Simulation Modeling and Application of Travel Mode Choice Based on Bayesian Network

Ruijing Chen[*]

Fujian Chuanzheng Communications College, Fuzhou, China

Abstract: In this paper, we study the travel mode choice of residents to determine the set of factors which can influence travel mode choice of residents and analyze the influence factor characteristics. Using Bayesian theory, we analyze the travel decision-making data of the residents, discrete them, and use them in Bayesian network structure learning and parameter estimation by K2 algorithm. We establish a Bayesian network simulation model to analyze the dependence probability relationship between the parent nodes and child nodes. Validation test was carried out for the building simulation model of Bayesian network. Data analysis results showed that the Bayesian network has a high accuracy prediction for actual travel mode choice of residents. This paper studies the Bayesian structure and parameters learning method for the actual travel behavior, and this method which provides a new method for studying the travel mode choice of residents can reveal the relationship between the various attributes associated with travel mode choice through a new perspective.

Keywords: Bayesian network, mode choice, multi-netlogit, simulation.

1. INTRODUCTION

Following the acceleration of urbanization, the contradiction between traffic supply and demand increased seriously, and this contradiction cannot be solved fundamentally just by infrastructure construction which can improve the traffic supply capacity. Traffic Demand Management (TDM) is the effective way to solve the contradiction between traffic supply and demand [1]. The core of the TDM is to ease traffic congestion in the city by inducing people to adjust their travel behavior. So people pay more and more attention to study the characteristics of the trip through the travel demand analysis and forecast method after TDM measures is carried out, and evaluates TDM measures feasibility. Travel choices constitute an important component in the travel demand modeling process. The consequence of travel choice modeling can be used to discover trip makers' underlying decision behavior and to provide insights for transportation policy analysis and travel demand forecasting [2]. Among a number of travel choice problems, mode choice modeling may have received the most attention, due to the fact that it has been an essential step to spatially distribute travel demand by transportation mode and then to model mode-specific travel behaviors in either trip-based or activity-based travel demand modeling framework. But mode choice involves many factors which have greater uncertainty [3]. The subjective factors such as the habit of travelers, preference and so on make the traditional random utility maximization modeling method difficult to accurately describe mode choice process.

Bayesian networks are a marriage between probabilistic theory and graph theory. In a Bayesian network, the graphical network topology specifies the model structure and the conditional probabilities provide the mechanism to represent the dependence relationships between variables. In this article, we propose a Bayesian network model to address travel mode choice problem and discover how trip makers' travel decision behaviors are affected by underlying socio-economic and level-of-service factors.

Bayesian networks are an important and currently the most active branch of graphical models, as specified as directed acyclic graph (DAG) models. In DAG, undirected or bidirectional arcs are not allowed and cyclic feedback loops are not permitted. Bayesian belief networks have a qualitative part and a quantitative part, represented by a graph of discrete probabilistic variables and tables with conditional probabilities for these variables, respectively. On the semantics, the Bayesian network decomposes the joint probability distribution. Bayesian network can provide great convenience for probabilistic reasoning. This is mainly because that on one side Bayesian network is a strict language of mathematics which fits for computer processing; on the other side, the intuitive and understandable features make it easy for discussion and to build model.

Currently, Bayesian network learning has two key aspects in the application of travel behavior analysis, structure learning and parameter learning. As for the structure learning, some scholars have built many kinds of causality diagram to study the relationships among travel behaviors and between travel behavior and impact factors [4-7]. These causality diagrams are based on the expression of Bayesian network probabilistic focusing on understanding the process of travel decision. As for the parameter

*Address correspondence to this author at the Fujian Chuanzheng Communications College, Fuzhou, China; Tel: 0086-591-83511366; Fax: 0086-591-83441953; E-mail: ruijing_chen@163.com

estimation, according to the research by some scholars [8-10], there will be big deviation when using MLE (maximum likelihood estimation) with limited sample size. Nevertheless, Bayesian parameter estimation could solve such problem effectively. Other scholars study points that the Bayesian methods could solve estimate value of different travel individuals [11,12]. The differences between individuals were ignored when using MLE to calculate the residents' estimate value which will lead to bias. So, using Bayesian network analysis to study the residents travel mode choice can better simulate the residents' real decision-making process, enhance the analysis accuracy. This paper uses a city in southeastern China residents' travel mode choice as case study to apply Bayesian methodology in real travel decision behavior analysis.

