Expert Systems with Applications 41 (2014) 3691–3702

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Tourism demand forecasting using novel hybrid system

Ping-Feng Pai [a], Kuo-Chen Hung [b], Kuo-Ping Lin [c,*]

[a] Department of Information Management, National Chi Nan University, 1, University Rd., Puli, Nantou 545, Taiwan, ROC
[b] Department of Computer Science and Information Management, Hungkuang University, Taiwan, ROC
[c] Department of Information Management, Lunghwa University of Science and Technology, Taoyuan, Taiwan, ROC

ARTICLE INFO

Keywords:
Forecasting
Tourism demand
Fuzzy *c*-means
Least-squares support vector regression
Genetic algorithms

ABSTRACT

Accurate prediction of tourism demand is a crucial issue for the tourism and service industry because it can efficiently provide basic information for subsequent tourism planning and policy making. To successfully achieve an accurate prediction of tourism demand, this study develops a novel forecasting system for accurately forecasting tourism demand. The construction of the novel forecasting system combines fuzzy *c*-means (FCM) with logarithm least-squares support vector regression (LLS-SVR) technologies. Genetic algorithms (GA) were optimally used simultaneously to select the parameters of the LLS-SVR. Data on tourist arrivals to Taiwan and Hong Kong were used. Empirical results indicate that the proposed forecasting system demonstrates a superior performance to other methods in terms of forecasting accuracy.

1. Introduction

Tourism is crucial to most countries due to the considerable intake of money for businesses in terms of goods, services and the opportunities for employment in service industries such as hospitality administration and transportation. Governments and industry consistently require accurate estimations of anticipated tourism demand for effective policy planning. Wander and Van Erden (1980) and Archer (1987) indicated that the availability of accurate estimates of international tourism demand has critical economic consequences and that the tourism industry must forecast accurately due to the perishable nature of many related products. Armed with accurate estimates of the demand for tourism, tourist authorities and decision makers in the hospitality industries will be better able to perform strategic planning. Therefore, more accurate forecasting methodologies are continuously being proposed. In recent years, a number of researchers have focused on forecasting tourism demand (Chen 2011; Chen & Wang, 2007; Chu, 2008a; Chu, 2008b; Chu, 2011; Claveria & Torra, 2014; Pai, Hong, Chang, & Chen, 2006; Shahrabi, Hadavandi, & Asadi, 2013; Song, Gao, & Lin, 2013; Tsui, Balli, Gilbey, & Gow, 2014; Wu, Law, & Xu, 2012). Among these studies, those by Chen and Wang (2007), Pai et al. (2006) and Chen (2011) used support vector regression (SVR) technology to forecast tourism demand. In experimental results, the SVR obtained a superior performance to those of the traditional prediction models, such as autoregressive integrated moving average (ARIMA).

Furthermore, Song and Li (2008) summarized and reviewed a number of the tourism demand modeling and forecasting studies completed since 2000. This paper presents the latest developments of quantitative forecasting techniques summarized into three categories, time-series models, the econometric approach and other emerging methods, such as the use of artificial intelligence techniques. This study also presents a model for further enhanced forecast accuracy through forecast combinations and forecast integration of quantitative and qualitative approaches. A number of recent studies have focused on developing combination methods of forecasting tourism demand. Shen, Li, and Song (2011) developed six combination methods to forecast UK outbound tourism demand in seven destination countries. The numerical results demonstrate that combination methods, in general, obtain superior performances. Andrawis, Atiya, and El-Shishiny (2011) used a combination of long term and short term forecasts for tourism demand. The combination method creates a yearly time series from a monthly time series, produces a forecast for both and then combines these forecasts. The combination method effectively improved forecasting accuracy in the forecasting of monthly tourism numbers for inbound tourism to Egypt. Chen (2011) combined linear and nonlinear statistical models to forecast Taiwan's outbound tourism. Empirical results reveal that SVR combination models have excellent directional change detectability. Shahrabi et al. (2013) developed a new hybrid intelligent model that is called Modular Genetic-Fuzzy Forecasting System (MGFFS) by a combination of genetic fuzzy expert systems and data preprocessing which includes *K*-means clustering and the Takagi–Sugeno–Kang (TSK)

* Corresponding author. Tel.: +886 2 82093211x6329; fax: +886 2 82093211x 6 302.
E-mail address: kplin@mail.lhu.edu.tw (K.-P. Lin).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.12.007

type fuzzy rule-based system. Results show that forecasting accuracy of MGFFS is relatively better than Classical Time Series models, Neuro-Fuzzy systems, and neural network.

