An Intelligent Fault Detection and Diagnosis Approaches using Parzen Density Estimation and Multi-class SVMs

Kwang-Kyu Seo*
*Dept. of Industrial Information and Systems Engineering, Sangmyung University

Parzen Density Estimation과 Multi-class SVM을 이용한 지능형 고장진단 방법

서 광 규*
*상명대학교 공과대학 산업정보시스템공학과

Abstract

본 논문은 상대적으로 새로운 기법인 Parzen Density Estimation과 Multi-class SVM을 이용한 지능형 고장 탐색과 진단 방법을 제안하고 있다. 본 연구에서는 롤링 베어링을 대상으로 고장을 탐색하고 진단하기 위한 방법을 제안하는데 Parzen Density Estimation과 Multi-class SVM은 고장 클래스를 잘 표현할 수 있다. Parzen Density Estimation은 새로운 패턴 데이터의 거절과 알려진 데이터 패턴의 밀도의 평가에 의해 새로운 패턴을 찾아낼 수 있고, Multi-class SVM 기반의 방법은 여러 클래스의 고장을 support vector로 표현하여 고장 패턴을 찾아낼 수 있다. 본 연구에서는 실제의 다중 클래스를 가지는 롤링 베어링의 고장 데이터를 사용하여 고장 패턴을 탐색하는 과정을 보여주는데, 커널함수의 적절한 파라미터의 선택에 의한 Multi-class SVM 기반의 방법이 multi-layer perceptron이나 Parzen Density Estimation 방법보다 우수함을 입증한다.
Keywords : Intelligent Fault Detection and Diagnosis, Parzen Density Estimation, Multi-class SVMs

1. Introduction

Rolling bearings play an important role as prime components of various machines due to their reliability. Although rolling bearings are reliable, the possible of unexpected faults is unavoidable. The issue of robustness and reliability is very important to guarantee the good operational condition. Therefore, condition monitoring of rolling bearings has received considerable attention in recent years. Early fault diagnosis and condition monitoring can reduce the consequential damage, breakdown maintenance and reduce the spare parts of inventories. Moreover it can increase the prolong machine life, performance, and availability of machine.

Many researchers have proposed the techniques and system for the diagnosis process. Various techniques have done by using motor current signature analysis [10], electromagnetic torque measurement [9], acoustic analysis [2], and partial discharge [7]. However, the most popular in techniques is using vibration analysis and stator current analysis because of their easy measurability, high accuracy and reliability. Recently, the application of the intelligent system for condition monitoring and fault diagnosis is used in many areas such as support vector machines (SVMs), as well as neural networks, have been extensively employed to solve classification problems [4, 5]. Detection of new patterns can also be done by estimating density of the known pattern data and rejecting new pattern data, which are below a probability threshold, in low probability areas [5].

† 교신저자: 서광규, 충남 천안시 안서동 산 98-20 상명대학교 공과대학 산업정보시스템공학과
M · P: 016-718-2682, kwangkyu@smu.ac.kr
2008년 10월 접수; 2008년 11월 수정본 접수; 2008년 11월 게재확정

The density estimation can be based on a model of the data, for instance a mixture of Gaussian distributions, or estimated by Parzen windows. This method requires large numbers of training data to make reliable probability density estimation, and has its drawbacks. In a Gaussian mixture the number of kernels has to be chosen beforehand. The assumption of Gaussian distributions can be a severe approximation. Parzen density estimation requires a width parameter that determines how smooth the resulting probability density distribution is. When large differences in density exist, the Parzen method will give poor results in low-density areas.

Another approach of new detection is to find bounded regions that contain all data, and use restricted shapes for their class boundaries like hyperspheres. Support vector data description which inspired on the support vector classifier and proposed by Tax and Duin [8], seems to give a flexible and tight data description among the boundary approaches. Since the support vector data description focuses on the boundary description and not on the complete data density, the required number of data is smaller than for, e.g., the Parzen density estimation.

In this paper, the multi-class SVMs and the Parzen density estimation are adopted to model known fault class samples, and to detect the new fault class samples that are different from those that have been modeled. Some fault class samples of the rolling bearing are taken to validate the proposed approaches.

2. Research Background

2.1 Parzen density estimation

Parzen density estimation is a non-parametric density estimation method. It does not require any assumption on the form of the probability density function, which is usually true for a real-world problem [1]. In this work, we use the standard Parzen density estimation. Here, the probabilities of the object in the training set give the value of the threshold on the probability.

$$\rho_{Parzen}(z) = \frac{1}{|X|} \sum_{x^k \in X} K(z, x^k) \qquad (1)$$

with a Gaussian Kernel $K(z, x^k)$:

$$K(z, x^k) = \frac{1}{\sqrt{(2\pi)^n s}} exp(-\frac{1}{2s}(z - x^k)^2) \qquad (2)$$

where n is the dimensionality of the data space and s is the smoothing parameter. We can use a leave-one-out method to estimate the optimal smoothing parameter. Parzen density estimation method are designed to search efficiently and robustly the largest cluster, which represents the pose information for the best match among different classes.

