Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

# A comparative study on ApEn, SampEn and their fuzzy counterparts in a multiscale framework for feature extraction*

Guo-liang XIONG†1, Long ZHANG†2, He-sheng LIU3, Hui-jun ZOU2, Wei-zhong GUO2

(1School of Mechatronic Engineering, East China Jiaotong University, Nanchang 330013, China)
(2School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)
(3Department of Physics, Shangrao Normal University, Shangrao 334001, China)
†E-mail: longzh@126.com; lgxcxx@ecjtu.jx.cn

Received June 19, 2009; Revision accepted Jan. 6, 2010; Crosschecked Mar. 10, 2010

**Abstract:** Feature extraction from vibration signals has been investigated extensively over the past decades as a key issue in machine condition monitoring and fault diagnosis. Most existing methods, however, assume a linear model of the underlying dynamics. In this study, the feasibility of devoting nonlinear dynamic parameters to characterizing bearing vibrations is studied. Firstly, fuzzy sample entropy (FSampEn) is formulated by defining a fuzzy membership function with clear physical meaning. Secondly, inspired by the multiscale sample entropy (multiscale SampEn) which is originally proposed to quantify the complexity of physiological time series, we placed approximate entropy (ApEn), fuzzy approximate entropy (FApEn) and the proposed FSampEn into the same multiscale framework. This led to the developments of multiscale ApEn, multiscale FApEn and multiscale FSampEn. Finally, all four multiscale entropies along with their single-scale counterparts were employed to extract discriminating features from bearing vibration signals, and their classification performance was evaluated using support vector machines (SVMs). Experimental results demonstrated that all four multiscale entropies outperformed single-scale ones, whilst multiscale FSampEn was superior to other multiscale methods, especially when analyzed signals were contaminated by heavy noise. Comparisons with statistical features in time domain also support the use of multiscale FSampEn.

**Key words:** Fault diagnosis, Bearing, Multiscale entropy, Feature extraction, Support vector machines (SVMs)
**doi:**10.1631/jzus.A0900360 **Document code:** A **CLC number:** TH17; TP18

## 1 Introduction

Rolling element bearings are an important, damageable element in rotating machinery. Bearing defects may result in huge economic loss and even catastrophic accidents if not detected in a timely fashion. Thus, the importance of bearing fault diagnosis has been widely appreciated in recent years. Fault diagnosis can be considered as a pattern recognition problem, in which data acquisition and feature extraction are of paramount importance. Vibration signals have been popularly adopted for diagnostic purposes due to their convenient access and rich information. Normally, vibration acceleration signals are collected by accelerometers mounted as close to the monitored elements as possible. The vibration signals are redundant and usually contaminated by heavy noise; thus, proper signal processing techniques are necessary for highlighting the characteristic features hidden in the data. To date, a wide variety of feature extraction techniques have been developed including time and frequency domain techniques, as well as advanced signal processing techniques such as time-frequency representation and wavelet decomposition. Due to factors such as clearances in bearings and nonlinear stiffness of rolling elements, however, vibrations of a bearing, especially under faulty conditions, are essentially governed by a nonlinear dynamic

* Project supported by the National Natural Science Foundation of China (Nos. 50875161 and 50821003), and the Natural Science Foundation of Jiangxi Province, China (No. 0450017)

model. As a result, these normally utilized feature extraction methods which are aimed mainly at linear vibration signals may all exhibit limitations. This study is therefore directed toward applying nonlinear dynamic theory to feature extraction for bearing fault diagnosis.

With the advancement of nonlinear dynamics, various nonlinear models have been constructed in order to examine nonlinear behaviors of bearings and many efforts have been undertaken to develop nonlinear dynamic parameters to characterize bearing vibration signals as well. Correlation dimension (CoD) is among these efforts for feature extraction (Logan and Mathew, 1996; Jiang *et al.*, 1999; Wang *et al.*, 2001). Though CoD has been gaining popularity in fault diagnosis, it has at least the following three deficiencies: (1) reliable estimation of CoD requires a large number of data points; (2) the typical estimation method (known as the Grassberger and Procaccia (G-P) algorithm) assumes signals to be stationary and noise-free; and (3) utmost attentions must be paid to the selection of corresponding parameters, otherwise under- or over-estimations are possible (Jiang *et al.*, 1999). Thus, the computation is complex (Yan and Gao, 2007).

Because of these major drawbacks of CoD, it is necessary to explore other nonlinear methods. Two of these alternatives are approximate entropy (ApEn) and sample entropy (SampEn), which have found many applications in processing physiological signals (Pincus, 1991; Richman and Moorman, 2000; Chen *et al.*, 2006; Kaffashi *et al.*, 2008; Liao *et al.*, 2008) as well as mechanical vibration signals ( Xu *et al.*, 2002; Yan and Gao, 2007). ApEn and SampEn quantify the regularity of time series by evaluating the conditional probability that two vectors of length $m$ similar within a tolerance $r$ will become dissimilar when a successive point is added to them. Therefore, a larger entropy indicates a greater irregularity of time series. By eliminating self-matches, the drawbacks associated with ApEn are removed from SampEn, these drawbacks including the bias in its estimation and poor relative consistence. In both ApEn and SampEn, however, the criteria $r$ used to determine the vectors' similarity is discontinuous, which may lead SampEn to be undefined for short records, and make both of them sensitive to noise and change in the parameters $m$ and $r$. In view of these deficiencies, Chen *et al.* (2009) proposed a fuzzy entropy, replacing this discontinuous criteria with a continuous Gaussian function of which the function shape is controlled by two parameters. Nevertheless, the physical meanings of these two parameters are not clear and the proposition is only integrated into ApEn. We will refer to Chen *et al.* (2009)'s fuzzy entropy as fuzzy ApEn (FApEn) so as to distinguish it from other types of fuzzy entropy. In this work, by firstly defining a new fuzzy membership function being of clear physical meaning, the idea of Chen *et al.* (2009) is carried over to SampEn resulting in a new entropy, named fuzzy SampEn (FSampEn).

