ARTICLE IN PRESS

JBR-09000; No of Pages 4

Journal of Business Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Business Research

Forecasting incoming call volumes in call centers with recurrent Neural Networks

Mona Ebadi Jalal [a,*], Monireh Hosseini [a], Stefan Karlsson [b]

[a] IT Department, Faculty of Industrial Engineering, K. N. Toosi University of Technology, Iran
[b] Faculty of Engineering (LTH), Lund University, Sweden

ARTICLE INFO

Article history:
Received 1 February 2016
Received in revised form 1 March 2016
Accepted 1 April 2016
Available online xxxx

Keywords:
Forecasting
Model prediction
Call center
Neural Networks

ABSTRACT

Researchers apply Neural Networks widely in model prediction and data mining because of their remarkable approximation ability. This study uses a prediction model based on the Elman and NARX Neural Network and a back-propagation algorithm for forecasting call volumes in call centers. The results can help determine the optimal number of agents necessary to reduce waiting time for customers, enabling profit maximization and reduction of unnecessary costs. This study also compares the performance of the Elman-NARX Neural Network model with the time-lagged feed-forward Neural Network in addressing the same problem. The experimental results indicate that the proposed method is efficient in forecasting the call volumes of call centers.

1. Introduction

Modern call centers are large and employ thousands of agents. Representing the front end of organizations, call centers are responsible for maintaining customer relationships and the overall public image of the organization. This study focuses on two important issues in call center operations:

First, call centers are labor-intensive, with the cost of agents generally comprising 60–80% of the overall operating budget (Aksin, Armony, & Mehrotra, 2007). Therefore, an effective management strategy must schedule a sufficient number of agents.

Second, call centers involve customer satisfaction factors such as perceived queuing time. Feinberg (Feinberg, Kim, Hokama, de Ruyter, & Keen, 2000) finds a statistically significant correlation between the length of queuing time and customer satisfaction, indicating that shorter queuing times have beneficial effects on customer satisfaction. Furthermore, according to Whiting and Donthu (2009), customer error in estimating queuing time affects their satisfaction. Consequently, shorter queuing times provide higher satisfaction than longer ones and avoid customer churn.

This study concludes that the first and most important step is to forecast future call volumes accurately.

Much literature exists on forecasting; however, few studies attempt to develop effective models to forecast incoming call volumes in call centers. This study proposes a Neural Network model for forecasting call volumes to reduce waiting times and determining the optimal number of agents.

Recently, the rapid development of artificial intelligence has contributed to the growth of several advanced algorithms that tackle a variety of analytical problems such as forecasting. An artificial Neural Network is a massively parallel-distributed processor of simple processing units, which has a natural propensity to store experimental knowledge and to make it available for use (Haykin, 1998). The main advantage of the Neural Network in comparison with classic linear methods is their ability to model functions characterized by non-linear dynamics. High adaptive ability, tolerance to various outer noises, and the influence of "heavy-tailed" distributions are other frequently mentioned advantages (Klevecka, 2011). Neural Networks have shown tremendous success in developing models that track the changes in the characteristics and parameters of a system (Adya & Callopy, 1998; Chen, Peng, & Wang, 2000; Dudul & Ghatol, 2004). Therefore, at any instant in time, Neural Network correctly simulates the given time-varying system, despite significant changes in the system's property, and perfectly mimics and identifies the actual physical system.

Considering these properties, this study addresses the problem of forecasting call volumes in a call center.

2. Neural Network models

Neural Network models developed as a generalization of the mathematical models of human cognition and neural biology. Some key attributes of the brain's information network include a nonlinear, parallel structure and dense connections between information nodes (Haykin, 1998). Neural Network models have proven successful in a variety of business fields such as accounting (Kuldeep & Sukanto, 2006; Landajo & de Andrés, 2007; Lenard, Alam, & Madey, 1995; Zhang, Cao, & Schniederjans, 2004), management information systems (Huang, Chiu, & Chen, 2008; Kuflik, Boger, & Shoval, 2006; Zhu, Premkumar, Zhang,

* Corresponding author.
E-mail addresses: m_ebadi_j@yahoo.com (M. Ebadi Jalal), hosseini@kntu.ac.ir (M. Hosseini), stefan.karlsson@plog.lth.se (S. Karlsson).

http://dx.doi.org/10.1016/j.jbusres.2016.04.035
0148-2963/

Please cite this article as: Ebadi Jalal, M., et al., Forecasting incoming call volumes in call centers with recurrent Neural Networks, *Journal of Business Research* (2016), http://dx.doi.org/10.1016/j.jbusres.2016.04.035

& Chu, 2001), marketing (Kim, Street, Russell, & Menczer, 2005; Thieme, Song, & Calanton, 2000), and production management (Al-Ahmari, 2008; Bhattacharyya, Sengupta, Mukhopadhyay, & Chattopadhyay, 2008; Das & Datta, 2007; Kaparthi & Suresh, 1994; Wang, Chen, & Lin, 2005; Wu, Chen, & Chian, 2006). However, when forecasting with Neural Networks, a major factor is that the goodness of results depends on its architecture.

2.1. Time-lagged feed-forward networks

Usually, researchers apply a special type of Neural Network known as a time-lagged feed-forward network in time series modeling and forecasting. A time-lagged feed-forward network is a powerful nonlinear filter consisting of a tapped delay memory and multilayer perceptron and the standard back-propagation algorithm is useful in training this type of Neural Networks (Haykin, 1998; Mozer, 1993).

2.2. Elman Neural Networks

The Elman Neural Network is a semi-recursive Neural Network that uses the back-propagation-through time learning algorithm to find patterns in value sequences. Elman Neural Network adds a layer to receive feedback from the network in the hidden layer as a step delay to the operator for purposes of memory; therefore, the system can adapt to time-varying characteristics, which directly reflects the dynamic process characteristics.

