International Journal of Intelligent Information and Management Science
ISSN: 2307-0692, Volume 7, Issue 2, April, 2018

Dynamic Trend Simulation Analysis of Tourism Big Data based on Feedback Constraint Association Rules

Ruiping Chen
Foshan Polytechnic, Foshan, 528137, China

Abstract: In order to improve the ability of tourism big data to predict and evaluate, realize the quantitative statistical evaluation of tourism information, and thus to guide tourism planning and decision-making, A dynamic trend prediction algorithm for tourism big data is proposed based on feedback constraint association rules. The statistical prior information of tourism big data is constructed by the autosimilar regression model. In the autosimilar regression model, the empirical modal decomposition of the tourism big data statistical distribution series is carried out. The feature extraction method of feedback constraint association rules is used to analyze the features of tourism big data, and the BP neural network classification model is used to deal with the feature information clustering and information fusion of tourism big data. The descriptive statistical analysis method is used to analyze the statistical characteristics of tourism big data, to construct the statistical characteristic quantity of the tourism big data distribution, and the principal component analysis is used to mine the association rules of tourism big data. The prediction of tourism big data's dynamic trend is realized. The simulation results show that the proposed method has high accuracy, good convergence and good prediction accuracy for the dynamic trend of tourism big data.

Keywords: tourism big data; feedback constraints; association rules; data mining; statistical analysis; feature extraction

1. Introduction

With the continuous improvement of people's living standards, people's enthusiasm for tourism is also rising, especially in recent years, in the national policy of macro-guidance and the implementation of the reform of holidays and holidays, for the tourism industry in all parts of the world. The tourism industry has been unprecedented development. As a tertiary industry, it has a great potential for development in the domestic and international market. In order to improve the ability of tourism big data to predict and evaluate, realize the quantitative statistical evaluation of tourism information, and thus to guide tourism planning and decision-making, it needs to study and prediction the changing and increasing situation of tourism big data scientifically and quantitatively, provide the scientific and effective reference index for the government and related department. It has great significance to establish the trend prediction model of tourism big data for the macro decision of tourism department.

The tourism big data prediction is constructed based on the comprehensive data of the total amount of tourism big data growth and other parameters, giving full consideration to the region, politics, and economy[1]. According to the value of tourism big data in the past and using certain mathematical model, the influence of culture can predict the trend of economic change in a certain period of time in the future. Due to the fact that there are a lot of non-lines such as human geography and so on, the tourism big data has been subjected to a lot of non-line data. The prediction algorithm needs to eliminate the information redundancy generated by many historical data, and thus effectively obtains a set of optimal and appropriate factors to capture the growth trend of tourism big data, and to obtain a correct tourism strategy[2]. The construction of tourism big data's prediction model is basically the process of predicting a group of time series. Traditionally, it is taken based on the linear model or equivalent approximate linear model to predict and study the tourism big data series. However, it is not accurate to approximate the statistical sequence of tourism big data's dynamic trend to artificial linear relationship, as a tourism big data model, it is influenced by local political, economic, cultural and other factors. The prediction mode is a real nonlinear model. Therefore, the traditional method cannot fit the prediction model of tourism big data effectively and accurately, and the prediction accuracy is poor. Thus, the related literature has improved the design of the model, in which, in reference [4], a dynamic trend prediction model of tourism large data based on support vector machine model is proposed. Based on the SVM model[3], the PCA analysis method is used to effectively detect and decompose the dynamic trend of the large tourist data, and improve the filtering and predic-