2. BAYESIAN NETWORK LEARNING

By studying the basic rules of residents activities travel behavior decision-making, we can establish Bayesian network for it. It is convenience to analyze the relationship between many travel behavior impact factors. The Bayesian network model not only models cause-effect or input-output relationships between independent and dependent variables, but also captures the correlation between independent variables simultaneously, which presents an overall picture demonstrating the interactions between all the variables. It can also provide the theory basis and the model base for travel forecast system based on Bayesian network.

Bayesian network construction steps include:

(1)   Using the residents travel RP and SP survey data, network simulation data, it is easy to analyze the

general rules of the residents' activities, travel time, mode and route choice based on travel-activity analysis theory. And then we generate alternative factors set, including travel-activities decision-making behavior set and impact factors set.

(2)   Initialize Bayesian network, and establish node set. Study the relationships among travel behaviors and between travel behavior and impact factors. Form a connect set by the application of heuristic approach and search scoring algorithm and constantly optimize the network structure to build a DAG.

(3)   Add conditional probability distribution for every connection which expresses probabilistic dependency relationship between the nodes, then format conditions probability set.

(4)   Generate Bayesian network based on Search scoring algorithm (Fig. **1**).

Bayesian network learning is a process to get Bayesian network through the data analysis, it includes two aspects, structure learning and parameters learning. Parameter learning means to determine the parameters when we know network structure; Structure learning is not only to make sure the network structure, but also to determine the parameters.

2.1. Structure Learning

Structure learning makes a discussion in two steps in general, namely model selection and model optimization. Model selection wants to answer that which kind of criteria is better to judge the difference between different model
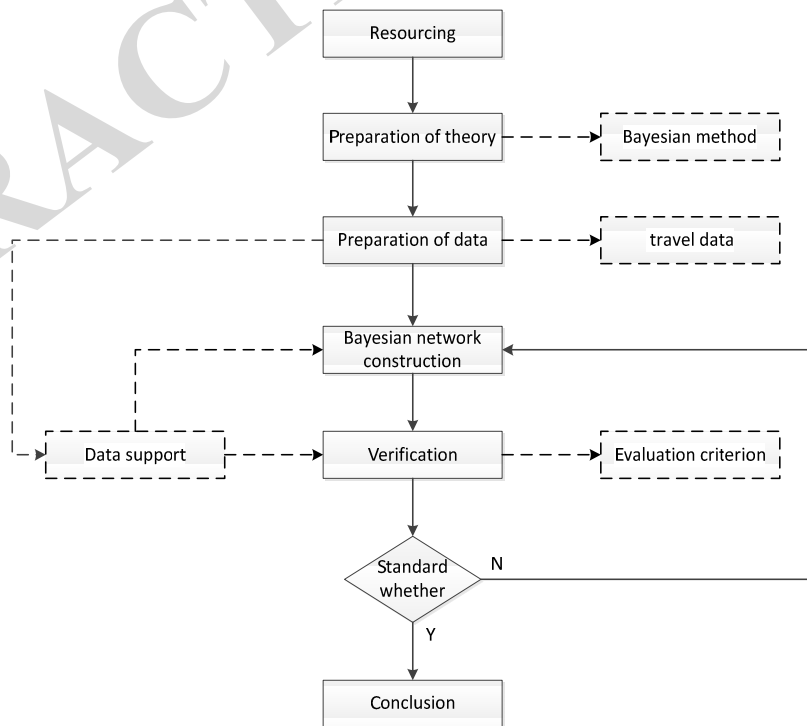

Fig. (1). Construction flow of Bayesian network.

structures, and the optimization model is to find out the best model structure. We need to follow the steps below to determine Bayesian network structure.

