Fault Diagnosis of Rotating Machinery using Improved Entropy Measures

UNIVERSITY OF
LINCOLN

A Dissertation Presented

by

Zhiqiang Huo

in

College of Science

School of Engineering

University of Lincoln

A thesis submitted in partial fulfillment for the degree of

Doctor of Philosophy

August 2020

To my parents, for their love, trust, encouragement, and their unwavering support.
Being your son is the luckiest and proudest thing in my life.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

A Dissertation Presented
by
Zhiqiang Huo
August 2020

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Yu Zhang, for her excellent guidance, great patience and support. Her enthusiasm and immense knowledge always inspire me, and her expertise, technical, and editorial advice was essential to my growth, completion of the dissertation and various publications.

Second, I want to express appreciation to Prof. Timothy Gordon and Prof. Lei Shu, for their unwavering support, suggestions and assistance. With their help, I have not only improved my research capabilities, but also my way of thinking and learning.

Third, I would like to convey gratitude to other faculty and staff members of the School of Engineering who broadened my horizon and enriched my knowledge during my PhD study. Mr. Daniel Stones, technical resource manager, thank you for your enthusiastic assistance in helping me with the experimental setup and solving technical problems. Dr. Dong Zhang, thank you for your care, patience, encouragement, and support during my hard times. Your wisdom and spirits will always courage me. Dr. Miguel Martínez-García, thank you for your support during my study as well as encouraging talks to me. Mr. Guangzhou Sui, Dr. Haishan Lian, and Dr. Xingzhao Ma, thank you for the joyful time we spent together during your visiting to the Lincoln. That pleasant and easygoing summer in 2017 with Dr. Dong Zhang is unforgettable. Mr. Xiaowen Liao, Mrs. Li Yan, thank you for joyful talks and company. Dr. Jarek Grebenik, thank you for inviting me to do climbing and scrambling in such marvelous and gorgeous national parks in the UK. That exciting experience and moments will be refreshing forever in my memories.

Last but not least, I would like to thank the teachers and students in China who engaged into my research: Mr. Shuiquan Lin, thank you for helping me set up experiments and collect measurements from industrial equipment. Mr. Yunrong Lv, thank you for support in organizing staff members to help with experiment design and data collection.

At last, I would like to take this chance to thank Miss. Xueting (Snow) Liu, for her unwavering care and company during the completion of this thesis. Her encouragement and support give me tremendous courage to finish my dissertation and to brave the new future.

Abstract

Fault diagnosis of rotating machinery is of considerable significance to ensure high reliability and safety in industrial machinery. The key to fault diagnosis consists in detecting potential incipient fault presence, recognizing fault patterns, and identifying degrees of failures in machinery. The process of data-driven fault diagnosis method often requires extracting useful feature representations from measurements to make diagnostic decision-making. Entropy measures, as suitable non-linear complexity indicators, estimate dynamic changes in measurements directly, which are challenging to be quantified by conventional statistical indicators. Compared to single-scale entropy measures, multiple-scale entropy measures have been increasingly applied to time series complexity analysis by quantifying entropy values over a range of temporal scales. However, there exist a number of challenges in traditional multiple-scale entropy measures in analyzing bearing signals for bearing fault detection. Specifically, a large majority of multiple-scale entropy methods neglect high-frequency information in bearing vibration signal analysis. Moreover, the data length of transformed multiple signals is greatly reduced as scale factor increases, which can introduce incoherence and bias in entropy values. Lastly, non-linear and non-stationary behaviors of vibration signals due to interference and noise may reduce the diagnostic performance of traditional entropy methods in bearing health identification, especially in complex industrial settings.

This dissertation proposes a novel multiple-scale entropy measure, named Adaptive Multiscale Weighted Permutation Entropy (AMWPE), for extracting fault features associated with complexity change in bearing vibration analysis. A new scale-extraction mechanism - adaptive Fine-to-Coarse (F2C) procedure - is presented to generate multiple-scale time series from the original signal. It has advantages of extracting low- and high-frequency information from measurements and generating improved multiple-scale time series with a hierarchical structure. Numerical evaluation is carried out to study the performance of the AMWPE measure in analyzing the complexity change of synthetic signals. Results demonstrated that the AMWPE algorithm could provide high consistency and stable entropy values in entropy estimation. It also presents high robustness against noise in analyzing noisy bearing signals in comparison with traditional entropy methods. Additionally, a new bearing diagnosis method

is put forth, where the AMWPE method is applied for entropy analysis and a multi-class support vector machine classifier is used for identifying bearing fault patterns, respectively.

Three experimental case studies are carried out to investigate the effectiveness of the proposed diagnosis method for bearing diagnosis. Comparative studies are presented to compare the diagnostic performance of the proposed entropy method and traditional entropy methods in terms of computational time of entropy estimation, feature representation, and diagnosis accuracy rate. Further, noisy bearing signals with different signal-to-noise ratios are analyzed using various entropy measures to study their robustness against noise in bearing diagnosis. Additionally, the developed adaptive F2C procedure can be extended to a variety of entropy algorithms based on improved single-scale entropy method used in entropy estimation. In the combination of artificial intelligence techniques, the improved entropy algorithms are expected to apply to machine health conditions and intelligent fault diagnosis in complex industrial machinery. Besides, they are suitable to evaluate the complexity and irregularity of other non-stationary signals measured from non-linear systems, such as acoustic emission signals and physiological signals.

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

λ	Time delay
μ	Mean value
ρ	Correlation coefficient
σ	Standard deviation value
τ	Scale factor
$C_{j,i}$	Wavelet decomposition coefficient
j	Wavelet decomposition level
m	Embedding dimension
N	Data length of time series
r	Tolerance
$R_{j,i}$	Reconstructed sub-signal based on wavelet coefficient
U_i	Selected reconstructed sub-signal
AI	Artificial Intelligence
AMWPE	Adaptive Multiscale Weighted Permutation Entropy
ANFIS	Adaptive Network-based Fuzzy Inference System
ApEn	Approximate Entropy
CWRU	Case Western Reserve University

CWT Continuous Wavelet Transform

DWT Discrete Wavelet Transform

F2C Fine-to-Coarse

F2CMPE Fine-to-Coarse Multiscale Permutation Entropy

FuzzyEn Fuzzy Entropy

GDUPT Guang Dong University of Petrochemical Technology

KS Kolmogorov-Sinai

MFEn Multiscale Fuzzy Entropy

MPEn Multiscale Permutation Entropy

MSEn Multiscale Entropy

PerEn Permutation Entropy

RWE Relative Wavelet Energy

SampEn Sample Entropy

ShanEn Shannon Entropy

SNRs Signal-to-Noise Ratios

SVM Support Vector Machine

WPerEn Weighted Permutation Entropy

WPT Wavelet Packet Transform

Chapter 1

Introduction

1.1 Research Background

Rotating machinery has been extensively employed in a variety of modern oil, power and process industries (shown in Fig. 1.1) to support power generation and transportation of material processing (e.g., oil and gas). While the machinery is designed to serve in a long-term in industrial settings, engineering machines are usually operated in complex and harsh environments, such as high temperature, moisture and variable operating conditions. Fatigue and damage to key rotating components will cause performance degradation and malfunctions in the mechanical system [1].

Fig. 1.1 An application of multi-stage centrifugal air pump in the petrochemical plant.

The existence of faulty components can result in unexpected energy waste and performance deterioration of the entire system. If the damaged component is not detected and repaired early, the fault will progress and cause subsequent damages to adjacent components. Moreover, severe conditions may lead to unplanned breakdowns or unexpected personnel

injury to first-line workers [2]. The performance deterioration of rotating machinery will further cause decreased reliability of the system as well as potential safety concerns. So, there is a necessity to ensure high reliability and safety in the operation of industrial mechanical systems [3].

According to the structure of the machinery, malfunctions can be attributed to different root causes. Structural damage is the most commonly seen fault in operation, which is usually linked to fatigue and overloaded operations. These damages can exist in any components in the machinery, such as motor, rotating shaft, rolling bearing and gearbox. Among these components, rolling bearing is the most widely used in mechanical systems. Its primary function is to support the mechanical rotating body on radial load and reduce friction during its movement. While its wide application in machinery, rolling bearing is prone to a myriad of malfunctions in operation.

The normal service life of the rolling bearing is often determined by material fatigue and wear at the running surfaces. In general, many reasons can lead to early bearing faults, such as faulty installation, corrosion, poor lubrication, and overload and overspeed during operation. The occurrence of early faults will accelerate the fatigue to the bearing, thus greatly reducing the bearing's service life. Some examples of these typical bearing failures are illustrated in Fig. 1.2.

Fig. 1.2 Some typical bearing failures: (a) fatigue and spalling, (b) corrosion and pitting, (c) melted balls with high-temperature.

More importantly, bearing failures not only cause performance degradation in the machinery but also increase maintenance costs and financial losses in manufacturing. For example, several surveys [4–6] with regard to the proportion of failure types, conducted by the IEEE Industry Application Society (IEEE-IAS), reported that bearing faults can account for 30-40% of the total number of failures in the machinery [7]. Moreover, related studies [8, 9] have revealed that motor and gearbox faults are attributed to bearing faults, which account for 40-50% of the total faults. With respect to the maintenance cost, it is reported that a faulty bearing worth of $1500 can result in a $100000 gearbox replacement and $70000 spent on

replacing the damaged components in wind turbines [10]. Consequently, it is paramount to monitor the health condition of the bearing and troubleshoot bearing failures in an early stage to ensure the reliable operation and save unnecessary financial losses for the enterprise. Therefore, how to detect and diagnose bearing health conditions in accurate, effective and intelligent manners have become a challenge for modern processing industries.

Recent advancements in computer, measurement and communication technologies have enabled intelligent health condition monitoring and fault diagnosis in the machinery [11]. Condition monitoring and fault diagnosis of rolling bearing has been a research frontier in the past decades [12], which have attracted considerable attention from both the academia and industrial fields. Given a system or process, be it natural or man-made, its evolution can be followed by a finite amount of measurements. Herein, a subject of interest is how to monitor the machinery and to analyze measurements to make diagnostic decision-making. There exist various condition monitoring technologies supporting the collection of different types of measurements (sensor data) in the operation of machinery. The most commonly used methods include vibration monitoring, acoustic emission monitoring, thermal imaging monitoring, temperature monitoring and current monitoring [13–15]. Among them, vibration monitoring is one of the most representative techniques for machinery health monitoring [16–18]. Under an invariable operating condition, sudden and dramatic changes in vibration (acceleration) data are often associated with the unexpected changes in the system's health condition. For example, when a fault is induced to the inner race in a rolling bearing, the rolling elements will strike the local fault and excite high-frequency resonance between the bearing and the acceleration sensor. Due to its applicability and convenience, vibration monitoring has been widely applied in varied machinery diagnostic systems [19].