tion ability of the data, but the method has a large overhead. The real-time performance is not good. Association rules mining and feature decomposition are used for the dynamic trend of large tourist data to improve the ability of data filtering and prediction. However, this method has a large overhead and poor real-time performance[4-6].
Aiming at the problems of traditional methods, this paper proposes a dynamic trend prediction algorithm for tourism big data based on feedback constraint association rules. The statistical prior information of tourism big data is constructed by the autosimilar regression model. In the autosimilar regression model, the empirical modal decomposition of the tourism big data statistical distribution sequence is carried out, and the feature extraction method of the association rules with feedback constraints is adopted. The method is used to analyze the characteristics of tourism big data, the BP neural network classification model is used to deal with the feature information clustering and information fusion of tourism big data, and the descriptive statistical analysis method is used to carry out the statistical features of tourism big data. This paper analyzes, constructs the statistical characteristic quantity of tourism big data distribution, uses principal component analysis to mine the association rules of tourism big data, and realizes the dynamic trend prediction of tourism big data. Finally, the experimental analysis shows the superiority of this method in improving the prediction ability of tourism big data's dynamic trend.

2. Model Construction and Statistical Sequence Analysis

2.1. Construction of tourism big data autosimilar regression model

The autosimilar regression model of the prior information of tourism big data is constructed, and the historical data of tourism big data is assumed to be $\{x_i\}_{i=1}^{N}$. In order to remove the dimension of the original data and the bad factors such as the excessive amplitude of the original data, the original model is given. Tourism big data can be regarded as a group of nonlinear statistical distribution series. The method of nonlinear statistical distribution sequence analysis is used to analyze the trend of tourism big data[7], carry on the tourism big data's unification. According to the analysis, a multivariate statistical characteristic equation is used to describe the fitting state model of tourism big data as follows:

$$\begin{pmatrix} X \\ P(X) \end{pmatrix} = \begin{Bmatrix} a_1, & a_2, & \mathbf{L}, & a_m \\ p(a_1), & p(a_2), & \mathbf{L}, & p(a_m) \end{Bmatrix} \quad (1)$$

Where, $0 \le p(a_i) \le 1 (i = 0,1,2,\mathbf{L},m)$,and $\sum_{i=1}^{m} p(a_i) = 1$, it represents tourism big data's autoregressive statistical characteristic parameter. By decomposing the solution vector of the statistical equation, the principal component a_{ii} of statistical characteristic information a is obtained, and the covariance matrix C of tourism big data is calculated:

$$C = \frac{1}{N}\left[X - \overline{X_l}\right]\left[X - \overline{X_l}\right]^T \quad (2)$$

Where

$$l = [1,1,\mathbf{L}\,1]_{1\times N} \quad (3)$$

$$\overline{x}_i = \frac{1}{N}\sum_{k=1}^{N} x_{ik} \quad (4)$$

$$X = [X_1, X_2, \mathbf{L}, X_m] \quad (5)$$

The association rule fusion method is taken[8], the characteristic equation of tourism data dynamic trend prediction is obtained:

$$(l\,l - S)U = 0 \quad (6)$$

According to the contribution degree of cumulative variance, the number of principal components is selected. Only when the cumulative contribution rate reaches a certain amount, the corresponding m principal components can be regarded as the principal components to be selected. Including all the state information parameters provided by the original data, the number of principal components is obtained to be m. in the autosimilar regression model, the empirical mode decomposition of tourism big data statistical distribution sequence is carried out to determine the principal component value. For the subsequent support vector machine SVM training input[9-11].

2.2. Statistical distribution sequence analysis of tourism big data

On the basis of constructing an autosimilar regression model for the statistical tourism big data prior information, the empirical mode decomposition of the tourism large data statistical distribution sequence is carried out in the regression model[12], so that the distribution characteristic information entropy of the tourism large data is obtained as follows:

$$H(X) = E(I(a_i)) = -\sum_{i=1}^{m} p(a_i)\log_2 p(a_i) \quad (7)$$

The random analysis model is used, and the expression of the cumulative quantity of tourism big data statistical distribution sequence x is obtained as follows:

$$C_{or3} = \frac{\left\langle \left(x_n - \overline{x}\right)\left(x_{n-d} - \overline{x}\right)\left(x_{n-D} - \overline{x}\right)\right\rangle}{\left\langle \left(x_n - \overline{x}\right)^3 \right\rangle} \quad (8)$$