(1)   Select a group of explaining random variables { $X_1,\ldots,X_n$ }, every random variable represents a node in Bayesian network.

(2)   Choose a variable order $\alpha =< X_1,\ldots,X_n >$.

(3)   Make an empty graph G, then let graph G accept every variable according the order α .

(4)   When we accept X_i , there are $X_1,X_2,\ldots,X_{i-1}$ in the graph G already.

Structure learning commonly used method includes K2 algorithm, climbing algorithm, SEM and so on. Due to the data dependence of K2 algorithm in structure learning process being low, this algorithm can be very good with real data analysis fitting. So we choose K2 algorithm. K2 algorithm was first derived by Cooper and Herskovits. They proved that each parameter vector is updated independently with unrestricted multinomial distributions, parameter independence, Dirichlet priors, and complete data. Consequently, the marginal likelihood of the data is just the product of the marginal likelihoods for each i-j pair. Given by Eq.(1).

$$P(D \mid B_s) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \left[\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \cdot \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})} \right] \quad (1)$$

where, $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \, \mathrm{d}t$

Node Xi has state r_i , and its father node has state q_i , where $q_i = \prod_{X_i \in Pa_i} r_i$. N_{ijk} is record number which means quantity of Xi is k and its father node is j, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. α_{ijk} and α_{ij} is priori information, and $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$, generally, $\alpha_{ijk} = 1$.

The main point of the network structure learning with K2 algorithm is: first, we define a quality measurement function to evaluate the network structure model, and then from an empty network, according to the determined node order, choose a node as the father node which can maximum the posterior probability of this structure. In turn to traverse all the nodes, we can add the other father node for each variable gradually. K2 algorithm is shown in Fig. (**2**).

2.2. Parameter Learning

Bayesian network parameter learning is essentially to study the probability distribution table on the known network structure. Early, in Bayesian network the probability distribution table is designated by the expert knowledge, but this method only with expert experience specified often makes a large deviation between observation data and
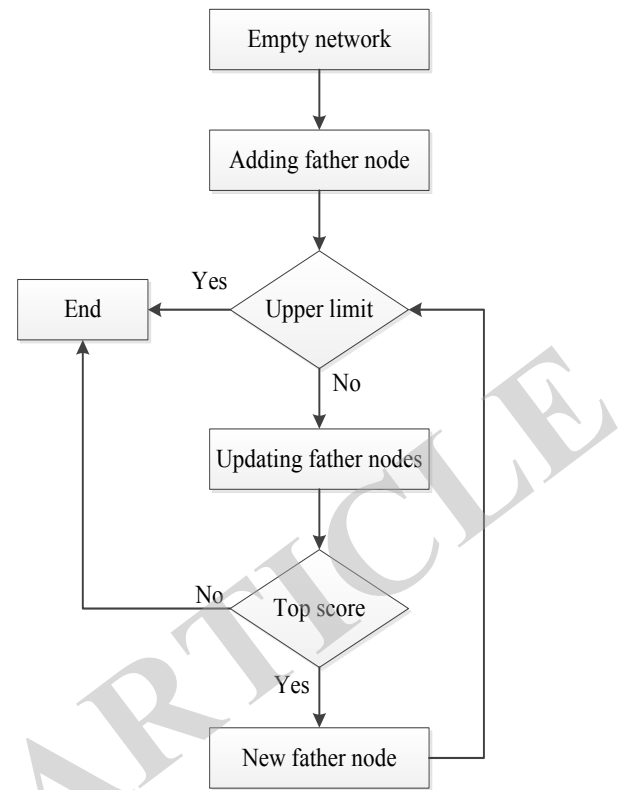

Fig. (2). Structure learning with K2 algorithm.

