choice decisions. This finding is in line with the finding (Rasouli and Timmermans, 2013) that the possessing of different transport modes in the Netherlands also does not show much variance across urban density categories.



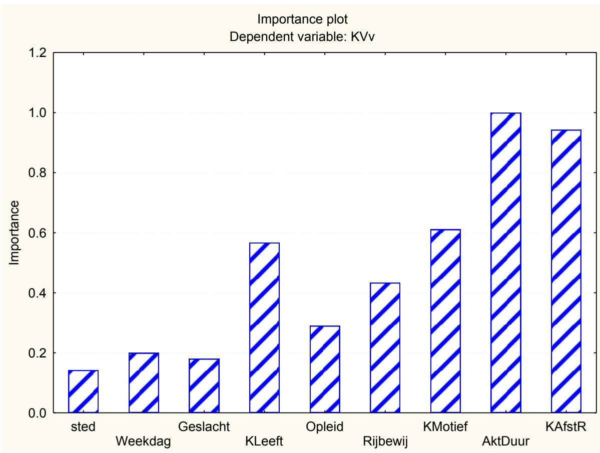
Number of nodes= 3



Number of nodes= 5



Number of nodes= 7



Number of nodes= 9

Figure 4 – Attribute importance

CONCLUSIONS AND DISCUSSION

In contributing to the scant literature on behavioral mixing, the aim of the present study has been to gain experience in allowing for different behavioural decision styles in predicting transport mode choice decisions, using the 2009 Dutch National Travel Survey. Technically, this exploration was based on boosted decision trees. The concept of boosting is based on the notion that an ensemble of relatively simple decision trees is used to predict or classify the dependent variable of interest. The approach involves extracting a sequence of decision trees, where each next tree is based on the prediction residuals of the preceding tree. Because each decision tree can be interpreting as a formalism of a decision style (which choice is made in a particular decision context by an individual of a particular socio-demographic profile), the results can be interpreted as reflecting different decision styles.

Operationally, the results of the decision tree induction may depend on decisions with respect to the complexity of the decision tree. To examine these possible effects, in the present paper, the maximum number of nodes was varied in the analyses. It should be emphasized that the maximum number of nodes is only one operational decision. The effects of other parameters were beyond the scope of the present analyses, but could be investigated in future research.

Results show the overall goodness-of-fit does not depend very strongly on the allowed maximum number of nodes. Overall, the solution with a maximum number of 5 nodes seems to provide the best results for this data set. Some evidence however was found for differences in the variability of the goodness-of-fit for increasing number of boosted decision trees. This finding suggests that further uncertainty analysis on the approach might be very beneficial.

The main implication of this result for policy development is that the application of ensembles of decision trees which depict different decision styles that may be hidden in the data is beneficial, at least if the focus concerns the accuracy of the prediction. The use of such ensembles does come, however, with the disadvantage that it takes more time to interpret all decision tables in a systematic and coherent manner. The development of more direct approaches to capture behavioral heterogeneity in decision-tree formalisms of choice behavior should therefore remain high on the research agenda.

REFERENCES

Arentze, T. A. and H. J. P. Timmermans (2004). A learning-based transportation oriented simulation system, Trans. Res. B, 38, 613-633

Brownstone, D. and K. Train (1998). Forecasting new product penetration with flexible substitution patterns, J. of Econ., 89, 109-129.

Brownstone, D., D. S. Bunch and K. E. Train (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles, Trans. Res. B, 34, 315-338.

Greene, W. H. and D. A. Hensher (2003). A latent class model for discrete choice analysis: contrasts with mixed logit, Trans. Res. B, 681-698.

Greene, W. H. and D. A. Hensher (2007). Heteroscedastic control for random coefficients and error components in mixed logit, Transp. Res. E, 43, 610-623.

Greene, W. H., D. A. Hensher and J. M. Rose (2006). Accounting for heterogeneity in the variance of unobserved effects in mixed logit models, Trans. Res. B, 40, 75-92.

Hensher, D. A., J. M. Rose and W. H. Greene (2008). Combining RP and SP data: biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects, J. of Trans. Geo., 16, 126-133.

Hess, S. and K. E. Train (2011). Recovery of inter- and intra-personal heterogeneity using mixed logit models, Trans. Res. B, 45, 973-990.

Janssens, D., G. Wets, T. Brijs, K. Vanhoof, T. A. Arentze and H. J. P. Timmermans (2000). Integrating Bayesian networks and decision trees in a sequential rule-based transportation model, Eur. J. of Op. Res., 175, 16-34.

Rasouli, S. and H. J. P. Timmermans (2013). Urban setting and transportation modes repositories: A multi-level mixture model specification. Proc. 92nd Annual Meeting of the Transportation Research Board, Washington D.C., USA (On-line: 8 pp).

Wen, Ch.-H., W.-Ch. Wang and Ch. Fu (2012). Latent class nested logit model for analyzing high-speed rail access mode choice, Trans. Res. E, 48, 545-554.

Wets, G., K. Vanhoof, T. A. Arentze and H. J. P. Timmermans (2000). Identifying decision structures underlying activity patterns: An exploration of data mining algorithms, Trans. Res. Rec., 1718, 1-9.