Wherein, x_n denotes the element of tourism big data information statistics, d denotes the sampling statistical delay term of tourism big data, $D = 2d$, $\bar{x}$ denotes the predicted principal component factor of tourism big data, and $\langle x(n)\rangle$ represents the mean value of $x(n)$:

$$\langle x(n)\rangle = 1/N\sum_{n=1}^{N} x(n) \tag{9}$$

For a continuous tourism big data statistical distribution sequence, the feature training subset $S_i(i=1,2,\mathbf{L},L)$ of each spatial solution vector for tourism big data meets the following conditions:

(1) $\mathbf{\Sigma} = diag(d_1,d_2,\mathbf{L},d_r)$, $d_i = \sqrt{l_i}$, $\forall i \neq j$

(2) $\cup_{i=1}^{L} S_i = V - v_s$;

(3) Set $x_{n+1} = mx_n(1-x_n)$ is the conjugate solution of a statistical model of tourism big data's statistical distribution sequence. It satisfies the eigenvalue decomposition condition $U = \{u(t)|u(t) \in X, \|u\| \le d, t \in I\}$, where $(I_i)_{i\in N} = \{x_1,x_2,\mathbf{L},x_m\}$. For a group of tourism big data statistical sequences with multivariate variables, the statistical characteristics of the fractional generalized integro-differential equation for describing tourism big data $x(n)$ are as follows:

$$\begin{aligned} &c_{1x}(t) = E\{x(n)\} = 0 \\ &c_{2x}(t) = E\{x(n)x(n+t)\} = r(t) \\ &c_{kx}(t_1,t_2,\mathbf{L},t_{k-1}) \equiv 0 \quad (k \ge 3) \end{aligned} \tag{10}$$

While $q = 2$, the tourism big data information vector satisfies the continuous functional condition of Bernoulli space for the constraint function of differential equation, where the constraint condition is:

$$\Psi_x(w) = \ln \Phi_x(w) = -\frac{1}{2}w^2 s^2 \tag{11}$$

Based on the factor analysis of tourism big data, the empirical modal decomposition formula of the statistical distribution sequence of tourism big data is obtained as follows:

$$X_{m+1}(m) = X_{k+1}(m) \pm \sqrt{(d_m(0)e^{l_1}+)^2 - \sum_{i=1}^{m-1}[X_{m+1}(i) - X_{k+1}(i)]^2} \tag{12}$$

In the above formula, $d_m(0)$ is the vector field of the tourism large data statistical distribution sequence in the global progressive steady state, and the $X_{m+1}(i)$ is the statistical quantity of the nonlinear component[14].

3. Feature Extraction and Prediction Algorithm Optimization

3.1. Feedback-constrained association rule feature extraction

On the basis of constructing the autosimilar regression model for the statistical prior information of tourism big data, empirical modal decomposition of the statistical distribution sequence is obtained, the dynamic state of tourism big data is carried out on the basis of the empirical modal decomposition of the statistical distribution sequence of tourism big data in the autosimilar regression model. This paper presents a dynamic trend prediction algorithm for tourism big data based on feedback constraint association rules[15]. Assuming that tourism big data is produced by linear correlation nonlinear statistical feature distribution series, the following association rule distribution model is obtained:

$$x_n = a_0 + \sum_{i=1}^{M_{AR}} a_i x_{n-i} + \sum_{j=0}^{M_{MA}} b_j h_{n-j} \tag{13}$$

Where, a_0 is the sampling amplitude of the initial tourism big data, x_{n-i} is the statistical characteristic distribution sequence of tourism big data with the same mean value and variance, and b_j is the oscillatory amplitude of tourism big data. Under the training of grey model, the association rules of tourism big data are shown as follows:

$$z(t) = x(t) + iy(t) = a(t)e^{iq(t)} + n(t) \tag{14}$$

In the formula, $x(t)$ is the real part of tourism big data's statistical characteristic distribution sequence, $x(t)$ is the imaginary part of tourism big data's statistical characteristic distribution sequence, $a(t)$ is phase randomization amplitude, $n(t)$ is interference vector.