product. The current popular method gets the probability distribution of these parameters from data learning, the data driven learning method has the very strong adaptability. The commonly used methods to parameters learning have EM algorithm and SEM algorithm, etc. Both structure learning and parameter learning can run in the Matlab environment with the Full-BNT toolbox.
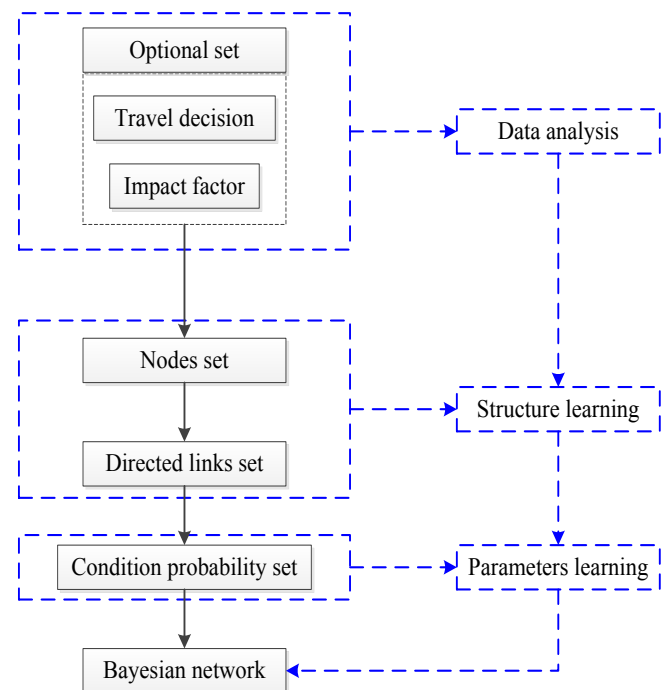

Fig. (3). Bayesian network construction process.

3. DATA ANALYSIS

Influence factors of travel mode decision can be divided into the macroscopical factors and micro factors. Macroscopical factors such as social economic development level, transportation policy, etc. decide the total travel mode structure. Micro factors such as professional property makers, income level, travel purpose and the travel time consuming decide individual choice. Combined with the trip survey data, we can use the data mining technology to find relationship between some factors and travel mode, and also we can make qualitative and quantitative description of the various attributes for residents which can impact travel mode. Residents of social attribute and economic attributes of its travel will affect decisions, such as the age of the residents, income, which will affect people travel mode.

We extract the basic information of 2006 families from 10 districts in a city in southeastern China to model travel mode decision with the third time integrated traffic survey data. First, the data has been screened, travel mode in this paper only refers to the internal traffic tools, and therefore, we delete 61 use inter-city traffic way travel records and relevant personnel and family information. Through sorting, we used 13016 travel article records for data modeling analysis. A survey of the object of age, gender, vocation, income, travel purpose, travel time consuming as inputs, and travel mode for decision variables.10000 travel records as training data set has been used to build travel mode Bayesian network.

Bayesian network algorithm requires the data node take discrete values, so we need to make a discrete pretreatment for residents travel data. The attribute range is divided into some intersection which can be presented by several son intervals, each interval corresponding to a discrete value, and finally, we change the original data into discrete values. Discretization results are shown in Table **1**.

According to the sample data, the influence factors of the travel model analysis are as follows:

3.1. Travel Behavior Analysis for Different Gender

From Fig. (**4**), in transit, rail transit and taxi which are public transportation, the ratio of mode choice is similar for male and female. But in private traffic, men turn even more

to choose car and non-motor vehicles; women more prefer to walk. Relative to the bus, subway belongs to public transportation, gender influence is not sensitive. In all travel modes, the order of the decision is walk, bike, electric bicycle, and bus.
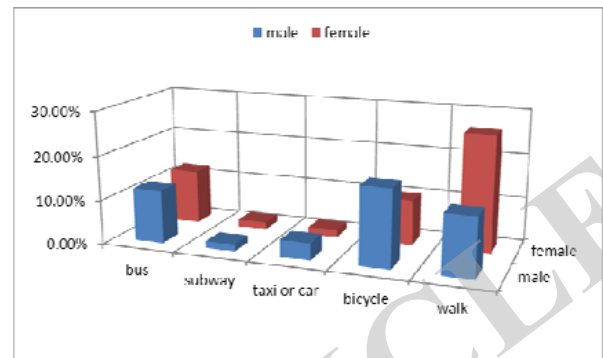

Fig. (4). Mode choice ratios for different gender.