2.2 Support Vector Machines (SVMs)

In this section, we give a brief overview of SVMs and then review two methods for dealing with multi-class SVMs [6].

2.2.1 Support Vector Machines

Given a set of M training set $\{(x_i, y_i)\}$, where x_i is data and $y_i = 1$ or -1 is the associated label, SVMs find the optimal linear hyperplane for good generalization performance by maximizing the margin which is the distance between the hyperplane and the nearest data point of each class in which the nearest data point is called support vector.

By SVMs learning, we can construct the following decision surface:

$$\begin{aligned} f(x) &= sgn(g(x))sgn(< w, x > + b) \\ &= \sum_{i=1}^{N_s} \alpha_i y_i K(x, x_i) + b = 0 \end{aligned} \qquad (3)$$

where N_s is the number of support vectors, α_i is coefficient weight, x_i is support vector, and K is a kernel function to transform input space into feature space. The output of the $g(x)$ gives an algebraic measure of the distance from x to the optimal hyperplane [6]. This measure can be understood as a confidence value of an input x for a given SVMs.

2.2.2 Multi-class SVMs

To get multi-class classifiers, it is common to construct a set of binary classifiers. One can construct M-class classifier using the following procedure [9]:

(1) one-versus-all strategy

A set of binary classifiers, $f_1, \cdots\cdots, f_M$, are trained to separate one class from the rest. Combined function $F(x)$ is obtained by finding the maximal output among the outputs of those M classifiers. This can be denoted as follows:

$$F(x) = \arg\max f_i(x),\ where\ f_i(x) = \sum_{k=1}^{m} y_k \alpha_k^i k(x, x_k) + b^i \tag{4}$$

(2) pairwise strategy

A set of binary classifiers is constructed for each possible pair of classes. For M class, this results in $M(M+1)/2$ binary classifiers. In this case, the winner can be decided in $M-1$ comparison times by tournament method in tree structure.

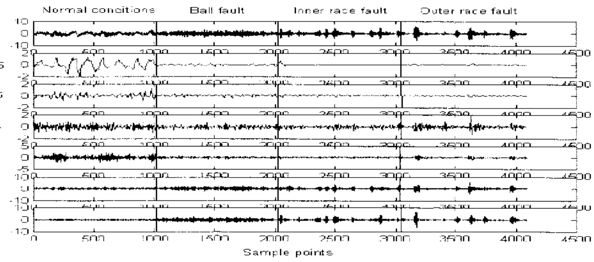
Though there are some differences in training and testing phase between one-versus-all strategy and pairwise strategy, the experiments on pattern recognition show similar classification performance. In our approach, the one-versus-all strategy is adopted.

3. Experiments using the Proposed Approaches

In this study, we consider four different operating conditions of the rolling bearing which are (i) normal conditions without faults; (ii) a ball fault; (iii) an inner race fault; (iv) an outer race fault. The ball bearing considered is installed in the drive-end of the motor; faults are introduced into it using the electro-discharge machining (EDM) method. An accelerometer is mounted on the motor housing at the drive-end of the motor to acquire the vibration signals from the bearing. The data recorder with a sampling frequency of 12 kHz is equipped with low-pass filters at the input stage for anti-aliasing.

Each set of vibration data used to the new detection experiments comes from two different loads (and rotation speeds) of the motor, i.e. 0.735kw (and 1,772rpm), 1.470kw (and 1,750rpm).

The experiments are carried out for each fault class considered a novelty, and the rests (including normal condition class) considered a target class, which are described by the support vectors.



<Fig. 1> Decomposition of vibration data under four different operating conditions

3.1 Feature Extraction of Rolling Bearings

The feature extraction procedure is used in the literature [3]. To make the signals comparable regardless of differences in magnitude, the first step is to preprocess the measured vibration data. The data of each class is normalized using its mean and standard deviation. Then, the discrete wavelet transform (DWT) is selected for the data analysis, which uses the Daubichies-2 wavelet by five levels.

