# Study on tourism crisis alarming model based on data mining

## Zhenhui Dong*, Fang Liu

*Gansu industry Polytechnic College Gansu tianshui 741025, China*

*\*Corresponding author's e-mail: dongjun@163.com*

**Abstract**

The article systematically analyzed the actuality and modality of Chinese listed corporations' tourism crisis, and discussed the reason of this problem appeared and the method how to resolve. The necessities of establish forewarning model in forecasting and avoiding tourism crisis were clarified; Tourism crisis forewarning models in existence were reviewed; several forewarning models using Enterprise Miner module in SAS were established. Compared with the existing models, such as discriminate analysis models, Logistic models and Neural Network models, the models in this article possessed many advantages and could be used more perfectly on forewarning purpose.

*Keywords:* data mining, tourism crisis, alarming model, monitor system

## 1 Current matters of Chinese tourism crisis

With the rapid development of tourism industry and the growth of risk that our human society have been up against, various crisis tampers with the working order of tourism industry and economy frequently. Because of the high industry-relationship character of the tourism industry and the influence of "butterfly effect", a tiny crisis case will bring on the large-scale decline of tourism industry and economy [1]. The tourism crisis management becomes one of the most popular research subjects. Based on the different character and effect system of tourism crisis in different tour destinations, a territorial tourism crisis management will make great significance to the development of the tourism and economy.

There are several definitions on tourism crisis in China and abroad, most of them give this concept from modality of tourism crisis, such as bankrupt, default preference stock dividend and incapacity to liquidate debt. Corporations are considered plunged in tourism crisis in case one of above phenomenon occurred. The humidity of tourism crisis is a violent eruption of tourism risks with the characters of extremely worsens tourism status, payment crisis and even bankruptcy [2].

Tourism status of Chinese listed corporations should be concerned. Such matters as blow might appeared in some degree if the tourism crisis happened: incapacity repaying debt, excessive cashing out, lacking cash payment, mint investment but no income product no market, merchandise keeping long in stock, main operation severely shrinking, involved in lawsuit payment, etc

The reason of Chinese listed corporations' tourism crisis happened included such factors as nation's macro management and macro-economy factor, security market limitation factor and corporate governance factor.

Chinese economic and account law system were imperfect and lacking of law basis in the process of necessary management and supervising. Sometimes nation's tourism policies and domain policies were even likely to solicit tourism crisis in related corporations. Depression and prosper of Macro economy would influence the status of corporations greatly.

In conclusion, management and supervise system of nation should be perfected and the active measures should also be taken by corporations according to the variation of nation's policy and macro-economical environment.

With the rapid development of Information Technology and significant improvement of management theory research, Information Technology has been paid more and more attention in the area of enterprise management decision. Confronting the fierce competition in the market, [3] the enterprise makes increasing requirements for risk management, therefore how to evaluate the existing tourism risks objectively and timely forewarning in the enterprise management process becomes the goal the enterprise is always seeking for. The traditional methods for tourism risk analysis and forewarning include statistical analysis method and neural network model, however with the enlargement of the enterprise in scale and more frequent information disclosure, the traditional statistical analysis methods have been unable to meet the requirement of massive data analysis; also the neural network model doesn't consider the time continuity of tourism data. In addition, the enterprise tourism risk analysis and forewarning research is influenced by a variety of internal and external factors with high uncertainty, however, the excellent performance of data mining technology in uncertainty theoretical study makes them closely linked. In order to solve the above problems existed in the traditional methods, after further research on the association rule mining methods, we present 3 improved algorithms based on new association rules, which improved the mining efficiency greatly.

Poor efficiency of the information in security market and the defect in the process of stock coming into the market resulted in many troubles such as falsehood action of listed corporations and failure of annex function between corporations [4]. These limitation factors in security market resulted that neither sides of the interest groups done their best to increase corporation's value.

Some measures to perfect Chinese security market should be taken as follows: perfecting stock coming system, establishing and perfecting stock law, setting up the exiting

mechanism based on the variation of market and the system, taking efforts to develop institution investor.

