Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 103 (2017) 28 – 35

Procedia Computer Science

XIIth International Symposium «Intelligent Systems», INTELS'16, 5-7 October 2016, Moscow, Russia

Prediction of engine demand with a data-driven approach

Hudson Francis, Andrew Kusiak*

The University of Iowa, Iowa City, IA 52242-1527, USA

Abstract

Models predicting volume of engine demand from historical data are developed. To accommodate seasonal effects, neural networks and autoregressive integrated moving average (ARIMA) approaches are considered. Previous research on the effectiveness of neural networks to model phenomena with seasonality and trend using raw data has been inconclusive. In this paper, four predictive models for a linear time series with seasonality are developed and their accuracy is studied. Performance of a dummy variable linear regression model, a seasonal ARIMA model, a neural network model using raw historical data, and a hybrid linear model is compared. The seasonal ARIMA and linear regression models are found to perform better than the neural network model. The hybrid linear model is found to outperform the three individual models.

Peer-review under responsibility of the scientific committee of the XIIth International Symposium "Intelligent Systems"

Keywords: manufacturing; neural networks; ARIMA model; seasonality; time series

1. Introduction

Prediction of demand is an important part of business management, especially in manufacturing. Accurate forecasts reduce cost through better inventory management. A data set was obtained from a company servicing engines in the US and Canada. This company sells replacement parts for failed engines. These failures cause delays, becoming expensive to the customer. It is important that the turnaround time between the sale of a part and the delivery to the customer is minimized. The goal of this paper is to reduce customer downtime by developing a predictive tool estimating future sales of engines. Predictions are made at two levels of granularity: aggregate level (all parts) and group level (a subset of all parts). Several time horizons are considered for making predictions: one year, six months,

* Corresponding author. Tel.: 319-335-5934; fax: 319-335-5669.
E-mail address: andrew-kusiak@uiowa.edu

1877-0509
Peer-review under responsibility of the scientific committee of the XIIth International Symposium "Intelligent Systems"
doi:10.1016/j.procs.2017.01.005

one quarter, and one month in advance. This paper compares the aggregate level accuracy of four prediction methods: dummy-variable linear regression, neural network (NN), seasonal auto-regressive integrated moving average (ARIMA), and a linear hybrid predictive model.

The ARIMA method has become a standard time series forecasting tool since it was introduced by Box and Jenkins[1] in 1970. It effectively models linear data with seasonality and trend. Since then, advances in computer processing have allowed data-driven models such as neural networks to gain popularity. Neural networks can handle highly nonlinear data. Hornik[2] found that a multilayer NN can approximate any function given enough hidden nodes. This versatility has led to the application of NN based sales prediction models in a wide range of industries, including electronics[3,4], food[5,6,7], clothing[8], and footwear[9].

Despite their success, NNs do not universally outperform ARIMA models. As Chatfield[10] pointed out, the "best" forecasting method is situational. Ho et al.[11] found that recurrent neural networks and ARIMA models outperformed NN models when predicting compressor failures at a Norwegian power plant. Khashei and Bijari[12] noted that neural networks do not always model linear data well. In addition, using NNs for data with seasonality has yielded mixed results. One option is to use a preprocessing method to deseasonalize the data. Nelson et al.[13] analyzed 68 times series and found that deseasonalized NNs performed much better than NNs that were not deseasonalized. However, Sharda and Patil[14] analyzed 75 times series and concluded that NN models performed at least as well as ARIMA models and did not need to be deseasonalized. Alon et al.[15] found that neural networks generally outperformed ARIMA and linear regression models for US retail sales predictions, but that the ARIMA method was a strong competitor. Furthermore, they concluded that NN models did not require deseasonalization or detrending for that data set. Taskaya-Temizel and Casey[16] claimed that deseasonalization is not necessary if the NN is properly specified, but detrending through differencing will increase the accuracy of the model. Zhang and Qi[17] found that both detrending and deseasonalization preprocessing provided the best NN forecasting results in a case study of retail sales. Chu and Zhang[18] found that deseasonalization improved NN accuracy and that the deseasonalized NN outperformed ARIMA and dummy variable regression models. Finally, Zhang and Qi[19] concluded that NNs are not able to model seasonality directly in a case study of nine data sets. They compared seasonal ARIMA, NN, detrended NN, deseasonalized NN, and detrended and deseasonalized NN models. They found that the deseasonalized and detrended NN model outperformed all other models. Their research indicates that NN models without detrending and deseasonalization may be inferior to seasonal ARIMA models. Zhang and Qi[19] claimed that "a trend time series does not meet the conditions for universal approximation" (p. 513), therefore, preprocessing was necessary.

Individual forecasting methods are best suited for specific data characteristics. For example, ARIMA models can handle seasonality and trend, but cannot handle nonlinear data. Zhang[20] claimed that real data is rarely only linear or nonlinear. In this case, individual models may not be appropriate. Indeed, Khashei and Bijari[21] wrote "if a time series exhibits both linear and nonlinear patterns during the same time interval, neither linear models nor nonlinear models alone are able to model both components simultaneously" (p. 480). Research[22] has shown that combined forecasting methods often outperform individual methods. These combinations do not need to be complex. Clemen[23] observed that simply averaging the results of multiple forecasts can sometimes improve the prediction accuracy. Combined forecasts generally have less variability in accuracy than individual forecasts[24]. Hybrid models are one way to handle seasonality, as well. Tseng et al. [25] created a hybrid seasonal ARIMA and back-propagation NN model and compared it to individual seasonal ARIMA, differenced NN, and deseasonalized NN models. They found that the hybrid model performed best, especially with limited history. Combined forecasts do not always outperform individual forecasts[16,24,26], however. This could be due to the assumed relationship between the linear and nonlinear structures in the data. For example, if a linear-nonlinear relationship is assumed to be additive but is actually multiplicative, the individual model may outperform the hybrid model. Khashei and Bijari[21] presented a hybrid ARIMA-NN model that improves on previous hybrid models because it does not assume an additive relationship. They guaranteed that this method would not be worse than individual NN or ARIMA models and illustrated the results with three empirical examples.

The goal of this paper is to predict future engine sales at several time horizons using four prediction methods. This paper compares those results to determine if the neural network model without deseasonalization or detrending can outperform the seasonal ARIMA model. It also examines whether a hybrid model can outperform the individual models.