

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Engineering 137 (2016) 729 – 737

**Procedia
Engineering**www.elsevier.com/locate/procedia

GITSS2015

Cluster-Based Logistic Regression Model for Holiday Travel Mode Choice

Juan Li^{a,*}, Jinxian Weng^a, Chunfu Shao^a, Hongwei Guo^b^a*MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, 3 Shangyuancun, Beijing 100044, China,*^b*Department of Transportation Engineering, Beijing Institute of Technology, 5 Zhongguancun South Street, Beijing 100081, China*

Abstract

With the rapid growth of holiday travel market, more and more attention has been paid to the analysis of holiday travel behavior, such as the holiday travel mode choice. This study presents a cluster-based logistic regression model for predicting the travel mode choice on holiday. At first, a regression and classification tree approach is employed to split the source data to clusters. Based on the data collected from the Beijing Fragrant Hills Park during the Qingming Festival (Tomb-Sweeping Day), an optimal tree with two levels and three leaf nodes is built and the collected data are divided into three clusters according to the tree structure. The three clusters are further considered as the dummy variables for the logistic regression analysis. Since the cluster based logistic regression model avoids the variable interaction effects, it significantly outperforms the logistic regression model in terms of prediction accuracy.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Department of Transportation Engineering, Beijing Institute of Technology

Keywords: holiday; regression and classification tree; logistic regression; cluster; mode choice

1. Introduction

Although the rapid growth of holiday travel market promotes the regions' development, it causes many problems such as traffic congestion and parking chaos during holidays. Transportation planners have to deal with this

* Corresponding author. Tel.: 86-010-51684591.
E-mail address: juanli@bjtu.edu.cn

challenge through distributing useful travel information to the tourists so that they can make practical travel arrangement and reasonable travel mode choice.

Many models have been proposed for the analysis of travel mode choice behavior. However, the major concern of these models is the commuters' work travel mode choice behavior [1-4]. Although there are limited researches on the modeling of holiday travel mode choice behavior, the focus was in the behavior for long travel [5-7]. In reality, with the growth of economic and aging population, the increase of individual leisure time could generate more short duration trips. It should be noted that the short duration trips have different features against the long duration trips, especially on holidays. Therefore, a thorough understanding of holiday travel choice behavior is needed.

This study presents a cluster based logistic regression model to describe the travel mode choice behavior on short-duration holidays. Firstly, a regression and classification tree approach is used to discover the variable interaction effects, which partitions the data into several smaller clusters recursively. Secondly, the logistic regression approach is employed to the clusters which are identified as the dummy variables. This cluster-based logistic regression model is developed based on the data collected from Beijing Fragrance Hills Park during Qingming festival using the revealed preference (RP) survey method.

2. Literature Review

Travel mode choice is a crucial part of travel demand forecasting. Two theories are often employed to develop travel mode choice models. The first one views travel mode choice as a discrete choice problem [8-14] which is on the basis of the principle of random utility maximization. The second one regards travel mode choice as a pattern recognition problem [2-4,15-20], and several data mining techniques are used to identify the complex relationships between explanatory variables and the travel mode choice.

The existing travel mode choice analysis is dominated by disaggregate discrete choice models which could analyze and predict travel choice based on the preference factors. Among these discrete choice models, logit models with different structures, such as multinomial logit (MNL) and nested logit (NL), have been widely used for the analysis of travel mode choice [10,14,29]. Over the last decade, data mining approaches have also been used for investigating travel behavior. Several data mining approaches, such as decision tree (DT), neural network (NN), naïve Bayesian classification (NBC), support vector machine (SVM), multilayer feed-forward neural networks (MLFNNs), class association rules (CARs), are employed to model the travel mode choice behavior. Xie et al. [3] conducted a comparison of the predictive performance of DT, NN and MNL model in predicting work travel mode choice behavior. Cantarella and Luca [19] applied MLFNNs for the travel mode choice analysis. Zhang and Xie [20] applied SVM to predict commute travel mode choices, and compared the prediction accuracy between MNL and MLFNNs. Lu and Kawamura [4] used CARs to estimate the work trip mode choice by using the data from Chicago Area Household Travel Survey. Although these data mining approaches can provide high prediction accuracy, their classification results cannot provide an elasticity analysis, which is important for transportation planners in designing the travel policies. Santoso and Tsunokawa [31] analyzed the performances of updating techniques in transferability of mode choice models using Naïve transfer methods.

