Chin. J. Chem. Eng., **15**(2) 233—239 (2007)

Fault Diagnosis Based on Fuzzy Support Vector Machine with Parameter Tuning and Feature Selection*

MAO Yong(毛勇)[a], XIA Zheng(夏铮)[a], YIN Zheng(尹征)[a], SUN Youxian(孙优贤)[a,] and WAN Zheng(万征)[b]**

[a] State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China
[b] School of Information Technology, Jiangxi University of Finance & Economics, Nanchang 330013, China

Abstract This study describes a classification methodology based on support vector machines (SVMs), which offer superior classification performance for fault diagnosis in chemical process engineering. The method incorporates an efficient parameter tuning procedure (based on minimization of radius/margin bound for SVM's leave-one-out errors) into a multi-class classification strategy using a fuzzy decision factor, which is named fuzzy support vector machine (FSVM). The datasets generated from the Tennessee Eastman process (TEP) simulator were used to evaluate the classification performance. To decrease the negative influence of the auto-correlated and irrelevant variables, a key variable identification procedure using recursive feature elimination, based on the SVM is implemented, with time lags incorporated, before every classifier is trained, and the number of relatively important variables to every classifier is basically determined by 10-fold cross-validation. Performance comparisons are implemented among several kinds of multi-class decision machines, by which the effectiveness of the proposed approach is proved.
Keywords fuzzy support vector machine, parameter tuning, fault diagnosis, key variable identification

1 INTRODUCTION

The impact of abnormal situations on the safety and economic aspects of process operations is enormous. Fault detection and diagnosis (FDD) has been recognized as an important aspect of process operations: the ultimate goal of FDD is to realize high level autonomy in a dynamic system together with more sophisticated control strategies.

In the past, extensive research efforts were accumulated by developing model-based methods, which were based on analytical redundancy (AR). The faults were indicated with the residual signal got by the difference between the measured output signal and the output value of a nominal system model[1—3].

But model-based FDD depends heavily on the system model, and it is very difficult to design these kind of algorithms for nonlinear or uncertain systems[4]. Knowledge-based methods have no need for an analytical model, and rely on data-driven and knowledge-based techniques to estimate the system dynamics[5]. In many of these techniques, different operating conditions including normal and abnormal ones are treated as patterns. Then, a specific classifier is applied to analyze the online measurement data and to map them to a known class label for fault or normal so that the current system condition is identified[4—6]. The major advantage of these direct FDD techniques is their superior capability in identifying/classifying the faults and versatility in white-box/black-box cases.

There have been many supervised classification techniques such as, K-nearest neighbor (KNN)[7], Fisher discriminant analysis (FDA)[5], artificial neural networks (ANN)[8] and SVM[9] can be used in this field. KNN and FDA are linear classifiers, by which only linear cases can be separated. ANN and SVM are apt to classify nonlinear/linear cases. But in applications, it is very hard to decide the number of hidden nodes in ANN, and another question is how to avoid trapping into the local minimum point. New theoretic directions are needed in solving these two problems. SVM is a set of universal feed-forward network-based classification algorithm combined with kernel techniques essentially. SVM is based on the statistical learning theory and structural risk minimization (SRM) principle developed by Vapnik[10]. On the basis of these two theories, there is only one minimum in the SVM algorithm, and the structure of the SVM network is fixed.

Application of SVMs to solve process engineering problems is relatively new[5,11]. In Ref.[5], pairwise SVM is used to classify a 3-class dataset in Tennessee Eastman Process (TEP). On the basis of the key variables identified by genetic algorithm combined with Fisher discriminant analysis (GA/FDA), pairwise SVM, achieved a relatively low misclassification rate. Still many aspects may be improved in the application. First, GA/FDA is a very time-consuming algorithm for identifying the key variables in 2-class cases. To find the exact key variables, GA will be run many times. The statistical results of GAs are outputs to evaluate which variables are more informative. And the time for searching for an optimal solution is a linear function of a prefixed number of key variables[5]. If auto-correlated data is considered, time lag should be incorporated; and the number of prefixed key variables increased will multiply with the time lag. So, a more efficient key variable identification algorithm is needed. In this article, accelerated recursive feature elimination, based on a SVM is adopted as a substitute for GA/FDA. Its performance has been proved com-

Received 2006-01-06, accepted 2006-09-18.
* Supported by the Special Funds for Major State Basic Research Program of China (973 Program, No.2002CB312200), the National Natural Science Foundation of China (No.60574019, No.60474045), the Key Technologies R&D Program of Zhejiang Province (No.2005C21087), and the Academician Foundation of Zhejiang Province (No.2005A1001-13).
** To whom correspondence should be addressed. E-mail: yxsun@iipc.zju.edu.cn

parable with GA/FDA in Ref.[11], and here the number of key variables is decided basically by 10-fold cross-validation. Secondly, pairwise and one versus all structural classifiers are discussed in Ref.[5]. But when utilizing these two structural classifiers, there are also unclassifiable regions. Here, a classifier named FSVM based on pairwise SVM classifier and fuzzy decision factor[12—14] is used to make up this blind point. Thirdly, the trial and error procedure is used for tuning the SVM parameters in Ref.[5]; such an approach, apart from consuming enormous time may not really obtain the best possible performance. In this study, sequential quadratic programming (SQP) combined with radius/margin bound of SVM leave-one-out errors are used for tuning the SVM parameters.