Based on the analysis of tourism big data's nonlinear statistical feature distribution sequence, the principal component feature extraction of tourism big data statistical feature distribution sequence information flow is carried out, and the mutual information of tourism big data dynamic trend distribution is obtained as follows:

$$I(t) = -\sum_{ij} p_{ij}(t)\ln\frac{p_{ij}(t)}{p_i p_j} \tag{15}$$

The descriptive statistical analysis method is used to analyze the statistical characteristics of tourism big data, the statistical characteristic quantity of the tourism big data distribution is constructed to provide an accurate data input basis for the dynamic trend prediction.

3.2. Mining association rules and realizing process of dynamic trend prediction

The feature extraction method of feedback constraint association rules is used to analyze the features of tourism big data, and the training function of m-dimensional phase space model is established as follows:

$$\boldsymbol{X}(n) = \{x(n), x(n+t), \mathbf{L}, x(n+(m-1)t)\} \tag{16}$$

Where, $n = 1,2,\mathbf{L},N$, the discriminant function for the error correction of tourism big data dynamic trend prediction is expressed as follows:

$$L_x = \begin{cases} |f(x)-y|-x & |f(x)-y| \ge x \\ 0 & |f(x)-y| < x \end{cases} \quad (17)$$

According to the contribution degree of cumulative variance, error correction and feature dimensionality reduction are carried out to reduce the complexity of the model, and the optimized model of tourism big data dynamic trend prediction is obtained as follows:

$$\min_{w,h,z_l,z_l^*} = \frac{1}{2}w^T w + c\sum_{i=1}^{l}(z_l + z_l^*) \quad (18)$$

Where, z_l, z_l^* denote redundancy and principal component feature, c represents cost coefficient of error correction of extreme learning machine. Of course, the bigger c value is, the better the prediction precision is. The descriptive statistical analysis method is used to analyze the statistical characteristics of tourism big data. Based on the analysis, the estimation function of the output statistical feature distribution series of tourism big data dynamic trend prediction is obtained as follows:

$$f(x) = \sum_{i=1}^{l}(a_i + a_i^*)k(x - x_i) + b \quad (19)$$

Based on the feature extraction of association rules with feedback constraints, the prediction error of tourism big data is corrected. At this time, the prediction error converges to:

$$d_m(0) = \|X_m - X_k\| \quad (20)$$

According to the principle of Lyapunov convergence, the prediction error will converge to zero, thus:

$$x(t_{n+1})' = X_{m+1}(m) \quad (21)$$

Above all, principal component analysis is used to mine the association rules of tourism big data to improve the dynamic trend prediction algorithm of tourism big data. The implementation flow of the improved algorithm is shown in figure 1.

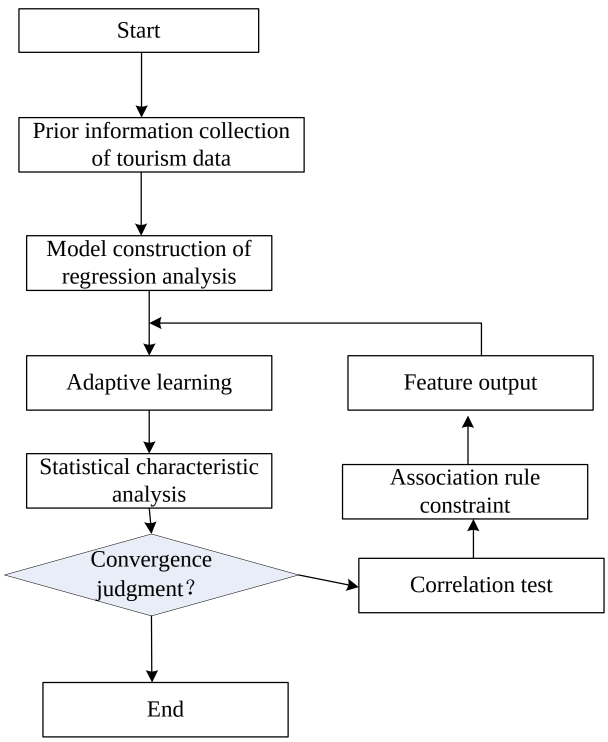
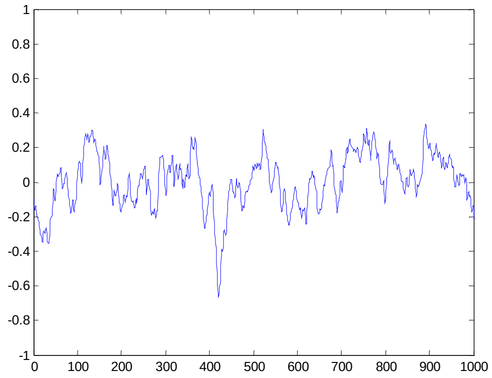