## 2 Governance factor

Defect of corporation governance factor is one of fundamental reasons in Chinese listed corporation tourism crisis. The structure of corporate ownership affected the formation, the operation and the performance of corporate governance. Differ corporate ownership structure cause differ corporate governance. The shortcomings of Chinese listed corporations' ownership structure were presented as follows: administrative decline possessed by national shares holders, excessive proportion of national shares which causes the biggest shareholders controlled the corporations, non-standard operation of directorate, lacking of inspiring and restricting mechanism. It's the main reason of

Chinese fisted corporations' tourism crisis that the irrationality of the ownership structure affected the value of corporation and the shareholders' benefit.
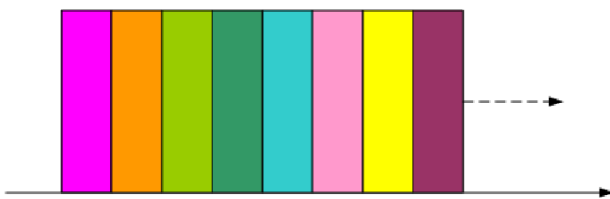


FIGURE 1 Incremental mining strategy based on time series

Incremental mining strategy based on time series is shown in Figure 1. Some tasks should be done to perfect Chinese listed corporations' ownership structure: reducing the proportion of national shares, setting up independent director system, establishing manager stock future system.

Tourism crisis forewarning models could forecast and control such matters as management failure and tourism misplaying, and it also provide important information's to decision-maker, investor and debate. It's necessary to establish tourism crisis forewarning models to perfect stock market.

*Data mining.*

Most tourism crisis forewarning models are based on Data Mining techniques. Data Mining can pick-up concealed, unknown but valuable information from those abundant imperfect, noise, fuzzy and random data, and conclude this data into structure models. Data Mining emphasizes particularly on the future instance and the reason of those happened. Data Mining is a cross subject, so it's combination of several speciality techniques. The task of Data Mining includes description, association analysis, classification and prediction, clustering and outlier.

*Data Mining Technology.*

Data mining techniques can be classified into Stat. Analysis Data Mining techniques, knowledge discovery Data Mining techniques and other Data Mining techniques.

1) Stat. analysis Data Mining techniques include regression analysis, time-series analysis and multi analysis, etc. Many Data Mining models are established based on Stat. Analysis. Stat. Analysis is a success Data Mining technique [5].

2) Knowledge discovery Data Mining techniques

include instance-based learning, Neural Network, decision tree, expert system, probability rule, fuzzy logic and evolutionary Computation, ete. Application process of Data Mining include confirming object, data preparation, founding model, data mining, analyzing result and knowledge application.

$$P(X_i, \beta) = F(\alpha + \beta X_i) \tag{1}$$

$$F(\alpha + \beta X_i) = l / (l + e^{-(\alpha + \beta X_i)}) \tag{2}$$

where $P(X_i, \beta)$ is the probability in Equation (1), $\alpha$ and $\beta$ are the indexing, respectively, of the sample, as shown in Equation (2).

$$L(\alpha, \beta) = \prod_{i=1}^{m} \left[ 1 - \frac{1}{1 + e^{(-\alpha - \beta X_i^{(1)})}} \right] \cdot \prod_{i=1}^{n} \left[ 1 - \frac{1}{1 + e^{(-\alpha - \beta X_i^{(2)})}} \right] \tag{3}$$

where $L$ is the sum of alarming value, is shown in Equation (3).

## 3 Tourism crisis model

Discriminate analysis distinguished which sort of the research objects belonged to, and it was used in the entire early tourism crisis forewarning models. Early tourism crisis forewarning models include single variable models and multi-variable models. Single variable models forecasted tourism crisis through analyzing single tourism ratios' trend, it was effect in some way but existed limitations. Multi-variable models are a multi member function that is adding powered multi-kind tourism ratios to produce overall score for forecasting the possibility of tourism crisis. The first multi-variable models for forecasting tourism crisis is Z-score model established by American scholar, Altman (1968) [6]. It's a much influenced model in tourism crisis forewarning. Exploiter bank of Japan model and F-score model are all based on Z-score model.

There are two unconquerable limitation in discriminate analysis tourism crisis forewarning models:

1) Fixed influence. Boundary effect of any explaining variables should be fixed. This problem existed in linear functions, but it's unconformity to reality.

