ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 2, Special Issue 1, December 2013

Proceedings of International Conference on Energy and Environment-2013 (ICEE 2013)

On 12th to 14th December Organized by

Department of Civil Engineering and Mechanical Engineering of Rajiv Gandhi Institute of Technology, Kottayam, Kerala, India

Mode Choice Modelling based on Work Trips – Artificial Neural Network Model

Rajalakshmi P

Affiliation Alappuzha, Kerala, 688004 India

ABSTRACT
The choice of transport mode is one of the most important models in transport planning. This is due to the key role played by public transport, walking and cycling in transport policy. Public transport makes more efficient use of road space compared to cars. The issue of mode choice is therefore one of the most important in transport modeling. It impacts on the efficiency to which people travel in urban areas, the amount of space devoted to transport and if a range of transport are available to people. It is important then to develop and use models which are sensitive to the attributes of travel that influence individual choices of modes. The factors that influence mode choice include characteristics of the trip maker, type of journey and characteristics of the transport facility.

An individual's intrinsic mode preference and responsiveness to level-of-service variables affects her or his travel mode choice for a trip. The current paper formulates a Artificial Neural Network model of travel mode choice that accommodates variations in mode preferences and responsiveness to level-of-service due to both observed and unobserved individual characteristics. The model was applied to examine sub urban work travel mode choice from a sample of workers from the Kottayam District.

1. INTRODUCTION

Urbanization has been one of the dominant contemporary processes as a growing share of the global population lives in cities. Considering this trend, urban transportation issues are of foremost importance to support the passengers and freight mobility requirements of large urban agglomerations. Transportation in urban areas is highly complex because of the various modes involved, the increase of origins and destinations, and the amount and variety of traffic.

The choice of a transport mode is probably one of the most important classic models in transportation planning. Transport modelling is used as an effective tool to manage sustainable development in most of the developed countries. Considerable investments have been made in transport planning and policy making in order to observe the travel behaviour and forecast the future demand of travel. This forecasting needs to incorporate the designing of transport systems, by making use of global infrastructure and understanding the travel behaviour of the residents of the study area, and develop a system that can accommodate the travel demands for the future.

The model specification developed for the study, for various trip lengths and trip purposes, considered all the commonly used travelling modes in the study area. Several level- of-service attributes of the modes and household parameters that were assumed to influence the travel behaviour of the targeted population, were tested in order to generate approximate model specification for each trip purpose

2. NEED OF STUDY

The history of transportation planning is quite long. For the past few decades, transportation infrastructure developments were justified by transportation planning models. But, still the problem of congestion is widely present in city areas and its outskirts. These are giving some evidences that, the conventional system of transportation planning has some drawbacks. The congestion may be due to the rapid urbanization, increased migration to cities from rural areas, and very high increase in population growth rate. These changes have to be incorporated in the conventional transportation planning procedure to rectify its drawbacks. This necessitates planning of transportation facilities in medium sized cities in developing countries like India, followed by the four stage transportation planning procedure, which include trip generation, trip distribution, mode split and traffic assignment stages. The changes in land use should also be considered in this procedure which may otherwise cause the total failure of the transportation system.

3. OBJECTIVES OF STUDY

- To study the travel behaviour and mode choice behaviour of commuters of different socio economic groups of a sub-urban town.
- To develop suitable mode choice models that reflects the current travel behaviour and mode choice behaviour of travellers for work journey.

4. LITERATURE REVIEW

Transportation modelling plays an important role in supporting transportation planning. One of the major roles of transportation modelling is to forecast travel demand based on changes in the transportation system. There are different types of models that have been developed to simulate actual travel patterns of people and existing demand conditions. The models are used to predict changes in travel and utilization of the transportation system in response to changes in land-use, demographics and socioeconomic conditions. The theoretical foundation of the logit approach is based on the theory of utility maximization incorporating a random factor for the representation of stochastic choice or decision elements [1]. A joint model of work mode choice and number of stops during the work commute provides an improved basis to evaluate the effect of alternative policy actions to alleviate peak-period congestion [2].GIS and multivariate regression analysis shows that each of the Five Ds of the built environment are statistically significant determinants of mode share for the journey to work, with the exception of pedestrian oriented design[3]. The basic idea for creating such an expert system was to define relations in modal split on the basis of parameters of transport system demand and supply in one city[4]. As far as mode choice is concerned, the results highlight the importance of contextual and individual factors besides mode characteristics (travel time, cost and comfort). The car ownership decision is found to be mostly related to income levels of the households, contextual constraints and location issues. The mode choice modelling to work has the central to the evaluation of the efforts to mitigate traffic congestion [5].