<Fig. 1> shows a combination of four classes of vibration data under the wavelet decomposition, i.e. the approximation ($a5$), and five levels of details ($d1-d5$). A DWT feature vector is defined for a given vibration data as $v=[v_1, v_2, ..., v_{10}]^T$ with its element defined as:

$$v_i = \sigma_i / \sigma_{ri},\ i = 1, 2, ..., 6, \tag{7}$$

where $i = 1, 2, ..., 6$ corresponds to $d1, d2, ..., d5$, $a5$, respectively and σ_i is the standard deviation of the i th decomposition, e.g. σ_1 is the standard deviation of $d1$; σ_{ri} is the standard deviation of ith decomposition of a reference signal (in this case we have chosen a data set acquired under normal operating condition and zero load).

v_7 and v_8 are Crest factor of $d5$, $a5$ respectively, v_9 and v_{10} are Impulse factor of $d1$, $d2$ respectively. The four time domain statistical parameters have been added in the feature vector except for $v_1 - v_6$ used in [3].

In our experiments, 400 10-dimensional feature vectors in each class are randomly extracted from each set of vibration data corresponding to one of four different operating conditions of the rolling bearing. The 240 feature vectors (60%) among them are used for training, and 80 feature vectors (20%) for the validation, the others (20%) for testing in each dataset.

3.2 Fault Detection Experiments using MLP, Parzen Density Estimation and Multi-Class SVMs

In this work, the multi-layer perceptron (MLP), one of the artificial neural network techniques, is used for a comparative algorithm to evaluate the performance of the proposed Parzen density estimation and multi-class SVMs. We briefly describe the experimental results of MLP and Parzen density estimation and focus on those of multi-class SVMs.

As mentioned before, the goal of this study is to identify the fault class from other three classes data that correspond to normal conditions without faults and the other faults.

First of all, we use a three-layer perceptron trained on the target data, used for the fault detection. This network has structures of 10-16-4 for each layer unit numbers, and has a total correct recognition rate 93.3% in training dataset. But the new fault detection rate of test dataset is 72.5% (58/80). Although the threshold setup is not rigorous, and the network structures are not optimal, we can find that the MLP used for new fault detection based on thresholding output of these networks seem to give unsatisfactory results.

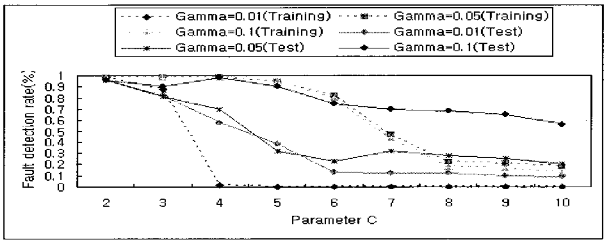
Secondly, we show the new fault detection results using Parzen density estimation method. Correspondingly used for three experiments above MLP such as outer race fault, inner race fault and ball fault. The Parzen density estimation method provides training dataset = 99.2%, test dataset = 65.0%; training dataset = 98.8%, test dataset = 80.0% and training dataset = 98.3%, test dataset 65.0% respectively on their testing dataset for s = 5 %. Obviously, it provides a slightly better new detection rate of training dataset, but a relatively lower recognition rate of test dataset in our experiments. The latter is caused because of the overtraining of the Parzen density estimation. Thus it cannot entirely meet the performance requirement.

Finally, we present the fault detection results using multi-class SVMs. The RBF kernel function is adapted in this study. The RBF kernel function has two parameters such as C and γ. We select γ = 1%, 5%, 10%, and C = 2-10 for RBF kernel parameter in all our experiments. As described before, the test results of training and test dataset provide respectively.

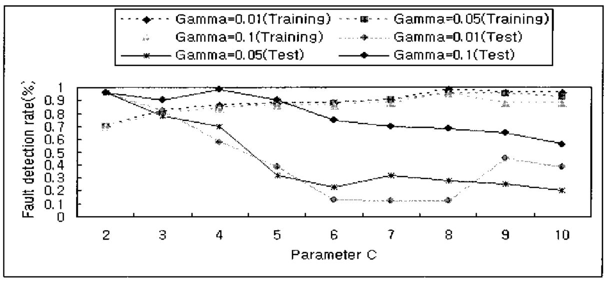
<Fig. 2> shows the new fault detection performance of ball fault by multi-class SVMs on the training and test dataset under varying γ and C. As shown in <Fig. 2>, the detection rates of both training and test dataset are very sensitive to the kernel parameter C, and decreases sharply as C increases, especially for γ = 1%, or 5%. In this case, the multi-class SVMs will be unable to find compact surrounding boundaries for data of each class when the γ is small. But, for γ = 10 % and C = 2-3 the multi-class SVMs can also give good results with the new detection rate in the range of 88.3%-98.3% of training dataset and in the range of 81.3 %-96.3 % of test dataset.

For the second experiment based on multi-class SVMs, the inner race fault is assigned to be a new fault class. As shown in <Fig. 3>, in contrast to the test success rate, the new detection rate of training dataset is very sensitive to the kernel parameter C, and decreases sharply as C increases, especially for γ = 1%, or 5%. In this case, the experiment results and trends are similar to those of the ball fault case.

