Measurement 46 (2013) 2726–2734

Contents lists available at SciVerse ScienceDirect

Measurement

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

Rolling element bearing fault detection using support vector machine with improved ant colony optimization

Xu Li [a], A'nan Zheng [b], Xunan Zhang [c], Chenchen Li [b], Li Zhang [b,*]

[a] The State Key Laboratory of Rolling & Automation, Northeastern University, Shenyang 110004, China
[b] School of Information, Liaoning University, Shenyang 110036, China
[c] Automation, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 14 December 2012
Received in revised form 21 March 2013
Accepted 29 April 2013
Available online 18 May 2013

Keywords:
Fault detection
Rolling element bearing
Ant colony optimization
Support vector machine
Meshing

ABSTRACT

In support vector machine (SVM), it is quite necessary to optimize the parameters which are the key factors impacting the classification performance. Improved ant colony optimization (IACO) algorithm is proposed to determine the parameters, and then the IACO-SVM algorithm is applied on the rolling element bearing fault detection. Both the optimal and the worst solutions found by the ants are allowed to update the pheromone trail density, and the mesh is applied in the ACO to adjust the range of optimized parameters. The experimental data of rolling bearing vibration signal is used to illustrate the performance of IACO-SVM algorithm by comparing with the parameters in SVM optimized by genetic algorithm (GA), cross-validation and standard ACO methods. The experimental results show that the proposed algorithm of IACO-SVM can give higher recognition accuracy.

1. Introduction

In the modern production, fault diagnosis technology of mechanical equipment is increasingly important. If a device failure is not discovered and eliminated timely, it will cause mechanical damage and serious death. Therefore, the status of fault diagnosis in the production line should not be neglected.

Rolling element bearings consist one of the most widely used industrial machine elements, and it is the interface between the stationary and the rotating part of the machine. It is important to give a fast and accurate detect of the existence of a fault in an installation during the operation process, since an unexpected failure of machine can lead to unacceptably long time maintenance stops [1]. Statistics show, about 30% of the rotating machinery faults are caused by the damage of the bearings. Therefore as one of the important parts of the running mechanical bearing, we have to improve the detecting ability of fault diagnosis.

The early rolling failure diagnostic methods include the use of hearing, shock pulse and resonance demodulation technologies. However, the accuracy and efficiency of these diagnostic methods cannot reach the standard of the industry level. With the continuous development of diagnostic techniques, artificial intelligence , such as expert system, artificial neural network (ANN), fuzzy logic, immune genetic algorithm have been widely used in machine fault diagnosis [2–9]. These methods are based on the empirical risk minimization principle, and there are some common shortcomings, such as relapsing into local minimum easily and slow convergence velocity and overfitting. Especially, generalization ability in a limited number of samples is too low. When using the artificial intelligence in the fault diagnosis, the lack of fault samples is the bottleneck problem. Low generalization ability can lead to wrong fault diagnosis results. Comparing with these methods, support vector machine (SVM) can perform significant well when facing with those limited factors.

SVM is a new machine learning method based on the structural risk minimization (SRM) principle, the purpose of which is to solve classification problems by maximizing

* Corresponding author. Tel.: +86 13390552858.
E-mail address: zhang_li@lnu.edu.cn (L. Zhang).

0263-2241/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.measurement.2013.04.081

the margin between the two opposing classes. SVM can solve the problem of model selection, over-fitting, nonlinear, the curse of dimensionality and local minimum in a better way [10,11]. It embodies the SRM principle that has achieved higher generalization performance with small number of samples and is shown superior to the empirical risk minimization that neural networks use. The introduction of the kernel function is a major advantage of the support vector machine. It endows with SVM the ability to deal with nonlinear classification problem by mapping the nonlinear feature space to high dimensional feature space to solve linear problem. The study found that different kernel functions have little effect on the performance of SVM, but the impact of the choice of parameters in different kernel functions on the performance of SVM is huge. Penalty factor C weighting the proportion between experience risk and confidence range is also a key factor affecting the performance of SVM. At present, there are many methods for optimizing these parameters, such as genetic algorithm, artificial immune algorithm, ant colony algorithm, particle swarm [12–20].

An ACO-based algorithm for parameter optimization of support vector machines was proposed by Zhang et al. [15]. The new state transition rule that ants build solution by applying a probabilistic decision policy to move adjacent states and the improved state updating rule that only applied to that subset of parameters were used to enhance the performance. The state transition rule increases the diversity of solutions as well as the possibility of non-optimal solution. The improved state updating caused that the pheromone on the line of local optimal solution increases so fast that the algorithm is prone to fall into local extremum.