Above-mentioned entropies, no matter ApEn, SampEn or their fuzzy counterparts (FApEn and FSampEn), all calculate entropies at only one time scale. This may lead to an inaccurate description of machine vibration signals, since the vibration signals of machinery are usually emitted from many adjacent sources (e.g., bearings, gears, rotors and macro-structural oscillation). These different sources will manifest themselves on different time scales. For this reason, it is preferable to extend the single-scale entropies to a multiscale framework. Similar to the vibration signals, physiological and physical signals also exhibit multiscale property. Based on this characteristic, Costa *et al.* (2002) proposed multiscale sample entropy (multiscale SampEn), which calculates SampEns across multiple time scales. When applied to a range of physiological and physical signals, this multiscale approach has achieved better and more meaningful results than single-scale ones (Costa *et al.*, 2002; 2005; Escudero *et al.*, 2006; Li and Zhang, 2008; Wang and Ma, 2008). Taking into account both the similar multiscale property and the effectiveness of multiscale SampEn showed by Costa *et al.* (2002), we propose to further extend ApEn, FApEn and the proposed FSampEn into the same multiscale framework for vibration signal processing, thus resulting in developments of novel concepts: multiscale ApEn, multiscale FApEn and multiscale FSampEn, respectively. The relationship between these single-scale entropies and their multiscale counterparts is shown in Fig. 1.

We illustrate the feature extraction ability of the entropies in Fig. 1 by processing vibration signals from twelve kinds of bearing conditions. Their classification performances are evaluated using support vector machines (SVMs). SVM is a classifier performing better on small sample-sized problems and has better generalization than the widely used artificial neural networks (ANNs) (Widodo and Yang, 2008).

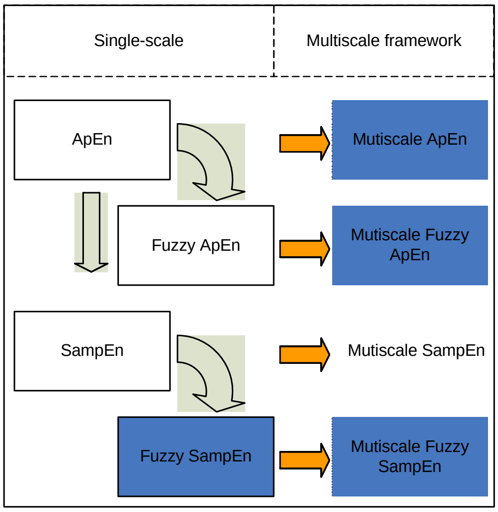


**Fig. 1 Relationship between various entropies**

## 2 Principle and methods

In this section, the theoretical background of ApEn, SampEn, FApEn as well as the multiscale framework will be briefly reviewed. In addition, the proposed FSampEn, multiscale ApEn, multiscale FApEn and multiscale FSampEn will be detailed.

### 2.1 ApEn and SampEn

The definitions and calculation steps of ApEn and SampEn are depicted in Figs. 2 and 3, respectively. There are three main differences between ApEn and SampEn: (1) self-matches are eliminated from SampEn, which is reflected by $j\neq i$ in Fig. 3; (2) in order for $\boldsymbol{X}$ to has equal number of pattern for both the pattern lengths $m$ and $m$+1, $\boldsymbol{X}$ has $N-m$ rows in SampEn instead of $N-m+1$ in ApEn in the case of pattern length $m$; and (3) to avoid ln0 caused by removing self-matches, addition operation is performed prior to the natural logarithm operator in SampEn (reflected in $\phi_m$ and the final output). These modifications are intended for SampEn to remove the bias in the estimation of ApEn.

Roughly speaking, in ApEn and SampEn, $B(i)$ presents the probability that the $i$th pattern finds a similar pattern in the pattern space $\boldsymbol{X}$. Hence, $\phi_m$ indicates the probability that for any pair of patterns of length $m$ to be similar with each other. As a result

Input: Data $x$ of length $N$, patter length $m$, tolerance $r$.
Procedure:

$$X=\begin{bmatrix} x(1) & x(2) & \cdots & x(m) \\ x(2) & x(3) & \cdots & x(m+1) \\ \vdots & \vdots & & \vdots \\ x(N-m+1) & x(N-m+2) & & x(N) \end{bmatrix};$$

For $i=1$ to $N-m+1$
For $j=1$ to $N-m+1$

$$d(i,j)=\max\left(|X(i,1)-X(j,1)|,|X(i,2)-X(j,2)|\right.$$
$$\left.\cdots,|X(i,i+m-1)-X(j,j+m-1)|\right);$$
$$D(i,j)=H\left(d(i,j)\right);$$

End

$$B(i)=\frac{1}{N-m+1}\sum_{j=1}^{N-m+1}D(i,j);$$

End

$$\phi_m=\frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\ln B(i);$$

Similar computations are performed for pattern length $m+1$ resulting in $\phi_{m+1}$;
Output: $\mathrm{ApEn}(x,m,r)=\phi_m-\phi_{m+1}$.