Work trips in any urban area are always at the centre of focus in urban transportation planning and policy analyses. Work trips in aggregation define peak versus off-peak-period traffic flow in the urban transportation network. Although activity based modelling practice extends beyond peak-period travel, commuting activities are always at the centre of all modelling approaches [6].

The importance of mode choice in transportation policy analysis and decision making has led to a variety of methods for predicting the effects of policy measures on travellers' mode choices [7]. Residential choice location is influenced by many variables including socio-economic characteristics, life cycle, location of work and other major activities such as schools, shopping, family and friend, real estate values, and characteristics of the residential and workplace area. Living close to the workplace reduces vehicle kilometre of travel and thus contribute to a more sustainable transportation system. [8].

5.METHODOLOGY

5.1 Study Area

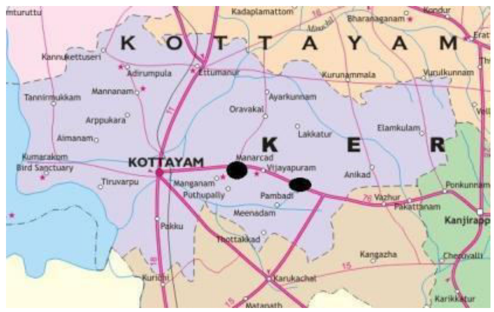

Fig.1 Kottayam District

The selection of the study area plays an important role in finding out the travel behaviour and mode choice behaviour of the commuters. The selected study area was sub-urban regions of Kottayam District. Kottayam is a city in the Indian state of Kerala. It is located in central Kerala and is also the administrative capital of Kottayam district. The areas selected for the study are spotted out in Fig 1 with black marking.

5.2 Data Collection

Questionnaire was designed in order to fit the objectives of this study. The choice sampling procedure was adopted which has been used for transportation, especially for development of mode choice. Samples were drawn at random and various strata of commuters based on sex, income groups and trip length.

The data required for the study was collected by home interview survey. Survey data was collected by directly interviewing the commuters working in various public and private sectors. Surveys were conducted at various sub urban regions of Kottayam district. The study area covered regions of Pampady and Manarcad of Kottayam district. Samples of about 500 were taken for the study avoiding the cases containing missing values from both the study areas. Pilot survey was carried out using 100 samples and analysis and model building was carried out to reveal the deficiencies in the design of the proposed procedure. The available modes of transportation in the study areas included bus, car, taxi, motorcycle, auto rickshaw and bicycle. Bus was the only means of public transport and other modes were grouped together as Personal transport modes. The study aimed at studying the travel and mode choice behaviour of the selected areas. Hence, work trip journeys of the commuters were only considered in the survey.

Choice riders were the only category included in the interview. The commuters were interviewed about their socio economic characteristics and trip details. The trip details included the origin, destination, mode of travel, travel time and travel cost.

The monthly income, age, gender and occupational category of the employee were considered as the case specific regressors. The age group 25- 40 was coded as 0 and 40-55 as 1. The gender was coded as 0 for males and 1 for females. Occupational category included government and private employees. Income also classified employees into two groups-Middle income groups and High income groups.

6.PRELIMINARY ANALYSIS

The data collected from the respondents through the questionnaire survey was fed into the computer and appropriate statistical analysis was carried out. The analysis
included coding and sorting of the input data**.** This sorted data was processed in the form of tables and charts. The sorting and processing of the coded data was done using Microsoft Excel. The important socio-economic characteristics such as age, gender, monthly income, vehicle ownership were analyzed. This data may identify current travel behaviour of employees in Kottayam and also reveals factors affecting the selection of modes.

The preliminary data analyses of the dataset are given in the following tables. Table I and II represents the descriptive statistics of the demographic profile of employees and the details of preference of mode choice for their work trips.

TABLE I DESCRIPTIVE STATISTICS ON THE DEMOGRAPHIC PROFILE OF EMPLOYEES IN KOTTAYAM DISTRICT

Demographic Variable			**Category**	**Frequency**	**Percentage (%)**
Age	0		25-40	332	66.4
	1		40-55	168	33.6
Gender	0		Male	375	75
	1		Female	125	25
Occupation	0		Private	293	58.6
	1		Government	207	41.4
Income Level	0	Medium	15000-30000	337	67.4
	1	High	>300000	163	32.6

The age level of employees was from 25 to 40 and 40 to 55. Of the two groups, the age range between 25to 40 occupied the highest proportion (66.4%), while the age group 40 to 55 occupied about 33.6%. Of the 500 employees, 75% of the total sample was males and the rest 25% was females. The employees were categorized into two categories on the basis of their work – Private and Government sectors. Income was also considered an important factor in employee's preference of mode choice. This included middle income group which constituted 67.4% and high income group with 32.6%.

