Applied Soft Computing 26 (2015) 435–443

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Review Article

Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm

Rong Chen[a,b,c], Chang-Yong Liang[a,c,*], Wei-Chiang Hong[d], Dong-Xiao Gu[a,c]

[a] School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China
[b] Department of Economic and Management, Bengbu University, Bengbu, Anhui 233000, PR China
[c] Ministry of Education Engineering Research Center for Intelligent Decision-making & Information Systems Technologies, Hefei, Anhui 230009, PR China
[d] Department of Information Management, Oriental Institute of Technology, Panchiao, New Taipei 220, Taiwan

ARTICLE INFO

Article history:
Received 15 March 2014
Received in revised form 12 July 2014
Accepted 16 October 2014
Available online 23 October 2014

Keywords:
Holiday daily tourist flow forecasting
Support vector regression
Adaptive genetic algorithm
Seasonal index adjustment
Back-propagation neural network

ABSTRACT

Accurate holiday daily tourist flow forecasting is always the most important issue in tourism industry. However, it is found that holiday daily tourist flow demonstrates a complex nonlinear characteristic and obvious seasonal tendency from different periods of holidays as well as the seasonal nature of climates. Support vector regression (SVR) has been widely applied to deal with nonlinear time series forecasting problems, but it suffers from the critical parameters selection and the influence of seasonal tendency. This article proposes an approach which hybridizes SVR model with adaptive genetic algorithm (AGA) and the seasonal index adjustment, namely AGA-SSVR, to forecast holiday daily tourist flow. In addition, holiday daily tourist flow data from 2008 to 2012 for Mountain Huangshan in China are employed as numerical examples to validate the performance of the proposed model. The experimental results indicate that the AGA-SSVR model is an effective approach with more accuracy than the other alternative models including AGA-SVR and back-propagation neural network (BPNN).

Contents

* Corresponding author at: School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China. Tel.: +86 551 62919150; fax: +86 551 62904962.
E-mail address: cylhfut@163.com (C.-Y. Liang).

http://dx.doi.org/10.1016/j.asoc.2014.10.022

1. Introduction

With the implementation of the new official holiday regulation from 2007 in China, the number of tourists in holidays has increased rapidly. The huge market has been boosting by the business of holiday tourism. According to reports of the National Holiday Office (NHO), in 2011 the total number of tourists reached about 153 million during Spring Festival and 302 million during National Day, with year-on-year growth of more than 22.7% and 18.8% respectively, and tourism income added up to 82.05 billion yuan and 145.8 billion yuan, increasing 27% and 25.1% respectively [1]. The rapid growth of tourists flow in the short term will be a heavy burden for sightseeing spots, airports, and hotels during holidays. Since various important decisions, such as tourism planning, transportation and accommodation have been made based on the results of holiday daily tourist flow forecasting, an optimal solution of the allocation of social resources in such areas is crucial [2]. Accurate forecasting in holiday daily tourist flow can provide direct basis for tourism decision-makers, which can help them make scientific decisions. Then human, financial and material resources of scenic spots can be well planned and allocated ahead of time, and the over-consumption of tourism resources can be avoided as well.

However, in China, there are seven legal holidays, including New Year, Spring Festival, Ching-Ming Festival, May Day, Dragon Boat Festival, Mid-Autumn Festival and National Day, as well as summer holidays. Each of them has its own fluctuation pattern from weather conditions and different holiday periods, which make holiday daily tourist flow present a complicated nonlinear characteristic and seasonal tendency. Because of their complicated nonlinearity and seasonality, currently existing methods cannot exactly deal with both issues. Accurate forecasting of holiday daily tourist flow remains a difficult task attracted attentions in the literatures, and it is greatly necessary to develop new forecasting techniques to obtain satisfied accurate level.