Recently, much attention has been paid to the relationship between tourists' mode choice and trip-tours on holidays [5-7, 21-25]. There are also two basic ways to modeling holiday travel mode choice: discrete choice models and data mining approaches. In particular La Mondia et al. [7] applied joint MNL to model tourists' holiday destination and travel mode with a European Union case study. An example of data mining approach can be found in Miller et al. [5] where they used an agent-based architecture to construct a tour-based model of travel mode choice. It was found that if an agent decides to take a vehicle, they are bound to use it for the entire trip.

This study aims to extend mode choice modeling into the analysis of holiday travel mode choice behavior. Data mining approaches are able to achieve high accuracy and precision while their classification results cannot give explicit interpretation of explanatory variables [26]. Although logistic regression model can give the relationship between a target variables and explanatory variable, it may have other shortcomings: one factor could display different exposure effects under different situations and some factors may hide effects in the full sample, which may result in low prediction accuracy. In order to avoid this issue, segmentation approaches can be employed to split the samples. Tree based methods are useful statistical techniques which are general employed to split the source data (Train, 1980). For example, a tree-based logistic regression approach is employed by Weng et al. [28] to assess work

zone casualty risk. However, their method can only be used with a very large data sample, because adequate data is needed to develop the corresponding regression model of each leaf node. If the clusters can share the same regression model, it would solve the problem of insufficient data. A cluster-based logistic regression model is presented in this study for predicting the holiday travel mode choice behavior.

3. Objectives and Contributions

The objective of this study is to develop a cluster-based logistic regression model to predict the holiday travel mode choice. At first, a regression and classification tree is employed to identify the interacting variables and split the entire data set into several clusters. Subsequently, these clusters are used as dummy variables for the development of the logistic regression model. Finally, the logistic regression model incorporating the cluster information could explicitly interpret the marginal effects of influencing factors on the holiday travel mode choice.

The contribution of this study is twofold. First, the cluster-based logistic regression model compensates for the weakness of the data mining approaches, which cannot interpret the marginal effects of influencing factors. Second, the cluster-based model can provide higher prediction accuracy than the pure logistic regression by avoiding the variable interaction effects and hidden effects of some influencing factors.

4. Methodology

There is a high-order interaction and nonlinear relationships between influencing factors and the travel mode choice. Moreover, some influencing factors could hide their effects in the full samples. In this case, a logistic regression model cannot describe the complicated effects of influencing factors of travel mode choice accurately.

In order to deal with this problem, a cluster-based logistic regression model is developed in this study, as shown in Fig. 1. At first, a regression and classification tree is employed to identify the interacting variables using the survey data. The full data sample is split into several smaller clusters according to the tree structure. Secondly, a logistic regression model is developed based on these clusters, which are used as dummy variables. This cluster-based logistic regression model could not only provide probability estimates, but also be able to interpret the effects of influencing factors on holiday travel mode choice.

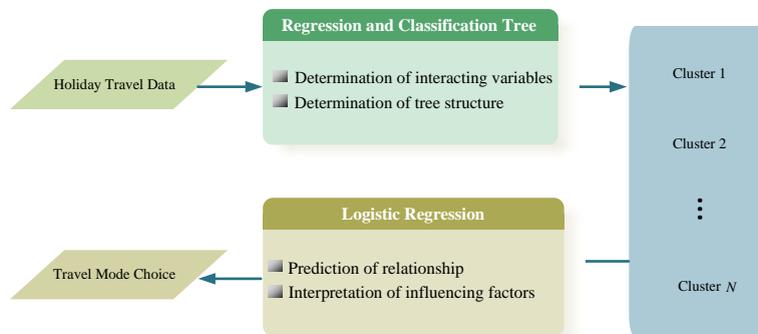


Fig. 1. Cluster-based logistic regression model

4.1. Regression and classification tree

Regression and classification tree (CART) is a useful partition technique, where each leaf represents class label, and each branch represents conjunction of features that lead to this class label. A path from the root to the leaf represents a classification rule.

