Advances in Engineering Software 71 (2014) 46–51

Contents lists available at ScienceDirect

Advances in Engineering Software

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

Force based tool wear monitoring system for milling process based on relevance vector machine

Guofeng Wang *, Yinwei Yang, Qinglu Xie, Yanchao Zhang

Key Laboratory of Mechanism Theory and Equipment Design of Ministry of Education, Tianjin University, Tianjin 300072, China

A R T I C L E I N F O

Article history:
Received 22 December 2013
Received in revised form 28 January 2014
Accepted 16 February 2014
Available online 16 March 2014

Keywords:
Tool wear monitoring
Relevance vector machine
Multinomial function
Milling process

A B S T R A C T

The monitoring of tool wear status is paramount for guaranteeing the workpiece quality and improving the manufacturing efficiency. In some cases, classifier based on small training samples is preferred because of the complex tool wear process and time consuming samples collection process. In this paper, a tool wear monitoring system based on relevance vector machine (RVM) classifier is constructed to realize multi categories classification of tool wear status during milling process. As a Bayesian algorithm alternative to the support vector machine (SVM), RVM has stronger generalization ability under small training samples. Moreover, RVM classifier results in fewer relevance vectors (RVs) compared with SVM classifier. Hence, it can be carried out much faster compared to the SVM. To show the advantages of the RVM classifier, milling experiment of Titanium alloy was carried out and the multi categories classification of tool wear status under different numbers of training samples and test samples are realized by using SVM and RVM classifier respectively. The comparison of SVM with RVM shows that the RVM can get more accurate results under different number of small training samples. Moreover, the speed of classification is faster than SVM. This method casts some new lights on the industrial environment of the tool condition monitoring.

1. Introduction

The monitoring of tool wear for milling process is paramount for guaranteeing the workpiece quality and improving the manufacturing efficiency. To realize more accurate recognition of the tool wear status, multi categories classification is preferred because it can not only judge whether the cutter is broken or not, but also recognize the tool wear scope [1]. In decades, many classifiers have been proposed such as feed forward neural network [2,3], hidden Markov model [4,5], and condition random field [6]. Recently, more attention is paid on the classification under small training samples. Because the tool wear is a gradual process and the tool wear status needs to be measured intermittently, the collection of training samples is time consuming and costive. So it is not realistic to collect large amount of samples to train the classifier. In addition, the complexity of tool wear profile also casts higher demand on the generalization ability of the classifier because it is impossible to collect all data within certain tool wear scope [7,8].

Support vector machine (SVM) is a novel machine learning method based on structure risk minimization principle, which can find global optimum solutions for problems with small training samples, high dimensions and non-linear features [9–11]. Therefore, in recent years many researchers used SVM as a classifier to realize the tool wear monitoring. Salgado and Alonso [12] used the least squares version of support vector machine to estimate the feed cutting force and realize the monitoring of the tool wear in turning. Elangovan et al. [13] analyzed the effect of SVM kernel functions on classification of the tool wear for a single point cutting tool based vibration signal. Shi and Gindy [14] realized the tool wear prediction using the combination of SVM and principal component analysis (PCA) [15]. Binsaeid et al. [16] used SVM to realize the classification of the tool wear status under different cutting conditions. These applications show that the SVM classifier has strong ability to realize the multi categories classification especially in the case of small training samples. However, some shortcomings also exist in the SVM classifier. The first is that the ability of the classifier depends on an important parameter–penalty parameter C [14]. However, up to now, the selection of C can only be realized by trial and error method, which makes it hard to get the optimum value. The second is that SVM makes unnecessarily liberal use of basis function, which results in the number of support vectors required typically growing linearly with the size of the training set. Therefore, the classification speed of the classifier is lowered obviously [17].

* Corresponding author.
E-mail address: gfwangmail@tju.edu.cn (G. Wang).

http://dx.doi.org/10.1016/j.advengsoft.2014.02.002
0965-9978/

In this paper, relevance vector machine (RVM) combined with multinomial function criteria is proposed to realize the multi categories classification under small training samples. RVM is introduced by Tipping [18] as a Bayesian treatment alternative to the SVM. In comparison with SVM, RVM gains more flexibility and convenience because there is no need to set the penalty parameter which usually requires cross validation and post optimization [7,8]. Moreover, the number of relevance vectors is much smaller than that of support vectors which yields a sparser representation than the SVM. Therefore, RVM classification can be carried out with higher decision speed and better generalization ability [19].

To show the effectiveness of RVM based tool wear monitoring system in the case of small training samples, milling experiment of Titanium alloy was carried out and the force signals during the machining process were collected as the sensory information to depict the variation of the tool wear status. Based on six features extracted in the time domain and frequency domain, the training data under different number of small training samples are utilized to train the RVM and SVM classifier respectively. In addition, both classifiers are adopted to realize the classification under different number of test samples. By comparison, it can be concluded that the generalization ability of RVM is higher than SVM in the case of small training samples. In addition, the classification speed of RVM is higher than SVM, which makes it more suitable for real time tool condition monitoring.

The structure of this paper is organized as follows. In Section 2, the principle of binary RVM is presented. Moreover, multinomial function is proposed to realize the classification of multi categories. In Section 3, online tool wear monitoring system is constructed based RVM classifier and the milling experiment was setup to testify the effectiveness of the monitoring system. Six features are extracted and normalized to denote the relationship between the tool wear status and the force signal. Moreover, different kinds of combination of training and test samples are firstly selected. Based on the collected dataset, RVM are utilized to realize the multi categories tool wear monitoring and compare with SVM classifier. The results show that the RVM can get higher classification accuracy with lower time consuming in the case of small training samples. In Section 4, some useful conclusions are drawn.