2) Entirely linear compensate matter. This matter exists in any linear functions. In order to overcome these limitations, logistic regression models were established and solved this problem [7].

$$R = \sum_{k=1}^{d-1} \left[ \binom{d}{k} \times \sum_{j=1}^{d-k} \binom{d-k}{j} \right] = 3^d - 2^{d+1} + 1 \tag{4}$$

Then, we use Equation (4) to draw the Apriori, IUA, HIUA and compared them with each other as shown in Figure 2 and Figure 3.
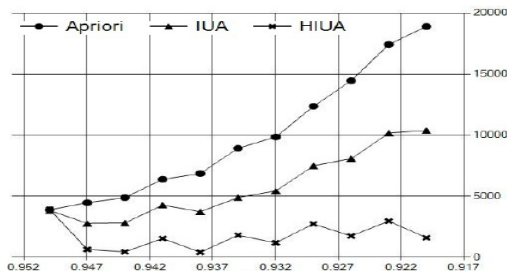
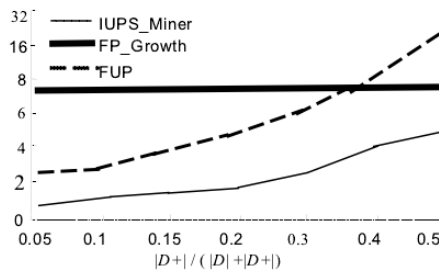FIGURE 2 The data set on the algorithm performance comparison



FIGURE 3 Comparison of three algorithms

## 4 Software

157 industrial corporations in Shanghai Stock Security Market were selected and 43 tourism indexes were calculated according to the tourism reports on five time spots. Data was input to SAS and analyzed, several models were established.

Single index only reflect the side face of tourism status but not panorama, so it's limited to judge those corporations befalling in tourism crisis depending only on single index or those corporations in "Special Treatment".

Principal Component Analysis is one of reduced dimension Stat. By using the least integration indexes, it can reflect the most information provided by original data. In this article, integration evaluation system of listed corporations' tourism status was established using Principal Component Analysis and every corporation' tourism status integration score was presented. This evaluation system is more precise than that system evaluating listed corporations' tourism status by certain tourism index or standard that weather the corporation is in Special Treatment.

$$r_{x,y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2} \cdot \sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}} \quad (5)$$

Then, we use Equation (5) to calculate the evaluation value as below [8].

$$E_{11} = \sum_{k=1}^{3} E_{11k} A_{11k} = 4\times 0.655 + 5\times 0.250 + 3.5\times 0.095 = 4.203$$

$$E_{12} = \sum_{k=1}^{2} E_{12k} A_{12k} = 4\times 0.75 + 5\times 0.250 = 4.250$$

$$E_{13} = \sum_{k=1}^{2} E_{13k} A_{13k} = 5\times 0.167 + 4.8\times 0.833 = 4.833$$

$$E_{21} = 4.7, E_{22} = 4.671, E_{23} = 1.5, E_{31} = 4.529, E_{32} = 4.604$$

The most influenced 19 tourism indexes were confirmed by Principal Component Analysis' characteristic vector and principal components' contribution ratios Five quarters integration scores shows that nearly half of the corporation's integration scores were negative in either quarter. A corporation has a negative integration score did not means definitely that the corporation was in tourism crisis, but it should be taken more attention.
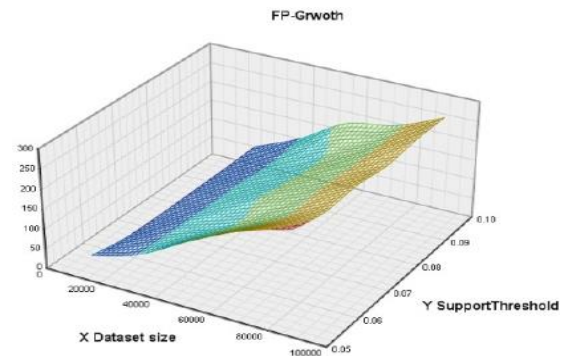


FIGURE 4 Run time

Run time of software is shown in Figure 4. Two sorts of listed corporations' tourism status in forecasting spot were marked based on the results of Principal Component Analysis. These two sorts were taken as forecasting target variables, and the 19 tourism indexes were taken as input variables. Logistic Regression and Decision Tree were used to forecast listed corporations' tourism status. Decision Tree is an inductive arithmetic based on instance, and it can deduced classification rules from a sort of cases without order and regulation. Three kinds of Decision Tree arithmetic were adopted in this article. Three kinds of arithmetic were all creditable because little difference of error rates could be found among them. Decision Tree arithmetic is a favourable method because it estimate listed corporations' tourism status expeditiously according to a few of tourism index. Those corporations might not in the status of tourism crisis definitely even though they were labelled poor tourism status, it's necessary to analyze further. In conclusion, the error rate of Decision Tree was less than that of Logistic, and the forecasting effect of Decision Tree was better than that of Logistic.