The various alternatives of modes available were car, bus, two-wheeler, auto and walk. The percentage for auto and walk as mode choice was very small and hence not taken into account. Bus was the only means of public transport considered and the rest of the modes are grouped under personalized mode of transport.

TABLE II PREFERENCE OF MODE CHOICE OF KOTTAYAM

Mode	Frequency	Percent age (%)
Car	107	21.4
Two wheeler	240	48
Bus	153	30.6

The modes were coded for making the analysis and modelling task easy. Cars were given a code of 1 and two wheelers were coded as 2 and bus as 3. From the above table two-wheelers were the prominent mode choice preferred for work trips and contributed about 48% of the total sample size. The two wheelers which constituted more share were considered as the reference category. The reason behind the preference of two-wheelers may be due to the type of work of the commuter, comfort and less travel time. The next preference was for bus which constituted 30.6% followed by car with 21.4%

7.MODEL FORMULATION AND VALIDATION

Several statistical techniques are available for developing mode choice models for work trips. Mode choice modeling can be regarded as a pattern recognition problem in which multiple human behavior patterns reflected from explanatory variables determine the choices between alternatives or classes. The variables that are used for model building includes travel mode, travel time, travel cost, sex, number of working members in the household, annual income, comfort, and safety. The dependent variables in this case are discrete and hence linear regression is not appropriate.

Modeling generally involves:

1) Specification of the model like Logit, Probit etc;
2) Identification of variables;
3) Considering the form in which the variable enter the utility function; and
4) Identifying individual's choice set.

The available data was divided into two parts- testing data and validating data. The amount of data that is to be used for the testing and validating purpose depends on the availability of the data. Any model has to be validated using some data. In general two-third of the full data is used for testing purpose and the remaining data is used for validation. The model predictions are compared with the information not used in the model estimation. The process is randomized for eliminating bias in sample selection for validation

The software used for modeling was SPSS (Statistical Package for the Social Sciences). SPSS

Statistics is a software package used for statistical analysis and is now officially named "IBM SPSS Statistics". SPSS is a comprehensive and flexible statistical analysis and data management solution. SPSS can take data from almost any type of file and use them to generate tabulated reports, charts, and plots of distributions and trends, descriptive statistics, and conduct complex statistical analyses. The software is used for developing models like:

1. Multinomial Logit models
2. Nested Logit models
3. Random parameter logit models
4. Probit models
5. Artificial Neural Network models

8.ARTIFICIAL NEURAL NETWORK MODEL

Neural networks used in predictive applications, such as the multilayer perceptron(MLP) and radial basis function(RBF) networks are supervised in the sense that the model predicted results can be compared against known values of the target variables. The SPSS neural networks option allows fitting MLP and RBF networks and saving the resulting models for scoring. Multilayer perceptron networks are used when parameters are non - linear and radial basis function networks are used if the parameters are linear. RBF gives predictions comparable in accuracy to those from MLP based models. This requires less time for model development since no repetition is required to reach the optimal model parameters. The choice of procedure will be influenced by the type of data and the level of complexity seeked to uncover.

The procedure adopted in this study was Multilayer Perceptron network since the parameters used for modeling are non-linear.

9.MULTILAYER PERCEPTRON NETWORK MODEL

The multilayer perceptron procedure produces a predictive model for one or more dependent (target) variables based on the values of the predictor variables. The dependent variables can be nominal, ordinal or scale. Multilayer perceptron procedure can find more complex relationships. If MLP network allows a second hidden layer, each unit of the second hidden layer is a function of the units in the first hidden layer, and each response is a function of the units in the second hidden layer. In this study two hidden layers for modeling using multilayer perceptron approach. A feed forward architecture of the perceptron model is given in Fig 2.