The principle in regression and classification tree building is to split the source set into subsets based on an attribute value test [28]. When the data enter the root node of the tree, all candidate splits among all variables are searched by a test with a splitting criterion, which is a central problem in the construction of the trees.

There are some general approaches, such as the Gini reduction, entropy reduction and Chi-square test splitting criteria. Since some explanatory variables are ordinal variables in this study, we chose the Chi-square test splitting criterion, which can be described as the procedure which determines the right variable to split the parent node in a most reasonable way. For every explanatory variable, the split with the smallest p-value is identified as the best split. The global best split is the one which has the global smallest p-value among all the best splits. The parent node is split if the global smallest p-value is less than the preset α split. When no more nodes at the bottom can be further split or the current tree depth reaches the preset maximum number, the tree will stop growing and the optimal tree is finally obtained.

4.2 Logistic regression model

Logistic regression model is a statistical method to predict the relationship between a set of independent variables (xs) and a dependent variable (y). The advantage of the logistic regression model is that it cannot only predict the probability of the mode choice but also estimate the marginal effect of each explanatory variable. Since this study focuses on the holiday travel mode choice, the outcome of a tourist is assumed to be dichotomous: car and noncar. A binary logistic regression model is used for each cluster of the regression and classification tree to estimate the likelihood of each outcome. The target variable can be set as: $y=1$ for car, while $y=0$ for noncar. The coefficient of the explanatory variable and Odds Ratio (OR) can be used to explain the relationship of explanatory variables (e.g., gender, annual household income, and travel time) and target variable (travel mode choice), which are estimated by the maximum likelihood estimation method. The OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure. The OR is a measure of effects caused by the xs on the y. For example, the explanatory variable “Gender (1=Male, and 2=Female)” has a coefficient of 1.41 and OR of 4.1, which suggests that the probability of a male tourist choosing to drive 3.1 times higher than a female tourist.

5. Case Study

5.1. Data collection

The data used in this study were collected using a revealed preference survey (RP Survey) method from Beijing Fragrance Hill Park during the Qingming Festival. The period for investigation was from 8:00 am to 8:00 pm on April 6 of 2013. The Qingming Festival is a statutory public holiday in mainland China and Beijing Fragrance Hill Park is approximately a distance of 10 miles to the west of Beijing. Many people, especially the families with elders and children, are likely to have one day travel to Fragrance Hill Park on holiday.

As shown in Table I, two different questionnaires were provided to tourists: Questionnaire A is about the Travel Tracker Survey for driver; Questionnaire B is about the Holiday Travel Survey for all tourists. 980 questionnaires were distributed and 891 questionnaires were returned. Finally, 731 effective questionnaires were selected from returned questionnaires. Among these 731 effective travel records, 305 travelers chose the model of car while the remaining 426 travelers choose the modes of taxi, bus, metro, bicycle and walk. Due to the relatively small samples, the modes of taxi, bus, metro, bicycle and walk were aggregated into the mode of noncar.

Table 2 shows the explanatory variables related to household properties, personal attributes and trip characteristics, which might affect the tourists' travel choice. It should be pointed out that personal attributes include gender, age, occupation and driver's license status. Age is split to three groups and occupation is divided to three categories. Household properties contain the information regarding the type of residents comprising resident and nonresident, household income, number of vehicles in the household ranging from 0 to 5. The trip characteristics contain the information regarding the number and relationship of people traveling together, travel time and stay time. The stay time (in the park) were categorized into four groups.

