Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 2 (2015) 82 – 86

Procedia MANUFACTURING

2nd International Materials, Industrial, and Manufacturing Engineering Conference, MIMEC2015, 4-6 February 2015, Bali Indonesia

A clustering-based sales forecasting scheme using support vector regression for computer server

Wenseng Dai[a], Yang-Yu Chuang[b], Chi-Jie Lu[b]*

[a]*International Monetary Institute, Financial School, Renmin University of China, Beijing 100872, China*
[b]*Department of Industrial Management, Chien Hsin University of Science and Technology, Taoyuan County 32097, Taiwan*

Abstract

In this study, a clustering-based sales forecasting scheme based on support vector regression (SVR) is proposed. The proposed scheme first uses k-means algorithm to partition the whole training sales data into several disjoint clusters. Then, for each group, the SVR is applied to construct forecasting model. Finally, for a given testing data, three similarity measurements are used to find the cluster which the testing data belongs to and then employee the corresponding trained SVR model to generate prediction result. A real aggregate sales data of computer server is used as an illustrative example to evaluate the performance of the proposed model. Experimental results revealed that the proposed clustering-based sales forecasting scheme outperforms the single SVR without data clustering and hence is an effective alternative for computer server sales forecasting.

Selection and Peer-review under responsibility of the Scientific Committee of MIMEC2015
Keywords: Sales forecasting; computer server; clustering algorithm; support vector regression

1. Introduction

Computer server industry is one of the most important segments in the internet age due to web search engines and internet application companies start to use cloud computing as their leading solution to achieve mass-data searching and accessing within a short period time. As the computer servers have long life cycle and premium price over other

* Corresponding author. Tel.: +886-3-4581196 ext 6712.
E-mail address: jerrylu@uch.edu.tw; chijie.lu@gmail.com

Selection and Peer-review under responsibility of the Scientific Committee of MIMEC2015
doi:10.1016/j.promfg.2015.07.014

computer products, how to create an effective sales forecast model becomes an essential topic for computer server providers.

Sales forecasting is one of the most important tasks in many companies since it can be used to determine the required inventory level to meet consumer demand and avoid the problem of under/over stocking. A number of studies have been conducted which focus on sales forecasting in various industries [1-5]. However, only few studies concentrate on sales forecasting in computer product industry [6-8]. To the best knowledge of authors, there is no study focus on developing sales forecasting model for computer server.

Many studies have proposed different kinds of clustering-based forecasting models to improve prediction performance [6,9,10]. For a given input data, the clustering-based forecasting model first uses a data clustering algorithm to partition the whole input data into several disjoint groups. Then, for each group, the prediction model is constructed to produce the output. Finally, for the prediction target, finding the group it belongs to and using the trained forecasting method corresponding to the cluster to generate final forecasting results. Since the partitioned groups have more uniform/stationary data structure than that of the whole input data, it will become easier for constructing effective forecasting model. However, most existing studies used single similarity measurement method to calculate the similarity of the prediction target and the clusters. This may make the clustering-based forecasting models cannot provide stable and effective results.

In this study, a clustering-based sales forecasting scheme by combining K-means algorithm with SVR and three similarity measurements is proposed. Although there are many different clustering algorithms have been proposed, but K-means algorithm is still one of the most effective clustering algorithms [11]. This study considers the SVR as the predictor due to its great potential and superior performance in practical applications [13]. SVR based on statistical learning theory is an effective neural network algorithm for solving nonlinear regression estimation problems and has been successfully used in sales forecasting[6-8].

In the proposed clustering-based sales forecasting scheme, first, the predictors are used as inputs of K-means algorithm to group the input data into several disjoint clusters. Then, for each cluster, the SVR forecasting model is constructed and the final forecasting results can be obtained. Finally, for a given testing data, three similarity measurements including MiniMax, Median and Mean methods are used to find the cluster which the testing data belongs to and then employee the corresponding trained SVR model to generate prediction result. A real monthly aggregate sales data of computer server collected from a computer server company in Taiwan is utilized as an illustrative example to evaluate the performance of the proposed model. The single SVR model without data clustering is used as baseline method in this paper.

2. Support vector regression

Support vector regression (SVR) is an artificial intelligent forecasting tool based on statistical learning theory and structural risk minimization principle [13]. It can be expressed as $f(x) = \left(\boldsymbol{z} \cdot \emptyset(x)\right) + b$, where $\boldsymbol{z}$ is weight vector, b is bias and $\emptyset(x)$ is a kernel function which use a non-linear function to transform the non-linear input to be linear mode in a high dimension feature space. Vapnik [13] introduced so-called ε-insensitivity loss function to SVR. When the predicted value falls into the band area, the loss is zero. Contrarily, if the predicted value falls out the band area, the loss is equal to the difference between the predicted value and the margin. After selecting proper modifying coefficient (C), width of band area (ε) and kernel function ($\emptyset$(x)), the optimum of each parameter can be resolved though Lagrange function. Since the most widely used kernel unction is the radial basis function (RBF), it is applied in this study as kernel function. The performance of SVR is mainly affected by the setting of parameters C, δ and ε (the parameter of RBF). This study employees grid search which uses exponentially growing sequences of the parameters to identify good parameters (for example, $\mathrm{C} = 2^{-15}, 2^{-13}, \ldots, 2^{15}$).

3. Proposed clustering-based forecasting scheme

In this study, a clustering-based sales forecasting scheme is proposed. The proposed model contains training and testing stages. In the training phase, the training dataset is divided into several disjoint groups and the SVR is used in each group to build forecasting model. The detailed procedure of the training phase is as following steps.