Further, the Elman Neural Network has better computing power than a feed-forward Neural Network; therefore, considering dynamic telephone traffic, Elman Neural Network is suitable to construct a nonlinear prediction model for the traffic time series data (Elman, 1990).

2.3. NARX Neural Networks

An important class of discrete–time nonlinear systems is the Nonlinear Auto-Regressive with eXogenous inputs (NARX) model (Chen, Billings, & Grant, 1990, Leontaritis & Billings, 1985, Ljung, 1987, Su & McAvoy, 1991, Su, McAvoy, & Werbos, 1992):

$$y(t) = f(y(t-1), \ldots, y(t-D_y), u(t-1), \ldots, u(t-D_u)) \quad (1)$$

where $u(t)$ and $y(t)$ represent input and output of the network at time t, D_u and D_y are the input and output order, and the function f is a nonlinear function. When the function f can be approximated by a Multilayer Perceptron, we call the resulting system a NARX recurrent Neural Network (Chen et al., 1990, Narendra & Parthasarathy, 1990).

3. Data and proposed model

3.1. Model specifications

Fig. 1 shows the proposed model that uses a combination of Elman Neural Network and NARX Neural Network. This study uses these networks because of their special abilities of forecasting. The Elman Neural Network can adapt to time-varying characteristics and find temporal patterns, which are important for forecasting call volumes. The NARX Neural Network stores information for long periods in the presence of noise; therefore, it has the ability to remove and ignore noises.

3.2. Data and forecasting accuracy procedure

The call volume data is from an Emdad Khodro Iran call center, which is a member of the Iran Khodro industrial group and complements IKCO's after sales services chain. Due to high call volumes and short-term forecasting, data collection was at random over a one-year period and a random selection is made of three months, and fifty days from these three months. Calculations for call volumes are for each

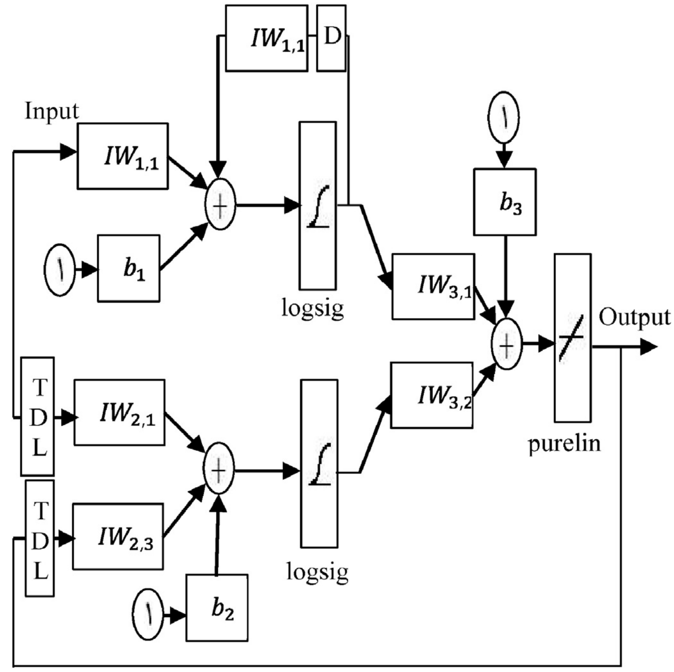

Fig. 1. Proposed model.

15-min case and 96 observations were made per day. Over fifty days, samples of 4800 observations were extracted from the raw data. Fig. 2 shows a scatter plot of call volumes per 15 min.

This study divides the data (4800 observations) into three parts (training, test, and check sets) to examine model performance. As per the standard method for train-and-test, this study removes 20% of the samples for testing and uses the remaining 80% for training (Thowmey & Smith, 1995). Further, the study uses 2% of the sample for checking, to prevent the problem of over-fitting. Therefore, 864 observations (test set) from the first 3840 (training set) are used to estimate the Neural Network weights of forecasting model and 96 observations (one day) to check it. After multiple iterations, the error is minimized thereby converging. The error signal can be given as:

$$e_i = d_i - y_i \quad (2)$$

where d_i is the targeted response and y_i is the actual output produced by the network response to input. The performance function used for the

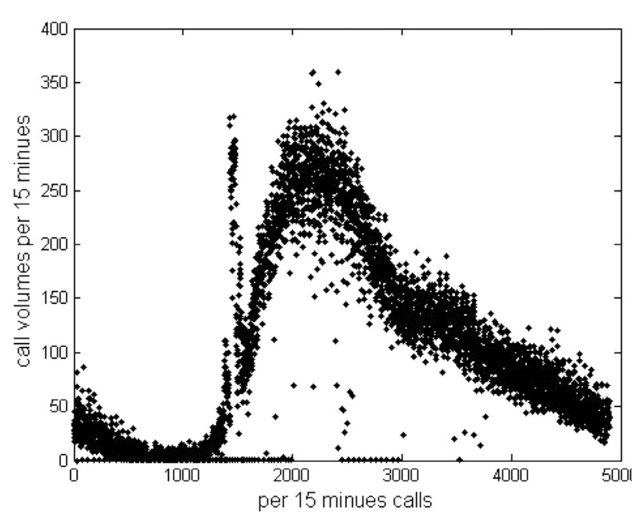

Fig. 2. The scatter plot of call volumes per 15 min.

Please cite this article as: Ebadi Jalal, M., et al., Forecasting incoming call volumes in call centers with recurrent Neural Networks, *Journal of Business Research* (2016), http://dx.doi.org/10.1016/j.jbusres.2016.04.035