Travel mode choice prediction: An empirical comparison of four models

Hichem Omrani, Omar charif, Xinhua Zhu, Anjali Awasthi, Philippe Gerber
GEODE dept. - CEPS/INSTEAD, 3 avenue de la Fonte L-4364,
Esch-sur-Alzette, Luxembourg

Abstract: Travel mode prediction of individuals is important for planning new transportation projects. In this paper, we apply four machine learning methods, namely, artificial neural net–MLP, artificial neural net–RBF, multinomial logistic regression and support vector machines, for predicting travel mode of individuals in Luxembourg City. The presented methods use individuals' characteristics, transport mode specifications and demographic and transportation network data related to places of work and residence. The dataset analyzed comes from a national survey. It contains information on the daily mobility (e.g., from home to work) of individuals who either live or work in Luxembourg. We compare the rates of successful prediction obtained by the four methods for the travel mode using cross-validation. The results show that artificial neural networks perform better than other alternatives. Our analysis can be used to support management decision-making and build predictions under uncertainty related to changes in people's behavior, economic context or environment and transportation infrastructure.

Keywords*:* Artificial neural networks, data mining, data split, logistic regression, machine learning, multinomial, support vector machine, validation.

1. Introduction

The globalization of the economy and the development of transport and telecommunication technologies have led to an increasing concentration of knowledge-intensive employment and global firms in metropolitan regions, such as Luxembourg City. Between 1985 and 2007, the labor force employed in Luxembourg has more than doubled, from 141700 (among 125600 residents of Luxembourg and 16100 cross-borders workers who commute daily across the borders) to 316500 workers (180250+136250, source: STATEC, Luxembourg). A consequence of this is a dramatic increase in road traffic and congestion in peak hours. Omrani et al. (2010) explored the spatial and temporal patterns of commuting to work in Luxembourg. The present study follows up on this by studying also the mode of transport, which is relevant to improve the sustainability of the Luxembourg transport infrastructure, and to further understand the travel behaviour of individuals and households.

We consider the following travel modes: private car, public transport (bus or train) and soft mode (walking or cycling). By modal split we mean the composition (percentages) of commuters who use each of these travel modes. We study and estimate the daily modal split for workers who reside in Luxembourg using several modelling techniques, and considering categories of social, economic, demographic characteristics, location (places of work and residence) and related variables, such as cost and availability and abundance of public transport. Prediction of the travel mode is a pattern recognition problem (supervised learning), in which several variables (e.g., human characteristics and geographical patterns) are used to explain the choices among the travel modes. We assess models and estimation procedures by the quality of their prediction.

Recently, new artificial intelligence models have been applied to predict individual travel mode. They have been introduced as alternatives to complex behaviour modelling and pattern recognition. The standard way of assessing the quality of prediction is by splitting the sample S into a learning and a testing dataset, denoted by L and T ($S = L \cup T$, $L \cap T = \emptyset$).

The model is fitted to L and its performance is evaluated by comparing the fit to the observed values on T. In cross-validation, the sample is split into K subsamples (folds), and a random subset of these subsamples forms L and the remainder forms T. Several splits of S to L and T are drawn at random, and prediction is evaluated on T. We denote this method by L.T (K, R), where K is the number of folds and R the number of replications.

The standard approach is L.T (2, 1); L.T (1, 1) corresponds to learning and assessment on the entire dataset, without splitting it. Zhang and Xie (2008) demonstrated that support vector machine (SVM) outperforms multinomial logit (MNL) in terms of prediction and generalization. They claim that the multilayer feed-forward neural network model, an adaptation of artificial neural network (ANN), is superior to MNL and SVM for fitting (L) but inferior for testing (T). Of course, the concern arises that this conclusion is specific to the dataset or its context. We assess this approach further by applying it to predicting travel mode. The ANN model is known for its high quality of prediction (Yamamoto et al. 2002; Li et al. 2008; Xianyu et al. 2008; Wang and Elhag 2007; Yang et al. 1993); we highlight its strengths for travel mode over the established models. We assess the performance of the method by cross-validation. Unlike several case studies, in which $L = T = S$ or $L \cap T = \emptyset$ with $R = 1$, we use many replications ($R = 100$) to redeem the effect of random sampling on the results. The results show that ANN predicts individual travel mode better on average than the alternatives.

The paper is organized in five sections. In the next section, we review the recent methodological developments in modeling travel mode, with an emphasis on the relevance and the motivation of ANN. Section 3 describes how travel mode is predicted by ANN using several explanatory variables. Section 4 presents the application and compares its results with several alternatives. The concluding section summarises the results and discusses a planned implementation of ANN in a user-friendly package implemented in R programming language (RDevelopment, 2009).

