International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Mode Choice Analysis: Conventional and Advanced Approaches

Athira C[1], Carol Varghese[2]

[1]*PG Student, Dept. of Civil Engineering, Jyothi Engineering College, Kerala, India*
[2]*Assistant Professor, Dept. of Civil Engineering, Jyothi Engineering College, Kerala, India*

Abstract – *Travel demand modelling is an important stage in a transportation planning project where future demand of travel is predicted by developing a travel demand model. It is a four step model with stages such as trip generation, trip distribution, mode split and trip assignment. Mode split is the stage of travel demand forecasting where the mode chosen for the trip is incorporated into the model. Mode choice models are developed to analyze the mode choice of trip makers and to predict the mode shares of available alternative modes. Mode choice models are formulated on the basis of utility maximization theory by employing mathematical techniques. Scientists have applied statistical as well as advanced soft computing techniques for developing mode choice models. Accurate prediction of mode shares helps in implementing new transport policies and evaluating existing transport system. This paper gives an overview of some conventional and advanced approaches that the scientists and researchers prefer for analyzing the mode choice of trip makers.*

Key Words: **Travel demand model, Mode choice model, Utility maximization theory, Statistical models, Soft computing techniques**

1.INTRODUCTION

Travel demand modelling is done to predict the future demand of travel and network needs. The conventional travel demand model consists of four distinct stages, i.e., trip generation, trip distribution, modal split and trip assignment. The third stage of travel demand modelling is the modal split, in which the choice of mode of travel is included. The position of mode split stage is usually at the third, but it may vary according to the nature of the trip and the studies. Mode split stage is one of the vital stages in transport modelling as it plays a major role in determining the shares of different modes of transport. Since the selection of mode has a greater impact on the entire transport system, mode choice has gained importance as an individual area of research. Mode choice models have been developed to predict the shares of different modes. The mode choice models are formulated on the basis of the utility based choice theory, which states that an individual chooses his/her alternative which maximizes his/her utility. If the mode choice is considered as a decision making process, four major components are to be considered. They are the choice maker, i.e., the individual who makes the choice; the alternatives which refers to the available modes to select from; attributes of alternatives, which refers to the various characteristics of each alternative considered while taking a decision and the decision rule, which is the rule or pattern followed by the decision maker to choose a mode. Researchers are using different mathematical techniques to develop mode choice models. Various approaches such as the conventional statistical methods and advanced soft computing methods are employed widely for formulating the mode choice models. The objective of this paper is to discuss some conventional and advanced methods in mode choice modelling.

2. TYPES OF MODE CHOICE MODELS

Different mode choice models are being developed by researchers which involves conventional statistical and advanced models. Some of them are discussed in the following sections.

2.1. Aggregate and disaggregate models

Mode choice models are classified primarily into aggregate and disaggregate models. Aggregate models attempt to represent the average behaviour of a group of travellers instead of a single individual. Examples of aggregate models include: Trend analysis models where past trends were extrapolated to estimate future travel; mathematical models like the direct demand models and sequential models are usually more difficult to implement, more time-consuming and costlier but provide more precision. The aggregate transportation planning models have been severely disapproved for their rigidity and imprecision. In disaggregate approach individual choice responses as a function of the characteristics of the alternatives available and sociodemographic attributes of each individual. It has a more causative nature and is thus more transferable to a different point in time and to a different geographic context, very well suited for pre-emptive policy analysis. Efficiency of disaggregate approach is more than the aggregate approach in terms of model dependability. Hence, disaggregate models are preferred over aggregate models for mode choice analysis.

2.2. Statistical mode choice models

Formulation of statistical mode choice models is the conventional approach followed by researchers in the discrete mode choice analysis. Statistical techniques are employed to predict the shares of different modes of transport by considering the statistical relation between certain independent and dependent variables. The

dependent variable is the mode choice, which is expressed as a utility function of the modal attributes. The independent variables usually include the socio economic characteristics of the trip makers and trip characteristics. Examples of statistical models include logit models, probit models and Generalised Extreme Value (GEV) models.