The dynamic response of a mechanical system, due to a change of state, is often reflected in the sensor measurements. The key to fault diagnosis of rolling bearing is to monitor the consistency between these measurements and the machine operational regime, so it is possible to predict the operating status of the machine and potential faults. In this case, operators can carry out corresponding maintenance schemes through the analysis of measurements, enabling cost-effective maintenance planning (such as repair and replacement of faulty components), rather than unnecessary and blind maintenance strategies (such as the replacement of some components just because of age). However, early failures, in the bearing diagnosis, often exhibit weak and transient failure characteristics that are hard to identify. Moreover, vibration measurements collected from the practical machinery often contain noises and become more complex to analyze. These factors will lead to non-linear and non-stationary characteristics in collected signals that contain periodic components caused by interactions between rotating components and transient components because of background

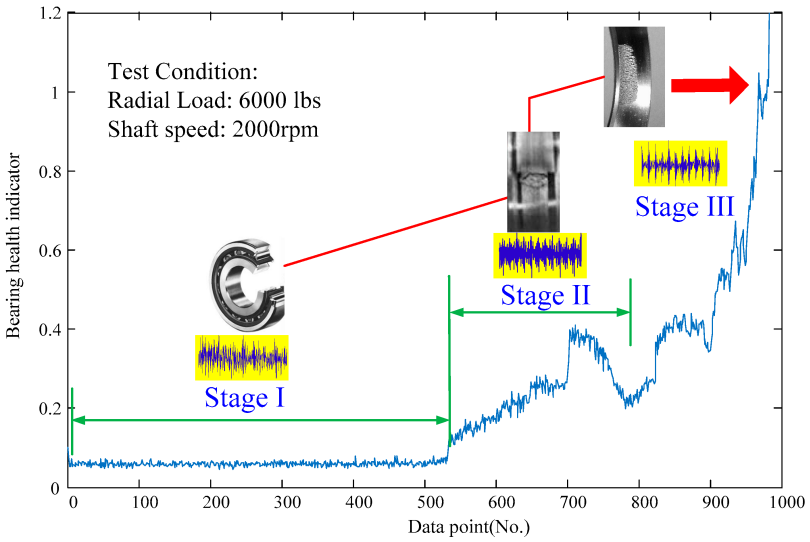


Fig. 1.3 Amplitude of the anomaly measure versus the time point for a real bearing of whole life [20]

noises. This has further added difficulty in analyzing non-linear and non-stationary signals and identifying fault patterns in industrial settings.

For instance, Fig. 1.3 shows a real service life of bearing in a run-to-failure experiment [20]. From Fig. 1.3, there are three stages: stage (I), the normal operation stage; stage (II), early fault stage; stage (III), severe fault stage. In the normal operating stage, the magnitude of vibration data collected from a healthy bearing is relatively low and steady. With the occurrence of a failure, the amplitude of vibration signals significantly increases in line with the progress of failure until it completely fails. Vibration monitoring has demonstrated its effectiveness in distinguishing between different health states of bearings. A special subject of interest is to construct a diagnostic decision model that enables inspecting the consistency between vibration measurements and bearing health conditions, so the fault type and severity can be qualitatively assessed.

1.2 Present State of Knowledge

Over the past decades, there is a fast-growing development in the field of fault diagnosis of rolling bearing, and fruitful results have been achieved. Fault diagnosis techniques can be broadly categorized into model-based methods, signal-based methods, and data-driven (knowledge) based methods. As early as the 1990s, model-based fault diagnosis was the mainstream of research [21]. The model-based methodology applies rigorous theoretical formulas and derives models to assess the health condition of bearings. For

instance, fundamental fault frequencies can be generated when rolling elements pass over a surface that has a local fault on either the rolling element or the raceway - inner race and outer race. For a stationary bearing, when its geometry (shown in Fig. 1.4) is known, the characteristic bearing defect frequencies can be derived from formulations as described in Table 1.1.

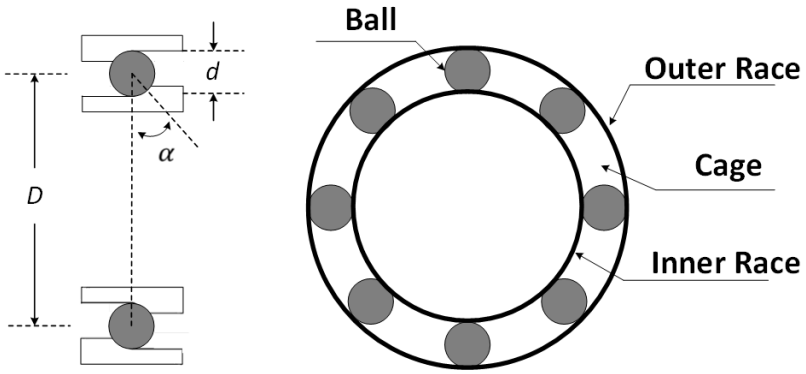


Fig. 1.4 Schematic diagram of a rolling element bearing.

Table 1.1 Formulae of characteristic bearing defect frequencies.

Bearing fault type	Characteristic frequency[1]
Ball-pass spin frequency (BSF)	$\mathrm{BSF} = \frac{D}{2d}\{1 - \left(\frac{d}{D}\cos\alpha\right)^2\}$
Ball-pass frequency at outer race (BPFO)	$\mathrm{BPFO} = \frac{nf_r}{2}\{1 - \frac{d}{D}\cos\alpha\}$
Ball-pass frequency at inner race (BPFI)	$\mathrm{BPFI} = \frac{nf_r}{2}\{1 + \frac{d}{D}\cos\alpha\}$
Fundamental train frequency (FTF)	$\mathrm{FTF} = \frac{f_r}{2}\{1 - \frac{d}{D}\cos\alpha\}$

[1] f_r is the shaft speed, n is the number of rolling elements, and α is the angle of the load from the radial plane, d is the diameter of the roller, and D is the mean diameter of the bearing as shown in Fig. 1.4.

These formulae are however the kinematic frequencies assuming no sliding contact, and it also assumes the rolling elements only pass over the raceway surfaces. In fact, there must virtually always be some sliding contact, and the resulting changes will give a fundamental change in the character of signals, thus leading to a deviation from the calculated fundamental frequencies. Also, the actual frequencies would have a stochastic variation as for different types of bearings and operating conditions. Therefore, the actual signals are often more complicated than simulated signals. Most importantly, model-based diagnostic methods often require sufficient prior knowledge and expert experience about the machinery. They

tend to outperform other methods when sufficient knowledge of physical mechanisms about the machinery is available. However, modern industrial mechanical systems are becoming larger and more complex; it is not always possible to fully understand the system's regimes and predict their behaviors. As a result, the model-based methods are often built case by case. Their applicability in knowledge transfer is limited, and therefore not suitable to generic diagnostic applications.

At almost the same period, the fast-growing improvement in digital signal-processing techniques significantly boosted the development of signal-based fault diagnosis methodology. Signals measured from machinery not only reflect system operating states but also contain rich statistical information that is closely associated with machine health conditions. The analysis of signals thus enable extracting fault-related features or statistical components from time- or frequency-domain rather than explicit input-output models in model-based method [22]. These extracted feature representations usually are useful to differentiate between healthy machine states or fault states, upon which a diagnostic decision is then made.

For example, a variety of signal-processing methods have been proposed and applied to machinery fault diagnosis [23]. In general, signal-processing methods can be broadly classified into three categories according to the signal analysis principles and signal types: time-domain signal analysis, frequency-domain signal analysis, and time-frequency-domain signal analysis. With respect to time-domain signal analysis, signals (time series) are considered as time-domain waveforms. The commonly used methods include autoregressive model, minimum entropy deconvolution, and spectral kurtosis. For instance, the autoregressive model is a suitable time-series analysis method assuming that time series are univariate and stationary. Nevertheless, the intrinsic dynamic nature during machine operation often leads to non-stationary signals, adding the difficulty in machinery fault diagnosis using time-domain analysis methods only. In this case, frequency analysis is a suitable tool to reveal the time-variant features from the spectrum.

In frequency-domain signal analysis, signals are transformed from time-domain to frequency-domain representations using spectrum analysis techniques. Some commonly used tools include Fourier transform and Hilbert transform [20]. Spectrum analysis detects the changes associated with status change or faults in the frequency domain. Feature representations can then be obtained from frequency components in the frequency domain. The spectrum analysis often requires the signals are stationary; otherwise, critical transient components underlying in non-stationary signals cannot be appropriately captured using traditional frequency-domain analysis methods. In this case, considering time-varying char-

acteristics of signals, several modern signal-processing methods - time-frequency signal analysis methods - have been developed by the researchers.

Time-frequency signal analysis methods take both time- and frequency-domain into account. They usually decompose signals into a set of transformed components with a good time- and frequency-domain resolution, thus enabling the examination of transient components in signals. Several commonly used methods include short-time Fourier transform, wavelet transform and empirical mode decomposition [23]. Time-frequency analysis methods have been widely applied to identify the constituent components of signals and their time variation, and thus to analyze non-stationary signals. Signal-based fault diagnosis methods are suitable tools to inspect the consistency between system measurements and machine health conditions. For signal-based analysis methods, while there are theoretical differences between above-presented techniques (i.e., time-domain, frequency-domain, and time-frequency domain analysis), they have the same goal. That is, they all aim to extract critical feature representations from original signals and to capture key transient components from time- or frequency-domain. These extracted features or components are often linked to certain machine operating conditions, revealing the underlying correlation between measurements and machinery health states.

In addition to the model- and signal-based fault diagnosis methods, recent advancements in the feature engineering and Artificial Intelligence (AI) methods have greatly promoted knowledge-based fault diagnosis, well known as data-driven based fault diagnosis, for intelligent machinery fault diagnosis [24]. Different from signal-based analysis methods, the basis of data-driven based fault diagnosis lies in building up the statistical model upon previously observed data with little or no prior expert knowledge. The complicated machinery behaviors can thus be understood, and machinery health conditions can be inferred and predicted, by checking the consistency between the constructed model and measurements. This methodology usually presents satisfactory diagnosis performance when there exist sufficiently abundant data sets and computationally intensive resources. The general framework of data-driven fault diagnosis methods (as shown in Fig. 1.5) incorporates three main processes: signal acquisition, feature extraction and representation, and fault pattern classification [25]. In the signal collection stage, various signals - such as vibration and current signals - can be measured from the machinery via condition monitoring techniques. Feature engineering methods are then needed to extract features carrying rich information in regard to machine health states. With these extracted features, advanced machine learning techniques (such as Support Vector Machine (SVM)) can be applied to learn, refine and summarize domain knowledge in the high-dimensional feature space [26, 27]. The acquired diagnostic model can finally be used to detect potential performance degradation and distinguish between

