Engineering Applications of Artificial Intelligence 55 (2016) 231–238

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server

Chi-Jie Lu [a,*], Ling-Jing Kao [b]

[a] Department of Industrial Management, Chien Hsin University of Science and Technology, Taiwan
[b] Department of Business Management, National Taipei University of Technology, Taiwan

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form
11 March 2016
Accepted 29 June 2016

Keywords:
Sales forecasting
Clustering
Ensemble learning
Linkage method
Extreme learning machine
Computer server

ABSTRACT

Sales forecasting has long been crucial for companies since it is important for financial planning, inventory management, marketing, and customer service. In this study, a novel clustering-based sales forecasting scheme that uses an extreme learning machine (ELM) and assembles the results of linkage methods is proposed. The proposed scheme first uses the K-means algorithm to divide the training sales data into multiple disjointed clusters. Then, for each cluster, the ELM is applied to construct a forecasting model. Finally, a test datum is assigned to the most suitable cluster identified according to the result of combining five linkage methods. The constructed ELM model corresponding to the identified cluster is utilized to perform the final prediction. Two real sales datasets of computer servers collected from two multinational electronics companies are used to illustrate the proposed model. Empirical results showed that the proposed clustering-based sales forecasting scheme statistically outperforms eight benchmark models, and hence demonstrates that the proposed approach is an effective alternative for sales forecasting.

1. Introduction

Sales forecasting is crucial for a company for financial planning, inventory management, marketing, and customer service. For example, sales forecasting has been used to estimate the required inventory level for satisfying market demand and avoiding the problem of over or under stocking. An effective sales forecasting model can reduce the bullwhip effect, thereby improving a company's supply chain management and sales management efficacy and, ultimately, increasing profits. Inaccurate sales forecasting may cause product backlog, inventory shortages, and unsatisfied customer demands (Luis and Richard, 2007; Thomassey, 2010; Lu et al., 2012; Lu, 2014). Therefore, it is important to develop an effective sales forecasting model which can generate accurate and robust forecasting results.

A number of sales forecasting studies have been proposed in the literature, while clustering-based forecasting models have been adopted to improve prediction accuracy (Tay and Cao, 2001; Prinzie and Van den Poel, 2006; Lai et al., 2009; Lu and Wang, 2010; Venkatesh et al., 2014; López et al., 2015). The fundamental idea of the clustering-based forecasting model is to utilize a clustering algorithm to partition whole training data into multiple disjoint clusters and construct a forecasting model for every cluster. The test data are assigned to a cluster by their similarity, and the forecasting model of a particular cluster is used to obtain forecasting outcomes for that cluster. Because data in the same cluster have similar data patterns, the clustering-based forecasting model can produce better forecasting accuracy than the forecasting model built upon a complete dataset.

Even though Venkatesh et al. (2014) found that the clustering-based approach yielded much smaller forecasting errors than the approach of direct prediction on the entire sample without clustering, the choice of the clustering approach, the similarity measurement, and the predictor will impact the performance of the clustering-based forecasting model. In the literature, a self-organizing map (SOM), the growing hierarchical self-organizing map (GHSOM), and the K-means clustering approach were applied to cluster data, while the support vector machine (SVM), support vector regression (SVR), case-based reasoning (CBR), neural networks, decision trees, and autoregressive integrated moving average (ARIMA) were used as predictors in the literature (Tay and Cao, 2001; Cao, 2003; Chang and Lai, 2005; Chang et al., 2009; Lai

* Correspondence to: Department of Industrial Management, Chien Hsin University of Science and Technology, Zhong-Li Dist., Taoyuan City 32097, Taiwan.
E-mail addresses: jerrylu@uch.edu.tw, chijie.lu@gmail.com (C.-J. Lu).

http://dx.doi.org/10.1016/j.engappai.2016.06.015
0952-1976/

et al., 2009; Huang and Tsai, 2009; Badge and Srivastava, 2010; Kumar and Patel, 2010; Lu and Wang, 2010; Zhang and Yang, 2012; Lu and Chang, 2014).

No matter which clustering approach is adopted, the linkage method must be selected to determine the similarity between objects so that a new observation can be assigned to the appropriate cluster. The single linkage, complete linkage, centroid linkage, median linkage and Ward's linkage methods are five well-known and frequently used linkage methods in clustering analysis, but different linkage methods have different characteristics and will generate different similarity measurement results (Palit and Popovic, 2005; Hair et al., 2006; Nandi et al., 2015).

Most of the specific clustering-based forecasting models mentioned above use only one linkage method to calculate the similarity of the prediction target and the clusters. However, using only one linkage method in the clustering-based forecasting model cannot provide a stable and effective outcome. Therefore, to solve this problem, this study proposed the use of ensemble learning to assemble the results of different linkage methods. Ensemble learning is a paradigm, where several intermediate classifiers or predictors are generated and combined to finally get a single classifier or predictor. It can be used to avoid the selection of the worst learning algorithm and improve the performance of classification or prediction (Dietterich, 2000; Polikar, 2006; Yang et al., 2010; Galar et al., 2012). Among various methods for the creation of an ensemble of classifiers, majority voting is the most widely used ensemble technique and considered a simple and effective scheme (Lam and Suen, 1997; Shahzad and Lavesson, 2013). Yeon et al. (2010) also proved majority voting is the optimal solution in the case of no concept drift. The majority voting scheme follows democratic rules, i.e., the class with highest number of votes is the outcome. Majority voting does not assume prior knowledge about the problem at hand, or classifiers, and does not require any parameter tuning once the individual classifiers have been trained (Lam and Suen, 1997).