Table 1. Summary of questionnaires

Questionnaire Type	Distributing Questionnaire	Recycling Questionnaire	Recycling Percentage	Effective Questionnaire	Effective Percentage
Questionnaire A	460	412	89.6%	313	76.0%
Questionnaire B	520	479	92.1%	418	87.3%
Total	980	891	90.9%	731	82.0%

Table 2. Data categories and variables

Variable	Definition	Notation	Descriptions
Target Variable			
	Outcome	y	0=Non-car; 1=Car
Personal Attributes			
GENDER	Gender	x1	1=Male; 2=Female
AGE	Age		1="<30 year-old"; 2="31~50 year-old"; 3=">51 year-old"
OCCUP	Occupation	x2	1= Manager & Staff; 2= Worker & Freelance; 3= Others (e.g., Student, Retired & Unemployed)
DLICENSE	Driver's license status	x3	1=Yes; 2=None
Household Properties			
RESID	Type of residents	x4	1= Resident; 2=Nonresident
HINCOME	Annual household income	x5	1="<50000 RMB"; 2="50000~100000RMB"; 3="100000~300000RMB"; 4=">300000RMB"
VEHS	Number of vehicles in the household	x6	Continuous
Trip Characteristics			
PEOS	Number of people traveling together	x7	Continuous
RELATION	Relationship of the people traveling together	x8	1=Families & Relatives; 2= Friends & Colleagues; 3=Classmates; 4=Clients; 5=Other
ELD&CHN	Number of elders and children	x9	Continuous
TRAVELTIME	Travel time	x10	Continuous
STAYTIME	Stay time	x11	1="< 0.5 hour"; 2="0.5~2 hours"; 3="2~4 hours"; 4=">4 hours"

5.2. Results

Model Structure

To build the regression and classification tree, the significance level is set to 0.15 (asplit) for the Chi-square test splitting criterion. One method to avoid the over-fitting problem is to set the maximum tree depth as a small value. Therefore, the maximum tree depth is set to be 3 given the number of variables. In order to guarantee that both the car and noncar classes in each leaf node have enough data for subsequent logistic regression analysis, the minimum number of data sets in each leaf node is set as 55. Because the Chi-square test statistic is insignificant for all the nodes, the tree stops growing at the second level. As a result, the optimal tree with three leaf nodes is constructed.

As shown in Figure 2, the variable of “number of vehicles in the household” (x6) is a significant interacting variable. According to the number of vehicles, the collected data samples are split into the following three clusters

for the three leaf nodes: x_6 is equal to 0, 1, and above 1 respectively. It can be found from Figure 2, the tourists with cars are more likely to choose the mode of car, comparing with those without cars. The more cars their families have, the bigger probability they choose the car mode. For example, a tourist in a household with at least two cars has a 5.3% higher probability of choosing the car mode, as compared with the one car household. The three clusters are used as the dummy variables in developing the logistic regression model. Due to the fixed value of interacting variable x_6 , the remaining variables (i.e., $x_1 \sim x_5, x_7 \sim x_{11}$) are available variables for model building in the clusters. The backward elimination method is employed to select the explanatory variables of the cluster-based logistic regression model. According to the significant level of the likelihood ratio, variables are tested and removed from the model successively. First, all explanatory variables are included in the logistic regression model. Second, the least significant variable, whose statistically significance is below the significance level of 0.15, is removed. This removal is repeated until the significance level of all the variables in the model is less than 0.15.

