
Prediction of Tourist Consumption Based on Bayesian Network and Big Data

Xu Cheng¹, Chenyuan Zhao^{2*}

¹ College of Economics and Business Administration, Chongqing University, Chongqing 400044, China

² School of Economics and Finance, Chongqing University of Technology, Chongqing 400054, China

Corresponding Author Email: cy.zhao@cqut.edu.cn

<https://doi.org/10.18280/isi.240505>

Received: 10 March 2019

Accepted: 19 July 2019

Keywords:

big data analysis, Bayesian network (BN), neural network (NN), air ticket price, hotel price, tourist consumption

ABSTRACT

The boom of tourism has generated a huge amount of data on tourist consumption. The consumption trends of tourists could be mined out of these data, promoting the development of tourism. Taking air ticket price as an example of tourist consumption, this paper designs a novel model to predict the trend of tourist consumption in the next moment based on the historical data, rather than the traditional approach to analyze the influencing factors of tourist consumption. Our prediction model was established based on the Bayesian network (BN). Only two variables were considered in the model, i.e. the number of the remaining seats and the ticket price. Three different scoring functions were tested with our model. The effectiveness of our model was confirmed through experiments and a comparative analysis with the neural network (NN). The results show that our BN model achieved an accuracy greater than 80% in predicting the air ticket price in the next moment, and outperformed the NN model on the same dataset; the fluctuation of the air ticket price in the next moment can be predicted accurately based on the price trends in the previous two days. The research results lay the basis for predicting the price volatility of the other product/service of tourism.

1. INTRODUCTION

The boom of tourism has generated a huge amount of data on tourist consumption. With the aid of information technology, airlines and hotels can mine the price sensitivity of tourists out of these data, and adjust the prices of air tickets and hotels in a flexible manner. In general, the prices of air tickets and hotels change dynamically with the market sales. Many factors have different degrees of impacts on the two prices, ranging from click-through rate (CTR), occupancy, season to weather. The price changes become more complex, as airlines and hotels may try to attract tourists through price competition. Therefore, it is very difficult for tourists to predict the prices of air tickets or hotels.

Many tourists are accustomed to booking air tickets and hotels in advance. They commonly believe that the closer to the departure time, the higher the prices. However, the previous research has shown that air ticket price fluctuates greatly 3~15 days before departure. Similar trends were observed in hotel price. Considering the importance of predicting the trends of tourist consumption, this paper designs a novel model to predict the trend of tourist consumption in the next moment based on the historical data. The mode was established based on the Bayesian network (BN), involving such two variables as the number of the remaining seats and the ticket price. The feasibility of the model was tested through experiments and compared with the neural network (NN).

2. RELATED WORK

The existing studies [1, 2] on commodity/service prices of tourism have mainly explored the pricing strategy for suppliers

(e.g. airlines and hotels) to maximize their profits, with only a few exceptions. This section briefly reviews some representative researches.

Magnini et al. [3] predicted the trend of hotel price in the next 30 days with a hybrid algorithm, which combines reinforcement learning, rule learning, and time series analysis. Vojinovic et al. [4] predicted the series of commodity/service prices of tourism in four steps: allocating the time-varying price series into clusters, analyzing the statistics in each cluster, identifying the variation amplitude and frequency of price series by point process theory, and outputting an accurate prediction of future prices.

Using the probability decision model, Lee et al. [5] predicted the number of seats occupied at different air ticket prices, and thus identified the number of effective seats at each price. Danziger et al. [6] developed a prediction method for hotel price at tourist attractions based on k-nearest neighbors (k-NN) algorithm, improved Q-learning algorithm, time series algorithm and subjective Bayesian integration algorithm. Some scholars collected air ticket prices from various social networks, and designed an air ticket price prediction system, which systematically learns the collected data through big data analysis and forecasts the future price of air tickets [7].

