Available online at www.sciencedirect.com

ScienceDirect

CAAI Transactions on Intelligence Technology 1 (2016) 30–42
http://www.journals.elsevier.com/caai-transactions-on-intelligence-technology/

Original article

Forecasting tourism demand by extracting fuzzy Takagi–Sugeno rules from trained SVMs

Xin Xu [a,*], Rob Law [b], Wei Chen [b], Lin Tang [c]

[a] College of Mechatronics and Automation, National University of Defense Technology, Changsha, 410073, China
[b] School of Hotel and Tourism Management, Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China
[c] College of Science, Hunan Agriculture University, Changsha, 410003, China

Available online 4 June 2016

Abstract

Tourism demand forecasting has attracted substantial interest because of the significant economic contributions of the fast-growing tourism industry. Although various quantitative forecasting techniques have been widely studied, highly accurate and understandable forecasting models have not been developed. The present paper proposes a novel tourism demand forecasting method that extracts fuzzy Takagi–Sugeno (T–S) rules from trained SVMs. Unlike previous approaches, this study uses fuzzy T–S models extracted from the outputs of trained SVMs on tourism data. Owing to the symbolic fuzzy rules and the generalization ability of SVMs, the extracted fuzzy T–S rules exhibit high forecasting accuracy and include understandable pre-condition parts for practitioners. Based on the tourism demand forecasting problem in Hong Kong SAR, China as a case study, empirical findings on tourist arrivals from nine overseas origins reveal that the proposed approach performs comparably with SVMs and can achieve better prediction accuracy than other forecasting techniques for most origins. The findings demonstrated that decision makers can easily interpret fuzzy T–S rules extracted from SVMs. Thus, the approach is highly beneficial to tourism market management. This finding demonstrates the excellent scientific and practical values of the proposed approach in tourism demand forecasting.

Keywords: Fuzzy modeling; Rule extraction; Support vector machines; Tourism demand forecasting

1. Introduction

Forecasting is an essential requirement for decision making and policy planning. Forecasting is widely conducted in various fields, including the tourism industry. Over the past decade, an increasing number of studies have focused on forecasting techniques for tourism demand [48,22,32]. The importance of accurate forecasting is basically attributed to the perishable nature of products and services in the industry. For instance, vacant airline seats, unoccupied hotel rooms, and unsold event tickets cannot be stockpiled for future use. Thus, accurate short-term and long-term forecasts of future demand are crucial [20,35,46]. Such forecasts are necessary for Hong Kong SAR, China, a key travel destination in Asia with an economically significant tourism industry. This major industry underwent substantial changes in market segments for inbound tourists. These changes demonstrate the urgent need to develop accurate methods for forecasting international demand for travel to Hong Kong SAR, China, which can be quantified by the number of tourist arrivals. Thus, tourism researchers continue to develop various techniques to predict the future demand for tourism.

In the tourism industry, accuracy and good comprehensibility of forecasting are required from policy makers and practitioners. As far as industrial applications are concerned, tourism practitioners can check the predicted values of tourist arrivals from different origins and plan for a change in demand from specific market segments by obtaining highly accurate estimates of such demand. By interpreting forecasting models, policy makers can analyze the key factors that contribute to

* Corresponding author.
E-mail address: xinxu@nudt.edu.cn (X. Xu).
Peer review under responsibility of Chongqing University of Technology.

http://dx.doi.org/10.1016/j.trit.2016.03.004

the increase or decrease in tourism demand from various regions. These practitioners can understand the underlying regularities according to the comprehensibility of tourism forecasting models. Policy makers can also plan tourism projects and related infrastructure development activities accurately and reasonably.