2. Principle of RVM for multi class classifier

2.1. Basic principle of RVM

Relevance vector machine is a kind of sparse learning classifier which is based on Bayesian principle. For the given dataset $\mathbf{D} = (\mathbf{X},\mathbf{T})$, $\mathbf{X} = (\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n,\ldots,\mathbf{x}_N)$ and $\mathbf{T} = (t_1,t_2,\ldots,t_n,\ldots,t_N)$, $t_n\{0,1\}$, “1” denotes the training sample belongs to this class and “0” refuses it. N is the number of training samples. A probability based decision model is built in which the target value of the classifier is depicted by a posterior probability distribution $P(\mathbf{T}|\mathbf{X})$, which is expressed as [18]

$$P(\mathbf{T}|\mathbf{X}) = \prod_{n=1}^{N}\sigma\{f(\mathbf{x}_n,\mathbf{W})\}^{t_n}[1-\sigma\{f(\mathbf{x}_n,\mathbf{W})\}]^{1-t_n} \tag{1}$$

where $\sigma(f)$ is a predefined logistic sigmoid link function which is utilized to generalize the linear model. The mathematical expression of this function is given as

$$\sigma(f) = \frac{1}{1+e^{-f}} \tag{2}$$

Here, the function f is the linear weighted sum of M nonlinear basis functions, which is expressed as

$$f(\mathbf{x}_n,\mathbf{W}) = \mathbf{W}^{\mathrm{T}}\Phi_n = \sum_{i=1}^{M} w_i\phi(\mathbf{x}_n,\dot{\mathbf{x}}_i) \tag{3}$$

where $(\dot{\mathbf{x}}_1,\dot{\mathbf{x}}_2,\ldots\dot{\mathbf{x}}_i,\ldots,\dot{\mathbf{x}}_M)$ are the relevance vector (RV) sequences which come from the training data $\mathbf{X}$ (as shown in Fig. 1). $\mathbf{W} = (w_1, w_2,\ldots,w_M)^{\mathrm{T}}$ is the corresponding weight of these vectors. $\mathbf{\Phi}_n$ is a $M\times 1$ vector which is represented as

$$\Phi_n = \Phi(\mathbf{x}_n) = \{\phi(\mathbf{x}_n,\dot{\mathbf{x}}_1),\phi(\mathbf{x}_n,\dot{\mathbf{x}}_2),\ldots,\phi(\mathbf{x}_n,\dot{\mathbf{x}}_M)\}^{\mathrm{T}}$$

where $\phi(\mathbf{x}_n,\dot{\mathbf{x}}_i)$ is a Gaussian kernel function whose expression is given as

$$\phi(\mathbf{x}_n,\dot{\mathbf{x}}_i) = \exp\left(-\frac{||\mathbf{x}_n-\dot{\mathbf{x}}_i||^2}{\beta^2}\right) \tag{4}$$

where β is the width of kernel function.

It can be seen that the computation of the posterior probability P can be realized when several “most relevant” training data and their weight value are known [20]. Moreover, these relevant vectors do not necessarily reside along the decision boundary (as shown in Fig. 1). Therefore, the RVM typically yields a sparser representation than the SVM [17], which reduce the complexity of the model and improve generalization performance of the classifier.

2.2. Training of RVM classifier

The training of RVM is to determine the relevance vectors and its corresponding weight value. However, direct estimation from Eq. (1) usually leads to severe over-fitting phenomenon. To avoid that, a hyper-parameter $\boldsymbol{\alpha}$ is proposed to restrict the scope of weight values $\mathbf{W}$ [18]. Therefore, $\mathbf{W}$ is viewed as a zero mean value Gaussian probability distribution, which is represented as

$$P(\mathbf{W}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} N(w_i|0,\alpha_i^{-1}) \tag{5}$$

For classification, the weights $\mathbf{W}$ cannot be integrated out based on Bayesian principle directly. So here an approximation procedure, which is based on Laplace’s method, is utilized to get the weights value iteratively. The whole process includes three steps [18].

Step 1: Fix $\boldsymbol{\alpha}$ and calculate the “most probable” weights $\mathbf{W}_{\mathrm{MP}}$ by calculating the maximum value of the penalized logistic log likelihood function L whose mathematical expression is given as

$$L = \log\{p(\mathbf{w}|\mathbf{T},\boldsymbol{\alpha})\} = \sum_{\mathbf{n}=\mathbf{1}}^{\mathbf{N}}[\mathbf{t_n}\log\mathbf{y_n} + (\mathbf{1}-\mathbf{t_n})\log(\mathbf{1}-\mathbf{y_n})] - \frac{\mathbf{1}}{\mathbf{2}}\mathbf{W}^{\mathrm{T}}\mathbf{AW} \tag{6}$$

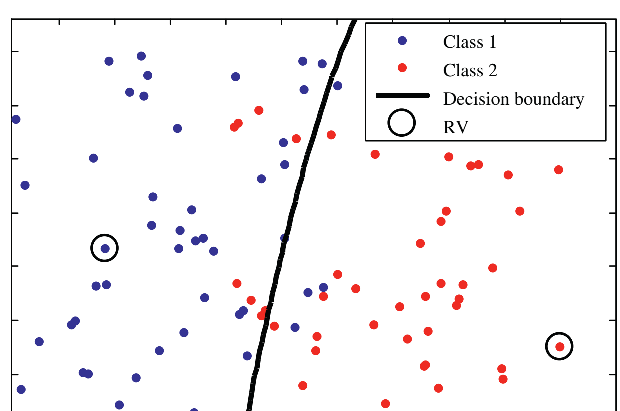

Fig. 1. RVM for binary classification.