An Intelligent Fault Diagnosis Method Based on Multiscale Entropy and SVMs

Long Zhang[1], Guoliang Xiong[2], Hesheng Liu[3], Huijun Zou[1], and Weizhong Guo[1]

[1] School of Mechanical Eng., Shanghai Jiaotong University, Shanghai 200240, China
[2] School of Mechatronic Eng., East China Jiaotong University, Nanchang 330013, China
[3] Department of physics, Shangrao Normal College, Shangrao 334001, China
longzh@126.com

Abstract. Sample entropy (SampEn) has been applied in many literatures as a statistical feature to describe the regularity of a time series. However, as components of mechanical system usually interact and couple with each other, SampEn may cause inaccurate or incomplete description of a mechanical vibration signal due to the fact that SampEn is calculated at only one single scale. In this paper, a new method, named multiscale entropy (MSE), taking into account multiple time scales, was introduced for feature extraction from fault vibration signal. MSE in tandem with support vector machines (SVMs) constitutes the proposed intelligent fault diagnosis method. Details on the parameter selection of SVMs were discussed. In addition, performances between SVMs and artificial neural networks (ANNs) were compared. Experiment results verified the proposed model.

Keywords: Fault diagnosis, Sample entropy, Multiscale entropy, SVMs.

1 Introduction

On line machine condition monitoring and fault diagnosis has been increasingly attracting attention from the research and engineering community worldwide over the past decades [1]. Generally, a simple condition monitoring system is approached from a pattern classification perspective. It can be decomposed into three general steps: data acquisition, feature extraction, and condition classification, among which the latter two are of significant importance.

Due to instantaneous variations in friction, damping, stiffness or loading conditions, mechanical systems are often characterized by non-linear behaviors that in turn make the vibration signals complex and non-linear. As such, commonly used signal processing techniques including time and frequency domain techniques, as well as advanced signal processing techniques, such as wavelet transform and time-frequency representation, may all exhibit limitations. Therefore, techniques for non-linear dynamic parameter estimation provide a good alternative to extracting defect-related features hidden in the complex and non-linear vibration signals [1, 2]. Hitherto, a number of non-linear parameter identification techniques have been investigated and introduced to fault diagnosis, among which correlation dimension is a typical one [3, 4, and 5].

W. Yu, H. He, and N. Zhang (Eds.): ISNN 2009, Part III, LNCS 5553, pp. 724–732, 2009.

Reliable estimation of correlation dimension, however, usually requires very long data set that is difficult or even impossible to be achieved especially in on-line, real-time monitoring and diagnosis. A brief review on non-linear dynamic parameters used for feature extraction and fault diagnosis can be found in literature [1], and in the same literature appropriate entropy (ApEn) was introduced and selected as a tool for rolling bearing health monitoring. Although ApEn has found its ways in fields of physiological signal and machine vibration signal processing [1,2,6], however due to the bias within its estimation, ApEn is heavily dependent on the data length and its estimated value is uniformly lower than that expected for short records, and lacks relative consistency as well [7]. In order to overcome the shortcomings of ApEn, Richman and Moorman [7] proposed a new kind of entropy, named sample entropy (SampEn) which seems much more promising and has attracted a lot of attention [7,8].

In a relatively recent paper [9], a new entropy based measure of complexity, which is the multiple scale entropy (MSE), was introduced. The authors applied their new complexity measure (i.e. MSE) to distinguish between young healthy hearts and congestive heart failure. Moreover, the MSE was able to distinguish atrial fibrillation from healthy hearts [9, 10, and 11]. The key to the MSE method lies in a multiscale approach [12]. Consider a machine composed of gears, bearings, shafts and other mechanical components [13]. Even a modest amount of machine complexity will result in measured vibration signals that contain multiple intrinsic oscillatory modes due to the interaction of the mechanical components, that implies non-linear dynamic parameters applied on single scale (such as ApEn and SampEn of original time series) may be insufficient for characterizing machine vibration signals. For this reason, the multiscale method was introduced and tried in the present study in the hope of improving performances of traditional non-linear dynamic methods on single scale within the context of machine fault diagnosis. To the best of the authors' knowledge, the MSE has not been applied in the field of fault diagnosis so far. Its advanced properties attract us for a trial of its use.