An improved ant colony optimization (IACO) algorithm for optimizing the parameters in SVM is presented in this paper. In the IACO, ant chooses several grid points with largest pheromones from the random selected points. While the way of updating pheromone trail density is changed, that is to say, increasing pheromone which is close to the optimal solution during the iteration, at the same time reducing pheromone which is close to the worst solution. Reasonable selection of meshing upper and lower bounds can improve the convergence rate. The IACO can reduce the blindness in the selection of parameters in SVM, and therefore obtains a higher diagnosis accuracy and convergence speed than other parameter optimization methods such as the cross verification trial and genetic algorithm.

The rest of this paper is organized as follows. In Section 2, we will give a brief introduction to the theory of SVM. Basic idea of ACO and the optimization procedure to the SVM parameters are presented in Section 3. In Section 4, we first describe how to choose the feature sets and the model parameters, and then the detailed comparative experiment is done to illustrate the performance of IACO-SVM by comparing with other optimization methods. Finally, this paper concludes with a summary in Section 5.

2. Support vector machine (SVM)

The support vector machine is a machine learning algorithm based on the structured risk minimization principle. There are many advantages in SVM including complete theory, strong adaptability, global optimization, insensitive to the dimension, shorter training time and good generalization performance. The basic idea of SVM to is construct optimal separating hyperplane and maps the training samples from the input space into a higher dimensional feature space via a mapping function φ.

Give a training set $\{(x_i, y_i)\}_{i=1}^{l}$, $x_i \in R^n, y_i \in \{1, -1\}$, where x_i is the input vector and y_i is the label of the x_i, and l is the number of the input vectors and n is the number of input dimension. Structure hyperplane is $w \cdot x + b = 0$, besides in order to meet the SRM principle the classification hyperplane should satisfy

$$y(w \cdot x + b) \geqslant 1 \tag{1}$$

where w is the normal direction of a separation plane, and b is the scalar. Since the distance between the closest sample points and a separation plane is $1/||w||$, the process of finding the optimal separation plan is to minimize$||w||^2$. Therefore the problem of constructing the optimal hyperplane is transformed into the following quadratic programming problem:

$$\begin{aligned} &\min_{x} \quad \tfrac{1}{2}||w||^2 \\ &s.t. \quad y_i(w \cdot x_i + b) \geqslant 1,\ i = 1, \ldots,\ l \end{aligned} \tag{2}$$

In most instances, problems could not be separated linearly, therefore some training sample points do not satisfy the constraint conditions. To obtain the optimal classifier, $|w|$ should be minimized under the following constraints

$$y_i(w \cdot x_i + b) \geqslant 1 - \zeta_i,\ i = 1, \ldots, l \tag{3}$$

The variables ζ_i are positive slack variables, which is necessary to allow misclassification. The objective function will produce the classification error, so we introduce a generalization parameter C. The objective function and the constraints transformed into the following problem:

$$\begin{aligned} &\min_{w\ b\ \zeta} \quad \tfrac{1}{2}||w||^2 + C\sum_{i=1}^{l}\zeta_i \\ &s.t. \quad y_i(w \cdot x_i + b) \geqslant 1 - \zeta_i \qquad \zeta_i \geqslant 0,\ i = 1, \ldots,\ l \end{aligned} \tag{4}$$

The greater we set the generalization parameter, the higher the misclassification error will be, as well as the heavier of the punishment.

According to Lagrangian principle, the above problem can be transformed to its dual form as follows:

$$\begin{aligned} &\min_{\alpha} \quad \tfrac{1}{2}\sum_{i=1}^{1}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{j=1}^{l}\alpha_j \\ &s.t. \quad \sum_{i=1}^{l} y_i \alpha_i = 0 \quad \alpha_i \geqslant 0 \quad i = 1, \ldots,\ l \end{aligned} \tag{5}$$

By introducing the mapping function, the sample space is mapped into a high dimensional feature space. The optimization problem can be rewritten in the form

$$\begin{aligned} &\min_{\alpha} \quad \tfrac{1}{2}\sum_{i=1}^{1}\sum_{j=1}^{l} y_i\ y_j \alpha_i \alpha_j k(x_i \cdot x_j) - \sum_{j=1}^{l}\alpha_j \\ &s.t. \quad \sum_{i=1}^{l} y_i \alpha_i = 0 \qquad \alpha_i \geq 0 i = 1, \ldots,\ l \end{aligned} \tag{6}$$