2. Related work

The prediction of the transport mode used by individuals has attracted much attention in recent years. Researchers used many methods to model the individual transport mode decision. These methods mainly belong to discrete choice models family. . Recently, several data mining methods (e.g. SVM, ANN, Decsion Tree) were successfully applied to solve this kind of problem. But it is not clear from the literature which model is the most appropriate for prediction. Xie et al. (2003) showed that the ANN is more robust for travel mode prediction than decision tree (DT) and multinomial logit (MNL) models. Zhang and Xie (2008) also demonstrated that SVM and ANN models outperform MNL model. These methods are briefly described in section 3 as we compare these four methods in this paper.

The previously mentionned models differ from methodological point of view. They have been used for pattern recognition problems and especially for classification. All have common characteristics related to model identification, definition of decision boundary etc.

Some models are more difficult to interpret and understand (e.g. SVM and ANN) than others (e.g. DT and MNL). All models use an optimisation routine to estimate the values of parameters. Some models (DT, MNL) are more flexible than others. SVM and ANN require a parameterisation (e.g. kernel function such as linear, quadratic, cubic or sigmoid, initial configuration, number of hidden layers, and number of units in them). The models entail the same problems with regard to convergence, complexity (NP-complete), stability, overfitting, and generalisation. To justify the model, a lot of papers dealing with the prediction tasks provide a quantitative comparison to several other methods (using cross-validation technique). In fact, some recent papers present an empirical study of travel mode analysis and they show that either ANN or SVM model lead to better results than logit or nested logit model (Hensher and Ton 2000). This practice of comparison is not very forceful and such comparisons are not sufficient. There is a need to supplement the theoretical evidence by empirical assessment. A profond study is needed to understand the reasons why one method is superior to another. Thus the main question is why a given model provides better prediction quality than other. Will a given model still be better if applied to other datasets? There is a weak justification given to answer this research question. We judge that this research question is not well studied and needs additional methodological exploration and further research. The travel mode choice modelling is a nonlinear regression problem which can be tackled using several approaches and methods. Hereafter we describe a set of methods used for predicting individuals travel mode.

3. Methods description

In this paper, we investigate four methods, namely, multinomial logistic regression, neural net-MLP (multi-layer perceptron), neural net-RBF (radial basis function) and support vector machines for travel mode choice predication. These methods are described as follows:

3.1. Multinomial logistic regression

The multinomial logistic regression is used to predict the probability of an outcome with more than two categories. It models the link between a set of independent (explanatory) variables and a categorical dependent variable. The outcome probabilities are defined as follows:

$$\begin{aligned} P(Y=1 \mid X) &= \frac{1}{1+\sum_{j=2}^{K}\exp(X\beta_j)} \\ P(Y=j \mid X) &= \frac{\exp(X\beta_j)}{1+\sum_{j=2}^{K}\exp(X\beta_j)} \end{aligned} \tag{1}$$

The parameter vectors βj are estimated by maximizing the likelihood function. Y has k states and X is a vector of independent variables which we assume to be related to the outcome Y.

3.2. Neural net–MLP

We used a conventional MLP with one hidden layer trained using back propagation method by minimizing the mean square error (MSE). The hidden layer is a set of simple nonlinear hidden neurons. According to Heaton (2005), neural networks without a hidden layer will not model any nonlinear function, but with a hidden layer they can approximate/estimate the relationship between the influencing factors and the outcome. This approach is known to provide good estimates of the output given the observed values of input variables. Let y be the output with a value in the set $(1, 2, \ldots, K)$. Its distribution is expressed as a function of the input $x = (x_1, x_2, \ldots, x_q)$ (see Fig. 1) as:

$$P(y_k \mid x) = \Psi\left\{\sum_{j=1}^{p} v_{jk}\Phi\left(\sum_{i=1}^{q} \omega i_j \cdot x_i + \omega_{0j}\right) + v_{0k}\right\}, \quad (2)$$

where the ω_{ij} and v_{jk} are weights assigned to the connections between the input layer and the hidden layer, and between the hidden layer and the output layer, respectively, ω_{0j} and v_{0k} are biases (or threshold values in the activation of a unit). Φ is an activation function, applied to the weighted sum of the output of the preceding layer (in this case, the input layer). Ψ is also an activation function applied, to each output unit, to the weighted sum of the activations of the hidden layer. This expression can be generalized to networks with several hidden layers. The output of the neural net–MLP is contained in (0, 1), but will not be exactly equal to 0 or 1. A suitable choice for Ψ is a function that maps the real axis $(-\infty, +\infty)$ to the interval [0, 1], such as a distribution function. The output of neural net–MLP is given in equation (2. Figure 1 shows the neural net–MLP architecture with an input layer, one hidden layer (with h units) and one output layer.

