Expert Systems with Applications 36 (2009) 6978–6985

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Supplier selection based on hierarchical potential support vector machine

Xuesong Guo [a,*], Zhiping Yuan [b], Bojing Tian [c]

[a] School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an, 710049, PR China
[b] School of Management, Xi'an Jiaotong University, Xi'an, 710049, PR China
[c] School of Architecture, Xi'an University of Architecture and Technology, Xi'an, 710055, PR China

ARTICLE INFO

Keywords:
Supplier selection
Potential support vector machine
Decision tree

ABSTRACT

Supplier selection is an important and widely studied topic since it has significant impact on purchasing management in supply chain. Recently, support vector machine has received much more attention from researchers, while studies on supplier selection based on it are few. In this paper, a new support vector machine technology, potential support vector machine, is introduced and then combined with decision tree to address issues on supplier selection including feature selection, multiclass classification and so on. So, hierarchical potential support vector machine and hierarchical system of features are put forward in the paper, and experiments show the proposed methodology has much better generalization performance and less computation consumptions than standard support vector machine.

0. Introduction

Supplier selection is one of the most critical activities of purchasing management in supply chain, because of the key role of supplier's performance on cost, quality, delivery and service in achieving the objectives of a supply chain.

Supplier selection is a multiple criteria decision making (MCDM) problem which is affected by several conflicting factors. Consequently, a purchasing manager must analyze the trade off among the criteria. And MCDM techniques support the decision-makers in evaluating a set of alternatives.

Among the methods supporting supplier selection, artificial intelligence (AI) based models play important role in the domains. AI based on computer aided systems can be "trained" by a purchasing expert or historic data. Subsequently, non-experts facing similar but new situations can consult the system. Examples of methods based on AI technologies that have been applied to supplier choice include neural networks and other new techniques.

One of the strengths of the methods is that they do not require formulization of the decision making process. In this respect, AI technologies can cope better with complexity and uncertainty than "traditional methods", because they are designed to be more like human judgment functioning.

The user of AI systems only has to provide the information on characteristics of current situation, e.g. performance of a supplier on the criteria The AI technologies subsequently make the actual trade off of the users, based on what they have "learned" from the experts or cases in the past. A decision support system based on neural networks is put forward in Albino and Garavelli (1998). Moreover, other technologies based on AI also have been applied in domains of supplier selection (Khoo, Tor, & Lee, 1998; Cook, 1997; Ng & Skitmore, 1995).

The support vector machine (SVM) method is a new and promising classification and regression technique. SVM, a development in statistical learning theory, is recently of increasing interests of researchers, though researches on issues of supplier selection are few (Sun, Xie, & Xue, 2005; Wen & Li, 2006). It is not only well founded theoretically, but also superior in practical applications. Moreover, SVM has been successfully applied in a wide variety of domains including handwriting recognition (Park & Woo Kim, 2005), speaker identification (Campbell, Campbell, & Gleason, 2007; Wan & Renals, 2005), face detection (Li & Tang, 2007), and text categorization (Lin & Peng, 2006). In most of these cases, the performance of SVM is either similar or significantly better than that of traditional machine learning approaches, including neural networks. Nevertheless, SVM has some problems to be solved when applied to practice.

Firstly, SVM is a binary classifier while many multiclass classifications are required to be accomplished in practice. In the domains of supplier selection, suppliers to be evaluated will usually be divided into more than two categories according to given criteria.

Secondly, feature selection is required to be performed while applied to classification. Feature selection is an important issue in building classification systems. It is advantageous to limit the number of input features in a classifier in order to have a good predictive and less computationally intensive model (Zhang, 2000). With a small feature set, the explanation of rationale for the classification decision can be realized more easily. In terms of supplier selection, supplier can be described by some attributes originally, which can be represented with features in the view of machine

* Corresponding author.
E-mail address: guoxues1@163.com (X. Guo).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.08.074

learning. And a subset of attributes by which suppliers are described should be selected from original ones so as to establish a system of criteria because some features are too noisy or not conveying correct information and can't be used in subsequent evaluation procedures. In order to accomplish feature selection when standard SVM is applied, many complex algorithms, i.e., GA (Huang & Wang, 2006), chaos optimization etc., have to be used only to improve computation consumption.

In this paper, a novel SVM approach, potential support vector machine (P-SVM) (Hochreiter & Obermayer, 2006) which can accomplish binary classifiers construction and feature selection simultaneously, is introduced to the study at first. And method proposed in the paper combines P-SVM and decision tree into a new algorithm, called as hierarchical potential support vector machine, which can accomplish multiclass classification and feature selection simultaneously. Therefore, a new system of criteria represented as hierarchical structure, which can be applied to supplier description, established with the method employed to solve problems of standard SVM mentioned above is put forward in the paper.