Wu et al. [8] analyzed the structure of the commodity/service price series of tourism, and ascertained the turning point and change law of the price series. Through parameter learning of Bayesian network (BN), Lai et al. [9] identified season and passenger flow as the leading factors affecting tourist consumption. Collum et al. [10] extracted eigenvalues from GPS data on travel, and conducted modelling by The BN to determine the travel mode and consumption pattern of tourists.

Mustelier et al. [11] studied the tourist evaluation system of

tourist attractions, and disclosed the relationship between tourist consumption and tourist satisfaction. Yu et al. [12] predicted the fluctuation of hotel price in tourist destinations, according to how tourists choose their destinations. López et al. [13] carried out big data analysis on tourist data with MapReduce, and made precise classification of tourists based on their consumption pattern.

In general, the data on tourist consumption are analyzed based on various influencing factors of commodity/service prices of tourism. However, many of these factors are not easy to measure accurately. In this paper, a novel approach is designed to predict the future trends of tourist consumption: big data analysis on the historical data in the previous days.

3. THE BAYESIAN NETWORK

The BN is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG) [14]. In each DAG, the nodes represent random variables, while the directed edges between the nodes represent the conditional dependencies between the variables. Each node has a conditional probability distribution (CPD) associated with the variable. If the variable is discrete, then the CPD is a list of conditional probabilities under the given state of the parent node. Here, the conditional probability reflects the strength of the relationship between variables.

Let $\{X_1, X_2, \dots, X_n\}$ be the set of random nodes (variables) in the DAG and $\text{Pa}(X_i)$ be the parent node of node X_i . Then, the joint probability distribution of the various variables in the BN can be defined as:

$$(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Pa}(X_i)) \quad (1)$$

Each state of the joint probability space can be expressed as the product of conditional probabilities. Then, the probability of each state of each variable can be calculated under each state.

When the BN is adopted to classify a sample, the class of the sample is set to $S = \{s_1, s_2, \dots, s_l\}$, and the feature space is set to $Y = \{Y_1, Y_2, \dots, Y_n\}$. For a sample $y = \{y_1, y_2, \dots, y_n\}$, the BN classification aims to determine the class that s belongs to through training the sample set D . The class of the sample can be determined based on the maximum posteriori probability $\max_{i=1, \dots, l} \{p(s_i | y)\}$:

$$p(s_i | y) = \frac{p(s_i) \times \prod_{j=1}^n p(y_j | s_i; \pi(y_j))}{p(y)} \quad (2)$$

where, $\pi(y_j)$ is parent node of node y_j other than the node set S . Therefore, the BN structure should be learned to identify the probability distribution functions $p(s_i)$ and $p(y_j | s_i; \pi(y_j))$ from training sample D .

The BN structure can be learned by two types of methods, namely, score-based learning and conditional independence test [15, 16]. The score-based learning consists of three steps: first, the scoring function is defined as a measure; then, all the structures of the model's structure space are scored; finally, the structure with the highest score is selected by the search algorithm as the final structure for the network.

In score-based learning, the scoring function can be established on Bayesian statistics or information theory. The

former type of scoring functions includes Bayesian Dirichlet-likelihood equivalent (BDe) and Bayesian information criterion (BIC) [17, 18]. The latter is represented by the minimum description length (MDL) [19]. Thus, both BDe and BIC scores are based on Bayesian statistics.

In BN structure learning, structure T and parameter ϑ are taken as random variables. The BN is composed of n variables $y = \{y_1, y_2, \dots, y_n\}$, where y_i is a variable with v_i values ($\vartheta = \{\vartheta_{ijk} | i = 1, \dots, n; j = 1, \dots, q_i; k = 1, \dots, v_i\}$). The parent node $\text{Pa}(y_i)$ of y_i has q_i values. Suppose structure T obeys a distribution with a prior probability $P(T)$. For a given dataset D , the maximum posteriori probability of the structure can be obtained by the Bayesian formula:

$$P(T|D) = \frac{P(D|T)P(T)}{P(D)} \quad (3)$$

where, $P(D)$ is independent of structure T . The logarithmic maximum of Eq. (3) $\log P(T, D) = \log P(D|T) + \log P(T)$ is defined as the Bayesian score of the structure.