After extracting MSE acting as feature vectors, one needs a classifier to fulfill automated fault recognition. A number of intelligent classification algorithms, such as artificial neural networks (ANNs) and support vector machines (SVMs) have been successfully applied to automated machine fault diagnosis [14]. The main advantages of SVMs lie in the fact that it can perform better in the processing of small-sample sized learning problems and has better generalization due to the replacement of Empirical Risk Minimization used in ANNs by Structural Risk Minimization. Due to these merits, SVMs has become a new research hotspot in recent years and have been applied successfully in many domains. Hence, SVMs was selected in the present study as a fault classifier.

2 Theoretical Background

2.1 Multiscale Entropy (MSE)

MSE was computed according to the procedure published by Costa et al [9, 10, and 11]. Given a one-dimensional discrete time series, $\{x_1, \cdots, x_i, \cdots x_N\}$, one can constructed

consecutive coarse-grained time series $\{y^{(\tau)}\}$ determined by the scale factor τ, according to the equation

$$y_j^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} x_i$$

where τ represents the scale factor and $1 \le j \le N/\tau$. In other words, coarse-grained time series for scale τ are obtained by taking arithmetic mean of τ neighboring original values without overlapping (Fig.1). The length of each coarse-gained time series is N/τ. For scale 1, the coarse-grained time series is simply the original time series. Then SampEn or ApEn is computed for the coarse-gained time series at each scale and plotted as a function of the scale factor.

SampEn is a refinement of traditionally used regularity measure ApEn statistics. Details on the SampEn algorithm can be found in many literatures [15]. Briefly, SampEn quantifies the regularity of time series. It reflects the condition probability that two sequences of m consecutive data points, which are similar to each other (within give tolerance r), will remain similar when one consecutive point is included [15]. The SampEn algorithm underlying the MSE computation requires setting two parameters: the tolerance level r and the pattern length m. According to previous studies, it has been chosen that $r = 0.15\times$ standard deviation of the time series to avoid distortion of SampEn values by changes in signal magnitude and $m = 2$.

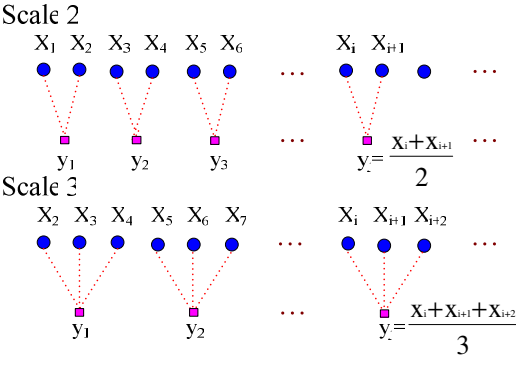

Fig. 1. The scheme illustrating the coarse-graining of an original time series for scales $\tau = 2$ and $\tau = 3$

2.2 Support Vector Machines (SVMs)

SVM is a classification method derived from Statistical Learning Theory (SLT) by Vapnik and Chervonenkis [16]. Its basic idea is to map the original data to a higher dimensional feature space and find the optimal hyperplane in the space that maximizes the margin between the classes, as illustrated in Fig.2. The essential difference between SVMs and ANNs lies in their requirements imposed on the hyperplane. In the case of SVMs, it is desired to find the hyperplane with maximal margin and minimal class error ratio on training data, whereas in ANNs, only the latter is necessary. According to the

SLT, for a trained classifier to predict unseen samples, the actual risk consists of two parts, i.e. empirical risk (R_{emp}) and confidence interval ϕ .

$$R \le R_{emp} + \phi$$

As such, in order to minimize actual risk, the only satisfaction with minimal class error of training data, i.e. minimal R_{emp} , is not enough. In ANNs, large number of training data is required to minimize ϕ . Training data, however, is usually limited, especially for fault samples of machinery. Therefore, the requirement on maximal margin is taken into account in SVMs to account for ϕ . According to the SLT, maximal margin will result in minimal ϕ . Hence, SVMs doesn't rely much on the amount of training data and possesses advantages over ANNs with respect to generalization performance. Maximal margin in conjunction with minimal error of training data is referred to as minimal structural risk.