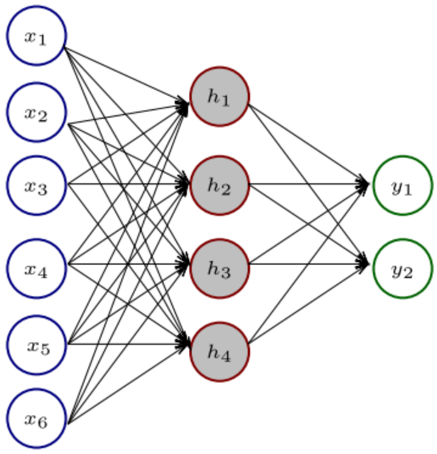

Figure 1: Architecture of neural net topology with 6 input variables, 4 hidden units and 2 outputs.

3.3. Neural net–RBF

A neural net–RBF is a model that has the structure of neural net–MLP (Lowe and Broomhead 1988). Neural net-MLP formulates the problem as a stochastic approximation while neural net–RBF as an interpolation one. A Neural net-RBF model assumes that the predicted value of an input x is likely to be equal to the output of other patterns (training) that have values close to x. A neural net–RBF model is generally composed of one input, one or more hidden and one output layers. The hidden layers contain a set of neurons, each neuron applies a nonlinear function (called radial basis function: RBF). Neurons of the output layer apply a linear function. The output is drived from input by the following equation:

$$y = \sum_{i=1}^{h} \omega_i \times \Phi(x, c_i), \quad (3)$$

where h is the number of hidden neurons, ωi and ci are respectively the optimized weights and the centre of hidden neurons number i and Φ(x, ci) is the radial basis function (defined later by equation 5). The two sets of values ω and c are optimized during the training procedure.

3.4. Support vector machine

Support vector machine (SVM), also called maximum margin classifier, is a supervised machine learning algorithm. It was introduced with all its features by Vapnik and his co-workers (Cortes and Vapnik 1995). This method presents an extension to nonlinear models of Vapnik (1963) for pattern recognition using generalized portrait method. The concept behind the SVM algorithm is to find a hyperplane which seperates two categories (classes) perfectly (e.g., all observations in one class above and all observations in other class below the hyperplane), in such a way to be as far as possible from the nearest members of both classes which are called the support vectors. Thus, SVM tries to find a hyperplane with the largest possible margin (see Fig. 2). This method assumes that new unseen input will often be close to the training patterns. The bigger the margin between the classes, the bigger the chances that the new input will be successfully classified and the higher is the ability to generalize.

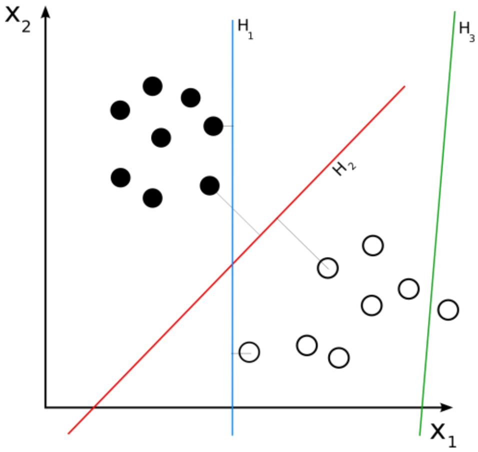

Figure 2: Classification problem using SVM with linear separable case: graphic showing three hyperplanes in two dimensions. H3 doesn't separate the two classes. H1 does, with a small margin and H2 with the maximum margin.

The separating hyperplane is given by: $\omega \bullet \Phi(x) + b = 0$, where ω is perpendicular vector to the hyperplanes ('•' is scalar product), b is a constant such that $b/||\omega||$ is the perpendicular distance from the hyperplane to the origin of the space and $\Phi(X)$ is the projection function that maps x to the hyperplane. This function will be used to define the kernel of SVM model given by: $K(xi, xj) = \Phi(xi) \bullet \Phi(xj)$ (4)

The kernel function is mainly used to project and remap the input vector to a higher dimension space in which this nonlinearly separable problem is transformed into linearly separable one in this new feature space. Thus, in case of a linearly separable problem, a linear kernel is used: $K(xi,xj)=xi \bullet xj$ with $\Phi(x)=x$ in (4). To successfully implement an SVM, it is enough to optimize ω and b to maximize the margin. The margin is given by: $2/||\omega||$. Optimisation is done using Lagrange multiplier and quadratic programming. In case of a binomial classification where the output is 1 or −1, each new input X is classified by evaluating $y=sgm(\omega \bullet \Phi(x)+b)$, where *sgm* is the sigmoid function having an *S* shape. Often, sigmoid function refers to the special case of the logistic function defined as: $sgm(x) = 1/1 + exp(-x)$.