**Fig. 2 Pseudocode of ApEn**

Input: Data $x$ of length $N$, patter length $m$, tolerance $r$.
Procedure:

$$X=\begin{bmatrix} x(1) & x(2) & \cdots & x(m) \\ x(2) & x(3) & \cdots & x(m+1) \\ \vdots & \vdots & & \vdots \\ x(N-m) & x(N-m+1) & & x(N-1) \end{bmatrix};$$

For $i=1$ to $N-m$
For $j=1$ to $N-m\,(j\neq i)$

$$d(i,j)=\max\left(|X(i,1)-X(j,1)|,|X(i,2)-X(j,2)|\right.$$
$$\left.\cdots,|X(i,i+m-1)-X(j,j+m-1)|\right);$$
$$D(i,j)=H\left(d(i,j)\right);$$

End

$$B(i)=\frac{1}{N-m-1}\sum_{j=1}^{N-m+1}D(i,j);$$

End

$$\phi_m=\frac{1}{N-m}\sum_{i=1}^{N-m}B(i);$$

Similar computations are performed for pattern length $m+1$ resulting in $\phi_{m+1}$, with

$$X=\begin{bmatrix} x(1) & x(2) & \cdots & x(m+1) \\ x(2) & x(3) & \cdots & x(m+2) \\ \vdots & \vdots & & \vdots \\ x(N-m) & x(N-m+1) & & x(N) \end{bmatrix};$$

Output: $\mathrm{SampEn}(x,m,r)=\ln\phi_m-\ln\phi_{m+1}$.

**Fig. 3 Pseudocode of SampEn**

of $\phi_m-\phi_{m+1}$ or $\ln\phi_m-\ln\phi_{m+1}$, the output entropy, ApEn or SampEn, represents the rate of creation of new information of the underlying system.

In both ApEn and SampEn, the function $H$ which determines vectors' similarity is a Heaviside step function, as formulated by Eq. (1) and illustrated in Fig. 4. Such a step function is discontinuous, which makes a binary decision on the similarity between two vectors according to their distance. Such a distance is defined as the maximum distance between the respective scalar elements of the vectors.

$$H=\begin{cases}1, & d\le r,\\ 0, & d>r,\end{cases} \tag{1}$$

where $H$=1 indicates that two vectors are similar with each other, otherwise they are dissimilar. The discontinuity of this function may give rise to the following problems: (1) for short dataset, $\phi_m$ or $\phi_{m+1}$ in SampEn may be zero, thus leading SampEn to be undefined; (2) little change of $m$ or $r$ or noise disturbance may result in a large change of entropy value, thus lacking robustness. Due to such shortcomings, Chen *et al.* (2009) suggested replacing $H$ with a continuous Gaussian type function, which is adopted to judge the level of similarity instead of a binary decision. This modification alleviates these deficiencies associated with ApEn. There are, however, still two points to be potentially enhanced for Chen *et al.* (2009)'s method. First, the parameters determining the Gaussian function could be defined with clearer physical meaning. Second, the idea of employing continuous function to replace the step function could be carried over to SampEn.

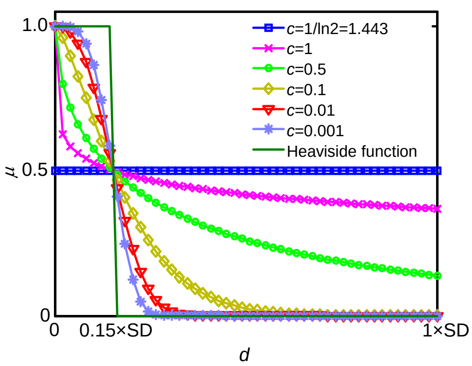


**Fig. 4 Heasivide step function and the formulated fuzzy membership function**
SD represents the standard deviation of original signals, $\mu$ represents the similarity level of two vectors, $c$ is the parameter controlling the steepness of the membership function, and $d$ indicates the distance between two vectors

## 2.2 Fuzzy SampEn (FSampEn)

Chen *et al.* (2009) replaced the function $H$ in ApEn with a Gaussian function defined as

$$\mu(d,r,n)=\exp(-d^n/r), \tag{2}$$

where parameters $n$ and $r$ control the shape of the function $\mu$. The way the function shape varies with these two parameters, however, is not intuitively clear. As such, we defined a new Gaussian type function shown in Eq. (3). This function was used for ApEn and SampEn, thus resulting in new entropies, i.e., fuzzy ApEn (FApEn) and fuzzy SampEn (FSampEn).

