European Journal of Operational Research 264 (2018) 967–977

Contents lists available at ScienceDirect

European Journal of Operational Research

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

The impact of special days in call arrivals forecasting: A neural network approach to modelling special days

Devon Barrow[a,*], Nikolaos Kourentzes[b]

[a] Faculty of Business, Environment and Society, Coventry University, Coventry, West Midlands CV1 5FB, UK
[b] Lancaster University Management School, Department of Management Science, Lancaster LA1 4YX, UK

ARTICLE INFO

Article history:
Received 15 September 2015
Accepted 9 July 2016
Available online 14 July 2016

Keywords:
Time series forecasting
Call centre arrivals
Outliers
Functional data
Neural networks

ABSTRACT

A key challenge for call centres remains the forecasting of high frequency call arrivals collected in hourly or shorter time buckets. In addition to the complex intraday, intraweek and intrayear seasonal cycles, call arrival data typically contain a large number of anomalous days, driven by the occurrence of holidays, special events, promotional activities and system failures. This study evaluates the use of a variety of univariate time series forecasting methods for forecasting intraday call arrivals in the presence of such outliers. Apart from established, statistical methods, we consider artificial neural networks (ANNs). Based on the modelling flexibility of the latter, we introduce and evaluate different methods to encode the outlying periods. Using intraday arrival series from a call centre operated by one of Europe's leading entertainment companies, we provide new insights on the impact of outliers on the performance of established forecasting methods. Results show that ANNs forecast call centre data accurately, and are capable of modelling complex outliers using relatively simple outlier modelling approaches. We argue that the relative complexity of ANNs over standard statistical models is offset by the simplicity of coding multiple and unknown effects during outlying periods.

1. Introduction

A key challenge in many call centres is the accurate forecasting of inbound call volumes required to support short, medium and long-term decisions. With 60–80 percent of the overall operating budget of a call centre resulting from staffing costs (Aksin, Armony, & Mehrotra, 2007; Brown et al., 2005), accurate and robust forecasting for workload management is an important issue. In recent years, consumer demand for call centre services has increased along with the dramatic expansion in the service industry (Aksin et al., 2007; Shen & Huang, 2005). With this increase in the number of call centres and the amount of call centre data, new challenges related to the handling and forecasting of such data are presented. Firstly, call centre arrivals data are typically high-dimensional and sampled at a high frequency, usually collected in daily or smaller time buckets (e.g. 15 or 30 minute periods). Under these conditions, conventional statistical models designed for low frequency time series may break down and be inappropriate (Kourentzes & Crone, 2010). Secondly, call centre arrivals data exhibit complex seasonal patterns (De Livera, Hyndman, & Snyder, 2011). The data often contains intraday, intraweek and even intrayear dependencies, meaning that call arrival volumes will typically exhibit multiple seasonal cycles which need to be modelled (for example, see Taylor, 2008a).

Call centre arrival data are quite context sensitive, meaning that the data can contain strong effects from holidays, special events, promotional activities and unexplained variations (Andrews & Cunningham, 1995). These will usually exhibit very different arrival characteristics from regular patterns of call arrivals, weakening or even destroying the correlation structures in the data. This may affect model specification and parameter estimation (Chatfield, 2013). This is distinct from the behaviour of electricity load data known to exhibit similar time series seasonal structure. Hence, modelling call centre data may require data cleansing. When outliers are due to holiday and promotional effects, these can often be captured through past experience and reference to market information. On the other hand, when outliers are due to systems errors and other unexplained events they become more difficult to identify and model.

Many forecasting models assume, as a standard approach, that this information is either available or that the forecaster has an external methodology for tackling outliers. Some methods developed specifically for call centre data acknowledge this assumption. For

* Correspondence to: School of Strategy and Leadership, Faculty of Business and Law, Coventry University, Coventry, West Midlands, CV1 5FB, UK.
E-mail address: devon.barrow@coventry.ac.uk (D. Barrow).