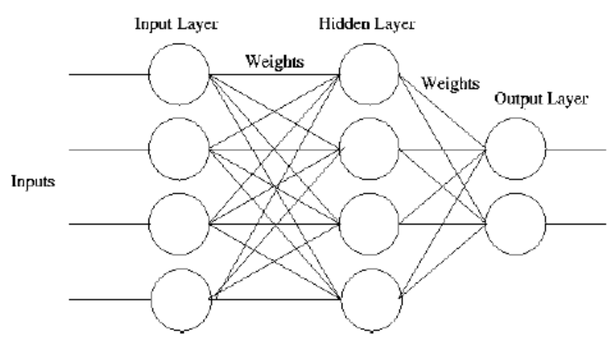

Fig 2 Feed Forward Architecture

10.MODEL OUTPUT USING TESTING SAMPLE

The universal choice set was identified as Car, Bus and Motorised Two-Wheeler. First a minimum acceptable base model was developed. The various covariates considered included the travel time and travel cost. Age, Gender and Income were considered as the factors that depend on the mode choice. The estimated results of the minimum acceptable model are given in the following tables.

11. CASE PROCESSING SUMMARY

TABLE III CASE PROCESSING SUMMARY

		N	Percent
Sample	Training	335	67.0%
	Testing	107	21.4%
	Holdout	58	11.6%
Valid		500	100.0%
Excluded		0	
Total		500	

The case processing summary given in Table III shows that 335 samples were assigned to the training sample, 107 to the testing sample and 58 to the hold out samples. None of the cases were excluded from the analysis.

12.NETWORK INFORMATION

TABLE IV NETWORK INFORMATION

Input Layer	Factor	1	AGE
		2	GENDER
		3	INCOME
Covariates	1		TIME
		2	COST
	Number Units		8
	Rescaling Method For Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		2
	Number of Units in Hidden Layers 1[a]		7
	Number of Units in Hidden Layers 2[a]		5
	Activation Function		Hyperbolictangent
Output Layer	Dependent variables	1	REGULAR MODE
	Number Of Units		1

	Rescaling Method For Scale Dependents	Adjusted Normalized
	Activation Function	Hyperbolictan gent
	Error Function	Sum of Squares

The network information in Table IV displays information about the neural network and is useful for ensuring that the specifications are correct. The number of units in the input layer is the number of covariates plus the total number of factor levels. Two hidden layers were requested and the procedure has chosen 7 units in the first hidden layer and 5 units in the second. A separate output unit was created for the scale- dependent variable. They are rescaled by the adjusted normalized method, which requires the use of the hyperbolic tangent activation function for the output layer. Sum-of-squares error was reported because the dependent variables are scale.

13.MODEL SUMMARY

TABLE V MODEL SUMMARY

Training	Sum of Squares Error	18.686
	Relative Error	.222
	Stopping Rule	1 consecutive step(s) with no decrease in error[a]
	Training Time	00:00:00.356
Testing	Sum of Squares	7.308
	Relative Error	.301
Holdout	Relative Error	.211

The model summary in Table V displays information about the results of training and applying the final network to the holdout sample. Sum-of-squares error was displayed because the output layer has scale-dependent variables. This was the error function that the network tries to minimize during training. Note that the sums of squares and all following error values are computed for the rescaled values of the dependent variables. The relative error for each scale-dependent variable was the ratio of the sum-of-squares error for the dependent variable to the sum-of-squares error for the "null" model, in which the mean value of the dependent variable is used as the predicted value for each case. There appears to be more error in the predictions of mode choice. The relative errors are fairly constant across the training, testing, and hold out samples, which gives some confidence that the model is not over-trained and that the error in future cases scored by the network will be close to the error reported in this table. The estimation algorithm stopped because the error did not decrease after a step in the algorithm.

14. INDEPENDENT VARIABLE IMPORTANCE

TABLE VI INDEPENDENT VARIABLE IMPORTANCE

	Importance	Normalized Importance
AGE	.030	5.4%
GENDER	.031	5.6%
INCOME	.045	8.2%
TIME	.350	64.3%
COST	.545	100.0%

Independent variable importance is given in Table VI and also in the form of bar chart in Fig 3.

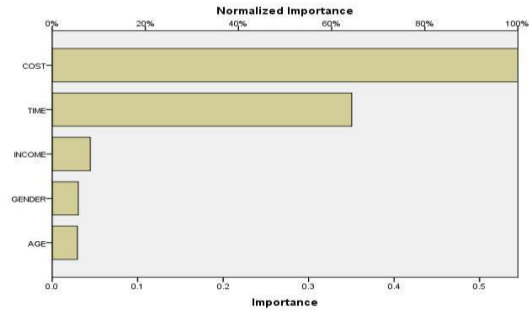

Fig. 3 Independent Variable Importance

The importance of an independent variable is a measure of how much the network's model-predicted value changes for different values of the independent variable. Normalized importance is simply the importance values divided by the largest importance values and expressed as percentages. The importance chart is simply a bar chart of the values in the importance table, sorted in descending value of importance. It appears that variables travel time and travel cost had the greatest effect on mode choice modeling and what could not be told was the "direction" of the relationship between these variables and the predicted probability of default.