From examining past research in tourism demand forecasting, a number of phenomena are clearly evident. These include the finding that SVR technology is suitable for a tourism demand forecasting problem, and that combination methods may obtain an optimal performance. SVR has been a widely used and effective forecasting model in recent years. SVR uses a structural risk minimization principle that attempts to minimize the upper bounds of the generalization error rather than minimizing the training error (Vapnik, Golowich, & Smola, 1996). The generalization error was defined as the expected value of the square of the difference between the learned function and the exact target (mean-square error), and training error is the average loss over the training data. SVR schemes were extended to manage forecasting problems and have provided several promising results (Gaoa, Bompard, Napoli, & Cheng, 2007; Hua & Zhang, 2006; Huang, Nakamori, & Wang, 2005; Kim, 2003; Levis & Papageorgiou, 2005; Lin, Pai, & Yang, 2011; Lu & Wang, 2005; Mohandes, Halawani, Rehman, & Hussain, 2004; Pai & Hong, 2005; Tay & Cao, 2002; Xi, Poo, & Chou, 2007; and Yu, Chen, & Chang, 2006). An alternative proposed method, least-squares support vector regression (LS-SVR) (Van Gestel et al., 2001), attempts to minimize the sum square errors (SSEs) of training data sets, while simultaneously minimizing margin error which is the amount of random sampling error. LS-SVR technologies were applied to such forecasting areas as financial time series (Hung & Lin, 2013; Van Gestel et al., 2001), electronics (Li et al., 2007; Yang, Gu, Liang, & Ling, 2010), plastic injection molding (Li, Hu, & Du, 2008), simultaneous modeling (Jayadva, Khemchandani, & Chandra, 2008; Qin, Liu, Liu, & Zhang, 2010), bioinformation (Cui & Yan, 2009; Quan, Liu, & Liu, 2010), Chemistry (Goodarzi, Freitas, Wu, & Duchowicz, 2010) and cost estimations (Deng & Yeh, 2011).

Past research has verified that SVR/LS-SVR can efficiently approach complex nonlinear/time serial problems and obtain superior performances. However, trends of raw data with time series cannot easily be grasped by using traditional SVR technology or SVR with heuristic methods. Furthermore, previous researches (Shahrabi et al., 2013; Shen et al., 2011) mentioned combination methods usually can obtain better performance. Therefore, this study attempts to more accurately forecast American tourist arrivals in Taiwan and Hong Kong based on air traffic. Using an overview of past research in order to capture the trend of raw data, this study combines the FCM clustering method with logarithm least-squares support vector regression with genetic algorithms (LLS-SVRGA). In first stage, the FCM can help to define the center points by using tourist arrivals data which can effectively grasp trends of raw data. In second stage, the LLS-SVRGA (Lin et al., 2011) can more accurately forecast tourism demand. This study firstly develops the new combination methods for tourism demand prediction. These are called FCM + LLS-SVRGA, FCM and LLS-SVRGA and are the data preprocessing and main mechanisms in our forecasting system, respectively. The remainder of this paper is organized as follows: Section 2 presents the proposed forecasting systems (FCM + LLSVRGA). Section 3 provides the experimental results of FCM + LLSVRGA in forecasting tourist arrivals, and Section 4 offers conclusions and suggestions for further research.

2. The proposed forecasting system

The proposed forecasting system combines FCM with the LLS-SVRGA model. In general, a number of crucial points can usually be selected from historical data points to represent the trends of data in a time series problem. Hence, our proposed forecasting system uses the center points of a cluster as crucial trend points to forecast value. The illustration of the forecasting system is shown in Fig. 1. In the first stage, FCM helps define the center points by using time series data. These center points approximately represent the trend of data, and based on the number of center points, the preprocessing data can present various trends of data in a time series problem. These meaningful center points do not automatically consider outlier points by FCM clustering methodology. In the second stage, the forecasting system uses the LLS-SVRGA model to forecast data points. The LLS-SVRGA model can train the tuned approaching value, which means meaningful center points from the first stage, and subsequently provide more precise forecasting points. Fig. 2 shows the basic concepts of the forecasting system.

2.1. Fuzzy clustering with time serial

Fuzzy clustering was considered for searching center points of clusters in this study. Clustering analysis involves the discovery of a data structure and partitions a data set into a number of subsets with correlated data. Clustering was widely applied in several fields, such as taxonomy, geology, business, engineering systems, medicine, and image processing (for example, Bezdek (1981), Yang (1993), Honda and Ichihashi (2004), Kohonen (1997), Liu and Wang (2007)). Among the various partitioning techniques, the

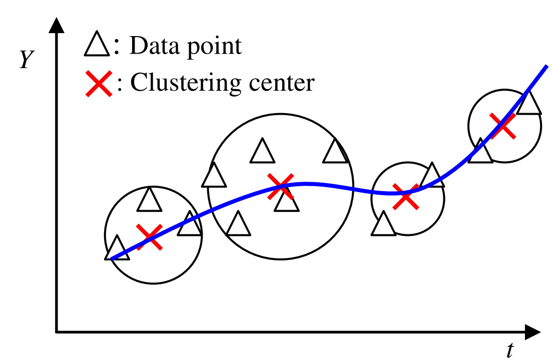

Fig. 1. Illustration of the forecasting system.

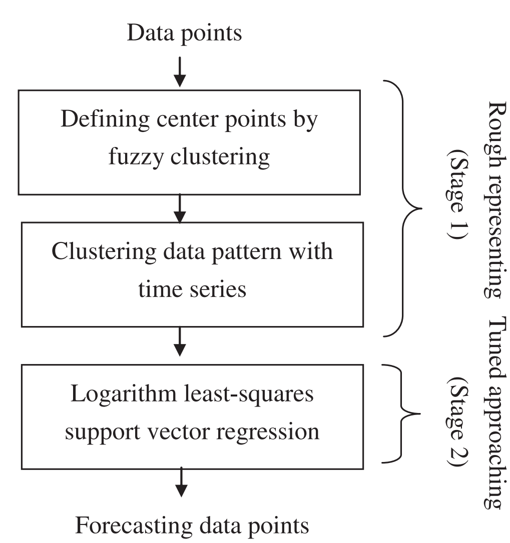

Fig. 2. The basic concepts of forecasting system.