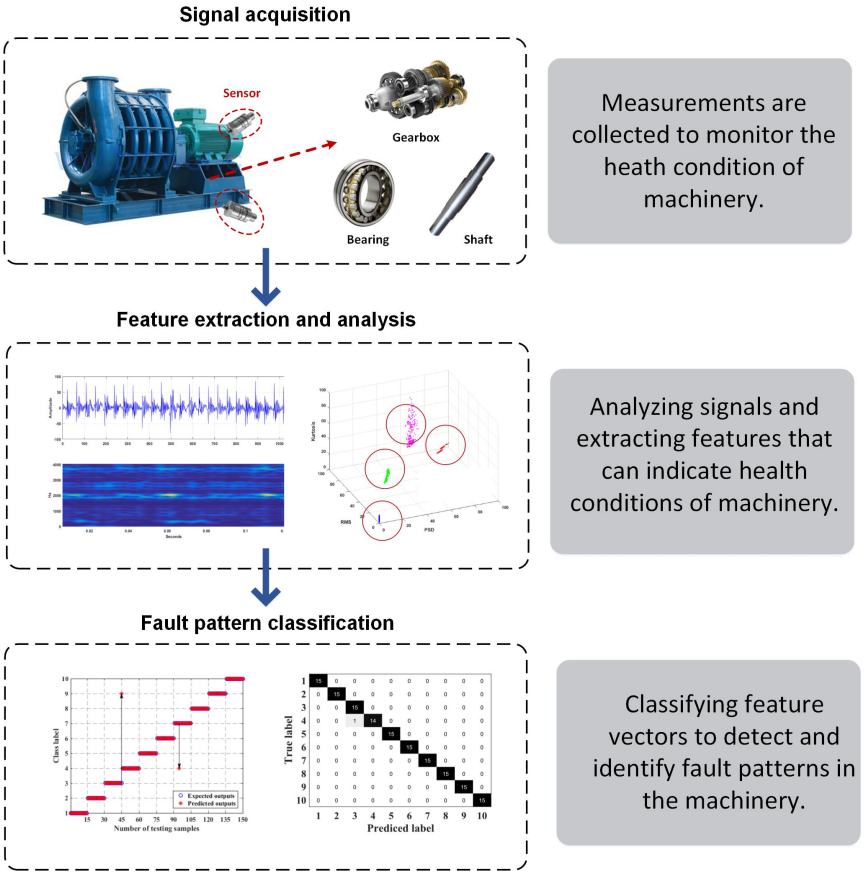


Fig. 1.5 General flowchart of data-driven based fault diagnosis methodology.

machine health conditions, supporting online diagnostic decision-making. In data-driven analysis methods, diagnostic performance is highly dependent on the richness of existing datasets and the type of feature representation. Proper design of fault features is fundamental to extract useful knowledge and information appropriately from measured signals. Extracted features can often represent the character of signals, by which the evolution of the state of the system can be traced and assessed. Therefore, one of the key challenges is to design and apply proper feature representations and to assess the intended model, so machine operating conditions can be distinguished accurately, and fault pattern type can be identified effectively.

There exist many statistical features that are suitable to extract characteristic features from vibration signals. Such features can be generally classified into time-domain and frequency-domain features [28]. Typical time-domain feature indicators include peak value, mean value, variance value, kurtosis value, and root-mean-square value. These feature values often reflect the statistical characteristics of waveforms of signals and are easy to implement. The time-domain waveform can either be original signals or transformed components via signal processing techniques. With respect to frequency-domain features, detecting the

changes in the frequency distribution or magnitude of components can differentiate between fault states and normal state. Some commonly used frequency feature indicators include frequency center value, mean-square frequency value, and energy value of the frequency spectrum.

One of the difficulties with these traditional methods is that they rely on linearity and signal stationarity assumptions - which may not appropriately extract signal symptoms, especially under complex environments with interacting components (systems of systems) and strong background noise. For example, even though sensors are mounted on the location near to monitored components, the interaction between structural and mechanical components - such as the interaction between rolling elements and the raceway - can result in interfered vibration signals by unrelated frequency components. In an actual example, the level of kurtosis was reduced as damage in the machine bearings increased; when the vibration pattern became more complex - due to the bearing damage, the kurtosis matched that of undamaged bearings [11]. In complicated industrial systems, measured signals often exhibit non-linear behavior due to instantaneous variations in friction, damping, or invariable load and speed conditions. As a result, quantifying such dynamic changes of system responses is significant to early fault detection [29].

Complexity analysis of complex systems has received ever-growing attention in the past decades [30]. Entropy, as a complexity measure, has been widely applied for time series analysis. The design of machinery condition monitoring and fault diagnosis systems is one preeminent example [31]. Entropy measures do not rely on linear assumptions, and are suitable for distinguishing regular, chaotic and random behaviors. Complex systems with non-linear dynamics often present larger response diversity and uncertainty; thus, it is sometimes easier to characterize underlying patterns in terms of dynamic changes, than to analyze the little knowledge base data available. One advantage of entropy measures is that they can directly measure dynamic changes and quantify the degree of complexity of a system, which would be challenging to assess by traditional statistical indicators [32]. Since the performance degradation of a machine will present more non-linear characteristics, the analysis of the complexity of the measurements has revealed that the change in the complexity value is related to the deterioration of the machine component [33]. Further, the extensive flexibility of entropy analysis methods is advertised by their all-encompassing applicability to the analysis of complex systems, be it natural or man-made; besides the subject of monitoring industrial machines, entropy analysis has been extensively applied for studying the complexity of dynamical systems in multiple fields. Such areas of research may be far more complex than mechanical systems, including language [34], biological [35], financial [36] and other complex systems [37–39].

Entropy measures can be generally classified into two categories based on their basic principles: single-scale entropy measures and multiple-scale entropy measures [31]. The most commonly used single-scale entropy measures include Shannon measure, Approximate Entropy (ApEn), Sample Entropy (SampEn), Fuzzy Entropy (FuzzyEn), Permutation Entropy (PerEn), and their variants. A greater entropy value often implies more irregularity and complexity observed in the measurement and system. By contrast, multi-scale entropy measures analyze a time series from a range of time scales [40], where the single-scale entropy measures provide the basis of entropy estimation. The concept of multiple-scale entropy measures is initially introduced in a modified SampEn method [41], where entropy values are calculated over a range of temporal scales. Some widely used methods include Multi-scale Entropy (MSE), Multi-scale Fuzzy Entropy (MFE), and Multi-scale Permutation Entropy (MPE), and their variants [40]. They have been successfully applied to the analysis of complex systems, such as in biological analysis [42] and wind turbine fault diagnosis [43].

Despite its wide application, the MSE algorithm, as well as their variants, however, encounters several limitations in entropy analysis for machinery fault diagnosis [42]. First, the coarse-graining procedure is similar to the finite-impulse response filter, which is a low-pass filtering operation. Thus, only low-frequency components are considered in the coarse-grained time series. In bearing fault diagnosis, the neglect of information in high-frequency components, however, corresponds to abandoning fault-related symptoms in high-frequency components. Second, the use of coarse-grained scale-extraction procedure will lead to dramatically reduced data length in time series, especially with an increasing scale factor in multiple-scale entropy measures. This may yield biased entropy values with a large variance when the scale factor is large. Though some improvements have been achieved, less research work focuses on addressing above-presented problems. There is still a strong need to investigate the applicability and reliability of fault diagnosis methods using improved entropy analysis methods. Consequently, this study mainly focuses on summarizing entropy measures in machine fault diagnosis from a systematic perspective, developing improved multiple-scale entropy measures for complexity analysis, and designing bearing fault diagnosis systems using improved entropy measures.

1.3 Research Hypotheses

The research of the dissertation concentrates on developing improved multiple-scale entropy measures for fault diagnosis of rolling bearing, which is founded on the following hypotheses:

1. The data-driven fault diagnosis method is feasible to detect early faults and assess fault severity levels in rolling bearings, where entropy measures extract complexity

features from vibration signals, and machine learning techniques classify fault patterns, respectively;

2. The improved multiple-scale entropy measures can characterize feature representations associated with complexity change in signals, providing satisfactory and reliable entropy analysis;

3. The improved scale-extraction mechanisms can generate multiple-scale time series with rich fault information, yielding consistent and appropriate entropy values via single-scale entropy analysis;

4. The proposed scale-extraction frameworks can be extended to new entropy measures where different single-scale entropy algorithms can be used for entropy estimation and complexity analysis;

1.4 Main Research Contributions

The dissertation systematically reviews the theoretical development of several fundamental entropy measures in the complexity analysis. In-depth insights into possible applications of entropy measures to machinery fault diagnosis are presented. This study proposes two novel multiple-scale entropy measures - Fine-to-Coarse Multiscale Permutation Entropy (F2CMPE) and Adaptive Multiscale Weighted Permutation Entropy (AMWPE) methods. They are applied to extract the complexity change from raw signals. Compared to traditional entropy methods, the proposed entropy algorithms have four advantages: 1) they take into account both low- and frequency-information in entropy estimation; 2) they reduce the bias in entropy values on adjacent scales with an increasing scale factor; 3) they present high computational efficiency in vibration analysis; 4) they exhibit high robustness against noise in analyzing noisy signals for bearing diagnosis. Also, new bearing diagnosis methods are developed based on the proposed new entropy measures and multi-class SVM classifier. The entropy measures characterize the complexity change in measurements, and then fault features are fed into the SVM for fault pattern recognition. The effectiveness of the proposed bearing diagnosis method is verified through real experimental examination. Three case studies are carried out to study the diagnostic performance of the proposed bearing diagnosis method under different fault types, severity degrees, and noise levels.

1.5 Related Publications

Related publications are listed as follows:

1. The research of entropy analysis is based on the following papers:

Zhiqiang Huo, Miguel Martínez-García, Yu Zhang, Ruqiang Yan, Lei Shu. Entropy Measures in Machine Fault Diagnosis: Insights and Applications. *IEEE Transactions on Instrument and Measurement*, 6:2607-2620, 2020.

Zhiqiang Huo, Yu Zhang, Gbanaibolou Jombo, Lei Shu. Adaptive Multiscale Weighted Permutation Entropy for Rolling Bearing Fault Diagnosis. *IEEE Access*, 8:87529-87540, 2020.

Zhiqiang Huo, Yu Zhang, Lei Shu, Michael Gallimore. A New Bearing Fault Diagnosis Method based on Fine-to-Coarse Multiscale Permutation Entropy, Laplacian Score and SVM. *IEEE Access*, 7:17050 - 17066, 2018.

Zhiqiang Huo, Yu Zhang, Lei Shu, Xiaowen Liao. Edge Permutation Entropy: An Improved Entropy Measure for Time-Series Analysis. In proceedings of In proceedings of 45rd Annual Conference of the IEEE Industrial Electronics Society (*IECON 2019*), Oct. 14-17, Lisbon, Portugal, 2019.

Zhiqiang Huo, Yu Zhang, Lei Shu. Fine-to-Coarse Multiscale Permutation Entropy for Bearing Fault Diagnosis. In Proceedings of IEEE Conference on International Wireless Communications & Mobile Computing Conference (*IWCMC 2018*), Limassol, Cyprus, June, 25-29, 2018.

Zhiqiang Huo, Yu Zhang, Lei Shu. A short survey on fault diagnosis of rotating machinery using entropy techniques. In proceedings of 3rd EAI International Conference on Industrial Networks and Intelligent Systems (*INISCOM 2017*), Ho Chi Minh, Vietnam, Sep. 4, 2017.

Zhiqiang Huo, Yu Zhang, Zhangbing Zhou, Jianfeng Huang. Crack detection in rotating shafts using wavelet analysis, Shannon entropy and multi-class SVM. In proceedings of 3rd EAI International Conference on Industrial Networks and Intelligent Systems (*INISCOM*

2017), Ho Chi Minh, Vietnam, Sep. 4, 2017.