In this paper, we are dealing with a nonlinearly separable problem, so we need a nonlinear kernel function and we choose the radial basis function (RBF), a real function that depends only on the norm of its arguments. The kernel we use is as:

$$k(x_i, x_j) = \exp\left\{ -\left(\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \right\}, \tag{5}$$

where x_i is the vector of variables *i*, σ is called the width of the radial base function and $||x_i - x_j||^2 = {}^t(x_i - x_j)\cdot(x_i - x_j)$ where ${}^t(x_i - x_j)$ is the transpose vector of $(x_i - x_j)$.

4. Model validation and assessment

4.1. Which criterion to use for judging model performance?

Criteria commonly applied in the literature for evaluating the performance of a model include the sum of square error (SSE), mean absolute error (MAE), mean absolute per- centage error (MAPE), root mean square percentage error (rMSPE), correlation co- efficient (R), Theil's inequality coefficient (U), sensitivity (overall prediction accuracy or percentage of cases correctly predicted), specificity, area under ROC curve (AUC), F-measure, and so on. The mean square error (MSE) is another good measure of performance because many methods minimize MSE and therefore, it is a standard measurement for performance of predicting methods. . Some of these criteria (i.e. sensitivity, specificity, AUC, F-measure) are based on the confusion matrixalso called classification table or tabulation of observed vs. predicted. . In the confusion matrix for trinomial outcomes, the element incell (k, h), with k≤3 and h≤3, is the number of individuals whose outcome is k and prediction is h. The accuracy rates for the outcomes are obtained as the within-row percentages of the diagonal elements. The conventional confusion matrix is obtained by replacing each triplet with the indicator of the highest probability. For example, (0.2, 0.5, 0.3) is converted to (0, 1, 0), just like (0.1, 0.85, 0.05) which indicates much less uncertainty about the prediction. Any criterion that ignores this difference is problematic. Other criteria (e.g. SSE, MAE, MAPE, rMSPE, R, U and MSE) are based on the set of predicted probabilities

In this paper, we use the average probability of correct assessment (APCA) for model performance. The APCA (also called the success rate) is defined as the mean of the fitted probabilities for the observed class. For example, if *p*= (0.6, 0.3, 0.1) and the observed class is 2, then this observation contributes to the success rate by 0.3. For a given dataset, the model with the highest APCA is considered to be the best. The APCA is also used for selecting an optimal split to construct learning and testing datasets. The APCA is defined as follows: $APCA = p = (p_1 + p_2 + p_3)/N_T$,

where $p_k = \sum_{i=1}^{N_T} P(\hat{y}_{ik} = k | X_i, y_{ik} = k)$; y_{ik} and $\hat{y}_{ik}$ are respectively observed and predicted values for individual *i* and mode *k*. p_k is the sum of the probabilities of correct assessment of class k, and the subscript T denotes restriction to the test dataset T.

5. Application

5.1. Descriptive analysis: daily mobility in Luxembourg and mode choices

In all analyses, we use data from the PSELL survey (Socio-Economic Panel survey Liewen zu Lëtzebuerg). This Survey was launched in 2003 with a representative sample of the resident population in Luxembourg. The sample size of the survey is around 3670 households (9500 individuals), which allows estimation of social, demographic and economic indicators for the whole population. The survey is carried out annually by CEPS/INSTEAD research center in collaboration with the Luxembourg Statistical Office (STATEC). It forms the Luxembourg's contribution to the European Union Statistics on Income and Living Conditions (EU–SILC). According to the PSELL (2007), 77% of workers residing in Luxembourg used their private cars for commuting to work (see Table 2). The proportion of people who use the public transport (PT) alone or combined with other modes was only 16%. Within public transport, bus was the dominant mode (80% of the commuters). Among soft modes, 9% of the workers used the bicycle (corresponding to about 0.5% of all cases). Men were more likely to use private cars and less likely to use public transport than women.

Table 2: Characteristics of sample (PSELL survey) and modal split (%)

	Car	PT	Soft mode
Luxembourg	83	11	6
European (EU)	70	24	6
Other EU	78	15	7
Non EU	59	33	8
Total	77	16	7

Source: PSELL, 2007, CEPS/INSTEAD, STATEC.