In this article, a brief introduction to the methods used in this study is given, including accelerated recursive feature elimination by the SVM and FSVM, the description of an estimate of generalization of errors and how to tune SVM parameters by minimizing this estimate. Finally, these three methods are combined in a framework. The results obtained by considering a 5-class problem from TEP[15,16] is discussed, and conclusions are summarized.

2 METHOD

2.1 Key variable identification by accelerated recursive feature elimination based on the support vector machine

Key variable selection is pivotal pre-processing for fault diagnosis. In Ref.[14], the classification performance is poor when all variables are used. It is in favor of achieving a low error rate by utilizing only key variables. An accelerated recursive feature elimination based on an SVM, (A-SVM-RFE), is adopted here to do variable selection. Recursive feature elimination based on SVM, (SVM-RFE), was first proposed by Guyon *et al.*[17], and has been applied in several gene-expression studies[11,14] for selecting important features. Here, it is used for finding key variables in an industrial process. SVM-RFE recursive feature elimination (RFE) is a circulation procedure for eliminating features by a criterion. It consists of three steps: (1) train the classifier; (2) compute the ranking criterion; (3) remove the features with smallest ranking scores. When an SVM is used as a classifier in this procedure, this method is called SVM-RFE, and the ranking criterion is relative to the realization of SVM.

In an SVM, a classification function is given by

$$f(\boldsymbol{x})=\sum\nolimits_{i=1}^{N}\alpha_i y_i K(\boldsymbol{x}_i,\boldsymbol{x})+b \tag{1}$$

where the coefficients $\boldsymbol{\alpha}=(\alpha_i)$ and b are obtained by training over a set of examples $S=\{(\boldsymbol{x}_i,y_i)\}_{i=1,\cdots,N}$, $\boldsymbol{x}_i\in\boldsymbol{R}^n$, $y_i\in\{-1,1\}$, and $K(\cdot,\cdot)$ is the kernel function. In the linear case, the SVM expansion defines the hyper-plane $f(\boldsymbol{x})=\langle\boldsymbol{w},\boldsymbol{x}\rangle+b$, with $\boldsymbol{w}=(w_k)=\sum\nolimits_{i=1}^{N}\alpha_i y_i\boldsymbol{x}_i$. The idea is to define the importance of a feature for an SVM in terms of its contribution to a cost function $J(\boldsymbol{\alpha})$. At each step of the RFE procedure, an SVM is trained on the given dataset, J is computed, and the feature contributing less to J is discarded. In the case of a linear SVM, the variation because of the elimination of the kth feature is

$$\delta J(k)=w_k^2\,. \tag{2}$$

In the nonlinear case,

$$\delta J(k)=\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{\alpha}-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Z}(-k)\boldsymbol{\alpha} \tag{3}$$

where $Z_{ij}=y_iy_jK(\boldsymbol{x}_i,\boldsymbol{x}_j)$. $\boldsymbol{Z}(-\mathrm{k})$ is $\boldsymbol{Z}$ with the kth feature removed, some improved techniques are analyzed in such cases[18,19]. The SVM is retrained after each elimination operation for possible important improvements of a feature of medium to low importance, by removing a correlated feature. By utilizing SVM-RFE, a ranked list of features is thus obtained.

Many industrial processes have hundreds of variables, if features are eliminated one by one, the same number of SVMs as that of variables will be trained in the whole procedure of SVM-RFE. On the other hand, to identify the key variables exactly, usually, several hundreds of samples are needed. Therefore the computation complexity is very heavy. In Ref.[11], an accelerated method of SVM-RFE (A-SVM-RFE) is proposed by eliminating r, lest informative features in a step of RFE procedure, where r is determined adaptively by a set of heuristic rules based on statistical index are extracted from contributions of each feature of J. The feature list ranked by A-SVM-RFE is similar to that ranked by SVM-RFE, but A-SVM-RFE is much faster than SVM-RFE, details are in Ref.[11].

By this ranked feature list, the number of the most important variables, τ, will be determined according to the predictive accuracy of cross-validation. By the first l, $(l=1,\cdots,n)$, the feature listed, the corresponding error rate of cross-validation on the training dataset is computed and $\boldsymbol{\tau}$ is set as the number of features corresponding to the lowest error rate. A 10-fold cross-validation is accurate enough to be used here.

2.2 Fuzzy support vector machine

Fuzzy support vector machine (FSVM) is a relatively new method, first proposed by Abe and Inoue[12,13]. It is proposed to deal with unclassifiable regions when using one versus rest or pairwise classification method. FSVM is an improved pairwise classification method with SVM. A fuzzy membership function is introduced into the decision function based on pairwise classification. For the data in the classifiable regions, FSVM gives out the same classification results as the pairwise SVM method; and for the data in the unclassifiable regions, FSVM generates better classification results than the pairwise SVM. In the process of being trained, FSVM is the same as pairwise SVM[14]. If there are K classes in a training dataset, there are $K(K-1)/2$ SVMs trained in FSVM. The deci-