2. The research of machinery fault diagnosis is based on the following papers:

Zhiqiang Huo, Yu Zhang, Lei Shu, Yunrong Lv, Shuiquan Lin. Bearing Fault Diagnosis using Multi-sensor Fusion based on weighted D-S Evidence Theory. In proceedings of Mechatronika 2018, December 5-7, Brno, Czech Republic, 2018.

Zhiqiang Huo, Yu Zhang, Pierre Francq, Lei Shu, Jianfeng Huang. Incipient Fault Diagnosis of Roller Bearing using Optimized Wavelet Transform based Multi-speed Vibration Signatures. *IEEE Access*, 5: 19442-19456, 2017.

Zhiqiang Huo, Yu Zhang, Lei Shu. A comparative study of WPD and EMD for shaft fault diagnosis. In proceedings of In proceedings of 43rd Annual Conference of the IEEE Industrial Electronics Society (*IECON 2017*), Oct. 29 - Nov. 2, Beijing, China, 2017.

Zhiqiang Huo, Yu Zhang, Richard Sath, Lei Shu. Self-adaptive fault diagnosis of roller bearings using infrared thermal images. In proceedings of 43rd Annual Conference of the IEEE Industrial Electronics Society (*IECON 2017*), Oct. 29 - Nov. 2, Beijing, China, 2017.

Zeyu Zhang, Amjad Mehmood, Lei Shu, **Zhiqiang Huo**, Yu Zhang, and Mithun Mukherjee. A Survey on Fault Diagnosis in Wireless Sensor Networks. *IEEE Access*, 6: 11349-11364, 2018.

1.6 Thesis Organization

The dissertation consists of six sections. The organization of the thesis is as follows:

Chapter 1 introduces the research background and present sate of knowledge of fault diagnosis of rolling bearings and briefly summarizes entropy analysis in bearing diagnosis, and concludes major contributions of the thesis.

Chapter 2 systematically reviews the fundamental concepts of entropy measures and their theoretical development and summarizes the roles and usages of entropy measures in machinery fault diagnosis.

Chapter 3 introduces the principles of some fundamental entropy measures for time series complexity analysis, including single-scale entropy and multiple-scale entropy methods.

Chapter 4 presents the improved multiple-scale entropy measures and new bearing diagnosis methods incorporating vibration monitoring, entropy analysis and fault pattern classification techniques. A numerical study is presented to investigate the performance of the proposed entropy method in analyzing the complexity change of synthetic signals.

Chapter 5 examines the effectiveness and performance of the proposed diagnosis methods on fault detection and identification in rolling bearings using laboratory test rigs and real industrial-scale machinery system. Comparative studies are also carried out to compare the diagnosis performance of different methods where conventional entropy and improved entropy algorithms are used for entropy analysis.

Chapter 6 finally concludes the dissertation and proposes suggestions for further research.

Chapter 2

Literature Review of Entropy Measures

Given a system, be it natural or man-made, its evolution can be followed by a finite amount of observations (measurements). A subject of interest is how to trace the evolution of such a system by evaluating the complexity change in the measurements. Entropy measures, as non-linear complexity metrics, are suitable to assess dynamic changes in the system. In the machinery, the occurrence of failures will often induce non-linear behaviors due to instantaneous variations in friction, damping, or load and speed conditions [44]. For instance, when localized failures are introduced to a bearing, sharp force impacts will excite high-frequency resonances in the bearing and its structure, resulting in the complexity change of vibration signals. For a roller defect, the defect will contact the surface of the raceways along with the rotation with shaft frequency, so measured signals are always non-linear and non-stationary.

One advantage of entropy measures is that they do not rely on linear assumptions, and are suitable for distinguishing regular, chaotic and random behaviors. It is sometimes easier to characterize underlying patterns in terms of dynamic changes than to analyze the little knowledge base data available. Therefore, assessing and quantifying dynamic changes provide the possibility of detecting faults in the rolling bearing [31]. In this chapter, the theoretical development of several fundamental entropy measures is first introduced, followed by a systematic summarization of potential roles and usages of entropy measures for machinery fault diagnosis.

2.1 Evolution of Entropy Measures

Entropy has been a transcendental and pervasive concept in numerous disciplines, ranging from logic and physics to biology and engineering. Although the notion of entropy can be traced back to the nineteenth century, it still attracts interest - due to its applicability

into different contexts, and to the multiple interpretations of its implications [45]. However, notions of entropy are defined differently in various contexts (e.g., information theory and dynamical systems theory), which may confound researchers in the applied sciences. In order to clarify the applicability of entropy in machinery fault diagnosis, this chapter reviews the theoretical development of some fundamental entropy measures in different contexts.

Historically, entropy arose after the invention of the heat engine, through pioneering research works towards clarifying thermodynamical processes [46]. The research led to the formulation of the *Second Law of Thermodynamics*, and entropy was defined as the ratio of energy over temperature [1] in equilibrium [47]. Their works reveal that entropy of an isolated system can never decrease over time.

Later, Ludwig Boltzmann [2] and Josiah Gibbs [3] independently interpreted the definition of entropy as a measure of the number of states that a physical system can adopt from a molecular perspective, giving rise to *statistical mechanics* [48]. They observed that macrostates with a higher number of possible microstates are more likely and exhibit larger entropy values - from a molecular perspective. More importantly, Gibbs revealed that entropy could be described in terms of statistical quantities, such as probabilities and their logarithms – setting the path towards the usage of entropy as a tool for non-linear time series analysis.

Subsequent research by Hartley, Wiener and Shannon resulted in the introduction of a parallel entropy formulation, which lies at the center of information theory – known as *information entropy* or *Shannon entropy* (ShanEn) [49]. ShanEn was proposed to quantify the amount of information content conveyed by messages from an information source [50]. Moreover, it interprets the uncertainty and randomness of the *system's events* – i.e., its behavior – from a probability viewpoint. That is, the examined system is understood as a random variable as well as its observations. Inspired by ShanEn, various concepts of entropy were later developed within complexity theory, particularly in the study of dynamical systems. One example is the Kolmogorov-Sinai (KS) entropy measure [51]. KS entropy occupies center stage in chaos theory - a mathematical theory of deterministic dynamical systems that may exhibit unpredictable and irregular behaviors. Specifically, in the phase space, the deterministic chaos is understood as the trajectories of a dynamic system whose initial conditions are extremely approximate, but their evolution will diverge exponentially fast. Such an exponential rate at which trajectories evolve - information is processed - is measured by the KS entropy. However, it was reported that KS entropy is not easy to

[1] $H = \frac{Q}{T}$ where Q is energy (heat), and T is temperature.

[2] Boltzmann entropy: $H = k \log W$ where k is known as Boltzmann's constant, and W is a measure of the possible states of nature.

[3] Gibbs entropy: $S = -k \sum_{i=1}^{N} p_i \log p_i$ where k is known as Boltzmann's constant, N is the number of states, p_i is the probability of the i^{th} states.

determine and to obtain from time-dependent measurements because KS entropy is hard to achieve convergence in the analysis of noisy signals, especially collected from real-world systems [52]. Though KS entropy is not applicable to short and noisy signals, since its introduction, many studies have attempted to estimate KS entropy for practical use in applied science.

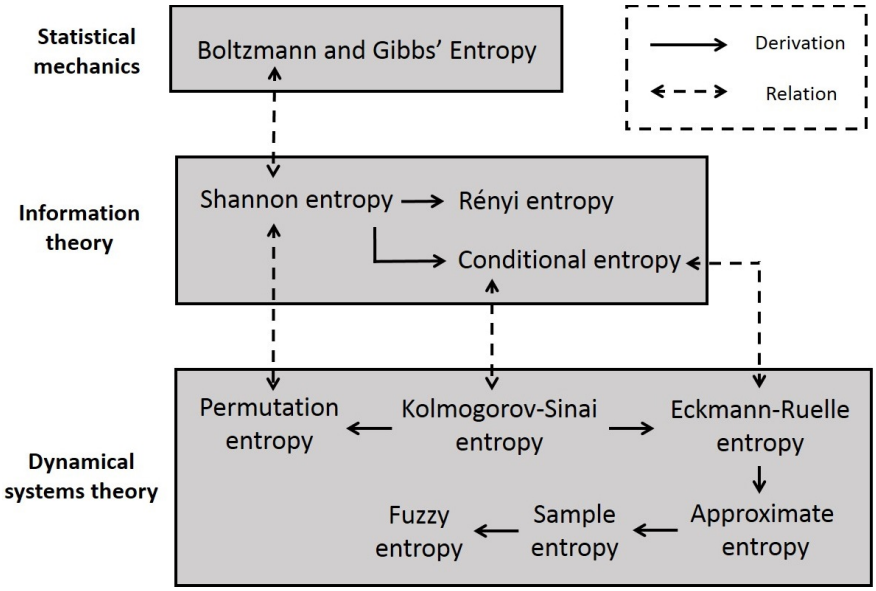


Fig. 2.1 Relations between the various entropy definitions found within the contexts of statistical mechanics, information theory, and dynamical systems (solid line arrows indicate direct mathematical derivations, while dashed arrows show conceptual association).

During the 1980s, several studies attempted to directly compute KS entropy, among which Eckmann-Ruelle entropy[4] [53] exhibits the greatest potential for practical implementation. A nonzero Eckmann-Ruelle entropy value assures the deterministic system is chaotic. Inspired by Eckmann-Ruelle entropy, some other fundamental entropy measures have been developed for time-series complexity analysis. Later, Pincus modified Eckmann-Ruelle entropy for the analysis of finite and noisy time series derived from experiments. For instance, Approximate Entropy (ApEn) was constructed to be thematically similar to KS entropy based on Eckmann-Ruelle entropy [54]. ApEn estimates dynamical changes of time series by quantifying the underlying deterministic or stochastic components. Later, Sample Entropy (SampEn) [55] and Fuzzy Entropy (FuzzyEn) [56] were proposed as improvements of ApEn for entropy estimation. Besides, Permutation Entropy (PerEn) was put forth by Bandt and Pompe to measure symbolic dynamic changes encoded in ordinal patterns underlying in time series [57].

[4]Eckmann-Ruelle entropy approximates the KS entropy as $\lim_{r\to 0}\lim_{m\to\infty}\lim_{N\to\infty}[\Phi^m(r)-\Phi^{m+1}(r)]$, and it is based on the work by Grassberger and Procaccia [52] and Takens

An interpretation of PerEn from an information theory perspective is PerEn quantifies the rate of generation of new permutation patterns in time series. Fig. 2.1 outlines the mathematical and conceptual interrelationships between different entropy definitions. All these complexity metrics are referred to as *single-scale* entropy measures because entropy values are calculated from the raw - one temporal scale - signal. Apart from the above-presented methods, there exist other concepts of entropy measures for time series complexity analysis, such as distribution entropy [58], increment entropy [59], symbolic entropy [60], frequency band entropy [61].

By contrast, *multiple-scale* entropy measures are derived from the above and lies in analyzing a time series from different time scales. The concept of multiple-scale entropy was initially introduced by Costa et al. [41]. A modified entropy definition, named Multiscale Entropy (MSEn), was proposed where entropy values are calculated over a range of scales by using a coarse-graining procedure [62]. Fig. 2.2 illustrates the calculation procedure of *multiple-scale* entropy algorithms.