The differences among the age groups and categories of educational level were much smaller. Purchase and maintenance of a private car is expensive for those with lower income, who tend to be in lower social-economic categories. In the bottom quartile of households according to equivalised income, only 61% of workers commute by private car, compared to 80% in the top quartile. Moreover, approximately 43% of households had at least two cars (source: PSELL, 2007). Another factor in the choice is availability (and abundance) of public transport. In (rural) areas not reached by public transport, private car is the only option for those who have a longer commute, especially when the car is essential also for other activities (such as shopping, entertainment, school). However, the choice of travel mode should not be regarded in total separation from the choice of location for residence and the associated life style. The private car has become more important with the peri-urbanisation, despite expansion of public transport. Simply, bus lines cannot cater for all the dispersed locations of those commuting to places of work, which themselves have become more dispersed.
The dataset extracted from the PSELL database is composed of 3670 observations and 15 relevant variables; details are given in Table 3. Only observations with complete records are included in the analysis (i.e. 3670 individuals out of 4138, thus nearly 11% of missing values). The outcome variable is the travel mode; it has three categories, private car, public transport (PT) and soft mode (Figure 3). We combine walking and cycling mode because their frequency is only 7%; see Table 2. The input variables are classified to the following groups:

i) C: cost
ii) D: income, age, gender, nationality, type of household, education

iii) A: car ownership, number of bus stops and train stations in the municipality of residence
iv) G: region of residence and type of work area.

Table 3. Explanatory variables and their description

	Group	Variables	Number of modalities	Description
1	D	Age of household	1	Quantitative
2	D	Standard of living	1	Quantitative
3	D	Household type	4	Single without children, couple without children, single v children, couple with children
4	D	Education	3	Primary/high school, higher non university degree, university
5	D	Nationality	4	Luxembourg, Portuguese, Other EU, non-EU
6	D	Gender	1	Binary (male, female)
7	A	Driving license	1	Binary (yes, no)
8	A	Number of train stations (in the municipality of residence)	1	Quantitative
9	A	Number of bus stations (in the municipality of residence)	1	Quantitative
10	A	Number of cars in the household	3	Without car, 1 car, 2 and more than 2 cars
11	G	Region of workplace	7	Centre north, centre south (without Luxembourg city), east region, north region, west region, south region, Luxembourg city
12	G	Topology of residence	6	Dense city, first-ring suburb, second-ring suburb, distant peri-urban, mining area, rural.
13	C	Travel cost by car	1	Quantitative
14	C	Travel cost by PT	1	Quantitative
15	C	Travel cost by walk	1	Quantitative

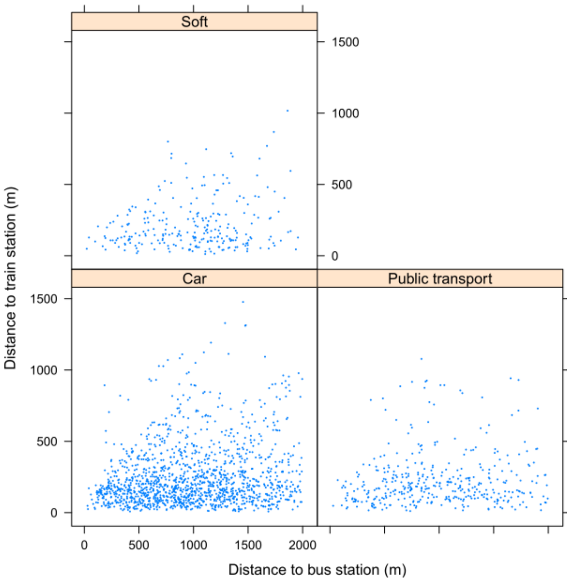

Figure 3: Travel mode according to distance to the closest bus station and distance to the closest train station at residence place.

5.2. Results

The observed composition of the modes is (77, 16, 7)% as presented in Table 2. For illustration, we give details of the confusion matrix for MLP applied to one replication (Table 5). To avoid vagaries, each model was run hundred times in each configuration. In Figure 4, we present the confusion matrix obtained from MLP after one hundred replications with a split of (60:40) of the overall sample S to subsamples L (60%) and T (40%). All computations and graphics have been obtained using R programming language (www.r-project.org). After estimating coefficients values, the predicted probabilities for car, public transport (PT) and soft modes are generated for each individual (as shown in Figure 7).

Table 5: Confusion matrix from MLP model

Observed mode	Predicted mode		
	Car	PT	Soft mode
Car	1 288	250	142
PT	277	72	31
Soft mode (walk, bike)	92	34	16

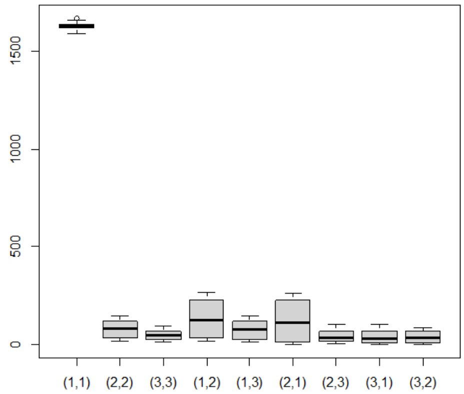

Figure 4: Boxplot of confusion matrices from MLP after 100 replications.