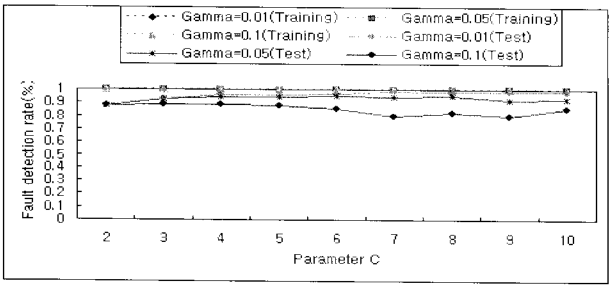
For the last experiment based on multi-class SVMs, the outer race fault is assigned to be a new fault class. <Fig. 4> shows the new fault detection performance of outer race fault by multi-class SVMs on the training and test dataset under varying γ and C.



<Fig. 2> The ball fault detection performance based on multi-class SVMs



<Fig. 3> The inner race fault detection performance based on multi-class SVMs



<Fig. 4> The outer race fault detection performance based on multi-class SVMs

As shown in <Fig. 4>, all new fault examples in the training dataset can be detected successfully, which is irrelevant to parameters γ and C in our chosen ranges. The fault detection rate for new examples of the test dataset also ranges from 89.2 % to 98.3 % for γ = 1%, C = 2-10, and 88.3 % to 95.0 % for γ = 5%, C = 2-10. The trend of results for γ = 1% is better than those of the others in last experiments.

Obviously, the multi-class SVMs gives better detection results for the new fault class than those of MLP and Parzen density estimation described previously.

4. Conclusion

In this paper, we applied the relatively new methods such as Parzen density estimation and multi-class SVMs besides MLP for intelligent fault detection and diagnosis of the rolling bearing. They were successfully applied for fault detection in training dataset, but the experiment results of test dataset is different. Especially, the experiment results of multi-class SVMs based detection and classification are satisfactory in case of choosing the optimal values of kernel parameters and that are very important in SVMs model selection. Therefore the proposed method, multi-class SVMs served to exemplify that kernel-based learning algorithms are very competitive on a variety problems with different characteristics and can be employed as an efficient method for intelligent machine fault detection, classification and diagnosis.

5. Reference

[1] Duda, R.O., Hart, P.E. and Stork, D.G., Pattern Classification, 2nd Ed., Wiley-Interscience, (2001)

[2] Lee, Y. S. et al., "An acoustic diagnostic technique for use with electric machine insulation", IEEE Transactions on Dielectrics and Electrical Insulation, 9(1994): 1186-1193

[3] Lou, X., Loparo, K. A., "Bearing fault diagnosis based on wavelet transform and fuzzy inference", Mechanical Systems and Signal Processing, 18(6)(2004): 373-390

[4] Markou, M., Singh, S., "Novelty detection: a review-part 1: statistical approaches. Signal Processing", 83(12) (2003): 2481-2497

[5] Markou, M., Singh, S, "Novelty detection: a review-part 2: neural network based approaches", Signal Processing, 83(12)(2003): 2499-2521

[6] Scholkopf, B., Smola, A.J., Learning with Kernels, MIT Press, Cambridge Massachusetts London, (2002)

[7] Stone, G. C., Sedding, H. G., & Costello, M. J., "Application of partial discharge testing to motor and generator stator winding maintenance", IEEE Transactions on Industry Application, 32(1996): 459-464

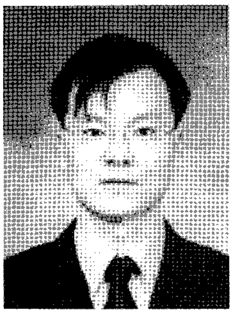
[8] Tax, M. J., Duin, P. W., "Support vector data description", Machine Learning, 54(2004): 45-66

[9] Thollon, F., Jammal, A., & Grellet, G., "Asynchronous motor cage fault detection through electromagnetic torque measurement", European Transactions on Electrical Power, 3(1993): 375-378

[10] Thomson, W. T., & Fenger, M, "Current signature analysis to detect induction motor faults", IEEE Transactions on Industry Applications, 7(2001); 26-34

저 자 소 개

서 광 규



Kwang-Kyu Seo is a professor of Industrial Information and Systems Eng. at Sangmyung University. Dr. Seo received a Ph.D. degree in Industrial engineering from Korea University and worked as a research scientist of division of system technology at Korea Institute of Science and Technology (KIST). He is interested in production/ operation management, Information system, Data mining and CRM, e-business and so on.

주소: 충남 천안시 안서동 산 98-20 상명대학교 공과대학 산업정보시스템공학과