Neural Network is information operation system that can imitate brain's structure and function. In Linear Regression and Neural Network models, every corporation's integration score was taken as target variables and the 19 tourism indexes were taken as input variables. Linear Regression model is simple and efficient, and it can estimate tourism status rapidly. The forecasting effect of Neural Network is better than that of Linear Regression.

## 5 Interactive mining algorithms

The present association rules mining algorithms are mainly based on the support and confidence framework. For the same database, different support and confidence threshold will generate different frequent item sets and different numbers of association rule with the same algorithm. Since the users cannot in advance know which support-confidence threshold is appropriate, they need to constantly test

different thresholds to get the ideal frequent item sets and association rule. The new algorithm is aimed to improve the association rule maintenance issues when the support threshold is changed. In other words, the previous algorithm involves scanning the database multiple times and repeated calculation issues while the users adjusted the threshold, so Hash based association rule interactive mining algorithm HIUA is proposed, which improves the pruning process of the original IUA algorithm and use Hash structure to quickly access to the support counting during the execution. By this way, the algorithm efficiency is improved.
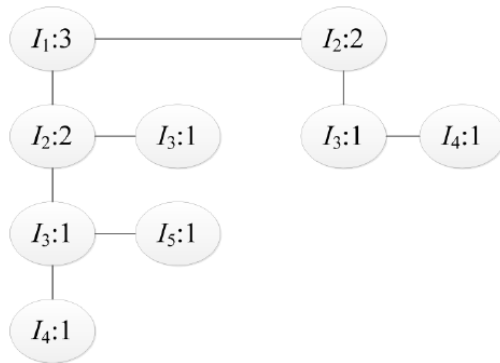


FIGURE 5 PS-Tree

In general, association rule mining algorithm assumes that the database is static, in the condition of specifying fixed threshold; it needs re-scan the database to compute new association rules once the database has been updated. Towards the above-mentioned association rule maintenance issues, we present an efficient association rule incremental updating algorithm based on PS-Tree (Figure 5) (IUPSES Miner) that only needs mining the new database. By merging the retained PS-Tree with the new PS-Tree to reduce the cost of scanning the original database and repeated calculation of associations, it efficiently maintained the previously discovered association rules and improved the algorithm efficiency.

The dynamic maintenance of association rules refers to the maintenance and update issues of association rules when both the database and support threshold are changed at the same time. The present mining methods usually have the problems that involve multiple scanning of database or repeatedly traversing the complex structure. In this paper, for the situations the database and support threshold are changed simultaneously, we present an association rule dynamic maintenance algorithm incremental mining, which uses the based on interactive mining already generated associations incremental mining and interactive mining. Basically the algorithm applies interactive mining to the original databases and then applies incremental mining to the new added databases using the new support threshold; furthermore, the efficiency is optimized and improved with Hash structure and pattern growth methods.

The objective of the enterprise tourism risk research is to constructing tourism risk index system, and then determines the high support patterns in the index system to help for the enterprise management decision. The traditional method for enterprise tourism risk is usually based on statistical analysis model, which drawback is many assumptions so that it cannot process mass data. Aim at the above-mentioned problems, association rule interactive mining is proposed in this paper, which choose multiple wider ranges of tourism indexes first, and then ultimately determine the most representative tourism risk indicator by mining the rules between all tourism indexes. The detailed steps can refer to the following: firstly establish tourism risk index system in which the selection of tourism index is based on variable correlation analysis; then build a risk conception hierarchical tree to find the rules between the finance risk indicators with interactive mining strategies of decreasing support threshold; finally, select the ST companies in the domestic listed companies for the empirical research of enterprise tourism risk analysis, and propose 10 key indicators that influence the enterprise tourism risk and suggestions for avoiding tourism risks.