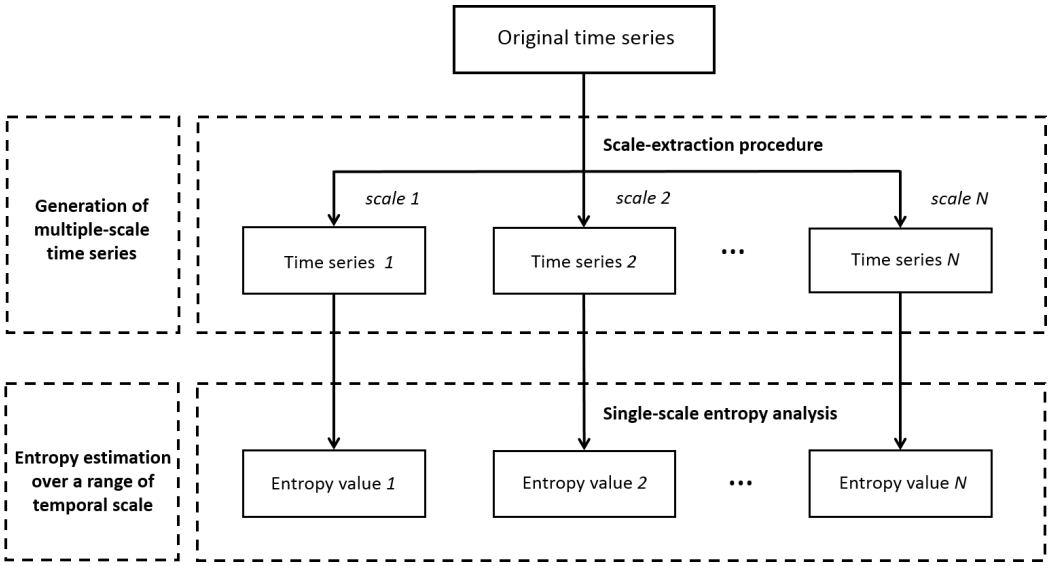


Fig. 2.2 Schematic diagram of the calculation procedure of multiple-scale entropy algorithms.

Since then, many modified scale-extraction procedures were put forth, corresponding to numerous new *multiple-scale* entropy measures [40]. In general, these methods improve from two aspects: one is the improved scale-extraction procedure, and another one is the improved *single-scale* entropy algorithm. Some fundamental *multiple-scale* entropy measures include Multiscale Fuzzy Entropy (MFEn) [63], Multiscale Permutation Entropy (MPEn) [64], Composite Multiscale Entropy (CMSEn) [65], Generalized Multiscale Entropy [66]. In these methods, single-scale entropy measures provide the basis of entropy estimation under the multiple-scale framework [42]. Continuing advancements in sensor networks, computing

systems, and AI techniques have led to a growing number of data-driven fault diagnosis systems. Such diagnostic systems are based on large amounts of sensor data and knowledge mining techniques [26].

2.2 Entropy Measures in Machinery Fault Diagnosis

Various reasons may cause the occurrence of faults and damages to mechanical and electrical components in the machinery. Such damages include fatigue, corrosion, excessive temperature, and lack of lubrication. For example, spalling or pitting are mostly seen in a bearing in its early operation stage. Although extensive research has been conducted on single-point failure models to calculate basic frequency components, the diagnosis of multiple-point failures or compound failures in bearings is still very difficult and challenging [67]. These signals are often non-stationary caused by instantaneous variances due to friction and damping.

Data-driven analysis of system performance has shown that changes in complexity are often linked to machine degradation and failure emergence. Entropy measures are suitable to detect and quantify underlying dynamic changes in system response. These changes in complexity allow for machine condition monitoring, and for distinguishing among various operational regimes. The entropy measures facilitate the usage of machine health condition monitoring in diagnostic applications. With advanced signal analysis and AI techniques, entropy measures have assisted in detecting early faults in machinery and in improving maintenance decision-making for fault diagnosis [31]. With these applications in mind, entropy analysis can be classified into three categories: entropy measure as a feature indicator, entropy criterion for parameter selection, and entropy usage in pattern recognition.

2.2.1 Entropy Measure as a Feature Indicator

In data-driven fault diagnosis, entropy measures are mostly employed as complexity indicators. Different from conventional time-frequency features, entropy measures are popularly used as non-linear feature indicators by directly estimating the complexity degree of time series. Since existing faults often introduce non-linear characteristics in the measurements, changes in the complexity of a system are correlated with its failure rate. Thus, entropy measures are suitable for machine condition monitoring and can detect performance degradation in machinery [68]. A schematic of the entropy-based feature extraction in machinery fault diagnostic systems is presented in Fig. 2.3.

Entropy measures are generally divided into two categories: single-scale entropy measures and multi-scale entropy measures. Specifically, several widely used single-scale entropy

metrics include ShanEn, ApEn, SampEn, FuzzyEn, PerEn, and their variants. Multi-scale entropy measure is a generalized entropy method under a specific scale extraction framework where the single-scale entropy metric is the basis. Their applications are introduced in the following, respectively.

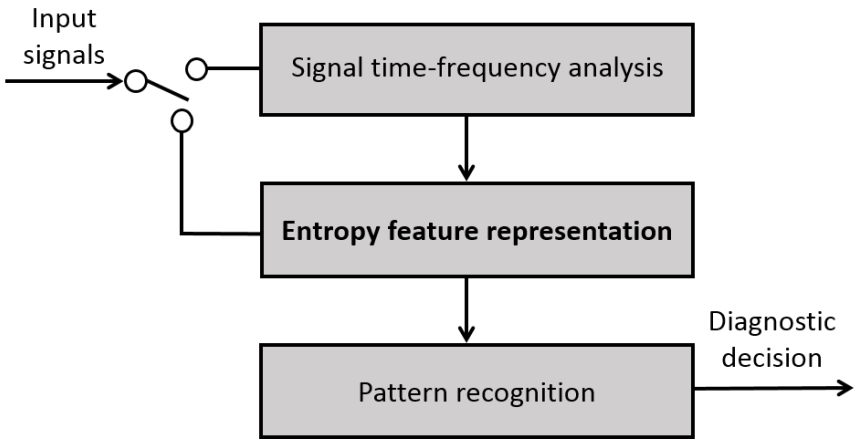


Fig. 2.3 Schematic of the usage of entropy measure as a feature indicator towards machine fault diagnosis.

Shannon Entropy based Feature Extraction

ShanEn was initially proposed to quantify the amount of information content conveyed by messages in information sources. It depicts the degree of surprise or uncertainty that information can give. As uncertainty is always linked to complexity and irregularity, ShanEn has been widely applied to complexity analysis in observations or measurements processed by a system. Various diagnostic systems have examined the applicability of ShanEn to machinery fault diagnosis. For example, Niu et al. [69] proposed a motor diagnosis method based on current signal analysis. ShanEn values are extracted from wavelet coefficients decomposed from current signals via wavelet analysis. Besides ShanEn, other statistical features are used to construct motor diagnostic models based on several machine learning methods. Tran et al. [70] applied ShanEn and statistical features to characterize fault features from vibration and current signals, respectively. Fault features are assembled using extracted features and are then input into Adaptive Neuro-Fuzzy Inference System (ANFIS) for fault pattern identification. Experimental results showed that vibration-based analysis performs better in diagnosing failures in the motor in contrast to current-based analysis. Apart from model diagnosis, ShanEn is also applied to bearing fault diagnosis. For instance, Elforjani et al. [71] conducted a fault diagnosis in slow-speed bearings with natural failures using Acoustic Emission (AE) monitoring. The real run-to-fail experimental results (shown in

Fig. 2.4) reveals that ShanEn indicator is more sensitive and representative to periods of high transient AE measurements than kurtosis and crest factor, especially for fault detection in natural degrading bearings.

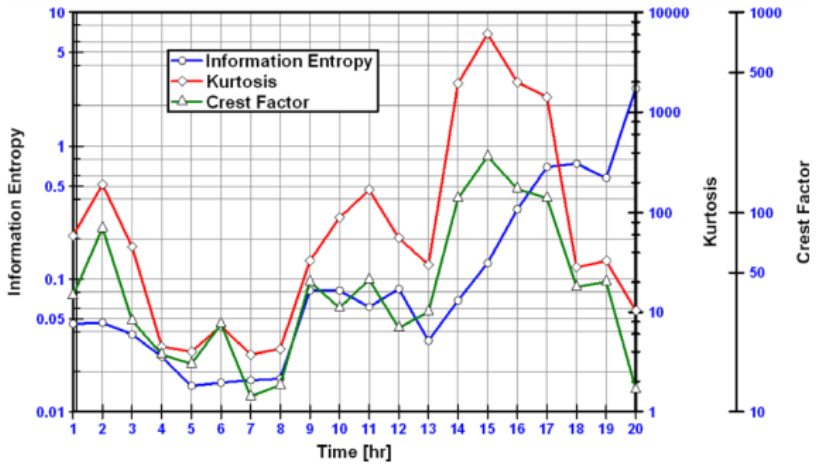


Fig. 2.4 Comparison between ShanEn, kurtosis and crest factor results related to the observation of AE signals collected from a bearing continuously running for 20 hours [71].

Additionally, there exist two generalized concepts of ShanEn based on time- and frequency-domain. That is, energy entropy and spectrum entropy, respectively. When failures exist in the machinery, instantaneous variances caused by faulty components will both change the waveforms and alter the spectrum of vibration signals. Energy entropy and frequency entropy are introduced to capture such changes in the time domain and frequency domain. These two indicators have been widely applied in various diagnosis systems. More specifically, energy entropy enables extracting the uncertainty properties based on the probability distribution of the power energy of signals. For instance, Yu et al. [72] applied energy entropy to capture energy changes in vibration signals for bearing diagnosis. In their study, vibration signals are first decomposed to Intrinsic Mode Functions (IMFs) through empirical mode decomposition. The energy entropy values of transformed components are fed into ANN for fault pattern identification. Similarly, Xie et al. [73] proposed a bearing diagnosis model where energy entropy features of IMFs are combined with several traditional time- and frequency-domain features. Assembled feature vectors are fed into Support Vector Machine (SVM) for fault detection in the bearing. Further, Yuan et al. [74] examined the performance of energy entropy in capturing weak and transient signatures to detect performance degradation in the bearing and gearbox components, respectively.

Different from energy entropy, spectral entropy extracts entropy values from the power spectrum of signals. For instance, Fei et al. [75] applied spectral entropy to motor fault diagnosis based on vibration signals via the Fourier transform. Results demonstrated that

spectral entropy features, combining with the SVM classifier, apply to distinguish between various motor states. Similarly, Jiang et al. [76] applied spectral entropy features and a probabilistic neural network for motor fault diagnosis. Moreover, Ai et al. [77] proposed a feature-level information fusion method using energy entropy and spectral entropy for bearing diagnosis. Various entropy features are extracted from vibration and AE signals. Afterwards, information entropy distances are calculated by fusing these obtained features. The pattern of the unknown fault can thus be determined by comparing the proximity between the known fault type and the unknown fault type. More specifically, an unknown fault's type is more likely to be the type of fault that has the smallest proximity to it.