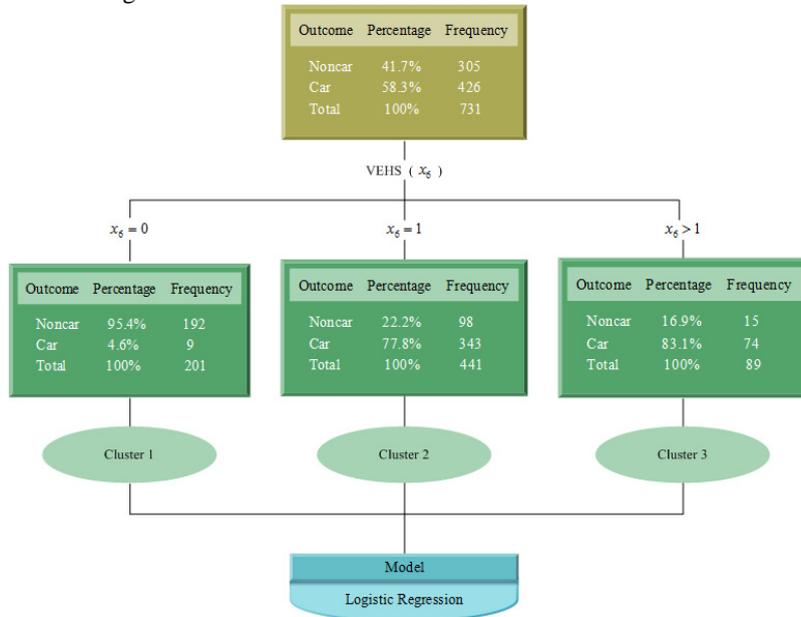


Fig.2. Selected tree structures for the cluster-based logistic regression model

Result Analysis

The statistic results of the cluster-based logistic regression model are presented in Table 3. There are only two influencing variables (i.e., household income and driver’s license status) which could cause significant effects on the travel mode choice in the whole sample.

The positive coefficient associated with the household income (OR=1.50) implies that a tourist with a high household income (i.e., >300,000RMB/year) has a 50% higher probability of choosing the car mode than the one with a low household income (i.e., <50,000RMB/year). One possible reason might be that tourists with high household income are more likely to own a car, which increases the likelihood of driving. In addition, this part of tourists may not care about travel expense. The negative coefficient associated with the driver’s license status (OR=-5.64) shows that tourists with driver’s license are more likely to choose the mode of car, which is consistent with the reality.

Due to data splitting, more influencing variables are found from the resulting clusters. For the families with one car, three additional influencing variables (i.e., gender, travel time and stay time) are found for cluster 2. And two additional influencing variables (i.e., gender and the number of elders and children traveling together) are found for cluster 3, in which families owned more than one car.

The positive coefficient of the factor of gender in cluster 2 shows that gender has a significant effect on the travel mode choice for the families with one car. The marginal effect of this factor is 4.11, which implies that men are

more like to drive for holiday travel than women in these families. The probability of choosing the mode of car for men is 3.11 times higher than women. The higher proportion of driving associated with men from one-car household mainly results from two factors: (1) more men hold driver’s license than women; (2) men are more likely to drive when travelling with others. Contrast to the cluster 2, the negative coefficient associated with the gender factor in cluster 3 shows that women are more likely to choose car for the families owning more than one car. A possible reason might be that the women are more likely to have driver's license and drive independently in these families, comparing with the women in the families with one car. In addition, women may spend more time with families, so women are more likely to travel with elders and children during holidays than men.

Table 3. Statistical results for the cluster-based logistic regression model

Variable	Coefficient	Standard Error	Wald Chi-square	Odds Ratio	p-value
Intercept	6.5060	1.1231	33.5542		<0.0001
GENDER (x1)* Cluster 2	1.4124	0.8417	2.8154	4.1058	0.0934
GENDER (x1) * Cluster 3	-0.4113	0.2569	2.5628	0.6628	0.1094
DLICENSE(x3)	-5.6460	1.0221	30.5160	0.0035	<0.0001
HINCOME(x5)	0.4059	0.1712	5.6217	1.5007	0.0177
ELD&CHN(x9) * Cluster 3	0.9964	0.5224	3.6387	2.7085	0.0565
TRAVELTIME(x10) * Cluster 2	1.1840	0.6963	2.8910	3.2674	0.0891
STAYTIME(x11) * Cluster 2	-0.9936	0.6616	2.2552	0.3702	0.1332

Tourists tend to drive when the travel time from home to the destination is much longer for the one car household families. Intuitively, the longer the travel time, the bigger probability of people choose to drive. Therefore, the probability of a tourist choosing the mode of car increases with the travel time (OR=3.27).

The negative coefficient associate with tourist behavior for cluster 2 indicates that the “stay time” could influence the mode choice. Tourists are unlike to drive if their stay time is longer than 4 hours (OR=0.37). This result may possibly be explained by the consideration of parking fee for the families with one car, which is relatively expensive during the holiday.

The number of elders and children traveling together is a factor that can significantly affect the travel mode choice for the families owing more than one car. This suggests that tourists tend to select the car mode when they are travelling with the elders and children (OR=2.71). One possible reason may be that the elders and children need relative comfortable travel environment. The more elders and children traveling together, the more likely tourists would choose to drive.