The MDL is a scoring function based on information theory, and the BIC score is a special case of the MDL. This function evaluates the description lengths of the data and the structure. The description length of the data reflects how similar the data is to the structure, and that of the structure shows how complex the structure is. The optimal structure should minimize the sum of the two description lengths. Generally, the structural complexity is penalized by the number of parameters:

$$C(T) = \frac{1}{2} \log d \sum_{i=1}^n (v_i - 1) q_i \quad (4)$$

where, d is the total number of samples in dataset D ; $\sum_{i=1}^n (v_i - 1) q_i$ is the total number of parameters in the network.

The compressed data length, i.e. the log-likelihood of dataset D and the model, can be defined as:

$$L_{MD}(T) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{v_i} d_{ijk} \log \frac{d_{ijk}}{d_{ij*}} \quad (5)$$

According to Eqns. (4) and (5), the corresponding MDL scoring function can be obtained as:

$$F_{MDL}(T|D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{v_i} d_{ijk} \log \frac{d_{ijk}}{d_{ij*}} - \frac{1}{2} \log d \sum_{i=1}^n (v_i - 1) q_i \quad (6)$$

Prior probability is not involved in the MDL scoring function. The MDL scoring function is the BIC scoring function, if the sample is sufficiently large and the dataset D obeys multinomial distribution.

During the learning of the BN structure, it is impossible to rate all the structures in the structure space. Therefore, the high-quality structures should be selected by a search algorithm according to the scoring function. In this paper, the hill-climbing algorithm is adopted for the search: First, the initial structure is modified by the search operator, creating a series of candidate structures. Then, the score of each candidate model is compared with that of the initial structure, and the structure with the highest score is taken as the initial structure for the next search. If the initial model has the highest score, the search will be terminated and the initial model will be returned as the result. The workflow of the hill-climbing algorithm is explained as follows:

Hill-climbing algorithm

1. An initial structure T is selected. T is often empty (but not necessarily empty).
 2. The score of T is calculated: $S_T = \text{Score}(T)$.
 3. $\max \text{Score} = S_T$ is executed.
 4. The following steps are repeated until $\max \text{Score}$ no longer increases.
 - A. A new structure T^* is obtained by adding, deleting, or replacing edges.The score of T^* is calculated: $S_{T^*} = \text{Score}(T^*)$.
If $S_{T^*} > S_T$, then $T = T^*$ and $S_T = S_{T^*}$.
 5. $\max \text{Score}$ is updated.
 6. The directed acyclic graph T is returned.
-

4. BAYESIAN NETWORK-BASED PREDICTION MODEL

This section applies the BN model to predict the trends of tourist consumption. Here, the air ticket is taken as an example for tourist consumption. The BN model was constructed after the collected data on air ticket had been preprocessed.

The raw data were collected from a special price information system. The dataset includes flight number, cabin status, collection data, departure date, departure city, and arrival city. Then, the author queried how the air ticket price of a flight changes in the next n days from the current date.

First, the flight to be predicted was determined. For example, the international flight CZ371 from Guangzhou to Hanoi was selected for prediction. The three codes of the departure city, the arrival city, and cabin status (e.g. J2 and KQ) were selected from the IBE interface. According to the cabin status, the lowest price of the air ticket price for the flight was obtained from the database of existing prices.

As mentioned before, the air ticket price fluctuated greatly 3~15 days leading to the departure date. This is a common rule for all flights, whichever the departure date. In this paper, the research period is set to February 17~June 15, 2019. In total, there are 43 departure dates. For simplicity, 13 days were selected as departure dates, with four measuring points per day. Hence, a total of 52 time points was obtained for further analysis.