15.CLASSIFICATION

TABLE VII CLASSIFICATION

Sample	Observed	Predicted			
		1	2	3	Percent Correct
Training	1	58	13	0	81.7%
	2	11	144	9	87.8%
	3	1	15	97	85.8%
	Overall percent	20.1%	49.4%	30.5%	85.9%
Testing	1	15	8	0	65.2%
	2	2	45	2	91.8%
	3	0	5	19	79.2%
	Overall percent	17.7%	60.4%	21.9%	82.3%
Hold out	1	8	5	0	61.5%
	2	1	25	1	92.6%
	3	0	0	16	100.0%
	Overall percent	16.1%	53.6%	30.4%	87.5%
Dependent Variable: Regular Mode					

15.1 Dependent Variable: Regular Mode

The classification table showed that using 0.5 as the pseudo-probability cut-off for classification, the network did considerably better at predicting the mode choice. The single cut-off value gave a very limited view of predictive ability of the network, so it was not necessarily very useful for comparing competing networks**.**

16. SUMMARY OF RESULTS

1. In the suburban regions of Kottayam the two wheelers are more because of the convenience and less travelling difficulty during the peak hours.
2. Cars are preferred next to two wheelers than buses in these areas.
3. Commuters are giving more importance to travel time and are likely to travel by the mode which takes less travel time.
4. Travel cost is another important parameter which affects the commuters mode choice. People living in the study area region do not give much importance to cost of travel.
5. Most of the commuters own both car and two-wheeler.

17. CONCLUSION AND FUTURE WORKS

The present study indicated that the work trip travel in the study areas was mainly by means of personalized modes. The universal choice set was identified as Car, Bus and Motorised Two-Wheeler. Two wheelers are preferred more compared to car and buses. There was a good evidence to show that the commuters prefer costlier mode. The perception of travel cost was either low or reasonable.

A model was also formulated with Multilayer perceptron procedure of Artificial Neural Networks. The model results obtained from this procedure were comparable to those results obtained using Logistic Regression. The time taken to obtain the solution is usually less than or equal to the time taken by traditional models. A physical equation was not obtained while modeling using neural network to predict the number of trips. Instead, a trained network was obtained which has weights stored in it. This was one of the criticisms of the researchers supporting logit and other traditional models that give a physical equation. But the trained network was as good as an equation and can be directly used on new data for predicting the output.

In this present study only time and cost were taken for building the model. There are several other factors which are affecting the choice of travel mode and the study can be extended by taking several other variables which are affecting mode choice. The choice of mode mainly depends upon various factors such as comfort, safety and convenience which are not considered in this model.

REFERENCES

[1] Akiva, B.M., Lerman, S.R. (1985). Mode Choice - Implications For Development." Aitpm 2011 National Discrete Choice Analysis: Theory And Application To Traveldemand."The Mit Press., Massachusetts, U.S.A.,

[2] C.R.Bhat, S.K. Singh, "A Comprehensive Daily Activity Travelgeneration Model System For Workers." Trans. Research Part A: Policy And Practice, Vol.34(1), Pp. 1-22, 2005.

[3] M. Mckibbin, "The Influence Of The Built Environment On Transit-Oriented Conference, 2011.

[4] Andr_E De Palma, Denis Rochat ,"Mode Choices For Trips To Work In Geneva: An Empirical Analysis", Journal Of Transport Geography 8, 43-51,2000.

[5] Khandker M. Nurul Habib" Modeling Commuting Mode Choice Jointly With Work Start Time", Transportation Research Part A 46 (2012) 33–47

[6] Joseph Prashker, Yoram Shiftan , Pazit Hershkovitch -Sarusi "Residential Choice Location,Gender And The Commute Trip To Work In Tel Aviv" Journal Of Transport Geography 16 (2008) 332–341.

[7] D.B Madan And R Groenhout ,"Modelling Travel Mode Choices For The Sydney Work Trip", Journal Of Transport Economics And Policy, May 1987.

[8] Hsin-Li-Chang And Shuen-Cheng Wu "Exploring The Mode Choice In Daily Travel Behaviour Of The Elderly In Taiwan", Journal Of The Eastern Asia Society For Transportation Studies, Vol. 6, Pp. 1818 - 1832, 2005 .