$$\mu(d,r,c)=\exp\left(-\frac{d^{\ln(\ln 2^c)/\ln r}}{c}\right). \tag{3}$$

Compared with those in Eq. (2), the parameters $r$ and $c$ in Eq. (3) have clearer physical meanings, wherein $r$ is a location parameter determining the location of the function (at which point the function takes a value of 0.5) and it corresponds to the tolerance $r$ in the function $H$. The parameter $c$ is a shape parameter deciding the steepness of the function as illustrated in Fig. 4. With $c$ ranging from 0 to 1/ln2, the function varies from the Heaviside function to a horizontal line of value 0.5. Therefore, the larger value the $c$ gets, the more relaxed the condition for vectors' similarity becomes, thus leading to more information loss. To maintain a balance between information loss and noise robustness, a value 0.01 was assigned to $c$. And $r$ is given a value of 0.15×SD locating in the range of [0.1, 0.25]×SD suggested in (Pincus, 1991; Richman and Moorman; 2000, Costa *et al.*, 2002; 2005), where SD represents the standard deviation of original signals.

Such a function in Eq. (3) can be viewed as a fuzzy membership function which measures the level of similarity between two vectors. Except the replacement of the $H$ by $\mu$, the definitions and calculation steps of FApEn and FSampEn are the same as those of ApEn and SampEn as shown in Figs. 2 and 3, respectively.

### 2.3 Multiscale framework

Essentially, multiscale entropy refers to calculations of corresponding entropy over a sequence of scales. Therefore, constructing the multiscale framework is of crucial importance. Costa *et al.* (2002; 2005) proposed an approach to obtain consecutive coarse-grained versions of an original signal, and defined multiscale SampEn to analyze physiological and physical signals.

As shown in Fig. 5, the coarse-grained version at scale $\tau$ is obtained by taking arithmetic mean of $\tau$ neighboring original data points without overlapping. For scale $\tau$=1, the coarse-grained version is simply the original time series. The length of each coarse-gained version is $N/\tau$, which decreases as the scale factor $\tau$ increases. Once those coarse-grained versions are built, multiscale entropy can be obtained by calculating conventional single-scale entropies on each of them. It is worth mentioning that the SD of the tolerance $r$=0.15×SD adopted to each scale is that of original signal, rather than that of each coarse-grained version at each scale.

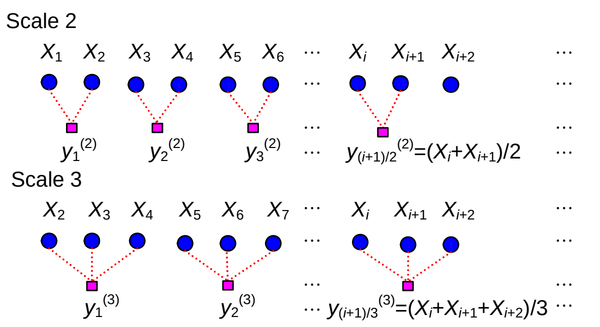


**Fig. 5 Illustration of how to built coarse-grained time series at different scales of $\tau$=2 and $\tau$=3**

## 3 Experimental

### 3.1 Experimental data

To validate and compare the feature extraction ability of various entropies for fault diagnosis, experiments were conducted on rolling element bearing vibration data acquired from Case Western Reverse Lab (Lou and Loparo, 2004; Loparo, 2005; Abbasion *et al.*, 2007). Vibration acceleration data were measured from a bearing mounted on the output end of a motor. The considered bearing conditions cover the following: normal, outer race fault, inner race fault, and rolling element fault. Each fault is a single-point fault induced artificially by electro-discharge machining, with defect sizes of 0.1778, 0.3556, 0.5334 and 0.7112 mm in diameter. The detail on the dataset is listed in Table 1. Each sampling collected 2048 data points, under a sampling frequency 12000 Hz and a shaft rotating speed 1797 r/min. Accordingly, the motor rotates about 5 revolutions over the time interval of a data sample.

### 3.2 Experimental scheme

As discussed above, one of the main advantages of FApEn and FSampEn is to improve the noise robustness of conventional ApEn and SampEn. Laboratory collected signals are often subjected to very small interference. Such signals, however, are too ideal compared to the field. Therefore, we designed two experimental schemes in order to highlight the noise robustness of the fuzzy entropies against that of the conventional ones.

Scheme I directly deals with the laboratory data shown in Table 1. In scheme II, the processed data are also those in scheme I, but with each data sample added with a white noise with noise power level located randomly between the range of [0.25, 1]×signal power level. For example, Fig. 6a shows the noise level added to each sample of bearing condition 2. Figs. 6b and 6d show the measured 12th and 32nd records, while Figs. 6c and 6e are the corresponding ones with noise intentionally added, respectively. Data samples from the same bearing conditions were added with different levels of noise instead of an

**Table 1 Description of bearing vibration dataset**

| Bearing condition | Defect size (mm) | Number of data samples | Label of bearing condition |
|---|---|---|---|
| Normal | 0 | 119 | 1 |
| Outer race fault | 0.1778 | 59 | 2 |
| | 0.3556 | 59 | 3 |
| | 0.5334 | 59 | 4 |
| Rolling element fault | 0.1778 | 59 | 5 |
| | 0.3556 | 59 | 6 |
| | 0.5334 | 59 | 7 |
| | 0.7112 | 58 | 8 |
| Inner race fault | 0.1778 | 58 | 9 |
| | 0.3556 | 59 | 10 |
| | 0.5334 | 59 | 11 |
| | 0.7112 | 58 | 12 |