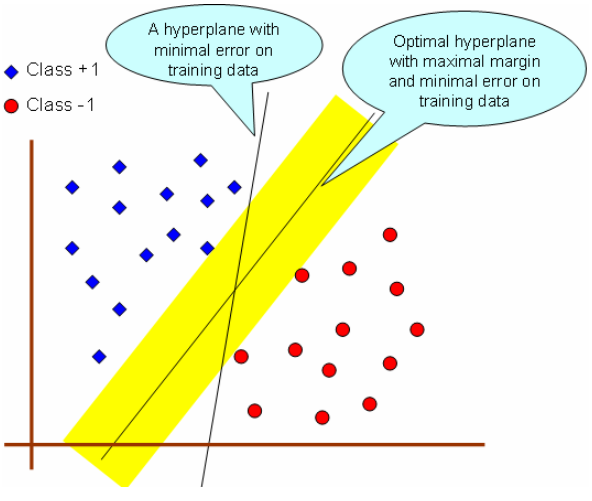

Fig. 2. The line with a yellow background illustrating an optimal hyperplane

3 Experimental Analysis

3.1 Experimental Setup

In order to validate the proposed fault diagnosis method, experimental analyses on rolling element bearings were conducted. All the bearing data and related system analyzed in this paper belong to Case Western Reverse Lab [17].

The test stand, shown in Fig.3, consists of a 2 hp, three-phase induction motor (left), a torque sensor (middle) and a dynamometer (right) connected by a self-aligning coupling (middle). The dynamometer is controlled so that desired torque load levels can be achieved. The test bearings support the motor shaft at both the drive end and fan end. Single point faults were introduced to the test bearings using electro-discharge

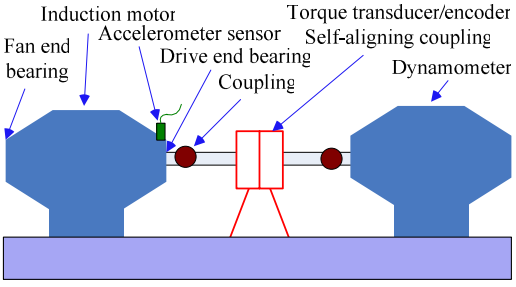

Fig. 3. Bearing fault test stand

machining with fault diameters of 7 mil, (1 mil=0.001 inches). Vibration data was collected using accelerometers, which were attached to the housing with magnetic bases [17, 18].

Vibration signals of drive end bearing under 0 hp load collected from good, outer race fault, inner race fault and ball fault condition were analyzed. For each condition, there are 25 samples and each sample contains 4096 data points. The sampling frequency is 12,000Hz, and the approximate motor speed is 1797 rpm. Hence, motor rotates about 11 revolutions over the time interval of 4096 data points. One sample of the four conditions was shown in Fig.4.

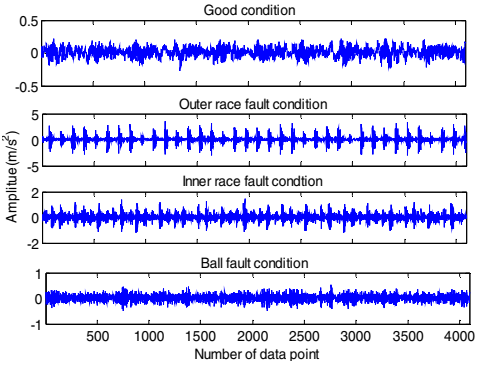

Fig. 4. Vibration signals in time domain of four different bearing conditions

3.2 Calculation of MSE

MSE, in essence, is to calculate sample entropy (or other type of entropy like ApEn) over a set of scales. For this purpose, prior to the calculation of MSE, there are three parameters to be defined, i.e. the tolerance level r, the pattern length m and the maximal scale factor. Values of r and m have been determined in section 2.1 according to previous studies. Maximal scale factor was selected as 50 by experiments. Figure 5 shows the MSE of the samples depicted in figure 4, from which the four conditions can't be separated linearly, so a nonlinear classifier is necessary.