Our research aims to predict the future volatility of air ticket price based on historical fluctuations. The price and number of remaining seats are both continuous data waiting to be discretized. Here, the price difference between the two time points is defined as the change trend. The three possible trends, namely, increase, constant and decrease, were represented by 0, 1 and 2, respectively. Then, the number of remaining seats

and the air ticket price at a time point were compared with those at the next time point to determine the specific change trend.

There is a clear similarity between the change trend of air ticket price and the above-mentioned fluctuations in 3~15 days leading to the departure date. The pre-sampled data were discretized into 20 time points, because the trend at the next time point is empirically affected by the data in the previous five days (20 time points). To determine the number of days that actually affect the price trend in future, the number of time points k was adjusted from 5 to 20. The adjustment to the k value led to changes in the sample size. Here, the sample size after the adjustment is obtained through experiments.

As mentioned before, the processed data contains 20 time points. With the decrease of time points, the data of departure dates will increase. The sample size can be computed by: $43 \times (20 - k) + 1,419$. Then, the data at the k -th time point was predicted based on the data of the previous $k-1$ time points.

The hill-climbing algorithm was adopted to search through the discrete sample data. For comparison, the score of the data structure at each time point was evaluated by the BID, BDe and k2 functions [20]. Finally, the optimal number of time points and structure diagram were outputted.

The 70% of all samples were taken as the training set, and the remaining 30% as the test set. A total of 1,419 samples were collected at the 20 time points. Thus, the training set and the test set respectively contain 993 samples, and 426 samples. In the test set, the data at the first $k-1$ time points serve as evidence variables. On this basis, the author predicted the probability of each change trend of the data at the k -th time point. The average of the results in ten experiments was taken as the final prediction.

The prediction results were analyzed from three aspects: the predicted number of remaining seats, the accuracy and stability of scoring functions, and the overall prediction accuracy. The prediction accuracies of different scoring functions on the number of remaining seats are listed in Table 1 below.

As shown in Table 1, the overall prediction accuracy of the number of remaining seats reached 78%. This means the number of remaining seats has little influence on the prediction accuracy of our model. In fact, the change trend of air ticket price could be predicted based on the price changes in the first few time points. The price changes already reflect the impacts from the number of remaining seats. Therefore, there is no need to include the number of remaining seats in the subsequent analysis.

Table 1. The prediction accuracy of the number of remaining seats

Remaining seats?	Scoring function	Overall accuracy	Accuracy of change trend 0	Accuracy of change trend 1	Accuracy of change trend 2
Yes	BIC	77.62	14.79	91.18	55.62
	BDe	78.12	17.55	91.62	56.93
	k2	78.30	16.49	92.02	56.15
No	BIC	77.96	15.02	90.32	55.22
	BDe	78.28	15.11	90.73	56.19
	k2	78.02	14.72	90.83	55.79

Table 2 compares the three scoring functions by the mean and variance of the data structure at each time point. The

structural score in the table reflects how well the data structure matches the actual data.

Table 2. The prediction accuracy of three scoring function

Standard	Scoring function	Overall accuracy	Accuracy of change trend 0	Accuracy of change trend 1	Accuracy of change trend 2	Structural score
Mean	BIC	83.06	10.28	93.75	69.11	-12338.69
	BDe	83.12	10.23	93.65	69.29	-12106.62
	k2	83.23	10.32	93.86	69.39	-11665.11
Variance	BIC	2.3291	4.2012	1.3221	5.4369	3,067.5731
	BDe	2.2017	4.2293	1.2293	5.3662	3,133.0162
	k2	2.1029	4.0106	1.1228	5.2172	3,188.2739

As shown in Table 2, k2 achieved the highest accuracy in predicting change trends 1 and 2, followed in turn by BDe and BIC. The accuracy ranking of the three scoring functions was exactly the opposite in predicting change trend 0. The accuracy difference between the three scoring functions fell between 0.1~0.2%. As for variance, the three scoring functions were ranked in ascending order as: k2, BD2 and BIC. The results are attributable to the penalty term in the BIC function and the sample size. The growing number of time points increases the sample size, leading to fluctuations in prediction accuracy.