Approximate Entropy based Feature Extraction

As a suitable complexity measure, ApEn has been widely applied to diagnose structural defects in the machinery. For example, Yan et al. [29] applied a bearing with a natural defect and systematically studied the diagnosis performance of ApEn in bearing health condition monitoring.

A run-to-fail experiment was continually carried out approximately 2.7 million revolutions from an early fault to a severe fault. When reaching 2.34 million revolutions, the defect size has enlarged to 5.5 mm from the initial 0.27 mm (as illustrated in Fig. 2.5). ApEn values are then calculated from bearing vibration signals. The results reveal that bearing degradation at a constant speed will lead to an increase in ApEn value. Similarly, an increase in the ApEn value is often consistent with the increase in the amplitude of the bearing vibration signal after it has a failure, as illustrated in Fig. 2.6. Moreover, the ApEn index shows a certain sensitivity to the speed change in variable speed bearing experiments, providing the possibility to detect failures under variable speeds.

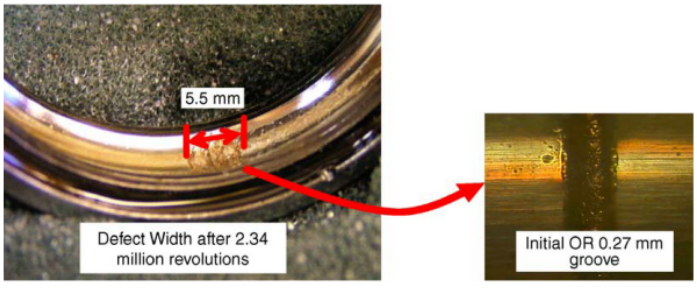


Fig. 2.5 Test bearing with a 0.27 mm outer race fault [29].

Further, He et al. [78] investigated the effectiveness of ApEn in the analysis of AE signals for bearing health monitoring. Their study measured various AE signals by using variable operating conditions, such as the variation of defect severity, rotating speed, and

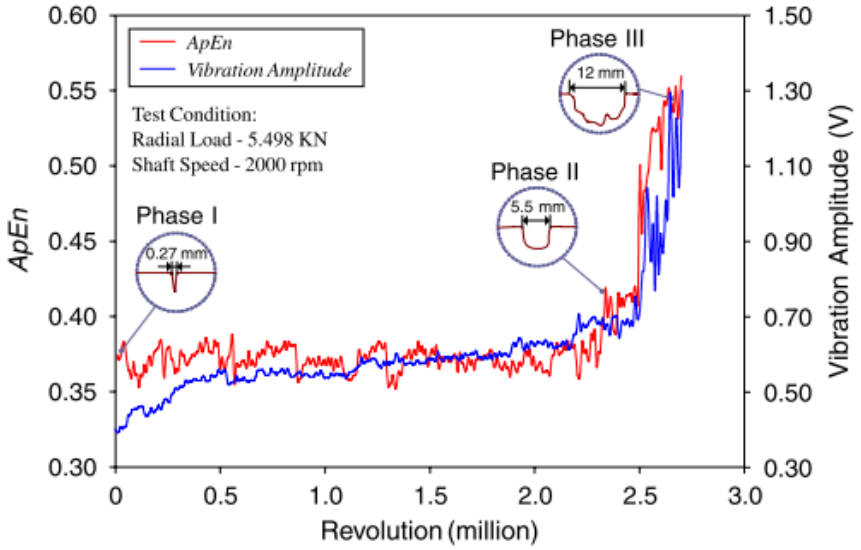


Fig. 2.6 ApEn value and vibration magnitude at various stages of the bearing life cycle test [29].

radial load. Experimental results demonstrated that varying load and speed conditions have a little influence on ApEn values; however, when a failure occurs in the bearing, the ApEn value will significantly increase along with the increasing speed. A larger ApEn value often corresponds to a more serious failure or a degraded bearing component.

Table 2.1 Typical characteristics in ApEn when applied for fault diagnosis of shaft faults [79]

Shaft defects	Typical characteristics of ApEn
Crack only	Peaks at the rotating speeds of $\Omega = \frac{1}{3}\omega_0$ and $\Omega = \frac{1}{2}\omega_0$
Misalignment only	No peaks at the rotating speeds of $\Omega = \frac{1}{3}\omega_0$ and $\Omega = \frac{1}{2}\omega_0$
Crack and misalignment	Single broadband peak in the frequency range of $\Omega = \frac{1}{3}\omega_0$ and $\Omega = \frac{1}{2}\omega_0$

In addition to bearing fault diagnosis, ApEn was also suitable for crack fault detection in the rotating shaft [79]. Vibration signals were collected in the vertical direction of machinery and then analyzed using ApEn measure. The authors studied the effects of different shaft fault - such as unbalance, damping and misalignment - on the calculation of ApEn values. Experimental results verified that if the crack depth is larger than 5% of the shaft diameter, the ApEn algorithm is suitable for detecting crack faults in the rotating shaft [79]. Moreover, certain relationships between various shaft defects and ApEn values are summarized, as described in Table 2.1. Besides, there are other related studies where ApEn was applied as the complexity indicator in machinery fault diagnostic systems, such as in [80–82].

Sample Entropy based Feature Extraction

As a refinement of ApEn, SampEn often exhibits consistent and unbiased entropy values for entropy analysis. SampEn has also widely used as a complexity indicator in machinery fault diagnosis. For example, Zhu et al. [83] investigated the performance of SampEn in distinguishing between bearing health conditions. In their experiments, bearing vibration signals are transformed to IMF envelope, upon which SampEn values are calculated. Then, SampEn values are input into SVM for bearing fault identification. Similarly, Han et al. [84] developed a bearing diagnosis method where SampEn and energy entropy are used to extract features related to complexity change in the system. In addition to bearing fault diagnosis, Kedadouche et al. [85] studied the application of SampEn in gearbox diagnosis. Their study considered gear meshing fault and modeled this fault to generate corresponding simulated vibration signals. The authors first investigate the calculation of SampEn values by changing different parameters, the values of which are shown in Fig. 2.7, where different data length and sampling frequency are considered. From the figure, it can be seen that SampEn has the

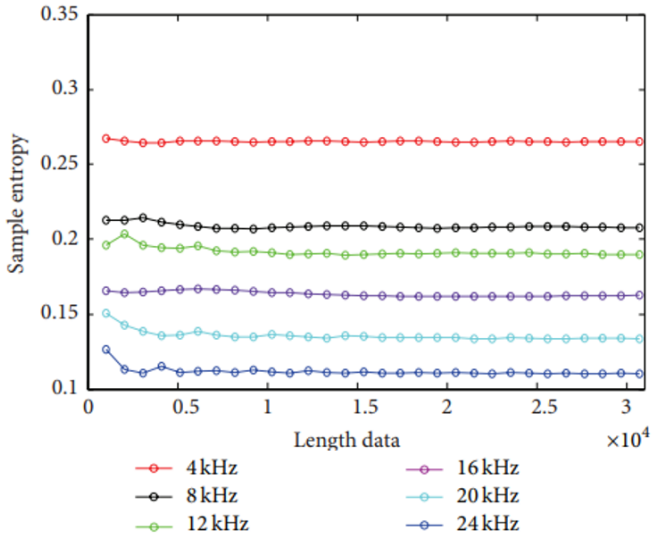


Fig. 2.7 SampEn values calculated from gearbox vibration signals using different data lengths and sampling frequency rates [85].

flexibility in data length selection for entropy analysis. Further, results indicate that, with a faster sampling frequency, its average SampEn value is relatively lower than that of a lower sampling frequency. The authors also pointed out that ApEn and SampEn are very sensitive to the strong noise in the machinery operation [85]. Feng et al. [86] developed a novel method for gearbox diagnosis under non-stationary operating conditions. Their work focuses on analyzing phase angle data extracted from planetary gearbox vibration signals to reduce

the complexity of original signals. SampEn values are then calculated from those phase angle data, and results showed that the SampEn indicator applies to distinguish between gearbox health conditions (such as crack, broken, wear failures), as illustrated in Fig. 2.8.

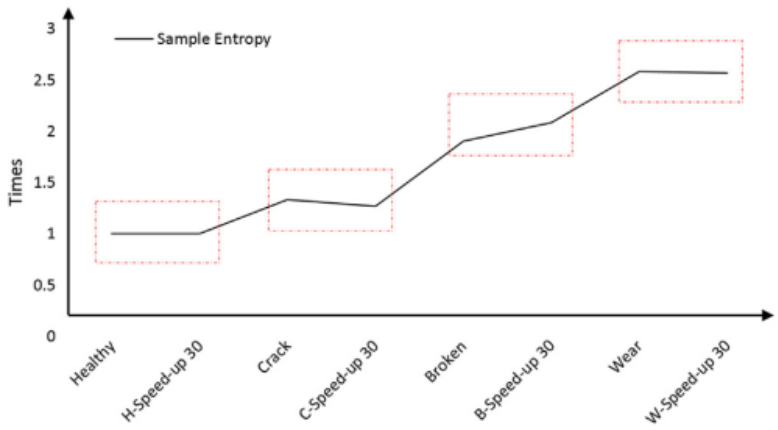


Fig. 2.8 Sample values of phase angle data with various fault scenarios for gearbox diagnosis [86].

Fuzzy Entropy based Feature Extraction

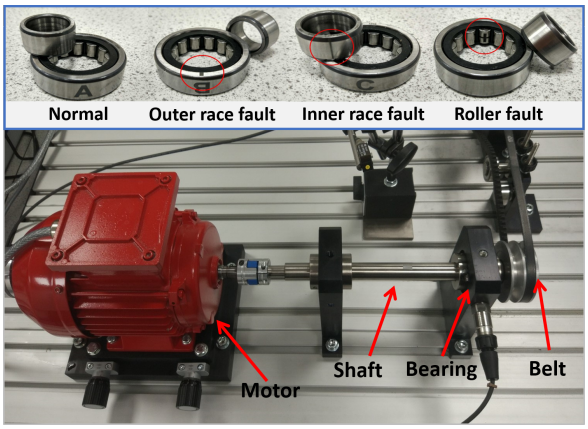
FuzzyEn is a suitable complexity indicator for analyzing signals towards machinery fault diagnosis. With respect to contrasting performance, Xiong et al. [87] applied FuzzyEn, ApEn and SampEn to the analysis of vibration signals. The authors investigated the applicability of FuzzyEn to fault pattern identification and compared its efficiency with ApEn and SampEn as well as their multi-scale extensions. The experimental results showed the improved diagnosis performance of FuzzyEn in analyzing vibration signals comparing with ApEn and SampEn for bearing diagnosis. Zheng et al. [88] presented a bearing diagnosis approach combining FuzzyEn and a signal processing method. Vibration signals are first decomposed into a set of time-frequency components. FuzzyEn values of these components are thus calculated and then input into the ANFIS for fault pattern identification. For gearbox fault detection, Chen et al. [89] proposed a combined method, incorporating the local mean decomposition, FuzzyEn and ANFIS. Similarly, FuzzyEn values of decomposed components are then used to construct the ANFIS-based diagnostic model. To detect faults in the motor bearing, Deng et al. [90] applied empirical wavelet transform and Hilbert transform to generate a series of amplitude modulated-frequency modulated components from vibration signals. FuzzyEn values of transformed components are then applied to indicate the intrinsic oscillation in signals, and a model is constructed using extracted features and the SVM. The results validated the effectiveness of their proposed method in distinguishing between bearing health conditions.