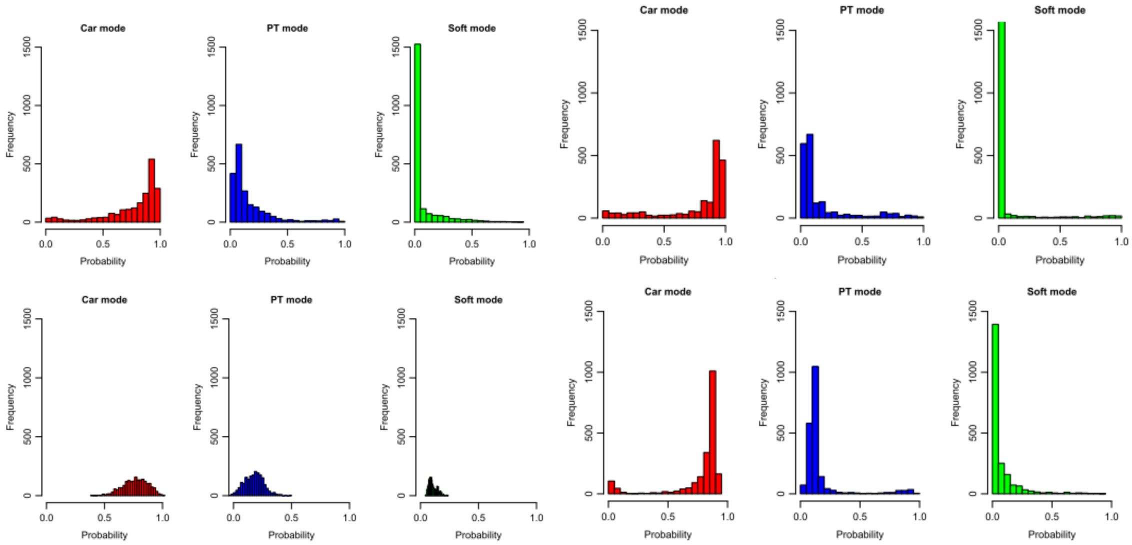

Figure 7: Histogram of predicted travel mode from MNL(top-left), neural net–MLP (top-right), neural net–RBF (bottom-left) and SVM models (bottom-right).

Predictions - ternary plot

The prediction for a case (individual i) is a triplet of probabilities, denoted by *pi=(pi1, pi2, pi3)*, one for each mode (component k). The estimated probabilities of belonging to the components (*C1*: Car, *C2*: PT and *C3*: Soft) are graphically summarised by a ternary (composition) plot (Aitchison 1982), as shown in Figure 8. Each individual is represented by a point. Vertices *C1*, *C2* and *C3* correspond to certainty that the individual belongs to the respective component 1 (Car), 2 (PT) and 3 (Soft). Proximity to the vertices reflects the probability p*ik* that the individual belongs to the corresponding component Ck. The graph confirms that most individuals can be allocated to a component with a high degree of certainty. Some points are on or near the two sides of the triangle, *C1−C2*, *C1−C3* or *C2−C3* . For these individuals, either components 1, 2 or 3 can be ruled out.

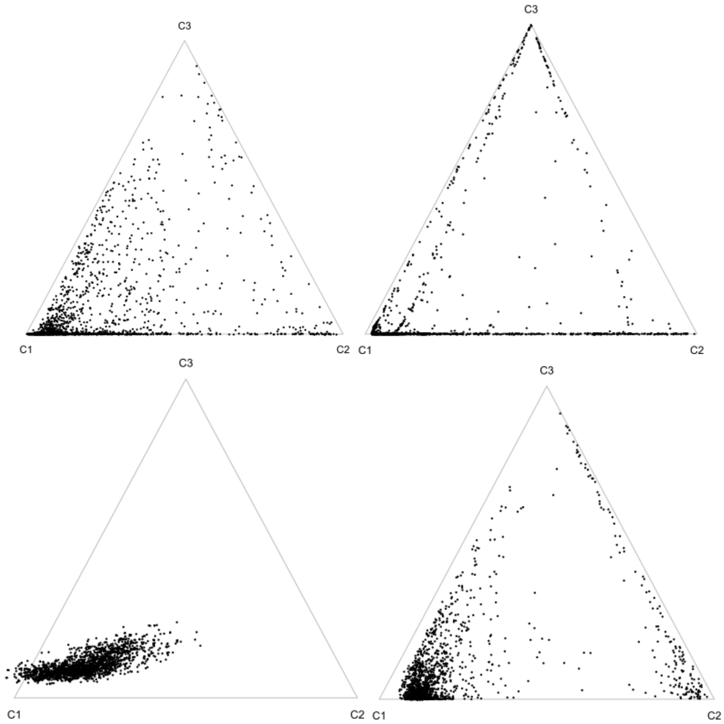

Figure 8: The ternary plot of the estimated probabilities from MNL, neural net–MLP, neural net–RBF and SVM, of belonging to the travel mode; three modes: Car(C1), PT(C2) and Soft(C3).

