Comparison of Different Mode Choice Models for Work Trips using Data Mining Process

Jayesh Juremalani* and Krupesh A. Chauhan

Civil Engineering Department, SVNIT, Surat – 395007, Gujarat, India; jayesh1575@gmail.com

Abstract

Objectives: An attempt made in this paper to compare the prediction ability of different mode choice behaviour models of commuters’ work trip. Methods/Statistical Analysis: The personal characteristics like age, income, family size, occupation of commuters and the trip characteristics such as trip time and trip cost were selected as independent variables. Four types of alternative modes were considered for developing the models such as private car, two wheelers, Shared Auto and bus. The behaviour predicting ability of three different models namely boost tree, MNL and SVM model compared using data mining process in Statistica software. Findings: We found that the boost tree model is superior among all. Applications/Improvements: Next step would be to find and to put the probabilities of each mode for each zone. Further to assign the route using all or nothing assignment method or capacity restraint assignment method.

Keywords: Boost Tree Model, Commuters’ Characteristics, Mode Choice Behavior, Trip Characteristics

1. Introduction

An attempt made in this paper to model the mode choice behaviour of commuters for their work trip. Total 224 respondent's behaviours have been used for the analysis purpose. The personal characteristics like age, income, family size, occupation of commuters and the trip characteristics such as trip time and trip cost were selected as independent variables. Four types of alternative modes were considered such as private car, two wheelers, Shared Auto and bus. The behaviour predicting abilities of three different models namely Boosted tree, linear and SVM model compared using data mining process in Statistica software.

2. Literature Review

Omer Khan estimated various nested logit models for different trip lengths and trip purposes, using the data from a Stated Preference (SP) survey conducted in the Shire. Pulugupta et al., concluded that fuzzy logic works better than MNL model as far as mode choice models are concerned. Koppleman and Chandra have written a book on a self instructing manual on mode choice behavior.

3. Data Collection

Total 251 respondents from Vadodara city of Gujarat state were asked to fill up the questionnaire. Out of which 224 were selected after filtering the data and removing the outliers. Among all 224 respondents 90 respondents were single and 134 were married. The personal characteristics like age, income, family size, occupation of commuters and the trip characteristics such as trip time and trip cost were selected as independent variables. Four types of alternative modes were considered such as private car, two wheelers, Shared Auto and bus. Table 1 shows the frequency of some of the parameters.

4. Data Analysis

The personal characteristics like age, income, occupation, marital status and gender are basically categorical data. So Chi test was used to check their effects of mode
Comparison of Different Mode Choice Models for Work Trips using Data Mining Process

Table 1. Frequency of Marital Status, Gender and Income of Commuters (Source: Author)

| Attributes    | Frequency | %   | Valid % | Cum % |
|---------------|-----------|-----|---------|-------|
| Marital Status|           |     |         |       |
| single        | 90        | 40.2| 40.2    | 40.2  |
| married       | 134       | 59.8| 59.8    | 100   |
| Gender        |           |     |         |       |
| Male          | 154       | 68.8| 68.8    | 68.8  |
| Female        | 70        | 31.3| 31.3    | 100   |
| Income        |           |     |         |       |
| Less than 5000| 12        | 4   | 4.1     | 5.5   |
| 5000-10000    | 37        | 16.5| 17      | 22.5  |
| 10000-20000   | 44        | 19.6| 20.2    | 42.7  |
| 20000-30000   | 41        | 18.3| 18.8    | 61.5  |
| 30000-40000   | 34        | 15.2| 15.6    | 77.1  |
| 40000-50000   | 25        | 11.2| 11.5    | 88.5  |
| Above 50000   | 28        | 10.7| 11      | 100   |