1. Proposed methodology

1.1. Introductions to P-SVM

In this section, a new technology of SVM called as potential support vector machine (P-SVM), which can perform feature selection and classification simultaneously is introduced. The main differences to previous approaches are just as follows:

Sphering: In order to judge the relevance of feature components, the variance should be normalized, that is, the data should be sphered (whitened). Therefore, an objective is formulated according to which the classifier is selected by maximizing the margin after sphering. It turns out that sphering has two advantages for support vector machine techniques. Firstly, the derived new support vector machine approach is invariant to linear transformation of the data, which are the margin bounds. Secondly, tighter margin bounds can be obtained.

New constraints: The constraints of the optimization problem are modified in order to ensure that the classifier is optimal with respect to the mean squared error between the classification function and the labels. In contrast to previous approaches where one constraint is associated with each of the m training examples, each constraint is now associated with one feature and the number of new constraints is equal to the number N of features.

Support features: The combination of the new objective with the new constraints allows assigning support vector weights to features, and the normal vector of the classification boundary is expanded in terms of these weights rather than in terms of support vector data points. This allows feature selection according to whether a feature is a support vector or not. As a side effect the dual optimization problem can be efficiently solved using a technique similar to sequential minimal optimization.

In summary, a classifier is selected from the set of all classifiers with minimal mean squared error which yields the largest margin after sphering the data. The new support vector machine removes irrelevant features, which lead to a minimal increase of the mean squared error when removed. More formally, feature selection is done by assigning support vector weights to features, the features which are support vectors are selected.

In the following subsections, we first briefly review the standard support vector machine (SVM). Then, we introduce a new objective for achieving scale-invariant support vector machine, present new constraints for correct classification, and combine the new objective and the new constraints into one framework. Finally, a summary of the new technique is given.

1.1.1. Standard support vector machine

Consider a set of m objects, which are described by feature vectors $\mathbf{x} \in R^N$, and let us represent the data set by the matrix $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$. We furthermore assume that every object belongs to one of two classes and the class membership is denoted by a binary label $y \in \{+1, -1\}$. The labels for the m objects are summarized by a label vector $\mathbf{y}$, where y_i is the label for $\mathbf{x}_i$.

The goal of standard support vector machine is to construct a linear classifier based on the feature vector $\mathbf{x}$. In standard support vector machine, the classifier is defined by taking the signs of the classification function shown

$$f(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i + b, \tag{1}$$

where the weight vector $\mathbf{w}$ has been normalized such that the margin ρ, that is the distance between the classification boundary and the closest data point, is $\rho = \frac{1}{\|\mathbf{w}\|}$.

Standard support vector machine constructs a classification function which maximizes the margin under the constraint that the training data is classified correctly, just as shown

$$\min_{w,b} \frac{1}{2}\|\mathbf{w}\|^2, \tag{2}$$

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geqslant 1. \tag{3}$$

Here, we assume that the data $\mathbf{X}$ with label vector $\mathbf{y}$ is linearly separable, otherwise slack variables have to be introduced. If the number of training samples m is larger than the VC dimensions h, then one obtains the following bound on the generalization error $R(f)$ of f, just as shown

$$R(f) \leqslant R_{\text{emp}}(f) + \sqrt{\frac{1}{m}\left(h\left(\ln\left(\frac{2m}{h}\right)+1\right) - \left(\frac{\ln(\delta)}{4}\right)\right)} \tag{4}$$

which holds with probability $1-\delta$. δ denotes the probability, that a training set $\mathbf{X}$ of size m has been randomly drawn from the underlying distribution, for which the bound equation does not hold. $R_{\text{emp}}(f)$ denotes the training error of f (also called " empirical risk of f"). For the set of all linear classifiers defined on $\mathbf{X}$, for which $\rho \geqslant \rho_{\min}$ holds, one obtain

$$h \leqslant \min\left\{\left[\frac{R^2}{\rho_{\min}^2}\right], N\right\} + 1, \tag{5}$$

where $[\bullet]$ denotes the integer part and R is the radius of the smallest sphere in the data space, which contains all the training data. The fact that the bounds become smaller for increasing ρ and decreasing N motivates the maximum margin principle and the concept of feature selection (Vapnik, 1995, 1998).

1.1.2. A scale invariant objective function

Both the selection of a classifier using the maximum margin principle and the values obtained for the bounds on the generalization error described in the last section suffer from the problem addressed in Hochreiter and Obermayer (2006); SchÄolkopf, Shawe-Taylor, and Smola (1999) that they are not invariant under linear transformations.

Here, we suggest scale the training data such that the margin ρ remains constant while the radius R of the sphere containing all training data becomes as small as possible. The scaling results derive a new sphere with radius $\tilde{R}$ which still contains all training data and which leads to a tight margin-based bound for the generalization error. Optimality is achieved when all directions orthogonal the normal vector $\mathbf{w}$ is scaled to zero and $\tilde{R} = \min_{t \in R} \max_i |\widehat{w} \cdot \mathbf{x}_i + t| \leqslant \max_i |\widehat{w} \cdot \mathbf{x}_i|$, where $\widehat{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$. Note that with offset b of the classification function the sphere must not be centered at the origin. Unfortunately, above formulation does not lead to an optimization problem. In this paper, the form of data points input to P-SVM is illustrated in Fig. 1.