The research of tourism crisis forecasting mainly focuses on tracking the tourism index fluctuations and trends, and the system is supposed to provide warning alert once the tourism index fluctuates beyond a certain range. The existing methods for tourism crisis forecasting are mainly based on artificial intelligence data mining models, which own the drawback that doesn't take the time continuity of the tourism index data into account. In this paper, considering the time series characteristic for the tourism index, we present a dynamic-maintained enterprise tourism crisis forecasting model based on time series. The concrete steps refer to the following: firstly construct the tourism data mining model based on time series; then based on time series incremental mining and interactive mining mechanism, find the rules between the tourism index and predict the development trend for the crisis enterprise with association rule dynamic maintenance mining; finally, we select the ST companies in the domestic listed companies for the empirical research of enterprise tourism crisis forecasting, and determine the key indicator to define different phase of crisis enterprise.

## 6 Conclusion

An integration score system of listed corporations' tourism status was established using Principal Component Analysis. Compared with the existing systems, the system adopted more tourism indexes and was more niceties.

The most influenced 19 tourism indexes were defined using characteristic vector matrix and Principal Components' contribution ratios in Principal Component analysis. The 19 tourism indexes involved the main seven aspects of listed corporation's tourism status, and included almost all indexes relating to cash flux and corporation grown.

Regression and Linear Regression forewarning model of tourism crisis were set up using the 19 tourism indexes. The veracity of the models was better than that of the existing models.

Decision Tree was used to forewarn tourism crisis, and three sorts of decision rule were defined using three decision tree arithmetic. This decision tree rules could be used conveniently in judging corporations' tourism status. A Neural Network model was established. This model is advanced to the existing models. The existing models' input variables were chosen by subjective method by contrast，the 19 tourism indexes were taken as input variables

objectively in this article. On the other hand, the existing models status between good and bad, but the model in only distinguished the future tourism the article forecasted tourism status' integrated scores.

## References

[1] Gui Y, Mei H, Chen Y 2010 An Analysis of the Causes and Enlightenment of the Us Subprime Crisis: An Asset Securitization Perspective *Proceedings of the Fourth International Conference on Management Science and Engineering Management Nov 15-17 2010 Taiwan*

[2] Mukuddem-Petersen J, Mulaudzi M P, Petersen M A, Schoeman I M. 2010 Optimal Mortgage Loan Securitization and the Subprime Crisis *Optimization Letters* **4**(1) 97-115

[3] Dias J 2012 Sovereign Debt Crisis in the European Union: A Minimum Spanning Tree Approach *PhysicaA: Statistical Mechanics and Its Applications* **391**(5) 2046-55

[4] Huang S-C 2011 Using Gaussian Process Based Kernel Classifiers for Credit Rating Forecasting *Expert Systems with Applications* **38**(7) 8607-11

[5] Wang Y, Wang S, Lai K K 2011 Measuring Tourism Risk with Generalized Asymmetric Least Squares Regression *Applied Soft Computing* **11**(8) 5793-800

[6] Sivapathasekaran C, Mukherjee S, Ray A, Gupta A, Sen R 2010 Artificial Neural Network Modeling and Genetic Algorithm Based Medium Optimization for the Improved Production of Marine Biosurfactant *Bioresource Technology* **101**(8) 2884-7

[7] Yin Y, Gong G, Han L 2011 Theory and Techniques of Data Mining in of Behavior Modeling *Science China: Information Sciences* **54**(4) 717-31

[8] Ablanedo-Rosas J H, Gao H, Zheng X, Alidaee B, Wang H 2010 A Study of the Relative Efficiency of Chinese Ports: A Tourism Ratio-Based Data Envelopment Analysis Approach *Expert Systems* **27**(5) 349-62

## Authors

**Dong Zhenhui, May, 1982, Lixian County in Gansu province, China.**

**Current position, grades**: lecturer in Gansu Industry Polytechnic College.
**University studies**: MSc degree at Tourism Management in Lanzhou University in 2007.
**Scientific interest**: tourism education and historical and cultural landscape research.
**Experience**: teaching tourism courses in Gansu Industry Polytechnic College in 2007, 3 scientific research projectsю

**Liu Fang, Jan, 1986, Tianshui in Gansu Province, China.**

**Current position, grades**: lecturer in Gansu Industry Polytechnic College
**University studies**: MSc degree at English Language and Literature in Northwest Normal University in 2009.
**Scientific interest**: English language literature and tourism English education.
**Experience**: teaching tourism English and tourism courses in Gansu Industry Polytechnic College in 2007, 2 scientific research projectsю

**Operation Research and Decision Making**