The results from different methods are shown in Table 6 (using the optimal split of 0.45). Note that no method is significantly better than neural net–MLP (only differences of more than 4% are significant at the level of 5% (Ripley 1994)).

Table 6: Results comparison: Means of hundred runs (with ± standard deviations) of the average probability of correct assessment (APCA), multiplied by 100.

	Overall	(Car)	(PT)	(Soft mode)
MNL	60.36	78.29	22.37	6.92
	±0.20	±0.83	±0.91	±0.47
Neural net–MLP	67.00	79.96	23.01	26.49
	±1.37	±2.16	±1.78	±0.57
Neural net–RBF	63.30	77.09	20.59	10.46

	±0.73	±0.72	±2.42	±0.34
SVM	61.08	76.22	9.16	15.20
	±0.35	±0.55	±2.25	±2.23

Notes: The methods are multinomial logit (MNL), multi-layer perceptron network (neural net–MLP), radial basis function network (neural–RBF) and support vector machine (SVM).

The results shows , in the context of Luxembourg, the following variables age, wage, public transport cost and walk cost can increase the probability to take the car. However, when the number of bus and train stations increase, the probability to take the car decreases. This is evident and that is way several cities control the quality of the transport service to decrease the car use and increase the use of public transport to foster sustainable mobility.

6. Conclusions and further research

In this paper, we apply four machine learning methods namely multinomial logistic regression, neural net–MLP, neural net–RBF and support vector machines for travel mode prediction of individuals in Luxembourg. A large set of variables extracted from PSELL data base related to transport cost, work and residence places, demography, individual characteristics (gender, nationality, education level, driving license, family status etc.) were used to model the mobility of workers in Luxembourg. The programming of the four models was done using R. By applying cross-validation technique and the efficiency criterion APCA (average probability of correct assessment), we find that neural net models outperform SVM and MNL modes in terms of prediction accuracy. The next step of our work involves using the results obtained from the neural net models for generating behavioural rules for modelling the transport behaviour of individuals.

Acknowledgements

The authors would like to thank anonymous referees for their helpful comments and suggestions to improve the presentation of this paper. This research work has been carried out under the framework of MOEBIUS project with the support of the National Research Fund of Luxembourg (project C09/SR/07).