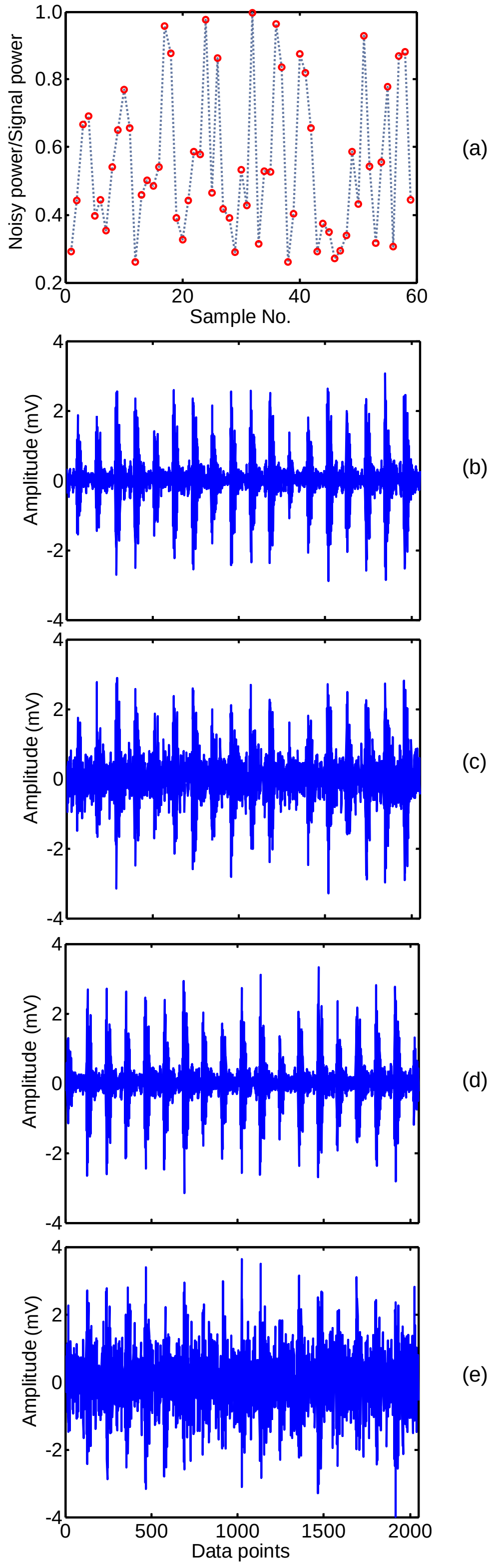


**Fig. 6 Taking the bearing condition 2 for example, shows (a) the power level of noise added to each data sample; (b) the 12th measured signal; (c) noise contaminated signal of (b); (d) the 32nd measured signal; (e) noise contaminated signal of (d)**

identical level. This reveal the ability of feature extraction method to extract characteristic features from vibration signals particularly in the case of the industrial practices filled with interference noise.

### 3.2 Parameter determination

Before calculating entropies, there are four parameters to be determined, i.e., $r$, $c$, $m$ and $\tau_{max}$. The values of $r$ and $c$ have been discussed above. The pattern length $m$ is usually fixed as 2 in (Costa *et al.*, 2002; 2005; Escudero *et al.*, 2006; Li and Zhang, 2008; Wang and Ma, 2008). This is because $m$=2 provides a precise result using less data points compared with $m$>2, and provides more information than $m$=1. Hence, $m$=2 is also adopted here. The value of maximum scale factor $\tau_{max}$ depends on the complexity of the system under consideration and the employed single-scale entropy. A larger $\tau_{max}$ is needed, when the underlying system is complex, or the feature extraction ability of the single-scale entropy is poor. For the considered bearing system, a maximum scale factor 20 was first chosen. The optimal $\tau_{max}$ is determined by a subsequent optimization procedure.

To quantify the feature extraction ability of various entropies, a classification algorithm is necessary. SVMs are increasingly exploited in many realms in contrast with ANNs (Hu *et al.*, 2005; Jiang *et al.*, 2005; Yuan and Chu, 2006; Yan *et al.*, 2007; Li and Wu, 2008). Due to the replacing of empirical risk minimization in ANNs by structural risk minimization, SVMs are reported to perform better on small-sample problems and have better generalization than ANNs. SVMs make use of kernel functions to map the original data to a higher dimensional space where the data may be linearly separated. There are various kernel functions used in SVMs, such as linear, polynomial, radial basis function (RBF) and sigmoid. Among them, RBF kernel is preferred because of fewer hyper-parameters and fewer numerical difficulties (Hsu *et al.*, 2009). For RBF kernel, there are two parameters to be determined, i.e., penalty parameter $C$ and RBF width parameter $\sigma$.