Logit model is most widely used conventional model in which prediction is done by employing simple mathematical techniques and also by the application of the utility maximization theory. The logit models can be further classified into multinomial logit models (MNL), nested logit models and mixed logit model. Multinomial logit model is the most basic member of the statistical model family. It predicts the mode choice by considering multiple independent variables by establishing a statistical relation between the dependent and independent variables. Since it is very easy to work with the modelling part, multinomial logit model is still preferred widely for mode choice analysis, even if it is a conventional method. The drawback of the MNL model is the Independence of Irrelevant Alternatives (IIA) property, which implies that any changes in the probability of a given alternative draw equally from the probabilities of all the other substitutes in the choice set. Hence, addition of new alternatives affects the probabilities of the selection of existing alternatives, which may not hold good in every case. Nested logit model estimates the mode choice by creating 'nests' of sub modes among the alternatives, for e.g.: car, two wheelers, etc. can be involved in the 'nest' of 'private vehicles' and bus, trains, etc. can be involved in the 'public transport' nest. The nested model provides a partial relaxation from the independence of irrelevant alternatives since the introduction of new alternatives will draw from corresponding nest only. The major drawbacks include the increase in complexity of the analysis as the selection of alternatives from nests is complex and also categorization of alternatives into mutually exclusive nests may not be possible in every cases. Mixed logit model is a highly flexible model which consists of components of both logit and probit models. These models are applicable to any random utility function and individual heterogeneity can also be taken into account.

Multinomial Probit (MNP) models finds application in mode choice modelling when there is complex correlation between the utilities of alternatives. In these cases, multinomial logit models may provide inaccurate predictions. The absence of the assumption of identically distributed error terms means that taste distinction and choice repetition can also be incorporated in probit models. Since the probit model involves the complex heterogeneities in modelling, it will lead to difficulties in interpretation, conceptual and statistical problems. Also, there will be difficulties in transferring models from one space time sampling frame to another.

Generalised Extreme Value (GEV) models are developed as a simplification of multinomial logit model. The model is formulated using the utility maximization theory. It is a closed form distribution that allows for various levels of correlation among the unobserved part of utility across alternatives.

The statistical mode choice models are formulated by establishing a statistical relationship between characteristics of trip makers and the choice of mode as utility functions. The statistical models predict the mode shares and also helps in determining the significant factors that influence the mode selection. The statistical models do not accommodate the uncertainties in the human behaviour attributes, which are complex in nature. If the uncertainties and approximations are taken into account for modelling the mode choice, accuracy of these models can be improved. Since it is difficult to do so in the conventional approaches, researchers started to consider advanced soft computing methods for mode choice analysis nowadays.

2.3. Soft Computing Mode Choice Models

Soft computing finds application in modelling the complex real life problems by employing approximations in the models. Unlike hard computing, i.e., conventional models, soft computing is tolerant of imprecision, uncertainty and approximations. Various soft computing techniques involving fuzzy logic, Artificial Neural Networks (ANN), genetic algorithm, random forest, etc. are considered for mode choice modelling. Also, the neuro- fuzzy or hybrid models, which is the combination of fuzzy logic and ANN techniques are also employed.

Artificial Neural Network (ANN) is an information processing system which exhibits high analogy to the biological neural system. ANN has basic units which are the simple processing elements called neurons; similar to the biological neurons. The neurons are interconnected by means of direct links. Each link has its own weight associated with it. Activation function or threshold function is a function which is used to map the input to specified output range. It is also known as squashing or transfer function. Different threshold functions involve linear, non-linear, sigmoid, hyperbolic and tangent. Sigmoid function is usually employed for mode choice analysis. The system has to be trained using training data with input and output and then the data to be tested has to be provided. Since the ANN system has a learning capacity, it will learn from the training data and a neural network will be created considering the input variables and the activation function and the output will be provided as the degree of significance of each input variable in the selection of mode, in the form of weighted graphs.