Table 2. Chi Square Test of Independence for Characteristics of Commuters with Mode Choice (Source: Author)

| Attributes        | Value Df | Assumption | Sig. (2-sided) |
|-------------------|----------|------------|----------------|
| Pearson Chi_square| 13.58 6  | 0.035      |                |
| L R               | 12.772 6 | 0.047      |                |
| Marital Status    |          |            |                |
| Value Df          |          |            |                |
| Assumption        |          |            |                |
| Pearson Chi_Square| 21.313 6 | 0.002      |                |
| L R               | 21.655 6 | 0.001      |                |
| Occupation        |          |            |                |
| Value Df          |          |            |                |
| Assumption        |          |            |                |
| Pearson_Chi_Square| 48.491 24| 0.002      |                |
| L R               | 53.505 24| 0          |                |
| Income            |          |            |                |
| Value Df          |          |            |                |
| Assumption        |          |            |                |
| Pearson_Chi_Square| 83.302 48| 0.001      |                |
| L R               | 75.194 48| 0.007      |                |
| Age               |          |            |                |
| Value Df          |          |            |                |
| Assumption        |          |            |                |
| Pearson_Chi_Square| 341.543 300| 0.049   |                |
| L R               | 275.932 300| 0.837    |                |

Table 3. Correlation coefficients

|                     | Travel_Time | Travel_Cost |
|---------------------|-------------|-------------|
| Pearson_Correlation | 1           | 0.272       |
| Sig. (2-tailed)     | 0.005       |             |
| N                   | 224         | 224         |
| Pearson_Correlation | 0.272       | 1           |
| Sig. (2-tailed)     | 0.005       |             |
| N                   | 224         | 224         |

Figure 1. Lift Chart of Three Models.

The p-value is exact 0.005 which means that both the variables have influences on mode choice behavior of commuters. After confirming the dependency all of independent variables with mode choice behavior of commuters, it is necessary to create a model which can predict the mode choice behavior of the other respondents too. There are so many models available which can predict the commuters’ behaviors reasonably well. Here in this paper, the predicting ability of three models
namely boost tree model, Multinomial logit model and SVM model were compared using data mining process in Statistica software.

5. Conclusion

Table 4 shows that errors of three models. It can be seen from the table 4 that boost tree model is having the lowest error only 15.98%. While MNL model is having 33.79% and SVM model is having 27.85% error. So predicting ability of boosted tree model is about 84%.

Table 4. Comparison of Errors among Three Models
(Source: Statistica Software)

| Model      | Boost Tree model | SVM model |
|------------|------------------|-----------|
| MNL Model  | 0.337900         | 0.278539  |
| Boost Tree | 0.159817         |           |
| SVM model  |                  | 0.278539  |

Figure 1 shows that highest value 2.3 is for boost tree model while generalized linear model and SVM model having 1.95 and 1.85 respectively. Now based on the results we can say that boost tree model can predict the mode choice behavior of commuters with highest accuracy and it is the most reliable model.

6. Future Scope of Work

After analysis of best suitability of model for predicting mode choice behavior of commuters’ next step would be to find put the probabilities of each mode for each zone. The population data can be collected from the census for each zone. By multiplying the zonal population with the probability we would have the predicted traffic volume on the urban roads. It can be verified by the actual classified volume count survey. The error in prediction can be found by using MAPE or RMSE method. The next step would be to assign the route using all or nothing assignment method or capacity restraint assignment method. The bottlenecking areas are to be found and appropriate solution can be worked out.

7. References

1. Khan O. Modeling Passenger Mode Choice Behavior Using Computer Aided Stated Preference Data [PhD Thesis]. Queensland University of Technology; 2007.
2. Bhat C, Koppleman F. A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. U.S. Department of Transportation Federal Transit Administration; 2006 Jan 31.
3. Pulugupta S, Arun A, Errampalli M. Use of Artificial Intelligence for Mode Choice Analysis and Comparison with Traditional Multinomial Logit Model. 2nd Conference of Transportation Research Group of India (CTRG). 2013 Dec 2; 104:583–92. Crossref