Figure 1 illustrates the overall prediction accuracy of each scoring function. As shown in Figure 1, with the growing number of time points, all three scoring functions exhibited a decline in the overall prediction accuracy. The trends of the BIC and the BDe were basically the same, and slightly more stable than that of k2.

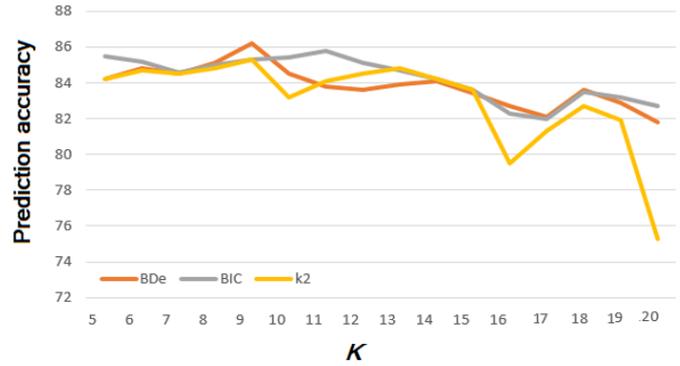


Figure 1. Overall prediction accuracy of each scoring function

Table 3 details the prediction accuracy of the BDe scoring function.

Table 3. The prediction accuracy of the BDe scoring function

Number of time points	Scoring function	Overall accuracy	Accuracy of change trend 0	Accuracy of change trend 1	Accuracy of change trend 2	Structural score
5	BDe	84.57	5.25	95.11	65.64	-6053.1
6	BDe	84.48	8.78	94.12	70.83	-7094.2
7	BDe	84.47	12.55	93.47	72.66	-8102.4
8	BDe	85.86	16.76	94.59	72.66	-9067.8
9	BDe	85.46	16.75	94.61	71.75	-9952.2
10	BDe	84.72	9.41	94.34	73.56	-10778.9
11	BDe	83.21	7.63	92.88	72.87	-11622.6
12	BDe	83.99	11.48	93.52	72.32	-12378.4
13	BDe	83.41	13.35	92.98	72.32	-13101.5
14	BDe	83.98	11.51	94.32	72.32	-13752.9
15	BDe	83.25	7.67	94.21	72.32	-14333.8
16	BDe	82.88	11.88	93.11	72.41	-14892.5
17	BDe	78.92	17.62	92.26	71.68	-15410.6
18	BDe	82.11	17.65	92.22	71.68	-15821.6
19	BDe	80.62	12.10	92.02	69.56	-16291.7
20	BDe	79.32	19.43	90.53	61.77	-16552.9

As shown in Table 3, the BDe scoring function achieved an overall accuracy of 82.30%. It correctly predicted 12.49% of change trends 0, 93.39% of change trends 1, and 69.42% of change trends 2. Considering each accuracy, it is recommended to select 8 time points (two days) to predict the change trend of air ticket price in the next time point. In other words, the change trend of air ticket price of the current day can be predicted accurately based on that of the previous 2 days.

5. COMPARATIVE ANALYSIS

This subsection introduces the NN model to predict the change trend of air ticket price of the said flight, and compares the prediction accuracy with that of the BN model. The NN

model was selected as a contrastive method, because it is an excellent tool for classification and prediction and a directed graph, too.

Before using the NN model for learning, the first step is to set the number of hidden layers. Experimental results show that, the more the number of hidden layers, the harder it is for the NN to converge to the global optimal solution. In addition, the NN with a single hidden layer has enough learning ability. Therefore, the NN model adopted for comparison contains only one hidden layer.

Next, the NN model was tested with different number of hidden layer neurons. It is found that the number of hidden layer neurons has little effect on the prediction results. With the growing number of hidden layer neurons, the number of iterations of the NN learning soared, while the error slowly decreased. The best prediction effect was achieved when there

were three neurons on the hidden layer.