3.2. Travel Behavior Analysis for Different Age

The residents' age can make certain effect to travel mode choice; the different age will produce preferences to certain way. Such as less than 20 young will be more inclined to bike or ride an electric bicycle, old people and children have short travel distance, therefore, their most choice is on foot or on public transportation. And middle aged people will prefer vehicles because of the physical abilities and vocation.
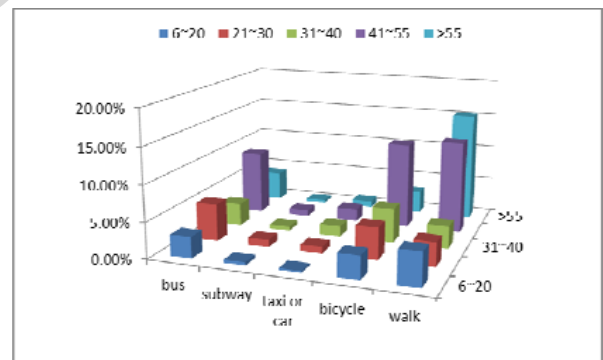

Fig. (5). Mode choice ratios for different age.

Table 1.   Discretization Results.

Variable Name	Discrete Value				
	1	**2**	**3**	**4**	**5**
Gender	Male	Female			
Age	6-20	21-30	31-40	41-55	55+
Vocation	Unit leader	Staff, Owner	Student	Retired	
Salary	0-800 RMB/M	801-2000 RMB/M	2001-5000RMB/M	5000RMB/M+	
Purpose	Commute	Life	Entertainment		
Duration	0-20minutes	21-50minutes	51-120minutes	120minutes +	
Mode	Bus	Subway	Taxi, Car	Bicycle	Walk

3.3. Travel Behavior Analysis for Different Incomes

Economic condition of the family is an important influence factor to the travel mode choice. Commonly, high income residents have strong travel ability, they have more choices. The low income residents have poor travel ability, they have fewer choices. High income residents will first consider comfortableness and convenient degree, and then consider travel cost, but low income residents will consider the cost first.

In the Fig. (**6**), low-income will choose bike, bus, no-motor bicycle or on foot. High-income levels will choose taxi or car as travel tools. Car or taxi users are mainly concentrated in the high-income levels.
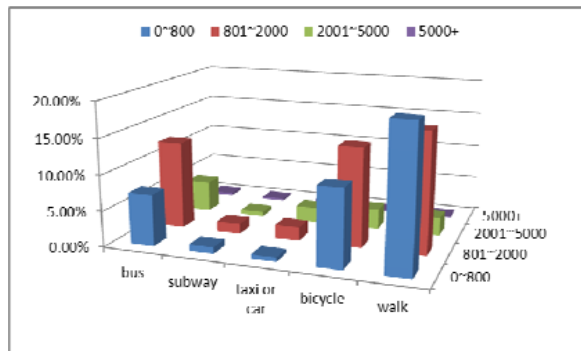

Fig. (6). Mode choice ratios for different income.

3.4. Travel Behavior Analysis for Different Purposes

Travel purpose is the cause of travel, but also can be used as travel behavior decision variable. It is an important factor to travel mode choice behavior. Travel purpose can be divided into: work and off duty, shopping, entertainment, business, life, etc. For different travel purposes, people have different travel mode choices.