Permutation Entropy based Feature Extraction

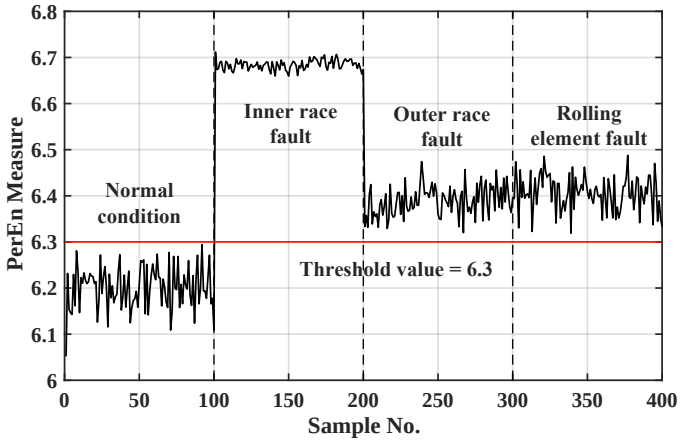
PerEn exhibits prospects in detecting early faults and differentiating between different machine health conditions. One example is the study in Yan [32], where a comparative study was performed on the usage of PerEn in bearing diagnosis. The authors carried out several experiments to investigate the effect of parameter selection on the calculation of PerEn values, such as data length, embedding dimension, time delay, and computational efficiency. A run-to-fail experiment was conducted where an early fault naturally progressed to a severe fault. The results demonstrated that PerEn applies to detect system's complexity change when the underlying failure occurs and grows. Benefiting from the complexity analysis based on PerEn, many works have applied PerEn to bearing signal analysis, where these signals are derived from signal decomposition using advanced time-frequency signal analysis methods. Such works refer to [91–93]. Further, Zhao et al. [94] studied the applicability of PerEn in gearbox diagnosis. PerEn values are calculated from decomposed time-frequency components and then fed into the SVM as fault feature vectors for identifying gearbox health conditions.

Due to its advantage in capturing complexity changes, PerEn has also been employed as a threshold indicator to distinguish between health conditions and fault conditions for bearing diagnosis. Usually, the occurrence of a fault will introduce coupling frequencies and therefore increase the amplitude of the bearing vibration signal. Compared to the normal bearing, vibration signals collected from a bearing with damages will exhibit more complexity degree, thus yielding a larger entropy value. An example of the usage of PerEn for detecting early bearing faults was given in Fig. 2.9. Fig. 2.9(a) shows a PT 500 series bearing test rig benchmark, composed of a motor, a shaft, bearing, and belt drive [31].

Four bearing states are considered, including normal bearing and faulty bearings with damage in the inner race, outer race, and roller element. Vibration data were collected with an operation at speed 2000 r.p.m and with a sampling frequency of 8 kHz. PerEn values are calculated from vibration signals with a data length of 1024. PerEn results are presented in Fig. 2.9(b); it demonstrates that machine faults can lead to higher complexity within the system. Also, entropy indicators apply to performance degradation detection and anomaly detection. More related studies are reported in the literature [95–97]. These studies collectively concluded that when diagnosing bearing failures in a laboratory environment, a threshold of around 0.7 PerEn value empirically selected can be used to distinguish between health and fault bearing states. Moreover, It is worth pointing out that the empirically selected PerEn value can be slightly different when the experimental subject and operation condition change. For instance, the PerEn threshold increases from 0.66 to 0.74 when the radial load increases from 0 to 2hp [98].



(a) PT 500 test rig



(b) PerEn measurement under different conditions.

Fig. 2.9 Comparison of PerEn values between four types of bearing vibration data (with embedding dimension $m = 6$ and time delay $t = 1$). The red line shows a threshold empirically tuned at PerEn $= 6.3$.

Multiple-scale Entropy Feature Extraction

The multiple-scale entropy algorithms often earn more satisfactory diagnosis performance compared to single-scale entropy methods. First, the basis of multiple-scale entropy methods lies in the scale-extraction procedure, which produces a set of temporal time series. Second, these multiple-scale time series contain spatial-temporal structure, thus providing more complexity analysis on hierarchical scales. Many studies have explored their capability as fault indicators for machinery fault diagnosis.

The concept of MSE was initially proposed by Costa et al. [41], which is an extended SampEn under the coarse-graining procedure. It was designed to analyze signals from non-linear systems based on multiple scales. The authors analyzed simulated white and $1/f$

noise signals using MSE and compared their entropy values over several scales (as presented in Fig. 2.10). It can be seen that on the first few scales, white noise time series has a higher

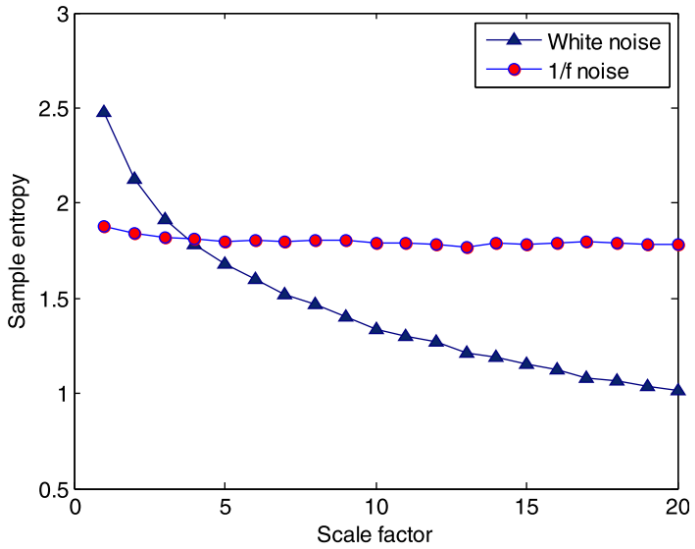


Fig. 2.10 SampEn as a function of the scale factor for the coarse-grained time series of white and 1/f noise [41].

entropy value than the $1/f$ time series, but when the scale becomes larger, it becomes smaller than that of the $1/f$ timer series. This result is consistent with the fact that unlike white noise, $1/f$ noise contains complex structures across multiple scales [41]. Due to its privilege in complexity analysis, many studies have applied MSE to a variety of diagnostic applications in machinery. Zhang et al. [99] investigated the application of MSE to the analysis of numerical noise signals and experimental bearing vibration signals. A diagnostic model is constructed based on MSE values and ANFIS to detect and classify fault patterns. Their results not only proved the effectiveness of the model in bearing diagnosis but also presented the superiority of MSE to bearing diagnosis compared to ApEn and SampEn [99]. Further, Wu et al. [100] also applied MSE to bearing diagnosis where MSE values of envelope signals were calculated based on vibration signatures. Their results proved that the MSE features give satisfactory diagnosis performance. With respect to motor diagnosis, Pan et al. [101] presented a diagnosis method using the MSE and SVM. Eight motor faults are considered in their study, and MSE values are calculated from vibration signals.

Inspired by MSE, many variants of MSE were later developed [42]. For instance, a modified MSE was applied for bearing fault diagnosis [102]. Li et al. [103] proposed an improved MFE algorithm where FuzzyEn is applied for time series complexity analysis, upon which a bearing diagnosis model was built. Zheng et al. [104] proposed the composite MFE algorithm for detecting failures in rolling bearings based on an ensemble SVM classifier.

They found that CMFE has anti-noise calculation, and the required data length is shorter for getting coherent value in comparison with the MFE method. Humeau-Heurtier et al. [105] presented a refined composite MPE to analyze vibration signals for bearing diagnosis, and Li et al. [106] also proposed an adaptive MFE concept for bearing diagnosis.

To analyze multi-channel bearing vibration signals, Zheng et al. [107] analyzed noise signals using a refined composite multivariate MFE method, the results of which were compared with conventional entropy approaches. Experimental results showed that their proposed method has improved for multi-channel time series analysis. Also, FuzzyEn exhibit better diagnosis performance compared to SampEn for bearing health monitoring [107]. More related works using improved multiple-scale methods refer to [108–110].

2.2.2 Entropy Criterion for Parameter Selection

Entropy measures bring up the possibility of specifying desired parameters that characterize time-frequency representations in signal processing techniques. In the machinery, the occurrence of defects in rotating components will excite characteristic amplitudes and frequencies in both the time- and frequency-domain. Usually, signal time-frequency analysis methods are used to transform raw signals into time-frequency representations, and then crucial fault symptoms of interest are characterized with statistical indicators from the obtained components.

Nonetheless, not all components are directly associated with fault symptoms, and some components contain redundant information. For example, the wavelet transform is capable of producing a series of time-frequency representations. By changing time scales, different wavelet components contain fine-grained resolutions of time and frequency information. Though obtained wavelet components carry the time and frequency information simultaneously, not all of them are closely related to fault patterns. Moreover, the redundant information in some components may lower the diagnosis efficiency and accuracy. Thus, the selection of prominent time-frequency components is a necessity. As larger entropy values usually indicate more irregularity, entropy measures can help to select salient components whose complexity degree may increase due to the existence of defects. Moreover, instead of specifying parameters according to prior knowledge, entropy measures facilitate the choice of the optimal parameters. Fig. 2.11 shows a schematic of entropy-criterion for parameter selection in signal time-frequency analysis.

One typical example of such entropy methods for parameter selection is *wavelet analysis*, that has been extensively applied for fault diagnosis – by transforming signals into wavelet coefficients in the time-scale domain. Examples of the studied wavelet analysis methods are Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), and Wavelet

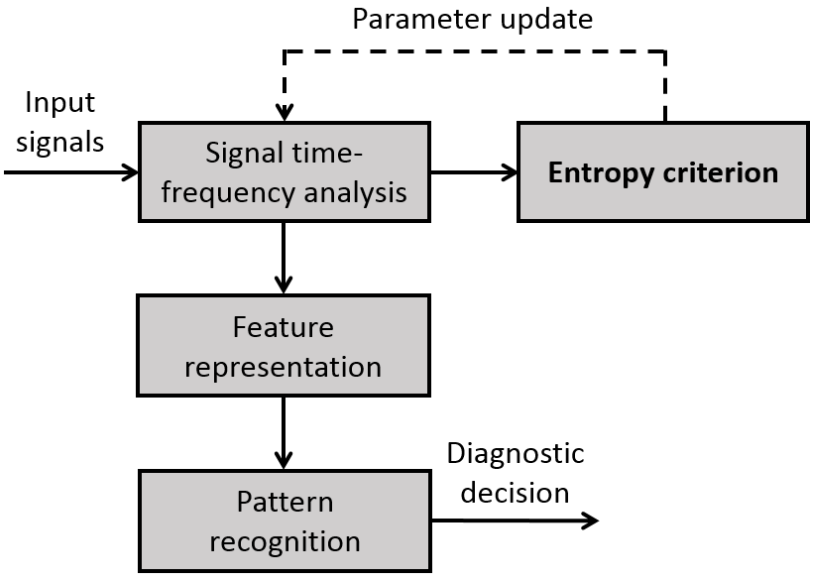
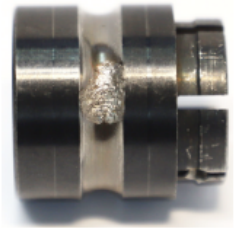


Fig. 2.11 Schematic of entropy-criterion for parameter selection in signal time-frequency analysis for fault diagnosis.