Thus, in order for the experiments to be carried out, there are still three parameters to be determined, i.e., $\tau_{max}$, $C$ and $\sigma$. This problem can be solved by grid-search and cross-validation. At first, we set a variable range for each of them, e.g., $\tau_{max}\in[1:1:20]$, $C\in[2^{-5:1:15}]$ and $\sigma\in[2^{-15:1:5}]$. Then, from each bearing condition, 20 data samples are randomly selected to constitute a database containing totally 240 samples. A 10-folds cross-validation is then performed on this database with each node $\{\tau_{max}, C, \sigma\}$. At last, the node $\{\tau_{max}, C, \sigma\}$ giving the best classification result is selected for subsequent applications. It is worthwhile noting that only 20 samples of each bearing condition are used in this stage. The purpose here is to examine the classification performance on unseen data for SVMs.

## 4 Results and discussion

### 4.1 Comparison among various entropies

Tables 2 and 3 show the results for schemes I and II, respectively. The feature extraction accuracies of all the entropies included in Fig. 1 are quantified by their classification rate over 100 tests. In each test, from each bearing condition, 20 samples were randomly chosen as training samples, the remainder for testing. The classification accuracies shown in the two tables are those for the testing dataset. The optimal $\tau_{max}$, average number of support vectors (SVs) and the optimal $(C, \sigma)$ are determined from the best results during the 'parameter determination' step. In that step, the 10-fold cross-validation was used, but different folds tend to yield different number of SVs. Therefore, the average number of SVs over 10 folds is likely not an integral. SVMs make classification decisions based on the SVs rather than all of the training samples. And a small optimal $\tau_{max}$ will lead to a small feature dimension. As a result, a small $\tau_{max}$ or a small average number of SVs implies that SVMs perform classification at a high speed. Consequently, the feature extraction ability should be evaluated from these two aspects of classification accuracy and classification speed.

For single-scale ApEn and SampEn, for either scheme I or scheme II, the differences between them

**Table 2 Results on scheme I (on directly measured data)**

| | | Accuracy over 100 tests (%) | | | Optimal $\tau_{max}$ | Average number of SVs | Optimal $(C, \sigma)$ |
|---|---|---|---|---|---|---|---|
| | | Average | Maximum | Minimum | | | |
| Conventional single-scale based | ApEn | 69.74 | 72.38 | 65.90 | 1 | 186.67 | $(2^{12}, 2^{-3})$ |
| | SampEn | 70.44 | 73.14 | 66.29 | 1 | 167.00 | $(2^{14}, 2)$ |
| | FApEn | 69.14 | 72.19 | 65.71 | 1 | 176.05 | $(2^{14}, 2)$ |
| | FSampEn | 70.76 | 74.86 | 66.29 | 1 | 169.24 | $(2^{13}, 2^{3})$ |
| Multiscale | ApEn | 96.91 | 99.05 | 94.86 | 16 | 151.15 | $(2^{15}, 2^{-7})$ |
| | SampEn | 95.93 | 97.52 | 93.90 | 15 | 143.91 | $(2^{13}, 2^{-4})$ |
| | FApEn | 97.42 | 99.17 | 95.81 | 6 | 112.86 | $(2^{15}, 2^{-1})$ |
| | FSampEn | 96.87 | 98.29 | 94.86 | 6 | 113.10 | $(2^{5}, 1)$ |

SVs: support vectors; $C$: penalty parameter of SVMs; $\sigma$: width of RBF kernel; for entropy calculation, $m$=2, $r$=0.15×SD of the analyzed signal, $c$=0.01

**Table 3 Results on scheme II (on data contaminated with noise)**

| | | Accuracy over 100 tests (%) | | | Optimal $\tau_{max}$ | Average number of SVs | Optimal $(C, \sigma)$ |
|---|---|---|---|---|---|---|---|
| | | Average | Maximum | Minimum | | | |
| Conventional single-scale based | ApEn | 19.07 | 24.57 | 16.19 | 1 | 212.94 | $(2^{11}, 2^{-1})$ |
| | SampEn | 20.75 | 24.95 | 17.33 | 1 | 213.76 | $(2^{15}, 2^{-7})$ |
| | FApEn | 23.77 | 27.81 | 19.62 | 1 | 212.94 | $(2^{12}, 2^{-1})$ |
| | FSampEn | 23.94 | 27.05 | 18.48 | 1 | 211.20 | $(2^{14}, 2^{-1})$ |
| Multiscale | ApEn | 69.99 | 73.90 | 65.71 | 17 | 199.24 | $(2^{7}, 2^{-2})$ |
| | SampEn | 68.94 | 72.95 | 64.19 | 18 | 206.48 | $(2^{14}, 2^{-10})$ |
| | FApEn | 73.94 | 76.76 | 70.29 | 20 | 202.14 | $(2^{10}, 2^{-5})$ |
| | FSampEn | 73.54 | 76.76 | 68.76 | 10 | 180.76 | $(2^{8}, 2^{-2})$ |

SVs: support vectors; $C$: penalty parameter of SVMs; $\sigma$: width of RBF kernel; for entropy calculation, $m$=2, $r$=0.15×SD of the analyzed signal, $c$=0.01

with respect to the accuracy and the average number of SVs are negligible. This may be due to the fact that SampEn is a modification of ApEn aimed to relieve the bias in the estimation of ApEn. That is to say that the entropy estimated by SampEn is closer to the theoretical one than that by ApEn. Here they have been used for feature extraction rather than a precise entropy value, so the advantage of SampEn over ApEn is not manifest. This reason holds for the negligible difference between single-scale FApEn and FSampEn in both schemes.