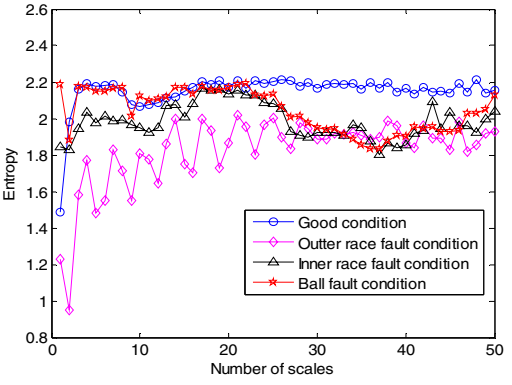

Fig. 5. MSE calculated over 50 time scales for the signals shown in Fig.4

3.3 Classification with SVMs

SVMs classify data in the form like a linear function. When linearly inseparable data are concerned, SVMs make use of Kernel trick to map the original data to a higher dimensional space where the data may be linearly separable. There are various kernel functions used in SVMs, such as linear, polynomial, radial basis function (RBF) and sigmoid kernel. Since RBF kernel has less hyperparameters and less numerical difficulties, it is a reasonable first choice [19].

Basic SVMs is developed for binary classification. In practice, however, there are many scenarios involving multi-class classification. To this end, a lot of methods have been developed such as one-against-rest and one-against-one. For the case of one-against-rest, there are possibly some data that can't be classified into any classes or will be classified into many classes. To avoid this deficiency, one-against-one paradigm was adopted.

After the determination of the type of kernel function and the multi-class method, there are still two parameters to be determined, i.e. penalty parameter C and RBF width parameter γ. This can be solved by cross-validation and grid-search [19]. As stated above, there are 25 samples for each bearing condition respectively. Among them are randomly selected 10 samples as training data, remainder 15 samples as testing data. Because of less training data, a two-fold cross-validation was implemented to determine the C and γ. For a given value of C and γ, the 10 training samples were split into two subsets each containing 5 samples. Then, the second subset was predicted by the SVMs trained with the first subset, and vice versa. The sum of the two prediction accuracy was used as a performance metric to evaluate the given C and γ. The values of C and γ within a prescribed range ($C \in 2^{-5:1:15}$, $\gamma \in 2^{-15:1:3}$ in this paper) achieving the highest prediction accuracy will be selected for future applications [19]. The validation prediction accuracies (testing accuracy) of the total $21 \times 19 = 399$ pairs of C and γ are shown in Fig. 6; where there are 373 cases reaching a testing accuracy of 100%. So many optimal values make it confused for the selection. In practice, if there are no other more training data available, any pair of the optimal C and γ is a possible candidate. In order to examine the performances of

all the possible 373 candidates, their classification rates on testing data are shown in Fig. 7. Among all the cases, there are 35.92% getting a rate of 100%, and 45.83% getting a rate of 98.33%, 11.26% getting a rate of 96.67, as well as 4.29% getting the lowest rate of 93.33%. As such, all of the 373 candidates produced very promising results in the classification of testing data. The details on the classification regarding the testing data of three couples of C and γ representing three kinds of classification accuracy are depicted in Table 1. For all the three cases, there are no samples of fault bearing misclassified into good condition, which implies a small risk of the proposed fault diagnosis method.