From Fig. (**7**), buses, bicycles and no-motor bicycles appear to be preferred for commuting purpose. And during off work, entertainment and shopping activities also have a high proportion in the walk choice.
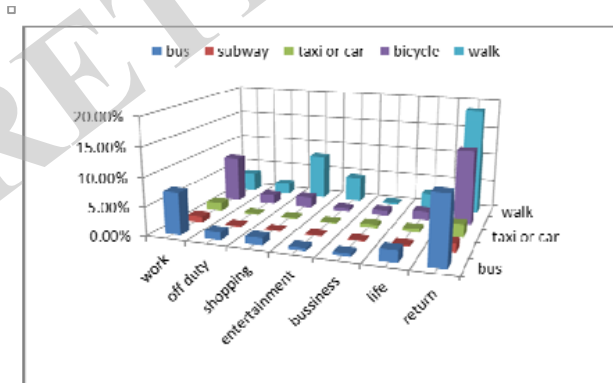

Fig. (7). Mode choice ratios for different purpose.

4. BAYESIAN NETWORK FOR TRAVEL MODE CHOICE

According to the characteristics of each attribute, input node is divided into connatural attribute and travel attribute. Gender, age, vocation, income will not change for a specific travel, known as the connatural attribute. And travel purpose and travel time cost will change with different travel, called travel attributes. The connatural attribute infers travel mode choice to reflect its preferences for different groups of residents. And the travel attribute infers travel mode choice objectively, because it ignores the residents' characteristics. We build Bayesian network by the two node types with training data set respectively. Discrete data is used for structure learning by K2 algorithm to get connatural attribute Bayesian network and travel attribute Bayesian network. They are shown in Figs. (**8**, **9**) respectively.
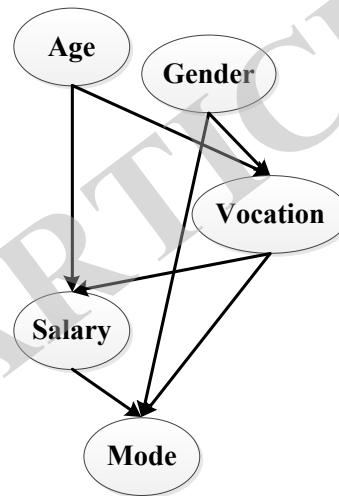

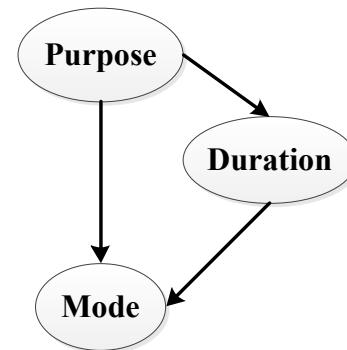**Fig. (8).** Connatural attribute bayesian network.

Fig. (9). Travel attribute bayesian network.

From Fig. (**8**), it is known that the gender, occupation, and income affect travel mode choice. For example, the statistics data can be seen that women are more likely to walk than men in all ages. This is according with our life common sense. Usually, more women prefer to give priority to walk a short trip than men. Income also has great influence on the travel mode choice, analyzing travel mode choice condition probability distribution of different income, as income increases, residents walk probability decreases, and the probability of choosing subway increases. The probability of choosing a car or a taxi has increased significantly. When the income is less than 800 Yuan a mouth, the probability of choosing a car or a taxi is 0.0153,

when income is in 800 ~ 2000 and 2000 ~ 5000, or 5000 Yuan or more, the probability become 0.0409, 0.1692, and 0.4638 respectively.

From Fig. (**9**), it is known that the purpose and the travel time consuming affect travel mode choice. For the purpose of travel to commute, the probability of choosing bus and subway is higher than others when the travel purpose is commuting. And the highest probability is choosing a car when the purpose is turning to entertainment. This is due to the characteristics of entertainment having more temporary, so there is low sensitivity in cost but high sensitivity in convenience and comfort. The influence of travel time consuming to travel mode choice is obvious, when the time consuming is more than 50 minutes, very few people choose walking, bike or no-motor bicycle as their travel tools