It is evident from both schemes that multiscale entropies overwhelm single-scale ones. And in the case of scheme I, all the multiscale entropies have a very high accuracy. Nevertheless, multiscale FApEn and multiscale FSampEn lead to a small $\tau_{\max}$ and a small average number of SVs in comparison with multiscale ApEn and multiscale SampEn, indicating a faster implementation of classification. Therefore, multiscale FApEn and multiscale FSampEn outperform multiscale ApEn and multiscale SampEn in experimental scheme I.

For the data in the presence of noise, the accuracy, $\tau_{\max}$ and the average number of SVs all degrade detrimentally for all the entropies as shown in Table 3. Noise of various power levels makes the dataset distribute much more dispersedly, and thereby the classification becomes more difficult. The degradation of the single-scale entropies makes them no longer useful for diagnosis purpose. However, the relatively high accuracies of FApEn and FSampEn over those of ApEn and SampEn can still reveal the merit of the replacement of the Heaviside function by the proposed fuzzy membership function. The accuracies of multiscale FApEn and multiscale FSampEn are better than those of multiscale ApEn and multiscale SampEn with an improvement around 5%. In addition, multiscale FSampEn has a smaller $\tau_{\max}$ and a smaller average number of SVs than multiscale FApEn. Hence, multiscale FSampEn is the best feature vector in scheme II where laboratory data are intentionally added with noise.

### 4.2 Comparison with time domain features

Statistical features in time domain are deemed to be able to extract plenty of information due to the fact that they do not transform the data into other domain like frequency or time-frequency domain, thus without information loss. There are many applications of these features in vibration signal processing (Nguyen *et al.*, 2008). The statistical features are used here for comparison purpose, the features including root mean square (RMS), skewness, kurtosis and maximum value (Max). The RMS and Max can indicate fault severity, while the skewness and kurtosis are able to identify the impulses induced by bearing fault generated shocks. As signal conditioner/amplifier variable gain, speed and load variations and other factors affect the measured signals, a preprocessing is carried out by dividing the signals by their SD (Fan *et al.*, 2007).

Table 4 shows the results of these time domain features on the dataset of the aforementioned scheme I and scheme II. It can be seen that for both schemes, either the accuracy or the numbers of SVs is inferior to that using multiscale FSampEn. This may be due to the fact that these statistical features do not take into account the nonlinearity of the underlying systems. It is also found that the results would be improved significantly if without the preprocessing, but these features will then become sensitive to the machinery working conditions like speeds and loads, as well as the signal acquisition device settings. In addition, a similar preprocessing was conducted in all the above mentioned entropies.

**Table 4 Results of statistical features in the time domain**

| Scheme | Accuracy over 100 tests (%) | | | Average number of SVs | $(C, \sigma)$ |
|---|---|---|---|---|---|
| | Average | Maximum | Minimum | | |
| I | 75.64 | 80.19 | 69.9 | 143.25 | $(2^{15}, 2^{-4})$ |
| II | 52.48 | 57.14 | 47.05 | 195.37 | $(2^{14}, 2^{-5})$ |

SVs: support vectors; *C*: penalty parameter of SVMs; $\sigma$: width of RBF kernel

## 5 Conclusion

In this paper, FSampEn is developed by formulating a fuzzy membership function with a physical meaning. Moreover, multiscale ApEn, multiscale FApEn and multiscale FSampEn are put forward.

The feature extraction ability of all those entropies were validated and compared by two experiment schemes. Scheme I deals with the directly measured data, while scheme II concerns the data tainted with noise. Experimental results demonstrate that multiscale entropies are absolutely superior to single-scale ones in both schemes. And multiscale fuzzy entropies (multiscale FApEn and FSampEn) outperform multiscale ApEn and multiscale SampEn. Moreover, multiscale FSampEn is advantageous over multiscale FApEn in terms of optimal $\tau_{max}$ and average number of SVs, especially in the case of processing noise corrupted data. Comparison with the time domain features also supports the multiscale FSampEn. In conclusion, the proposed multiscale FSampEn provides a favorable feature extraction method with certain noise robustness in the context of bearing vibration signal processing.

## References


Abbasion, S., Rafsanjani, A., Farshidianfar, A., Irani, N., 2007. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing*, **21**(7): 2933-2945. [doi:10.1016/j.ymssp.2007.02.003]

Chen, W., Wang, Z., Ren, X., 2006. Characterization of surface EMG signals using improved approximate entropy. *Journal of Zhejiang University-SCIENCE B*, **7**(10):844-848. [doi:10.1631/jzus.2006.B0844]

Chen, W., Zhuang, J., Yu, W., Wang, Z., 2009. Measuring complexity using FuzzyEn, ApEn, and SampEn. *Medical Engineering & Physics*, **31**(1):61-68. [doi:10.1016/j.medengphy.2008.04.005]

Costa, M., Goldberger, A.L., Peng, C.K., 2002. Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, **89**(6):068102. [doi:10.1103/PhysRevLett.89.068102]

Costa, M., Goldberger, A.L., Peng, C.K., 2005. Multiscale entropy analysis of biological signals. *Physical Review E*, **71**(2):021906. [doi:10.1103/PhysRevE.71.021906]

Escudero, J., Abasolo, D., Hornero, R., Espino, P., Lopez, M., 2006. Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy. *Physiological Measurement*, **27**(11):1091-1106. [doi:10.1088/0967-3334/27/11/004]

Fan, X.F., Liang, M., Yeap, T.H., Kind, B., 2007. A joint wavelet lifting and independent component analysis approach to fault detection of rolling element bearings. *Smart Materials & Structures*, **16**(5):1973-1987. [doi:10.1088/0964-1726/16/5/056]

Hsu, C., Chang, C., Lin, C., 2009. A Practical Guide to Support Vector Classification. Available from http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf [Accessed on May 20, 2009].