Fuzzy logic is another artificial intelligence or soft computing technique, which can be effectively applied in the

mode choice analysis. Based on the concept of relative graded membership, fuzzy logic predicts the chances of an event that can be true to a certain degree between 0 and 1. In the conventional models, the prediction is of crisp nature, i.e., there are two chances of an event occurring or not. In case of mode choice modelling, statistical models provide the crisp output, i.e., whether a mode is selected or not. If fuzzy logic technique is applied to mode choice analysis, the prediction will be in such a way that the chances of selection of each mode from the alternatives are graded between 0 to 1. Hence, it is closer to the complex human behaviour pattern, which is vague or fuzzy in nature. So, fuzzy logic is widely preferred for mode choice analysis nowadays. The major steps involved in the fuzzy logic analysis are fuzzzification, in which the crisp input variables are converted to fuzzy sets with linguistic variables, for e.g. "high", "many", "low", etc. Also, each variable has a membership function which represents the degree of truth in graphical form. Usually, triangular membership function is selected for mode choice analysis as it is easy for analysis. Similarly, output variables are also defined in this step. In mode choice analysis, the output variables are the utilities of each mode. The second step is the introduction of fuzzy inference system, in which the fuzzy rule base is formed with several IF-THEN rules connecting input and output variables. The general form of the rule is "IF x is A, THEN y is B" [1]. Similarly, in mode choice analysis, rules connecting the trip maker characteristics and the mode utilities will be formed. The last step is the process of defuzzification, in which the output values, which are also in the form of fuzzy sets are to be converted into interpretable form or crisp form. It is done by several methods and the most widely used method is centroid method, in which the crisp output values are derived from the centroid of the surface plots formed by considering the membership functions and rule base. It is impossible to obtain a well-defined relation between the input and output variables in the fuzzy logic analysis as it provides the defuzzified predicted values directly as the output. Fuzzy logic provides means to model the uncertainty associated with vagueness, imprecision, and lack of information.

Hybrid models or neuro-fuzzy models is formed by combining the fuzzy logic and ANN techniques, to make use of the advantages of both the techniques. The Artificial Neural Network has a very good learning capacity and can capture the complex relationship between the variables if good amount of data is available, while fuzzy logic can capture the linguistic and vague expressions that are more close to human behaviour but without any learning abilities. Fused Neuro Fuzzy (NF) architecture let ANN learning algorithms to determine the parameters of Fuzzy Inference System (FIS). So on combining both these techniques a powerful model can be developed that has a learning ability as well as can grasp the vagueness involved in human decision making procedure. The other fused NF models that can be used are cooperative and concurrent fused neuro-fuzzy models. In the cooperative NF model, the ANN decides the membership function or the fuzzy rules to be implemented based on the data and then go to background. In concurrent NF model ANN dynamically assists FIS to set its parameters.

'Random Forrest Decision Trees are another advanced technique used for mode choice analysis. A Random Forrest Decision Tree is a tree constructed randomly from a set of possible trees with random features at each node. "At random" implies that in the set of trees each tree has an equal chance of being sampled, i.e. trees have a "uniform" distribution. Random trees can be generated efficiently and the combination of large sets of random trees generally leads to accurate models. Random Forest is one of the most efficient methods for classification and regression in data mining. It can classify an object or an instance to a predefined set of classes based on their attributes values such as age or gender and other characteristics of the trip maker. A Decision Tree starts from the root and moves downward. The starting point of the tree is called a root node while where the chain ends is known as the "leaf" node' [2]. A node represents a certain characteristic while the branches represent a range of values. The Random Forrest models exhibit higher accuracy than conventional statistical models in mode choice analysis.

3. COMPARISON BETWEEN THE MODELS

MNL model has been the most widely used in mode choice analysis but its IIA assumption is misleading as it is irrational. Nested logit model with its tree and branched structure gives some relaxation from the IIA property of MNL but the nested logit structures are either inconsistent with utility maximization theories or are not considerably better than the multinomial logit model. Multi nomial logit models are applied in mode choice analysis where there are more than two alternatives to choose from. Probit mode choice model which provides greater flexibility, is rarely used in travel demand modelling because these models pose the complexity of introducing several additional parameters in the covariance matrix which require high dimensional integrals which is difficult. Ghareib (1996) [3] compared the application of logit and probit models in mode choice analysis of two modes of transit in different cities of Saudi Arabia and recommends the application of logit models over probit models as logit models are analytically easier and provide satisfactorily good results but probit models have a higher degree of analytical complexity, which makes it tedious to work with. Neural Network based mode choice models gives higher accuracy than the logit and probit models because these models are capable in mapping the independent and explained variable but its "black box" image does not help when it comes to easy interpretation of the results. Neural network models are formulated to estimate the degree of influence of each variable on the mode choice. Dongwoo Lee et. al (2018)[4] compared the

application of ANN and MNL techniques in mode choice analysis of people in Chicago city and found that ANN models are way ahead of MNL models in terms of accuracy and hence recommends it for mode choice analysis. Fuzzy logic based mode choice models employed in modelling does associate closely to human linguistic expression but deciding on the best suited membership function is very time taking and tiring process. Sarada Pulugurtaa et.al. (2013) [5] compared the MNL mode choice model with a Fuzzy logic mode choice model for mode choice analysis of Port Blair city and concluded that fuzzy logic models perform better than MNL models as they were better to incorporate the human reasoning into mode selection behaviour. Random Forrest models creates several decision trees and chooses the best decision tree, considering the mode selection as a decision making process.