Connatural attribute and travel attribute variables reveal the relationship between the travel mode choice and themselves from the general and special Angle respectively. It is needed to combine connatural attribute and travel attribute together, when analyzing a particular trip. Based on this criterion, we build Bayesian network with two attributes. It is shown in Fig. (**10**).
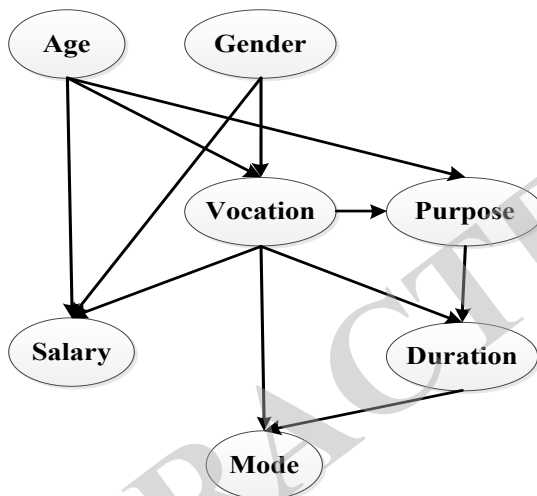

Fig. (10). Bayesian network for travel mode with two attributes.

Comparing Figs. (**8-10**), we can find 3 sides are missing in Fig. (**10**): Gender→Mode, Salary→Mode, and Purpose → Mode. This is because the global optimal solution is not simple joining of local optimal solution in the K2 algorithm. In the comprehensive consideration of the various attributes, nodes are connected with higher correlation with K2 algorithm. For example, after adding travel time consuming variables, the influence of income to the travel mode choice is not significant. At the same time, relative to income, the vital effect to travel mode choice is vocation. Similarly, for gender and travel purpose, their dependent probability relationship comparatively is not significant. Therefore, we can accept the network structure shown in Fig. (**10**).

5. ACCURACY VERIFICATION OF THE NETWORK STRUCTURE

Individual choice hitting rates are defined as follows:

First of all, we put parameters into utility function to figure out the probability of every choice branch, Pm.

$$\hat{\delta}_m = \begin{cases} 1, & p_m \text{ is max,} \\ 0, & \text{others,} \end{cases} \tag{2}$$

$$S_{in} = \begin{cases} 1, & \delta_{in} = \hat{\delta}_{in}, \\ 0, & \delta_{in} \neq \hat{\delta}_{in}, \end{cases} \tag{3}$$

$$\begin{cases} HitR = \sum_{n=1}^{N} \sum_{i \in A_n} S_{in} \Big/ \sum_{n=1}^{N} J_n, \\ HitR_i = \sum_{n=1}^{N} S_{in} \Big/ N_i, \end{cases} \tag{4}$$

where HitR and HitRi denote the whole hitting rate and hitting rate for branch i respectively, Jn means choice branch number for individual n; Ni is the sample number of branch i. Hitting rate for Bayesian network and MNL model are shown in Table **2**, where BN means Bayesian network model.

Usually, it is a very good result, if the hitting rate is more than 80%. So, Bayesian network model has high prediction accuracy for travel mode choice, and it has a better result to MNL model.

6. CONCLUSION

This article first makes a discretization for a city in southeastern China inhabitants travel survey data. We analyze the influential factors set affecting travel mode choice, and then build residents travel mode choice Bayesian network structure by the K2 algorithm. Bayesian method is used to model the parameters of the Bayesian network. The Bayesian network modeling has its advantage in travel behavior analysis when making a comparative analysis with MNL model. It can make prediction of travel behavior by its network structure and conditional probability. We can see that how various attributes influence each other directly by the network structure. For example, from Fig. (**10**), it is obvious that the length of travel time is affected directly by vocation and travel purpose factors. Retired staff and students tend to short time travel, and enterprises and institutions staffs are more likely to adopt long travel time. The trip will take a short time if its purpose is to work, but if the purpose is enjoyment or entertainment, it will give priority to the convenience and short time. It is can be seen directly from Fig. (**10**) that age is influential factor of vocation and income. And gender also affects travel purpose and travel mode, the probability of travel for life (shopping, family) in women is also different from men.

