Journal of Air Transport Management 37 (2014) 20–26

Contents lists available at ScienceDirect

Journal of Air Transport Management

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

Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches

Gang Xie [a,*], Shouyang Wang [a], Kin Keung Lai [b]

[a] Center for Forecasting Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS), Beijing 100190, China
[b] Department of Management Sciences, City University of Hong Kong, Hong Kong, China

A R T I C L E I N F O

Article history:
Available online 22 February 2014

Keywords:
Hybrid approach
Short-term forecasting
Air passenger
Seasonal decomposition
Least squares support vector regression

A B S T R A C T

In this study, two hybrid approaches based on seasonal decomposition and least squares support vector regression (LSSVR) model are proposed for short-term forecasting of air passenger. In the formulation of the proposed hybrid approaches, the air passenger time series is first decomposed into three components: trend-cycle component, seasonal factor and irregular component. Then the LSSVR model is used to predict the components independently and these prediction results of the components are combined as an aggregated output. Empirical analysis shows that the proposed hybrid approaches are better than other time series models, indicating that they are promising tools to predict complex time series with high volatility and irregularity.

1. Introduction

As incomes and populations have increased and the structure of industry has changed, air transportation has grown considerably around world. The gradual freeing of trade across the globe has added to this growth (Alekseev and Seixas, 2009). With this overall expansion of demand, the patterns of traffic have also changed and become more complex. For example, there is competition between high-speed railroad service and air transport (Park and Ha, 2006). Obviously, it is important to forecast and meet the exact demand of air transportation in airport management.

Forecasting is the most critical area of airline management. An airline predicts demand so as to plan the supply of services required to meet that demand (Doganis, 2009). To an airport, short-term forecasting of air passenger provides a key input into decisions of daily operation management, including aircraft scheduling decisions, maintenance planning, advertising and sales campaigns, the opening of new sales offices, etc. (Smyth et al., 2012; Scarpel, 2013). Therefore, we focus on short-term forecasting of air passenger at an airport in this study.

In the field of air passenger forecasting, there are a number of empirical models, which can be classified as judgemental, causal, and time series. Expert judgement is based on the experience and intuition of specialists, who have a deep knowledge of the problem under consideration (Profillidis, 2012). Causal models are popular. Grosche et al. (2007) presented two gravity models for the estimation of air passenger volume between city-pairs. Sellner and Nagl (2010) used an econometric endogenous growth model to estimate the impact of air accessibility on GDP and investment growth. Dupuis et al. (2012) examined a logical analysis for improving the accuracy of passenger show rate predictions for airlines.

Also, there are lots of time series models, such as second-degree polynomial (Profillidis, 2000, 2012) and autoregressive integrated moving average (ARIMA) model (Samagaio and Wolters, 2010). In particular, Alekseev and Seixas (2009) developed a hybrid approach based on decomposition and back-propagation neural network (BPNN) for air transport passenger analysis. The results showed that forecasting performance was improved when data preprocessing of decomposition was fully adopted.

However, BPNN often suffers local minima and over-fitting, and it is sensitive to parameter selection (Xie et al., 2013). Support vector machine (SVM) has played an important role in the pattern recognition, machine learning and prediction, by minimizing an upper bound of the generalization error (Vapnik, 1995). SVM can be applied to classification and regression, i.e. support vector classification (SVC) and support vector regression (SVR). Since it adopts the structural risk minimization (SRM) principle, SVR can alleviate

* Corresponding author. Tel.: +86 10 62610229; fax: +86 10 62541823.
E-mail address: gxie@amss.ac.cn (G. Xie).

http://dx.doi.org/10.1016/j.jairtraman.2014.01.009
0969-6997/

the over-fitting and local minima issues and its solution is more stable and globally optimum (Xie et al., 2013).

Moreover, in order to reduce the computational complexity of SVM, Suykens and Vandewalle (1999) proposed least squares support vector regression (LSSVR) model, which solves a system of equations instead of a quadratic programming (QP) problem and leads to significantly improved speed of calculations. Though LSSVR model is an efficient method for function estimation problem, its solution is prone to outliers (Chen et al., 2012). To deal with the outliers caused by SARS epidemic in 2003 and Great Recession in 2008, we add a dummy variable which has value of 1 in the month during SARS epidemic and Great Recession and 0 in the remaining months in this study.