Hence, an NN model with three neurons in one hidden layer was adopted to learn the data on air ticket price collected at 20 time points, and used to predict the number of remaining seats in the next time point. The average of the results in ten experiments was taken as the final prediction.

The prediction abilities of the NN model and the BN model were compared from three aspects. In terms of structure, the

BN clearly illustrates the relationship between variables, with the connection parameter between nodes being conditional probability. The structure of the NN presents the relationship between neurons, with the connection weight being the coefficient of the connection between neurons. However, the NN structure cannot reflect the relationship between variables.

The prediction accuracy of the number of remaining seats is shown in Table 4.

Table 4. The prediction accuracy of the NN model for the number of remaining seats

Remaining seats?	Overall accuracy	Accuracy of change trend 0	Accuracy of change trend 1	Accuracy of change trend 2
Yes	74.18	16.26	88.36	42.39
No	76.30	20.86	82.19	51.07

As shown in Table 4, the NN model achieved an accuracy of 74.18% and 76.30%, when there are seats remaining and no seat remaining, respectively. Obviously, the overall accuracy, accuracy of change trend 0, and accuracy of change trend 2 were higher in the case with no remaining seat than the case with remaining seats. This confirms our conclusion based on the BN model that remaining seats have no positive effect on the prediction results, but will increase the model complexity.

Finally, the BN model and the NN model were compared in prediction accuracy. Based on eight time points, 70% of the sample data were taken as the training set and the remaining 30% as the test set. Ten experiments were conducted for each model and the average was taken as the final prediction. The prediction accuracies of the two models are compared in Table 5.

Table 5. The prediction accuracies of the BN model and the NN model

Model	Overall accuracy	Accuracy of change trend 0	Accuracy of change trend 1	Accuracy of change trend 2
The NN model	82.36	16.72	90.52	70.72
The BN model	84.72	14.35	94.02	73.37

As shown in Table 5, the BN model outperformed the NN model in all cases, except the accuracy of change trend 0. This means our model is better than the NN model in predicting the change trend of air ticket price.

6. CONCLUSIONS

Taking air ticket as an example of tourist consumption, this paper establishes a prediction model for air ticket price based on the BN model. The model forecasts the change trend of air ticket price in the next moment through analyzing the historical data. Only two variables were considered in the model, i.e. the number of the remaining seats and the ticket price. Experimental results show that the BN model achieved a desirable accuracy of 80%. According to big data analysis, the fluctuation of the air ticket price in the next moment can be predicted accurately based on the price trends in the previous two days. The research results lay the basis for predicting the price volatility of the other product/service of tourism.

REFERENCES

[1] Saayman, A., Cortés-Jiménez, I. (2013). Modelling Intercontinental tourism consumption in south Africa: A systems - of - equations approach. *South African Journal of Economics*, 81(4): 538-560. <http://dx.doi.org/10.1111/saje.12018>

[2] Pathompituknukoon, P., Khingthong, P., Suriya, K. (2012). Can rising tourism income compensate fading agricultural income? A general equilibrium analysis of

income distribution and welfare in a rural village in Northern Thailand. *The Empirical Econometrics and Quantitative Economics Letters*, 1(1): 5-16.

[3] Magnini, V.P., Honeycutt Jr, E.D., Hodge, S.K. (2003). Data mining for hotel firms: Use and limitations. *Cornell Hotel and Restaurant Administration Quarterly*, 44(2): 94-105. [http://dx.doi.org/10.1016/S0010-8804\(03\)90009-7](http://dx.doi.org/10.1016/S0010-8804(03)90009-7)

[4] Vojinovic, Z., Kecman, V., Seidel, R. (2001). A data mining approach to financial time series modelling and forecasting. *Intelligent Systems in Accounting, Finance & Management*, 10(4): 225-239. <http://dx.doi.org/10.1002/isaf.207>