1.1. Traditional tourist flow forecasting approaches

In recent years, various researches in tourist flow forecasting have resulted in the development of numerous forecasting approaches, which in general can be classified into two types, classical linear methods and nonlinear methods [3]. The most popular classical linear methods include autoregressive moving average (ARMA) [4,5], autoregressive integrated moving average (ARIMA) [6,7], and exponential smoothing (ES) [8,9], etc. These methods usually employ historical data to forecast the future tourist flow by a univariate or multivariate mathematical function, which mostly depend on linear assumptions. However there are some drawbacks in linear methods, for example, inability to capture the seasonal and nonlinear characteristics [10]. Thus, the nonlinear methods have been paid much more attention in recent years, for example artificial neural network (ANN) method. These methods emulate the processes of the human neurological system to process self-learning from the historical tourist flow patterns, especially for nonlinear and dynamic variations. And now the ANN methods have been widely used in tourist flow forecasting [11,12]. Empirical evidence shows that the ANNs generally outperform the classical linear methods (ES, ARIMA) in tourism forecasting [13]. But the methods lack a systematic procedure for model building, therefore, obtaining a reliable neural model involves selecting a large number of parameters experimentally through trial and error [14]. Additionally, ANN cannot successfully capture the changes of seasonality or trend [15].

1.2. Support vector regression with evolutionary algorithms in tourist flow forecasting

In mid 1990s, Vapnik [16] developed a statistical learning algorithm – support vector machines (SVM), which adheres to the principle of structural risk minimization seeking to minimize the upper bound of the generalization error, rather than minimize the training error. SVM has been extended to solve nonlinear regression estimation problems, i.e., the so-called SVR.

As a statistical theory-based method, SVR overcomes shortcomings of traditional tourist flow forecasting approaches. At present, SVR has been applied to financial time series forecasting [17–19], electric power load forecasting [20,21] and traffic flow forecasting [22,23]. It has also successfully been applied to tourist flow forecasting [24–27] which have been proved superior to ARIMA, ES and BPNN etc.

However, the previous research results show parameter optimization in SVR plays an important role in building a prediction model with high prediction accuracy and stability [28]. And poor forecasting accuracy is mainly suffered from lacking knowledge of the selection of three parameters (C, σ^2 and ε) in a SVR model. But till now, no general rules are feasible to determine suitable parameters [29]. Therefore, the determination of optimal parameters is a critical procedure in the SVR research fields.

In recent years, GAs have already been used to select optimization parameters of SVR in nonlinear forecasting successfully [30,31]. As an auto-adaptive stochastic search techniques, the core of this class of algorithms lies in the production of new genetic structures along selection, crossover and mutation operators, thereby providing innovations to solutions for the problem at hand [32]. However, fixed parameters of the crossover probability p_c and the mutation probability p_m in the GAs often affect performance directly, its parameter settings without tuning often lead to some questions, such as premature convergence and local optima [33]. In order to overcome these drawbacks, various adaptive techniques have been suggested to adjust parameters such as mutation probability and crossover probability in the process of running Gas [34]. Adaptive genetic algorithm (AGA) is one of them. In the AGA, the probabilities of crossover p_c and mutation p_m are adaptively varied depending on the fitness values of the solution. High fitness solutions are 'protected', while solutions with subaverage fitnesses are totally disrupted [35]. Recently, the AGA combined with SVR approach has been concerned in literatures [36]. But there are few articles in tourism forecasting which refer to this combined method. Therefore, SVR with AGA will be employed in this paper, by which, forecasting performance in capturing nonlinear and searching for parameters will to be expected to greatly improved.

1.3. Seasonal adjustment in tourist flow forecasting

Seasonality is a notable characteristic of tourist flow forecasting, it affects the accuracy of tourist flow forecasting and cannot be ignored in the modeling process. Therefore, how to handle seasonal variations of tourism data has always been an important and complex issue in tourist flow forecasting analysis [37]. The common models used for dealing with seasonal variations are to eliminate seasonal variations by some seasonal adjustment approaches, and then the models are scaled back through the estimated seasonal effects for prediction. Generally speaking, the common seasonal adjustment method is to eliminate the seasonal factors, filtering the original data by differencing before forecasting. Seasonal ARIMA (SARIMA) method is the most widely used seasonal adjustment forecasting methods. As another most popular and important seasonal adjustment methods, X-11-ARIMA and X-12-ARIMA have also been applied to adjust seasonality in time series [38–40]. But all models mentioned above have often been criticized. Firstly,