To the best of our knowledge, the application of hybrid seasonal decomposition and LSSVR approaches for air passenger forecasting has not been studied in the literature. In this study, LSSVR model is integrated with two seasonal decomposition methods X-12-ARIMA and TRAMO/SEATS to form hybrid approaches for air passenger forecasting. Based on time series of air passenger at Hong Kong International Airport, empirical analysis is implemented to compare the proposed hybrid approaches with other time series methods in terms of measurement criteria on the forecasting performance. Finally, some related issues are discussed and conclusions are drawn.

The remainder of the paper is organized as follows. The hybrid approaches for short-term forecasting of air passenger are proposed in Section 2. Section 3 illustrates the problem by using empirical analysis with experiments. Then, some related issues are discussed in Section 4. Section 5 draws conclusions and suggests some directions for future investigations.

2. Hybrid approaches based on seasonal decomposition and LSSVR model

In this section, the overall formulation process of hybrid approaches is presented. Firstly, LSSVR model and seasonal decomposition techniques are briefly introduced. Then hybrid approaches are formulated and corresponding steps involved in their implementation are described in details.

2.1. Least squares support vector regression model

In a least squares support vector regression (LSSVR) model, the regression problem can be transformed into an optimization problem, as follows.

$$\text{Min } \left(w^T w\right)/2 + \left(\gamma \sum_{i=1}^{l} e_i^2\right) \Big/ 2 \tag{1}$$
$$\text{s.t. } y_i = w^T \phi(x_i) + b + e_i,\ (i = 1, 2, \ldots,\ l)$$

where e_i is the error variable and γ is the penalty parameter. γ is used to control the minimization of estimation error and the function smoothness.

In order to solve the optimization problem, the Lagrange function is developed as

$$\begin{aligned} L(w,b,e,\alpha) = {} & \left(w^T w\right)/2 + \left(\gamma \sum_{i=1}^{l} e_i^2\right) \Big/ 2 \\ & - \sum_{i=1}^{l} \alpha_i \left[w^T \phi(x_i) + b + e_i - y_i\right] \end{aligned} \tag{2}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_l)$ is the Lagrange multiplier. Differentiating L with respect to variables w, b, e and α, we obtain

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{l} \alpha_i \phi(x_i), \tag{3}$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{l} \alpha_i = 0, \tag{4}$$

$$\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i, \tag{5}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \phi(x_i) + b + e_i - y_i = 0. \tag{6}$$

After solving the above functions, we can obtain the solution of the problem in the following form:

$$f(x) = \sum_{i=1}^{l} w_i K(x, x_i) + b \tag{7}$$

where $K(\cdot)$ is the kernel function. Here, the usual Gaussian RBF $K(x, x_i) = \exp[\|x - x_i\|/(2\sigma^2)]$ with a width of σ is employed.

In the following subsections, we introduce two seasonal decomposition methods: X-12-ARIMA and TRAMO/SEATS.

2.2. Seasonal decomposition methods

In order to capture seasonal characteristics of observations in different years, we use two popular seasonal decomposition methods X-12-ARIMA and TRAMO/SEATS. X-12-ARIMA is the Census Bureau's latest seasonal adjustment programme (Findley et al., 1998), and this method decomposes time series y_t into three components, i.e. trend-cycle component tc_t, seasonal factor sf_t and irregular component ir_t, which can be combined into the original data in additive and multiplicative forms, as follows:

$$y_t = tc_t + sf_t + ir_t, \tag{8}$$

$$y_t = tc_t \times sf_t \times ir_t. \tag{9}$$

Comparing the two forms of seasonal decomposition, the multiplicative decomposition is a more suitable choice for most seasonal time series. The main reasons for the priority are summarized into two aspects: On one hand, the seasonal factor of the multiplicative form is a relative value of the original series; On the other hand, most seasonal time series with positive values has the characteristic that the scale of seasonal oscillations increases in the level of original time series (U.S. Census Bureau, 2011). As a consequence, the multiplicative form is employed for seasonal decomposition via X-12-ARIMA program in this study.

TRAMO (Time series Regression with ARIMA noise, Missing values, and Outliers) is a program for estimation and forecasting of regression models with errors that follow (in general) nonstationary ARIMA processes, when there may be missing observations in the series, as well as contamination by outliers and other special (deterministic) effects. SEATS (Signal Extraction in ARIMA Time Series) is a program for decomposing a time series into its unobserved components (i.e., for extracting from a time series its different signals), following an ARIMA-model-based method (Maravall, 2005). When used together, TRAMO preadjusts the series, and SEATS decomposes the linearized series into its stochastic components (Maravall, 2006). In this study, for the purpose of comparison with X-12-ARIMA method, TRAMO/SEATS method is also used to decompose time series y_t into trend-cycle component