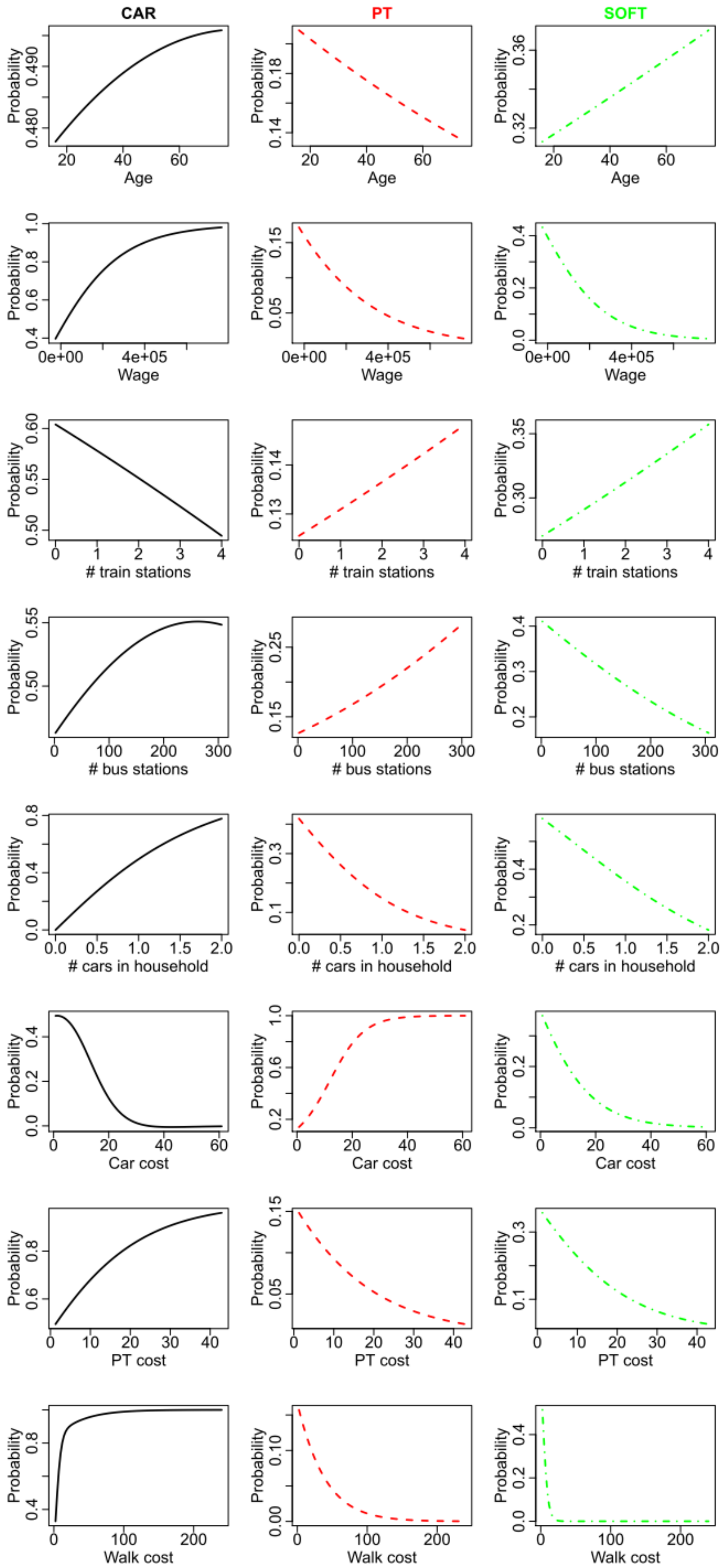

Figure 10: Impact or partial effects of significant variables on travel mode choice: dark color (solid line) for CAR mode, red color (dashed line) for PT and green color (dot-dash line) for SOFT mode.

References

Aitchison, J. (1982). The statistical analysis of compositional data. Journal of the Royal Statistical Society. Series B (Methodological), pages 139–177.

Cohen, S. and Intrator, N. (2004). On different model selection criteria in a forward and backward regression hybrid network. International Journal of Pattern Recognition and Artificial Intelligence, 18(5):847–866.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3):273–297. Heaton, J. (2005). Introduction to neural networks with Java. Heaton Research Inc.

Hensher, D. and Ton, T. (2000). A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. Transportation Research Part E: Logistics and Transportation Review, 36(3):155–172.

Li, X., Lord, D., Zhang, Y., and Xie, Y. (2008). Predicting motor vehicle crashes using support vector machine models. Accident Analysis & Prevention, 40(4):1611–1618.

Lowe, D. and Broomhead, D. (1988). Multivariable functional interpolation and adaptive networks. Complex systems, 2:321–355.

Omrani, H., Charif, O., Klein, O., Gerber, P., and Trigano, P. (2010). An approach for spatial and tem- poral data analysis: Application for mobility modeling of workers in luxembourg and its bordering areas. In Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, pages 1437–1442. IEEE.

Omrani, H., Charif, O., Gerber P., Awasthi A. Prediction of individual travel mode using evidential neural network model, Transportation Research Board, Washington, January 13-17, 13-0284, 2013.

RDevelopment, C. (2009). Team, 2006: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www. R-project. org, ISBN, pages 3–900051.

Ripley, B. (1994). Flexible non-linear approaches to classification. Nato Asi Series F Computer and Systems Sciences, 136:105–105.

Samandar, A. (2010). Ranking water quality variables using feature selection algorithms to improve generalization capability of artificial neural networks. Scientific Research and Essays, 5(11):1254–1259.

Souche, S. (2010). Measuring the structural determinants of urban travel demand. Transport Policy, 17(3):127–134.

Vapnik, V. (1963). Pattern recognition using generalized portrait method. Automation and Remote Control, 24:774–780.

Wang, Y. and Elhag, T. (2007). A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks. Expert Systems with Applications, 32(2):336–348.

Weigend, A., Huberman, B., and Rumelhart, D. (1990). Predicting the future: A connectionist approach. International Journal of Neural Systems, 1(3):193–209.

Xianyu, J., Juan, Z., and Gao, L. (2008). Knowledge-based neural networks and its application in discrete choice analysis. In Networked Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on, volume 1, pages 491–496. IEEE.

Xie, C., Lu, J., and Parkany, E. (2003). Work travel mode choice modeling with data mining: Deci- sion trees and neural networks. Transportation Research Record: Journal

of the Transportation Research Board, 1854(-1):50–61.
Yamamoto, T., Kitamura, R., and Fujii, J. (2002). Drivers' route choice behavior: Analysis by data min- ing algorithms. Transportation Research Record: Journal of the Transportation Research Board, 1807(-1):59–66.
Yang, H., Kitamura, R., Jovanis, P., Vaughn, K., and Abdel-Aty, M. (1993). Exploration of route choice behavior with advanced traveler information using neural network concepts. Transportation, 20(2):199–223.
Zhang, Y. and Xie, Y. (2008). Travel mode choice modeling with support vector machines. Transportation Research Record: Journal of the Transportation Research Board, 2076(-1):141–150.