In the activities-travel behavior study, application of Bayesian network theory can make a full expression in travel behavior prediction with residents' personal attributes, travel habits, priori experience and experts' knowledge of modeling. It can make more comprehensive and accurate

Table 2.　Hitting rate for Bayesian network and MNL model (%).

Hitting Rate	Mode=1		Mode=2		Mode=3		Mode=4		Mode=5	
	BN	MNL	BN	MNL	BN	MNL	BN	MNL	BN	MNL
HitR	89.22	81.56	91.37	79.34	92.08	80.09	87.59	77.34	90.88	78.02
HitR$_i$	88.30	77.25	89.61	75.21	90.17	77.28	86.61	75.92	89.17	76.97

prediction. According to the cause and effect relation, the travel decision and its influencing factors are embodied in Bayesian network, and by constantly learning, updating our knowledge, we can analyze interaction response behavior between the travel decision and policy.

However, in constructing the Bayesian network model, the learning process itself will cause the defects of model building method. In order to improve the precision of the model, the following issues are still to be further researched: (1) The discretization of the residents' travel data mainly rely on the frequency method and experience, this affects the structure of network learning and parameter learning process. (2) K2 algorithm needs the travel attribute nodes order which lacks effective algorithms. (3) How to connect the Bayesian network model and other prediction model that can improve efficiency and precision of Bayesian parameters learning also needs further discussion.

CONFLICT OF INTEREST

This article content has no conflict of interest.

REFERENCES

[1]　T.A. Arentze, and H.J.P. Timmermans, "Capturing the Role of Awareness and Information Search Processes on Choice Set Formation in Models of Activity-Travel Behavior", *Transportation Research Board Annual Meeting*, 2004.

[2]　M. Ben-akiva, and S. Lerman, "Discrete choice analysis: Theory and application to travel demand", Massachusetts: MIT Press, pp.128-135, 2000.

[3]　C. Bhat, A. Govindarajan, and P. Vamsi, "Disaggregate attraction-end choice modelling: formulation and empirical", *Transportation Research Record*, vol. 11, pp. 60-68, 1998.

[4]　G. Cooper, and E. Herskovits, "A bayesian method for the induction of probabilistic networks data", *Machine Learning*, vol. 9, pp. 309-330, 1992.

[5]　A. Domencich, and D. Mcfadden, "Urban travel demand: A behavioral analysis", *Mt. Pleasant*, Michigan: Blackstone Company, 1996, pp. 75-77.

[6]　T. Hagerstrand, "What about people in regional science", *Papers of the Regional Science Association*, vol. 24, pp. 7-21, 1970.

[7]　J. Heckman, and B. Singer, "A method for minimizing the impact of distributional assumptions in econometric models for duration data", *Econometrica*, vol. 52, pp. 271-320, 1984.

[8]　D. Heckerman, "Bayesian networks for data mining," *Data Mining and Knowledge Discovery*, vol. 11, pp. 79-119, 1997.

[9]　D. Heckerman, D. Geiger, D. Chickening, "Learning Bayesian Networks: the combination of knowledge and statistical data," *Machine Learning*, vol. 20, pp. 197-243, 1995.

[10]　D. Janssens, G. Wets, T. Brijs, K. Vanhoof, and H.J.P. Timmermans, "Identifying behavioral principles underlying activity patterns by means of bayesian networks," *Transportation Research Board Annual Meeting*, 2003.

[11]　D. Janssens, G. Wets, T. Brijs, K. Vanhoof, "Improving the performance of a multi-agent rule-based model for activity pattern decisions using bayesian networks," *Transportation Research Board Annual Meeting*, 2004.

[12]　M.G. Scuderi, "*Bayesian approaches to learning from data how to untangle the travel behavior and land use relationships*," University of Maryland, Doctor Dissertation, 2005.