Previous studies on tourism demand analysis and forecasting produced two types of results in terms of forecasting quality. One type focused on error magnitude accuracy and the other aimed to improve understandability based on symbolic rules or knowledge bases. Despite the consensus on the accuracy of forecasting and the clear understanding of the advantages of accurate forecasts, few tourism forecasting methods have been developed that outperform other methods both in terms of accuracy and comprehensibility. According to Walle [57], qualitative models may exhibit good interpretability, but these approaches usually lack generalization ability, which substantially limits their applications. Traditional statistical regression models, such as auto-regressive integrated moving-average (ARIMA) approaches, also have a certain degree of comprehensibility. However, the prediction accuracy of these techniques may be unsatisfactory when nonlinearity and noise exist in tourism demand data. Artificial neural networks (ANNs) and support vector machines (SVMs) have recently gained significant interest because of their generalization ability and forecasting accuracy. SVMs have been demonstrated to perform better than ANNs and ARIMA. For example, Chen and Wang [11] combined a genetic algorithm with a support vector regression (SVR) to model tourist arrivals in China from 1985 to 2001. Their study shows that this approach outperforms ANNs and ARIMA models based on the normalized mean square error and mean absolute percentage error (MAPE). SVMs are a class of machine-learning algorithms based on the structural risk minimization principle [11]. The generalizability of SVMs can be optimized by controlling structural complexity, which makes SVMs superior to other machine-learning and data-mining algorithms. Moreover, the SVM training procedure is a convex quadratic programming process through which a global optimal solution can be obtained. Thus, compared with previous approaches, SVMs provide a more accurate and flexible forecasting technique for tourism forecasting.

Although SVMs are computationally accurate and exhibit satisfactory performance in tourism studies, both SVMs and ANNs are basically "black-box" techniques with poor explanatory capability and comprehensibility. The knowledge gleaned from such techniques is difficult to understand. Thus, the rule extraction from SVMs or ANNs was recently studied [6,10] and applied to various domains, such as credit scoring and fraud detection [16]. By extracting the rules from SVMs and ANNs, the comprehensibility of these black-box models can be enhanced, and a compromise between forecasting accuracy and interpretability can be achieved. Although forecasting accuracy and good comprehensibility are essential for policy makers and practitioners in the tourism industry, existing studies do not incorporate symbolic rule extraction and SVMs into tourism forecasting.

To fill this research gap, a novel tourism demand forecasting method is proposed in this paper, which is based on support vector machines with rule extraction (SVMRE). This method can extract fuzzy T−S rules from SVMs trained on tourism demand data. The aim of the present study is to incorporate extracted fuzzy rules from highly accurate SVMs into tourism demand forecasting. The fuzzy T−S rules generated from the outputs of SVMs can verify the information encoded in these models. Thus, the fuzzy rules for tourism demand extracted from SVM models exhibit high forecasting accuracy and easy comprehensibility to industry practitioners. The tourism demand forecasting problem in Hong Kong SAR, China was studied as an application case. It is demonstrated that that the proposed approach performs comparably with SVMs and can achieve better prediction accuracy than other forecasting techniques in most cases. In addition, decision makers can easily interpret the fuzzy T−S rules extracted from SVMs. Thus, the approach is highly beneficial to tourism market management. This finding shows the excellent scientific and practical values of the proposed approach in tourism demand forecasting.

The rest of this paper is organized as follows. In Section 2, some research background is introduced. In Section 3, SVMRE approach is presented for Tourism Demand Forecasting. In Section 4, the tourism demand forecasting problem in Hong Kong SAR is used as a case study and the performance of the SVMRE is evaluated and compared with other popular techniques for tourism demand forecasting.

2. Research background

2.1. Tourism forecasting techniques

Accurate forecasts are crucial because of the unique nature of the tourism industry [19,29,31,34]. Tourism demand forecasting employs qualitative and quantitative approaches [20,28,48]. Qualitative approaches depend on substantial information and human experiences. Walle [57] criticized these techniques for their lack of generalizability. As a result, tourism researchers do not primarily use qualitative forecasting methods. Formal scientific techniques that unambiguously represent the relationship between demand for travel and its underlying factors are more useful than qualitative forecasting methods in helping tourism decision makers understand the travel demand for a given destination.

Quantitative tourism demand forecasting models adopt mathematical functions to form the relationships of certain phenomena using numeric data [15,38]. These models are used to estimate future values based on past performance. Quantitative tourism forecasting approaches include causal relationship (regression) and time series techniques [1,26,30,50,58]. Although these approaches have achieved a certain degree of success, one fundamental problem is their inability to predict changes associated with other determining factors.