

Frontiers

Modeling a combined forecast algorithm based on sequence patterns and near characteristics: An application for tourism demand forecasting

Wang Jun^a, Luo Yuyan^{b,c,*}, Tang Lingyu^d, Ge Peng^e^aBusiness School, Sichuan Normal University, Chengdu, 610101, China^bCollege of Management Science, Chengdu University of Technology, Chengdu 610059, China^cPost-doctorate R & D Base of Management Science and Engineering, Chengdu University of Technology, Chengdu 610059, China^dMathematic School, Sichuan Normal University, Chengdu, 610101, China^eBusiness School, Sichuan University, Chengdu, 610065, China

ARTICLE INFO

Article history:

Received 24 March 2017

Revised 4 November 2017

Accepted 19 January 2018

Keywords:

Artificial neural network

Combined forecasting method

K-means clustering

Tourism demand forecasting

ABSTRACT

Tourism demand forecasting is essential for forward tourism planning. To develop appropriate public policies and ensure sound business investment decisions, both government administrations and private sector businesses use basic tourist demand forecasting to plan future operations and assess the need for facilities and infrastructure investment. Therefore, forecasting has become indispensable to tourism management. This study proposes a combined tourism forecasting model using an artificial neural network (ANN) and a clustering algorithm, which considers two aspects of the given data series: sequence patterns and near characteristics, which embody structural changes and time series correlations. Training data were clustered into homogenous groups, and for each cluster, a dedicated forecaster was employed. Several neighboring samples were then selected to capture the current changes in the data series trends. Finally, the two prediction results derived from the sequence patterns and near characteristics were combined to determine the final forecast results. To verify the superiority and accuracy of the proposed model, it was compared with three other ANN-based models and the most popular ARIMA model using three non-linear, non-stationary tourist arrivals data series. Experimental cases studies demonstrated that the proposed combination method consistently outperformed the other related methods.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Forecasting is vital to the tourism industry. However, due to the complex, evolutionary nature of the tourism market, tourist demand series tend to be inherently noisy, conditionally non-stationary, and in some situations, deterministically chaotic. Because modeling these dynamic, non-stationary demand series is challenging, there is a need for a system that allows for more accurate forecasts with less noise and complexity.

Over the last three to four decades, many studies have been undertaken in this area [1]; consequently, many tourism forecast models have been proposed that have sought to explore the underlying tourist volatility mechanism and develop appropriate predictions for tourist demand or arrivals. Most of these studies have

adopted quantitative approaches, which can be subdivided into (1) causal econometric models, (2) time series models, and (3) artificial intelligence (AI) models, as described in [2] (Fig. 1).

Causal econometric models have sought to establish the relationships between variables such as tourism demand (as measured by tourist arrivals at a destination) and tourist spending using a set of hypothesized explanatory factors [3], from which tourism forecasts can then be made based on the functions of the values predicted by these explanatory variables. Time series methods predict future trends by extrapolating from no more than one previous data series. When testing the accuracy of the different tourist arrival forecasting models, it has been found that time-series models can generate acceptable forecasts at a low cost and with reasonable benefits. Generally, time series models have been dominated by autoregressive integrated moving average models, such as in [1,4–8]. Although time series approaches have proven to be useful for tourism demand forecasting, they have some major limitations such as their construction is not based on any economic theory that includes tourist decision-making processes. Therefore,

* Corresponding author at: College of Management Science, Chengdu University of Technology, 610059, China.

E-mail addresses: luoyuyan13@mail.cdut.edu.cn, wangjun.sicnu@qq.com (L. Yuyan).

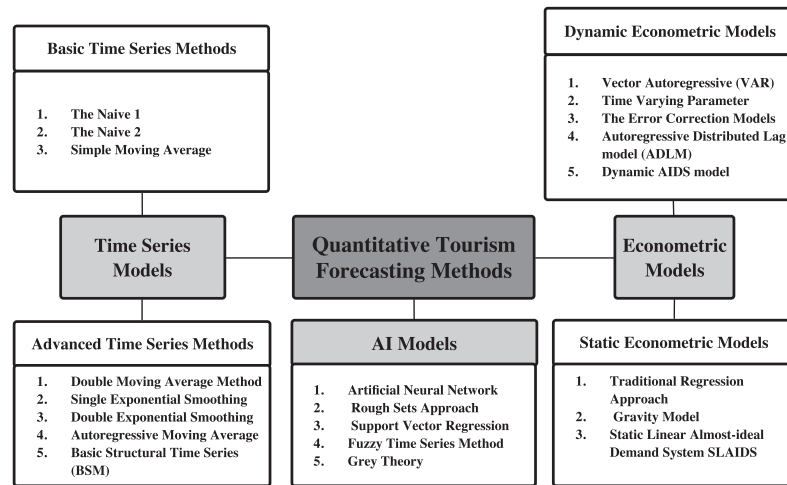


Fig. 1. The three main quantitative forecasting method categories from [2].

they are unable to analyze tourist behavior and are incapable of evaluating the effectiveness of specific tourism focused strategies and policies [2].