5.3 Remarks

The cluster-based logistic inherits the advantage of logistic regression model, which compensates for the disadvantage of data mining approaches. It can provide interpretation on the marginal effects of influencing factor. The cluster-based logistic approach deals with variable interaction problem by splitting the entire data into several clusters. As shown in Table 4, the prediction accuracy is increased for the travel mode of noncar and car (i.e., 77.1% and 98.4%).

However, the splitting criterion for the tree building can affect the prediction performance of this cluster-based logistic approach. The different choice of splitting criterion will alter the structure of tree and the size of leaves, which significantly change the clusters of regression and classification tree.

Table 4. Comparison results on prediction accuracy

Method	Outcome	Field Data	Predicted Data	Accuracy
Logistic Regression	Non-car	305	230	75.4%
	Car	426	417	97.8%
Cluster-based Logistic Regression	Non-car	305	235	77.1%
	Car	426	419	98.4%

6. Conclusion

This paper presented a cluster-based logistic regression model for the holiday travel mode choice prediction, which is an improvement of logistic regression model. The data used in this analysis were collected in Beijing Fragrance Hill during the Qingming Festival using a RP survey method. The results indicated the important effects of influencing factors (e.g., gender, driver's license status, the number of car household, household income, number of elders and children traveling together, travel time and stay time) on holiday travel mode choice.

The current studies of travel mode choice are dominated by discrete choice analysis. However, the models often provide low prediction accuracy, due to the interaction effects and hidden effects of influencing factors. Therefore, this study adopted the regression and classification tree approach to identify the interacting variables and split the samples into three clusters according to the tree structure. The three clusters were used as dummy variable in developing the logistic regression model as dummy variables.

This cluster-based logistic regression model outperforms the logistic regression mode. It provides higher prediction accuracy for the data, especially in predicting the mode of car. Due to the limitation of sample size, this study only takes into account two travel mode choices: car and noncar. Future research should take into account more travel mode choices (e.g., bus, metro, bicycle and walk) with large amount of samples on the basis of the current study.

Acknowledgments

This project is supported by National Nature Science Foundation of China (No. 51308038 & 71301010), Ministry of Education, Humanities and Social Sciences Youth Fund (No. 13YJCZH082).

References

- [1] Baltas, G., 2007. Econometric models for discrete choice analysis of travel and tourism demand. *Journal of Travel & Tourism Marketing*, 21(4), 25-40.
- [2] Ben-Akiva, M. E., Lerman, S., 1985. *Discrete choice analysis: theory and application to travel demand*. The MIT Press, Cambridge, MA, USA.
- [3] Biagioni, J. P., Szczurek, P. M., Nelson, P. C., Mohammadian, A., 2009. Tour-based mode choice modeling: using an Ensemble of conditional and unconditional data mining classifiers. The 88th Annual Meeting of the Transportation Research Board (CD-ROM), Washington, D.C., USA.
- [4] Brey, E. T., Lehto, X. Y., 2007. The relationship between daily and vacation activities. *Annals of Tourism Research*, 34(1), 160-180.
- [5] Brownstone, D., 2001. *Discrete Choice Modeling for Transportation*. Travel Behavior Research: the Leading Edge, D.A. Hensher, ed., Pergamon, Amsterdam, the Netherlands, 97-124.
- [6] Cantarella, G.E., Luca, S.D., 2005. Multilayer feedforward networks for transportation mode choice analysis: an analysis and a comparison with random utility models. *Transportation Research Part C*, 13(2), 121-155.
- [7] Chung, P.J., 2006. The determinants of consumers' outbound tourism mode. National Sun Yat-Sen University, Department of International Business. Master's Thesis.
- [8] Forsey, D., Habib, K. N., Khandker, M., Miller, E. J., Shalaby, A., 2009. Evaluation of impacts of introducing new transit system on commuting mode choice and transit ridership: case study on VIVA transit in Toronto, Canada. The 88th Annual Meeting of the Transportation Research Board (CD-ROM), Washington, D.C., USA.
- [9] Grigolon, A.B., Kemperman, A., Timmermans, H. J., 2013. Mixed multinomial logit model for out-of-home leisure activity choice. The 92th Annual Meeting of the Transportation Research Board (CD-ROM), Washington, D.C., USA.
- [10] Hensher, D.A., Ton, T.T., 2000. A Comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transportation Research Part E*, 36(3), 152-172.