4. SUMMARY

Selection of mode is a vital stage in the travel demand modelling which always needs in- depth analysis. Mode choice analysis is done by predicting shares of available modes by formulating mode choice models. Since mode choice modelling is inevitable in almost all the transportation planning studies, several techniques are developed to build accurate mode choice models. It involves conventional statistical techniques such as logit and probit models and also, the advanced approaches like soft computing or artificial intelligence methods. Of the conventional statistical methods, multinomial logit models are still preferred by scientists for simple mode choice predictions; as it models by establishing statistical relationship between the explanatory and output variables. The soft computing techniques such as fuzzy logic, ANN, hybrid models and Random Forrest are nowadays preferred by researchers as these techniques models the real world complex situations considering the uncertainties and approximations. Since these techniques considers the uncertainties, they are close to human behaviour and thus can model mode choice accurately as selection of mode is influenced by human behaviour also. Researchers are working on several advanced artificial intelligence and soft computing methods for performing mode choice analysis accurately as the development of precise models of mode choices will help in studying and analysing the effectiveness of new transport policies even before their implementation i.e., by the prediction of mode shifts accurately.

REFERENCES

[1] https://en.wikipedia.org/wiki/Fuzzy_rule
[2] Ch. Ravi Sekhar, Minal, and E. Madhu ,'Mode choice analysis using Random Forrest decision tree', Transportation Research Procedia 17, 644 – 652, 2016
[3] Ahmed Hamdy Ghareib, 'Evaluation of logit and probit models in mode-choice situation' Journal of Transportation Engineering, 122:282-290, 1996
[4] Dongwoo Lee, Sybil Derrible, and Francisco Camara Pereira, 'Comparison of four types of artificial neural network and a multinomial logit model for travel mode choice modeling' Transportation Research Record, 1-12, 2018
[5] Sarada Pulugurtaa, Ashutosh Arun, Madhu Errampalli, 'Use of artificial intelligence for mode choice analysis and comparison with traditional multinomial logit model' Procedia - Social and Behavioral Sciences ,104, 583 – 592, 2013
[6] Amrita Sarkar, G.Sahoo and U.C.Sahoo, 'Application of fuzzy logic in transport planning" International Journal on Soft Computing (IJSC) Vol.3, No.2, 2012
[7] Ashu Shivkumar Kedia, Krishna Bhuneshwar Saw & Bhimaji Krishnaji Katti, 'Fuzzy logic approach in mode choice modelling for education trips: a case study of Indian metropolitan city', Transport, 30:3, 286-293, 2015
[8] Dogan Ibrahim, 'An overview of soft computing' Procedia Computer Science,34-38, 2016
[9] Frank S. Koppelman and Chandra Bhat, 'A self-instructing course in mode choice modeling: multinomial and nested logit models', U.S. Department of Transportation, Federal Transit Administration, 2006
[10] Joon-Kyu Lee, Kwang-Eui Yoo, Ki-Han Song, 'A study on travellers' transport mode choice behaviour using the mixed logit model: A case study of Seoul-Jeju route' Journal of Air transport management,1-7, 2016
[11] Lijun Yu, Bin Sun, 'Four types of typical discrete choice models: which are you using?' IEEE, 2012
[12] Minal, Ch. Ravi Sekhar ,'Mode choice analysis: the data, the models and future ahead' International Journal for Traffic and Transport Engineering, 2014
[13] Ramanuj P. S. and Varia H. R., 'Application of soft computing techniques in mode choice analysis' Pramana Research Journal, Volume 8, Issue 11,357-365,2018
[14] https://www.geographynotes.com/articles/4-main-stages-of-the-transportation-planning-process-explained/219