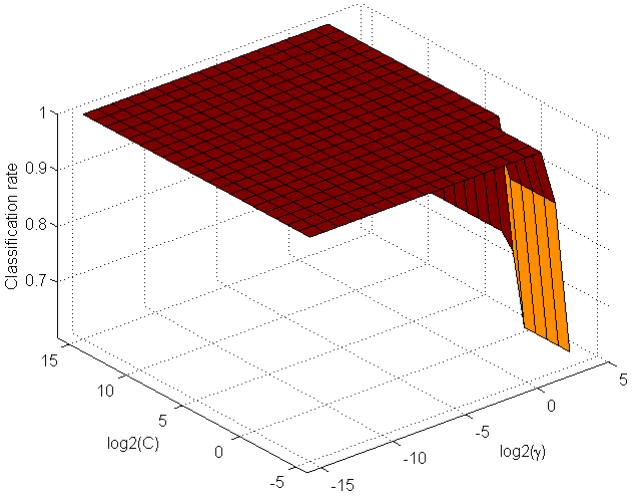

Fig. 6. Classification rate of various pairs of C and γ in cross-validation

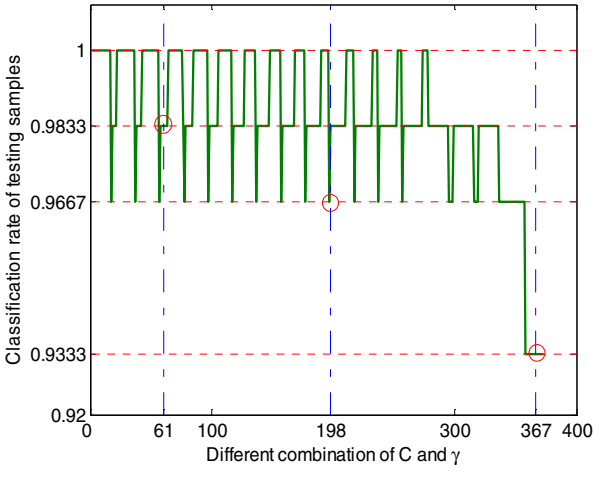

Fig. 7. Classification rate on testing data with various pairs of C and γ from the optimal values determined by cross-validation

A three-layer BP neural network with 80 and 4 nodes in middle and output layer respectively was also investigated. As showed in Table 2, a total of 11 samples were misclassified with a classification rate of 81.67%, among which 6 fault samples were treated as good condition. This will lead to a rather larger risk than SVMs. Proper increase of the node number of middle layer will give a possible raise to the accuracy. Nevertheless, due to the high dimension of feature characteristics (i.e. MSE over 50 scales), too many nodes in middle layer will render the training and testing speed very slow, which doesn't suit on-line and real-time application.

Table 1. Confusing Matrix of SVMs with different values of C and γ indicated by lines vertical to the horizontal axis in Fig.7 at points 61, 198 and 367

	$C=2^{13},\gamma=2^{-13}$				$C=2^{3},\gamma=2^{-6}$				$C=2^{9},\gamma=2^{2}$			
Condition	A	B	C	D	A	B	C	D	A	B	C	D
A	14	0	1	0	13	0	2	0	12	1	2	0
B	0	15	0	0	0	15	0	0	0	15	0	0
C	0	0	15	0	0	0	15	0	0	0	15	0
D	0	0	0	15	0	0	0	15	0	0	1	14
Accuracy	98.33%				96.67%				93.33%			

A--Good condition, B--Outer race fault, C--Inner race fault, D--Ball fault.

Table 2. Confusing Matrix of a three-layer BP network

Condition	A	B	C	D
A	15	0	0	0
B	0	15	0	0
C	2	2	11	0
D	4	0	3	8
Accuracy	81.67%			

A--Good condition, B--Outer race fault, C--Inner race fault, D--Ball fault.

4 Conclusions

Experiments verified the effectiveness of the combination of multiscale entropy (MSE) and SVM. MSE can extract the nonlinear information hidden in vibration signals over multiple scales. SVMs are superior in terms of good generalization performance as well as less dependence on the amount of training data. In comparison with SVMs, the accuracy rate of BP network is slightly lower. The rather higher accuracy of both SVMs and BP in turn demonstrated the effectiveness of the features extracted by MSE. How to determine the maximal scale factor to which MSE will be calculated is an open question. In this work, it's selected as 50 by trials.