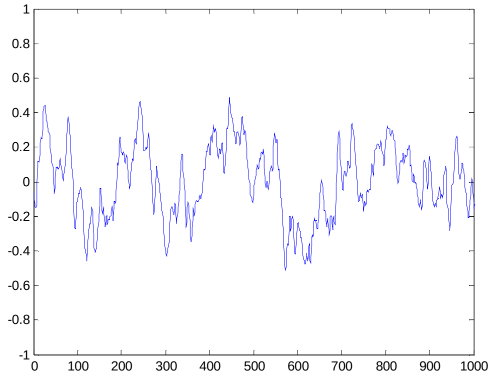
Figure 1. Flow chart of algorithm implementation

4. Simulation Experiment

In order to test the practical application performance of this algorithm in the dynamic trend prediction of tourism big data, the simulation experiment is carried out. The experiment is designed with Matlab simulation tool. The simulation environment is: IntelCore3-530 1 GB memory, and the operating system is Windows 10. Tourism big data collected time range from June 2015 to March 2018. The sampling data are taken from the National Tourism Administration and major tourism companies. The sample size is 2000, the training sample set is 100, the initial frequency is 12KHz, and the interval between sections is 0.23, the training step size is 2.5, and the iteration number is 500. According to the simulation environment and parameter setting above, the dynamic trend prediction simulation analysis is carried out, and two groups of tourist data samples are obtained as shown in figure 2.



(a) Test sample set 1



(b) Test sample set 2

Figure 2. Tourism big data test sample

The samples shown in figure 2 are taken as the test set, the dynamic trend prediction is carried out, and the prediction results are shown in figure 3.

Figure 3 shows that this method is used to analyze tourism big data, it can accurately predict the dynamic distribution trend of tourism big data and the prediction results are accurate. The accuracy of different methods for dynamic forecasting of tourism big data is tested. The results of comparison are shown in figure 4. Figure 4 shows that the accuracy and convergence of using this method to carry out the dynamic trend of tourism big data is higher, and the convergence is better. It has good prediction accuracy.