Another important forecasting branch has been AI forecasting methods, which include artificial neural networks [9–11], support vector regression (SVR) [12,13], rough sets theory [14,15], fuzzy time series theory [16–18], chaos theory [19,20], and expert systems [21–24]. In general, these methods have tended to perform better than traditional forecasting methods. The ANN models have been the most widely used tourism demand forecasting AI method due to their ability to capture subtle functional relationships within the empirical data, even when the underlying relationships are unknown or difficult to describe [9]. In addition, as ANN has the ability to overcome any multiple regression analysis restrictions, the computation of nonlinear threshold functions can be enabled.

At present, to improve forecasting accuracy, clustering techniques have been widely introduced for forecasting. Sina et al. [25] pointed out that one of the most primitive human actions was grouping objects into different categories based on their similarity; likewise, one of the most important and useful concepts for analyzing large quantities of data has been clustering, as data clustering techniques are able to discover structures in time series. Then, more recent research has investigated a combination of forecasting and clustering, with several hybrid forecasting models having been proposed that involve a clustering algorithm and a forecaster. Using the clustering, the training set is apportioned into homogenous clusters, which display high similarity compared with the other training set patterns. Then, different forecasters are trained with the sets that correspond to the different clusters; i.e. SVRs or ANNs. These approaches have been found to lead to better-trained forecasters as they are trained using better correlated data. Because of the encouraging performances of these cluster-based forecasting models, cluster forecasting frameworks have been developed for a wide range of applications such as sales forecasting, electricity load forecasting, and export trade value forecasting.

Hadavandi et al. [26] presented a sales forecasting approach that integrated genetic fuzzy systems and data clustering to construct a sales forecasting expert system. All data records were first categorized into k clusters using the k -means model, after which all clusters were fed into an independent genetic fuzzy system to construct the rule base and adjust the shape of the membership functions. Next, the data was assigned to an appropriate cluster and the forecast value generated using the related genetic fuzzy system.

Cheng et al. [27] proposed a day-ahead direct time series forecasting method for electricity loading and price based on clustering and next symbol prediction. In the clustering step, pattern sequences and their topological relationships were extracted from self-organizing map time series clustering. In the subsequent symbol prediction step, each cluster label in the pattern sequence was represented as a pair of its topologically identical coordinates, and the artificial neural network used to predict the topological coordinates for the following day by training the relationships between the previous daily pattern sequence and the following day pattern.

Pedro and Llanos [28] used clustering and classification techniques to forecast hourly global solar radiation. First, a clustering algorithm was proposed to identify the type of days, after which decision trees, artificial neural networks, and support vector machines were employed to estimate the parameters. Two procedures were proposed to analyze the data and estimate the models.

Kuo and Li [29] developed a three-stages forecasting model to value export trade in Taiwan. A wavelet transformation was first utilized to reduce the data preprocessing noise, after which a firefly algorithm-based k -means algorithm was employed for the cluster analyses, and a forecasting model built for each individual cluster.

Generally, these clustering-forecasting research has considered the similarities and heterogeneous characteristics of the data series and selected different forecasters for the different clusters to improve forecasting accuracy. However, insufficient attention has been placed on the structural changes in the data series. Key problems in time series forecasting have been the structural changes caused by economic recession or growth, politics, or other environmental factors that are inevitable and unmanageable, as well as the non-stationary nature of the time series, which can lead to gradual changes in the relationships between the independent and dependent variables [30]. Numerous studies, such as [31–35] have demonstrated that structural changes are inherent characteristics of non-stationary time series, with recent data always providing more information than more distant data.

Therefore, it is necessary to develop a data driven system that considers the structural changes in the data that bring in as much important information as possible into the model. The present study focuses on both sequence patterns and the near characteristics of the given data series, and proposes a novel 3-stages prediction framework that uses the ANN technique; the C-C-ANN model. The proposed forecasting framework follows a regime-switching strategy that accounts for the structural changes by using a given model for a limited time and then constructing a new model using

Download English Version:

<https://daneshyari.com/en/article/8253992>

Download Persian Version:

<https://daneshyari.com/article/8253992>

[Daneshyari.com](https://daneshyari.com)