- [11]Ho, C., Mulley, C., 2013. Tour-based mode choice of joint household travel patterns on weekend and weekday. The 92th Annual Meeting of the Transportation Research Board (CD-ROM), Washington, D.C., USA.
- [12]Karlaftis, M. G., Vlahogianni, E.I., 2011. Statistical methods versus neural networks in transportation research: differences, similarities and some insights. *Transportation Research Part C*, 19(3), 387-399.
- [13]Kim, M., 2009. Two-stage logistic regression model. *Expert Systems with Applications*, 36(3), 6727-6734.
- [14]Koop, G., Poirier D.J., 1993. Bayesian analysis of logit models using natural conjugate priors. *Journal of Econometrics*, 56(3), 323-340.
- [15]Koppelman, F.S., Wen, C.H., 1998. Alternative nested logit models: structure, properties and estimation. *Transportation Research Part B*, 32(5), 289-298.
- [16]LaMondia, J., Snell, T., Bhat, C. R., 2009. Traveler behavior and values analysis in the context of vacation destination and travel mode choices: european union case study. The 88th Annual Meeting of the Transportation Research Board (CD-ROM), Washington, D.C., USA.
- [17]Lu, Y. D., Kawamura, K., 2010. Data-mining approach to work trip mode Choice Analysis in the Chicago Area. The 89th Annual Meeting of the Transportation Research Board (CD-ROM), Washington, D.C., USA.
- [18]McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, P. Zarembka, ed., Academic Press, New York, USA.
- [19]Miller, E.J., Roorda, M.J., Antonio, J. 2005. A tour-based model of travel mode choice. *Transportation* 32, 399-422.
- [20]Mohammadian, A., Miller, E.J., 2002. Nested logit models and artificial neural networks for predicting household automobile choices: comparison of performance. In *Transportation Research Record*, 1807, 92-100.
- [21]Moons, E., Wets, G., Aerts, M., 2007. Nonlinear models for determining mode choice: accuracy is not always the optimal goal. In *Progress in Artificial Intelligence, Lecture Notes in Computer Science*, Vol. 4874, Springer Berlin/Heidelberg, 183-194.
- [22]Omrani, H., Charif, O., Gerber, P., Awasthi, A., 2013. Prediction of individual travel mode using evidential neural network model. The 92th Annual Meeting of the Transportation Research Board (CD-ROM), Washington, D.C., USA.
- [23]Park, D., Lee, S., Kim, C., Choi, C., Lee, C., 2010. Access mode choice behaviors of water transportation: a case of Bangkok. *Journal of Advanced Transportation*, 44, 19-33.
- [24]Poirier, D.J., 1996. A Bayesian analysis of nested logit models. *Journal of Econometrics*, 75(2), 163-181.
- [25]Santoso, D.S., Tsunokawa, K., 2010. Comparison of updating techniques in transferability analysis of work trip mode choice models in developing countries. *Journal of Advanced Transportation*, 44, 89-102.
- [26]Sirakaya, E., Woodside, A. G., 2005. Building and testing theories of decision making by travelers. *Tourism Management*, 26, 815-832.
- [27]Train, K. 1980. A structured logit model of auto ownership and mode choice. *Review of Economic Studies*, 47(2), 357-370.
- [28]Weng, J. X., Meng, Q., Wang, D. Z., 2013. Tree-based logistic regression approach for work zone casualty risk assessment. *Risk Analysis*, 33(3), 493-504.
- [29]Xie, C., Lu, J., Parkany, E., 2003. Work travel mode choice modeling with data mining: decision trees and neural networks. In *Transportation Research Record: Journal of the Transportation Research Board*, 1854, 50-60.
- [30]Yang, C., Tsai, M., Chang, C., 2014. Investigating the joint choice behavior of intercity transport mode and high-speed rail cabin with a strategy map. *Journal of Advanced Transportation*, DOI:10.1002/atr.1264.
- [31]Zhang, Y., Xie, Y., 2008. Travel mode choice modeling using support vector machines. In *Transportation Research Record: Journal of the Transportation Research Board*, 2076, 141-150.