[5] Lee, T.C., Hersh, M. (1993). A model for dynamic airline seat inventory control with multiple seat bookings. *Transportation Science*, 27(3): 252-265. <https://doi.org/10.1287/trsc.27.3.252>

[6] Danziger, S., Israeli, A., Bekerman, M. (2006). The relative role of strategic assets in determining customer perceptions of hotel room price. *International Journal of Hospitality Management*, 25(1): 129-145. <http://dx.doi.org/10.1016/j.ijhm.2004.12.005>

[7] Talluri, M., Kaur, H., He, J.S. (2015). Influence maximization in social networks: Considering both positive and negative relationships. In *2015 International Conference on Collaboration Technologies and Systems (CTS)*, 479-480. <https://doi.org/10.1109/CTS.2015.7210473>

[8] Wu, L., Zhang, J. (2013). Inbound tourism based on China's inbound tourism receipts that eliminated the direct price effect. *Tourism Tribune*, 28(3): 29-37. <https://doi.org/10.3969/j.issn.1002-5006.2013.03.004>

[9] Lai, T.M., To, W.M., Lo, W.C., Choy, Y.S. (2008).

- Modeling of electricity consumption in the Asian gaming and tourism center-Macao SAR, People's Republic of China. *Energy*, 33(5): 679-688. <http://dx.doi.org/10.1016/j.energy.2007.12.007>
- [10] Collum, K.K., Daigle, J.J. (2015). Combining attitude theory and segmentation analysis to understand travel mode choice at a national park. *Journal of Outdoor Recreation and Tourism*, 9: 17-25. <http://dx.doi.org/10.1016/j.jort.2015.03.003>
- [11] Mustelier-Puig, L.C., Anjum, A., Ming, X. (2018). Interaction quality and satisfaction: An empirical study of international tourists when buying Shanghai tourist attraction services. *Cogent Business & Management*, 5(1): 1470890. <http://dx.doi.org/10.1080/23311975.2018.1470890>
- [12] Yu, F.L., Huang, Z.F., Lu, L. (2016). Characteristics and influence mechanism of rural households' tourism destination choice behavior in developed regions: A case study of South Jiangsu. *Acta Geographica Sinica*, 71(12): 2233-2249. <https://doi.org/10.11821/dlxb201612012>
- [13] López, V., Del Río, S., Benítez, J.M., Herrera, F. (2015). Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data. *Fuzzy Sets and Systems*, 258: 5-38. <https://doi.org/10.1016/j.fss.2014.01.015>
- [14] Borsuk, M.E., Stow, C.A., Reckhow, K.H. (2004). A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, 173(2-3): 219-239. <http://dx.doi.org/10.1016/j.ecolmodel.2003.08.020>
- [15] Scutari, M., Brogini, A. (2012). Bayesian network structure learning with permutation tests. *Communications in Statistics-Theory and Methods*, 41(16-17): 3233-3243. <https://doi.org/10.1080/03610926.2011.593284>
- [16] de Morais Andrade, P., Stern, J., de Bragança Pereira, C. (2014). Bayesian test of significance for conditional independence: The multinomial model. *Entropy*, 16(3): 1376-1395. <http://dx.doi.org/10.3390/e16031376>
- [17] Yang, S., Chang, K.C. (2002). Comparison of score metrics for Bayesian network learning. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 32(3): 419-428. <http://dx.doi.org/10.1109/ICSMC.1996.565479>
- [18] Posada, D., Buckley, T.R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53(5): 793-808. <https://doi.org/10.1080/10635150490522304>
- [19] Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J. C., Taylor, C.J. (2002). A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5): 525-537. <http://dx.doi.org/10.1109/TMI.2002.1009388>
- [20] Peng, H., Jin, Z., Miller, J.A. (2017). Bayesian networks with structural restrictions: parallelization, performance, and efficient cross-validation. In 2017 IEEE International Congress on Big Data (BigData Congress), 7-14. <http://dx.doi.org/10.1109/BigDataCongress.2017.11>