Acknowledgments. The work was supported by the Natural Science Foundation of Jiangxi Province under the grant No. 0450017.

References

1. Yan, R.Q., Gao, R.X.: Approximate Entropy as a Diagnostic Tool for Machine Health Monitoring. Mechanical Systems and Signal Processing 21, 824–839 (2007)
2. Yan, R.Q., Gao, R.X.: Machine Health Diagnosis Based on Approximate Entropy. In: ICMT 2004 Instrumentation and Measurement Technology Conference, pp. 2054–2059. IEEE Press, Italy (2004)
3. Jiang, J.D., Chen, J.: The Application of Correlation Dimension in Gearbox Condition Monitoring. Journal of Sound and Vibration 224, 529–541 (2004)
4. Logan, D., Mathew, J.: Using the Correlation Dimension for Vibration Fault Diagnosis of Rolling Element Bearings-I. Basic Concepts. Mechanical Systems and Signal Processing 10(3), 241–250 (1996)
5. Logan, D., Mathew, J.: Using the Correlation Dimension for Vibration Fault Diagnosis of Rolling Element Bearings-II. Selection of Experimental Parameters. Mechanical Systems and Signal Processing 10(3), 251–260 (1996)
6. Xu, Y.G., Li, L.J., He, Z.J.: Approximate Entropy and Its Applications in Mechanical Fault Diagnosis. Information and Control 31(6), 547–551 (2002)
7. Richman, J.S., Moorman, J.R.: Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy. Am. J. Physiol. H. 278, 2039–2049 (2000)
8. Haitham, M.A., Alan, V.S.: Use of Sample Entropy Approach to Study Heart Rate Variability in Obstructive Sleep Apnea Syndrome. IEEE Trans. Bio. Eng. 50, 1900–1904 (2007)
9. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale Entropy Analysis of Complex Physiologic Time Series. Phys. Res. Lett. 89(6), 68–102 (2002)
10. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale Entropy to Distiguish Physiologic and Synthetic RR Time Series. Computers in Cardiology 29, 137–140 (2002)
11. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale Entropy Analysis of Biological Signals. Phys. Rev. E. 71, 21906 (2005)
12. Ranjit, A.T., Georg, A.G.: On Multiscale Entropy Analysis for Physiological Data. Physica A 366, 323–332 (2006)
13. Fan, X.F., Zuo, M.J.: Machine Fault Feature Extraction Based on Intrinsic Mode Functions. Meas. Sci. Technol. 19, 245105(12pp) (2008)
14. Lu, S., Yu, F.J., Liu, J.: Bearing Fault Diagnosis Based on K-L Transform and Support Vector Machine. In: 3rd IEEE International Conference on Natural Computation, pp. 522–527. IEEE Press, New York (2007)
15. Trunkvalterova, Z., Javorka, M., Tonhajzerova, I., Javorkova, J., Lazarova, Z., Javorka, K., Baumert, M.: Reduced Short-Term Complexity of Heart Rate and Blood Resssure Dynamics in Patients with Diabetes Mellitus Type 1: Multiscale Entropy Analysis. Physiol. Meas. 29, 817–828 (2008)
16. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, Berlin (1995)
17. Case Western Reserve University Bearing Data Center Website, `http://www.eecs.case.edu/laboratory/bearing/welcome_overview.htm`
18. Lou, X.S., Loparo, K.A.: Bearing Fault Diagnosis Based on Wavelet Transform and Fuzzy Inference. Mechanical Systems and Signal Processing 18, 1077–1095 (2004)
19. Hsu, C.W., Chang, C.C., Lin, C.J.: A Practical Guide to Support Vector Classification (2008), `http://www.csie.ntu.edu.tw/~cjlin`