http://dx.doi.org/10.1016/j.ejor.2016.07.015
0377-2217/

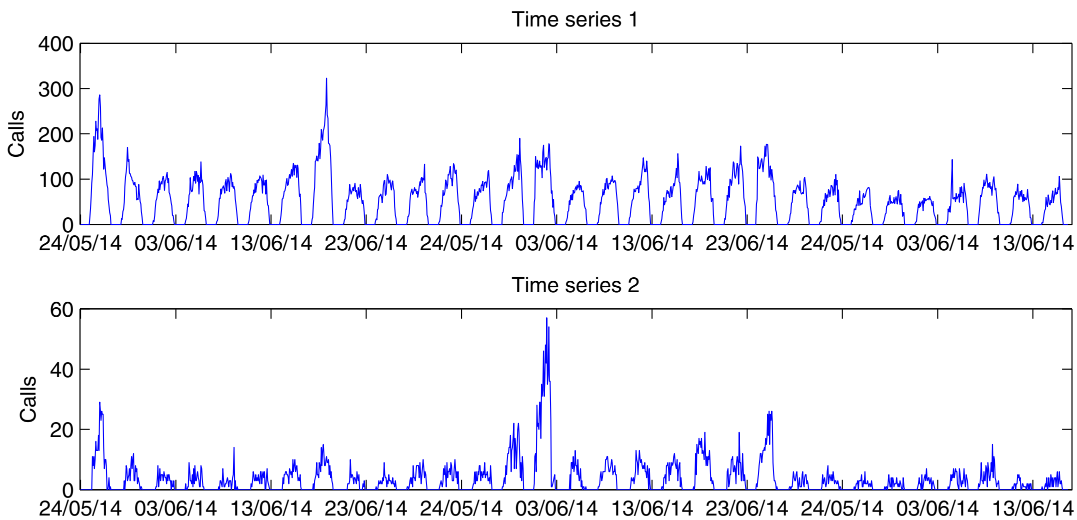

Fig. 1. Examples from the two time series from Sunday 25th May to Tuesday 24th June 2014.

example, Taylor, De Menezes, and McSharry (2006) take the approach of removing such days, while Shen and Huang (2005) describe the application of singular vector decomposition for outlier detection but provide no empirical evaluation. Many others, however, avoid this problem by assuming the data pre-cleansed (for examples see: Jongbloed & Koole, 2001; Avramidis, Deslauriers, & L'Ecuyer, 2004; Taylor, 2008a; Pacheco, Millan-Ruiz, & Velez, 2009; Taylor, 2010b). Conejo, Contreras, Espínola, and Plazas (2005) in forecasting electricity prices try to automatically correct outliers using conventional time series modelling approaches, but have limited success as the high frequency nature of the data makes such approaches inappropriate. Nonetheless, in the context of electricity load forecasting Kourentzes (2011) demonstrates that there are substantial accuracy benefits to be had from modelling irregular load patterns.

This study has two aims. First we evaluate a variety of univariate time series forecasting methods that are in theory capable of modelling data exhibiting characteristics of intraday call centre arrivals. This evaluation is central to identifying which method is most fit for purpose (Petropoulos, Makridakis, Assimakopoulos, & Nikolopoulos, 2014). We identify empirically artificial neural networks (ANNs) as having the best forecasting performance and argue that this is due to their modelling flexibility. This is a useful finding given the limited research in ANNs for forecasting call centre data, despite evidence which shows them capable of handling the complex seasonal structure of this data (Lee Willis & Northcote-Green, 1983; Temraz, Salama, & Chikhani, 1997). We augment the best performing method to model special events and outlying days. We introduce and evaluate multiple alternative methodologies, ranging from including the outlier information in the forecasts to cleaning the data prior to fitting the forecasting model. This addresses a gap in research of practical significance, considering the difficulty and cost associated with manual exploration and treatment of high frequency data by experts, and limited advancements in automatic outlier identification for such data. Note that we restrict our attention to time series methods that forecast call arrivals volume. We do not investigate models for call arrivals rate. For example by considering other classes of methods e.g. those which incorporate stochastic models (Avramidis et al., 2004) and those based on assumptions of a Poisson process (Aldor-Noiman, Feigin, & Mandelbaum, 2009; Ibrahim & L'Ecuyer, 2013).

The rest of the paper is organised as follows: first we describe the nature of call centre data by introducing the dataset for the empirical results in Section 2. Then Section 3 presents the evaluation of numerous forecasting methods and assesses the impact of outliers on the performance of these methods. Subsequently, Section 4 presents and evaluates various alternatives to modelling of outliers using the best performing forecasting method. The final section provides a summary and concluding comments.

2. Call centre arrival data and outliers

2.1. Case study dataset

To illustrate the data properties of call arrival time series, we describe the two time series that are used in the empirical evaluations. These recorded calls are received by a large call centre of one of Europe's leading entertainment companies representing inbound sales calls.

The data is sampled at half-hourly intervals and covers a period of 103 weeks and 3 days from 29 June 2012 to 23 June 2014 inclusive, including bank holidays and weekends. Fig. 1 provides examples of the two time series. Observe that from day to day there are some hours with zero values, which represent the hours that the call centre is not operational. Furthermore, it is evident that there are some days with extraordinary levels of calls, which we refer to as outlying days.

Fig. 2 provides a weekly seasonal plot of the first time series. To make the plot less cluttered, we provide various percentiles. The difference between the outer bands and the median provides some insight into the variability of each half-hourly interval of the time series. Two seasonal cycles are evident: an intraday and an intraweek. Monday to Friday look relatively similar, although Monday, Wednesday and Friday exhibit slightly increased afternoon/evening variability. The weekend days are substantially different. Also note that the starting time for each day is not always at 08:00, reflecting slight variations in the starting time that the call centre services calls. The second series exhibits similar seasonal properties.

For these time series, we are provided with periods of extraordinary activity. These were identified by experts in the organisation that the data originate from, using domain knowledge. In total there are 87 such outlying days for the first time series and 115 for the second.