Hu, Z., Cai, Y., Li, Y., Xu, X., 2005. Data fusion for fault diagnosis using multi-class Support Vector Machines. *Journal of Zhejiang University-SCIENCE A*, **6**(10):1030-1039. [doi:10.1631/jzus.2005.A1030]

Jiang, J.D., Chen, J., Qu, L.S., 1999. The application of correlation dimension in gearbox condition monitoring. *Journal of Sound and Vibration*, **223**(4):529-541. [doi:10.1006/jsvi.1998.2161]

Jiang, Z., Fu, H.G., Li, L.J., 2005. Support Vector Machine for mechanical faults classification. *Journal of Zhejiang University-SCIENCE A*, **6**(5):433-439. [doi:10.1631/jzus.2005.A0433]

Kaffashi, F., Foglyano, R., Wilson, C.G., Loparo, K.A., 2008. The effect of time delay on Approximate & Sample Entropy calculations. *Physica D Nonlinear Phenomena*, **237**(23):3069-3074. [doi:10.1016/j.physd.200 8.06.005]

Li, Z.W., Zhang, Y.K., 2008. Multi-scale entropy analysis of Mississippi river flow. *Stochastic Environmental Research and Risk Assessment*, **22**(4):507-512. [doi:10.1007/s00477-007-0161-y]

Li, Z.Y., Wu, W.L., 2008. Classification of power quality combined disturbances based on phase space reconstruction and support vector machines. *Journal of Zhejiang University-SCIENCE A*, **9**(2):173-181. [doi:10.1631/jzus.A071261]

Liao, F.Y., Wang, J., He, P., 2008. Multi-resolution entropy analysis of gait symmetry in neurological degenerative diseases and amyotrophic lateral sclerosis. *Medical Engineering & Physics*, **30**(3):299-310. [doi:10.1016/j.medengphy.2007.04.014]

Logan, D., Mathew, J., 1996. Using the correlation dimension for vibration fault diagnosis of rolling element bearing-I. basic concepts. *Mechanical Systems and Signal Processing*, **10**(3):241-250. [doi:10.1006/mssp.1996.0018]

Loparo, K.A., 2005. Bearing Vibration Data: Case Western Reserve University Bearing Data Center Website. Available from http://www.eecs.case.edu/laboratory/bearing/welcome_overview.htm [Accessed on June 23, 2008].

Lou, X.S., Loparo, K.A., 2004. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing*, **18**(5):1077-1095. [doi:10.1016/S0888-3270(03)00077-3]

Nguyen, N.T., Lee, H.H., Kwon, J.M., 2008. Optimal feature selection using genetic algorithm for mechanical fault detection of induction motor. *Journal of Mechanical Science and Technology*, **22**(3):490-496. [doi:10.1007/s12206-007-1036-3]

Pincus, S.M., 1991. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences USA*, **88**(6):2297-2301. [doi:10.1073/pnas.88.6.2297]

Richman, J.S., Moorman, J.R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, **278**(6):H2039-H2049.

Wang, J., Ma, Q.L., 2008. Multiscale entropy based study of the pathological time series. *Chinese Physics B*, **17**(12):4424-4427. [doi:10.1088/1674-1056/17/12/015]

Wang, W.J., Chen, J., Wu, X.K., Wu, Z.T., 2001. The application of some non-linear methods in rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, **15**(4):697-705. [doi:10.1006/mssp.2000.1316]

Widodo, A., Yang, B.S., 2008. Wavelet support vector machine for induction machine fault diagnosis based on transient current signal. *Expert Systems with Applications*, **35**(3):307-316. [doi:10.1016/j.eswa.2007.06.018]

Xu, Y.G., Li, L.J., He, Z.J., 2002. Approximate entropy and its applications in mechanical fault diagnosis. *Information and Control*, **31**(6):547-551 (in Chinese).

Yan, R.Q., Gao, R.X., 2007. Approximate Entropy as a diagnostic tool for machine health monitoring. *Mechanical Systems and Signal Processing*, **21**(2):824-839. [doi:10.1016/j.ymssp.2006.02.009]

Yan, Z.G., Wang, Z.Z., Ren, X.M., 2007. Joint application of feature extraction based on EMD-AR strategy and multi-class classifier based on LS-SVM in EMG motion classification. *Journal of Zhejiang University-SCIENCE A*, **8** (8):1246-1255. [doi:10.1631/jzus.2007A1246]

Yuan, S.F., Chu, F.L., 2006. Support vector machines-based fault diagnosis for turbo-pump rotor. *Mechanical Systems and Signal Processing*, **20**(4):939-952. [doi:10.1016/j.ymssp.2005.09.006]