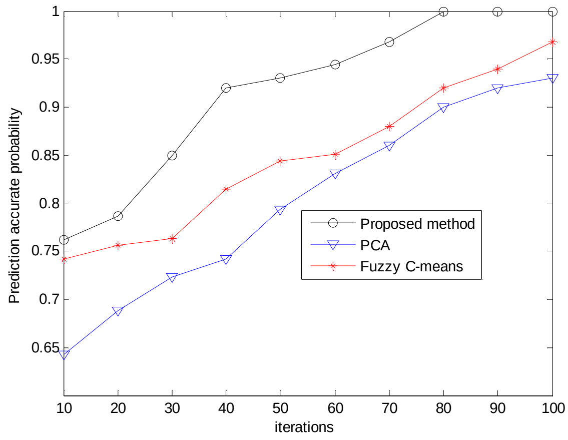


Figure 4. Comparison of prediction accuracy of tourism big data's dynamic trend

5. Conclusions

In this paper, a dynamic trend prediction algorithm for tourism big data based on feedback constraint association rules is proposed. The statistical prior information of tourism big data is constructed by the autosimilar regression model. In the autosimilar regression model, the empirical modal decomposition of the tourism big data statistical distribution sequence is carried out, and the feature extraction method of the association rules with feedback constraints is adopted. The method is used to analyze the characteristics of tourism big data, the BP neural network classification model is used to deal with the feature information clustering and information fusion of tourism big data, and the descriptive statistical analysis method is used to carry out the statistical features of tourism big data. Principal component analysis method is used to mine the association rules of tourism big data, and it realizes the dynamic trend prediction of tourism big data. The simulation results show that the proposed method has high accuracy, good convergence and good prediction accuracy for the dynamic trend of tourism big data. It can improve the performance of tourism information processing and data analysis.

References

[1] SUZUKIT and KUDO H. Two-dimensional non-separable block-lifting structure and its application to M-channel perfect reconstruction filter banks for lossy-to-lossless image coding[J]. IEEE Transactions on Image Processing, 2015, 24(12): 4943-4951.

[2] RAJAPAKAHA N, MADANAYAKE A, and BRUTON LT. 2D space-time wave-digital multi-fan filter banks for signals consisting of multiple plane waves[J]. Multidimensional Systems and Signal Processing, 2014, 25(1): 17-39.

[3] XIAO Wen, HU Juan. Performance analysis of frequent itemset mining algorithms based on sparseness of dataset. Journal of Computer Applications, 2018, 38(4): 995-1000.

[4] JIA D Y, ZHANG F Z. A collaborative filtering recommendation algorithm based on double neighbor choosing strategy[J]. Journal of Computer Research and Development, 2013, 50(5):1076-1084.

[5] POLATIDIS N, GEORGIADIS C K. A multi-level collaborative filtering method that improves recommendations[J]. Expert Systems with Applications, 2016,48:100-110.

[6] YANG Huanhuan, LI Tianrui, CHEN Xindi. Visualization of time series data based on spiral graph. Journal of Computer Applications, 2017, 37(9): 2443-2448.

[7] FANG H, BAO Y, ZHANG J. Leveraging decomposed trust in probabilistic matrix factorization for effective recommendation[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Menlo Park, CA:AAAI Press, 2014:30-36.

[8] ZHOU S, FENG Y, WU W Y, et al. A novel method based on the fuzzy C-means clustering to calculate the maximal Lyapunov exponent from small data[J]. Acta Physica Sinica, 2016, 65(2):42-48.

[9] MA Bin and XIE Xianzhong. PSHO-HF-PM: an efficient proactive spectrum handover mechanism in cognitive radio networks[J]. Wireless Personal Communications, 2014, 79(3): 1-23.

[10] PUSHP M and AWADHESH K S. A survey on spectrum handoff techniques in cognitive radio networks[C]. 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, 2014: 996-1001.

[11] SEO D W , LEE J H , LEE H S . Optimal coupling to achieve maximum output power in a WPT system[J]. IEEE Transactions on Power Electronics, 2016,31(6): 3994-3998.

DAI H , HUANG Y , LI C ,et al. Energy-efficient resource allocation for device-to-device communication with WPT[J]. IET Communications, 2017,11(3): 326-334.

[12] LI D , SHEN C , QIU Z . Two-way relay beamforming for sum-rate maximization and energy harvesting[C]// IEEE International Conference on Communications (ICC). 2013: 3115-3120.

[13] LIU Y . Joint resource allocation in SWIPT-based multi-antenna decode-and-forward relay networks[J]. IEEE Transactions on Vehicular Technology, 2017,66(10): 9192-9200.

[14] DU C , CHEN X , LEI L . Energy-efficient optimisation for secrecy wireless information and power transfer in massive MIMO relaying systems[J]. IET Communications, 2017,11(1): 10-16.

[15] FENG W , WANG Y , LIN D ,et al. When mm wave communications meet network densification:a scalable interference coordination perspective[J]. IEEE Journal on Selected Areas in Communications, 2017,35(7): 1459-1471.