Instead of using conventional predictors like ARIMA or artificial neural networks, this study used extreme learning machine (ELM) as the predictor due to its great potential and superior performance in practical applications (Huang et al., 2015). ELM is a novel learning algorithm for single-hidden layer feedforward neural networks (SLFNs), which randomly selects the input weights and analytically determines the output weights of SLFNs (Huang et al., 2006). Different from traditional gradient-based learning algorithms for neural networks, ELM not only tends to reach the smallest training errors but also the smallest norm of output weights. Thus, the ELM algorithm provides much better generalization performance with much faster learning speed and avoids many issues faced with the traditional algorithms, such as stopping criterion, learning rate, number of epochs and local minima, and the over tuned problems (Yeon et al., 2010). ELM has attracted much attention in recent years and has become an important forecasting method (Sun et al., 2008; Wong and Guo, 2010; Chen and Ou, 2011; Lu and Shao, 2012; Wang and Han, 2015).

In this study, the clustering-based sales forecasting scheme is implemented as follows. First, the K-means algorithm is used to partition the whole training sales data into multiple disjoint clusters. We adopted the K-means algorithm because it is one of the most popular methods (Nandi et. al, 2015) and is effective and efficient in most cases (Jain, 2010). Then, the ELM is applied to the construct forecasting model for each cluster. Next, for a given testing dataset, the ensemble learning based on the majority voting scheme is utilized to combine the results of the five linkage methods, including single linkage, complete linkage, centroid linkage, median linkage, and Ward's linkage, to find the cluster which the testing data set belongs to. Finally, the ELM model corresponding to the identified cluster is used to generate the final prediction result.

Two real, monthly aggregate sales data sets of computer servers collected from two multinational electronics companies were utilized as an illustrative example to evaluate the performance of the proposed model. The forecasting accuracy of the proposed approach was compared with three single forecasting models, i.e., simple naïve forecast, seasonal naïve forecast, and pure ELM models, and five clustering-based forecasting models with different linkage methods. The model comparison shows that the proposed approach provides much more accurate predictions. This study contributes to the literature by proposing ensemble linkage to avoid the problem caused by choosing a single linkage method as well as by providing an application of the ELM model.

The rest of this paper is organized as follows. Section 2 gives a brief introduction about extreme learning machine. The proposed clustering-based sales forecasting model is thoroughly described in Section 3. Section 4 presents the experimental results. The paper is concluded in Section 5.

2. Extreme learning machine

ELM is one kind of single hidden-layer feedforward neural networks (SLFNs). It has a three layers structure, including the input layer, the hidden layer, and the output layer. It endeavors to conquer the challenging issues of the traditional SLFNs such as slow learning speed, trivial parameter tuning and poor generalization capability (Huang et al., 2015).

One key feature of ELM is that a researcher may randomly choose input weights and hidden node parameters. After the input weights and hidden nodes parameters are chosen randomly, SLFNs become a linear system where the output weights of the network can be analytically determined using a simple generalized inverse operation of the hidden layer output matrices (Huang et al., 2006).

Consider N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{y}_i)$ where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ is the input data and $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{im}]^T \in R^m$ is the target output. If the SLFNs with η hidden neurons and activation function vector $\theta(x)$ can approximate N samples with zero error, this means $\sum_{i=1}^{N} \|\mathbf{q}_i - \mathbf{y}_i\| = 0$, where $\{\mathbf{q}_i$, for $i = 1, 2, \ldots, N\}$ is the output values of the SLFN. It can then be written compactly as:

$$\mathbf{HB} = \mathbf{Y} \tag{1}$$

where $\mathbf{H}_{N\times\eta} = [\theta(\mathbf{w}_i \cdot \mathbf{x}_j + b_i)] = \begin{bmatrix} \theta(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & \theta(\mathbf{w}_\eta \cdot \mathbf{x}_1 + b_\eta) \\ \vdots & \ddots & \vdots \\ \theta(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & \theta(\mathbf{w}_\eta \cdot \mathbf{x}_N + b_\eta) \end{bmatrix}_{N\times\eta}$ represents the hidden layer output matrix of the neural network. The ith column of $\mathbf{H}$ is the ith hidden node output with respect to inputs $\mathbf{x}_i$. $\mathbf{B}_{\eta\times m} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_\eta]$ is the matrix of output weights and $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the ith hidden node and the output nodes. $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector connecting the ith hidden node and the input nodes; $\mathbf{w}_i \cdot \mathbf{x}_j$ denotes the inner product of $\mathbf{w}_i$ and $\mathbf{x}_j . b_i$ is the threshold (bias) of the ith hidden node. $\mathbf{Y}_{N\times m} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ is the matrix of targets.

Huang et al. (2006) has proven that the input weights $\mathbf{w}_i$ and the hidden layer biases b_i of SLFNs need not be adjusted and can be given arbitrarily. Under this assumption, the input weights $\mathbf{w}_i$ and hidden biases b_i are randomly generated in ELM algorithm and the output weights can be determined as simple as finding the least-square solution to the given linear system. The minimum norm least-square solution to the linear system (i.e. Eq.(1)) is

$$\hat{\mathbf{B}} = \mathbf{H}^{\psi}\mathbf{Y} \tag{2}$$

where $\mathbf{H}^{\psi}$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$. The minimum norm least-square solution is unique and has the smallest norm among all the least-square solutions (Huang et al., 2006).