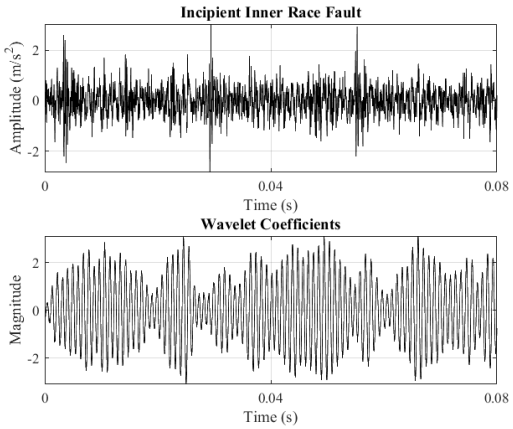
Table 2.2 Entropy-based criteria for optimal parameter selection in wavelet analysis.

Criterion	Description	Application
Minimum-entropy	A node is decomposed if and only if entropy of its two child nodes is no larger than that of their father node.	Optimal tree selection, suitable for DWT and WPT [111]
Minimum ShanEn	Energy content of a few wavelet coefficients is high with the occurrence of characteristic frequency components, resulting in decreased entropy values.	Optimal coefficient selection, suitable for CWT [112]
Maximum energy to ShanEn ratio	Desired wavelet usually extracts maximum amount of energy while minimizing the ShanEn of corresponding wavelet coefficients.	Optimal coefficient selection, suitable for CWT [113]

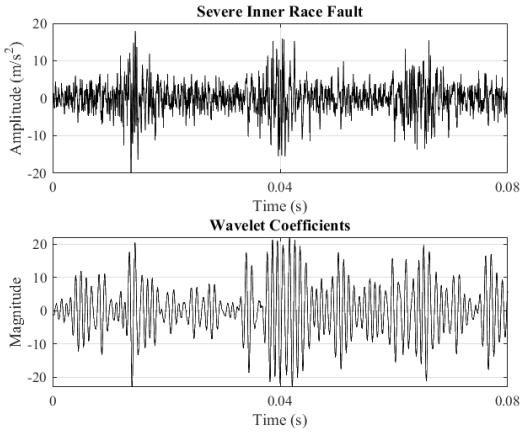
Packet Transform (WPT) [114]. In wavelet analysis, the selection of appropriate mother wavelet and decomposition scale is the key to capture crucial features from signals; however, it usually requires prior knowledge to fine-tune these wavelet parameters for any signal. The most common criteria include minimum ShanEn criterion [112], minimum-entropy criterion [111], and maximum energy to ShanEn ratio criterion [113]. Some related works refer to [115–119]. In these works, wavelet analysis using selected wavelet parameters based on entropy-based criteria are investigated, and its effectiveness is verified in extracting key time-frequency representations in signal analysis. Table 2.2 summarizes the description and applicability of three typical ShanEn-based criteria for wavelet analysis.



(a) Bearing with wear on inner race



(b) Incipient inner race fault



(c) Severe inner race fault

Fig. 2.12 (a) Bearing with inner race fault, (b) raw bearing signal with incipient fault and its wavelet coefficients, and (c) raw bearing signal with severe fault and its wavelet coefficients based on CWT analysis using maximum energy to ShanEn ratio criterion.

An example of CWT analysis for extracting fault features from the fault-deduced transient vibration signals is presented (in Fig. 2.12), where an appropriate mother wavelet is selected using the maximum energy to ShanEn ratio criterion. In this case, a bearing with wear damage on the inner race was studied with vibration data contributed by the Xi'an Jiaotong University [120]. The bearing is tested under 2400 r.p.m and continually operated with 25 hours and 15 mins until the vibration amplitude achieves stopping threshold. Fig. 2.12 (a) shows the tested bearing with inner race fault. For comparison, Fig. 2.12 (b) and Fig. 2.12 (c) present raw signals and transformed signals of bearings with incipient and severe inner race faults, respectively. It is apparent that the bearing signal with severe inner race fault has much greater amplitudes and has more transient components. To better assess fault severity level, the CWT is applied to extract important fault components. To select optimal wavelet kernel, the maximum energy to ShanEn ratio values are calculated to generate appropriate wavelet coefficients. For this purpose, the vibration sensor data is decomposed into 64 sub-signals

using five different mother wavelets: Meyer, Morlet, Mexican, Daubechies 4, and Haar. The results showed that the Morlet wavelet performs best and corresponding coefficients are shown in Fig. 2.12 (b) and Fig. 2.12 (c). The figure suggests that characteristic fault symptoms that are related to the successive periodic pulses, caused by the inner race defect frequency in the bearing. This has shown the possibility of entropy-based criterion in wavelet analysis for machinery fault detection.

In summary, several entropy-based criteria are available for specifying appropriate parameters in multi-resolution signal analysis. Through maximizing the total amount of extracted information, fault detection performance can be enhanced via optimal transformation of raw signals – and the extraction of characteristic fault features.

2.2.3 Entropy Usage in Pattern Recognition

Various entropy-based methods can be employed for pattern classification and model optimization. In pattern recognition, designing reliable and optimized data-driven models [121] is the key to guarantee accurate diagnostic decision-making. As ShanEn evaluates the uncertainty in the variables of a system, based on an empirical probability distribution, it can be used to describe the closeness of two probability distributions - the ground-truth and prediction probability distribution. This is done via a generalization of ShanEn known as cross-entropy. Smaller cross-entropy values indicate that the probability distribution of a model is closer to the empirical distribution in the data. Fig. 2.13 presents a schematic of entropy-based pattern recognition techniques towards machine fault diagnosis.

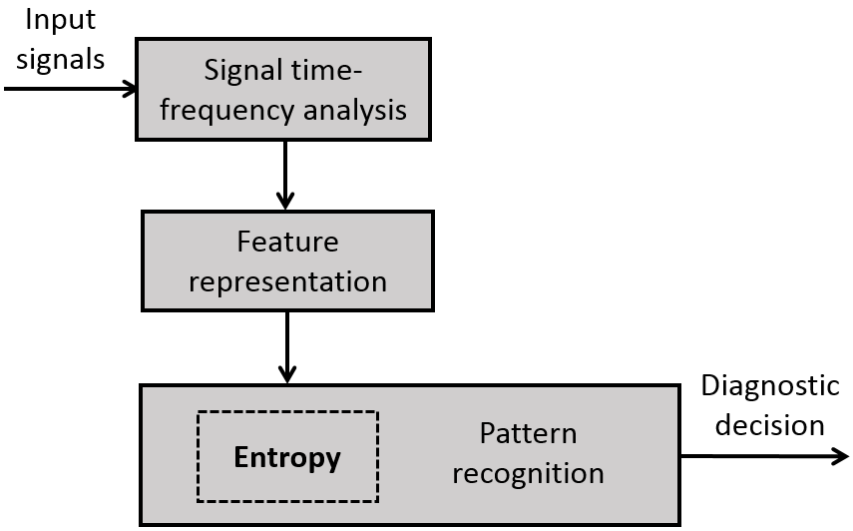
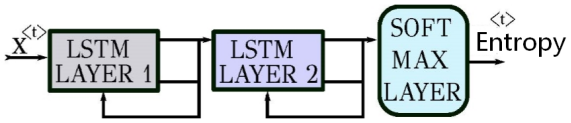
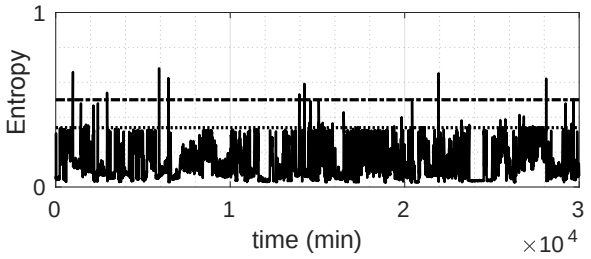


Fig. 2.13 Schematic of entropy measure-based model optimization in pattern recognition for fault diagnosis.

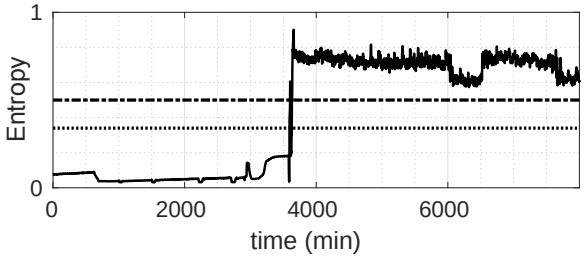
For these reasons, cross-entropy has been the most commonly used loss function that has been commonly used for training and evaluating the performance of artificial neural network classifiers [122]. Regarding probabilistic classification, the estimation of the effectiveness of the acquired models is usually required [123], by which hyper-parameters are fine-tuned through minimizing the cross-entropy over development and a test set not used during the training phase. The cross-entropy function [124] is expected to perform better at improving the efficacy of training models compared with traditional square error objective functions. Related works where cross-entropy is used for the construction of *deep learning models* refer to [125–129].



(a) Deep recurrent entropy adaptive model



(b) Entropy of healthy gas turbine engine



(c) Entropy of gas turbine engine with compressor fault

Fig. 2.14 Illustration of ShanEn in neural networks for industrial gas turbine fault diagnosis. The dashed lines indicate entropy-based thresholds for warning and a faulty system. LSTM: Long Short-Term Memory.

Because ShanEn is always smaller or equal than cross-entropy, minimizing cross-entropy can be understood as estimating ShanEn. The usage of cross-entropy for industrial gas turbine compressor fault diagnosis is explored in [130]. A regressor *recurrent neural network* model was converted into a classifier by bucketing the outputs. An example of this model is presented in Fig. 2.14. The model consists of two *long short-term memory* layers, incorporating a gating

mechanism to control the memory retention operation. The classifier – once trained through a cross-entropy approach – yields ShanEn estimates, indicating the degree of uncertainty in the system. After that, the adaptive entropy model is capable of distinguishing between typical dynamics, corresponding to healthy engines, and anomalous behaviour from faulty engines. Also, it indicated that changes in the uncertainty values often correspond to machine health conditions in industrial gas turbine diagnostic systems.

In summary, entropy measures facilitate optimizing the pattern recognition models - especially for training deep learning neural networks - in data-driven machinery fault diagnostic systems.

2.3 Summary

The concepts of entropy are defined in different contexts - such as dynamical systems and information communication system - and exist in a wide range of research fields. Entropy measures are suitable metrics for time series complexity analysis. They can be broadly divided into single-scale entropy measures and multiple-scale entropy measures. By extending single-scale entropy approaches, multiple-scale entropy measures enable extracting more underlying information from time series under the multiple-scale framework. Many works have studied the the effectiveness of entropy measures in complexity analysis for machinery fault diagnosis. Their potential usages and roles in fault detection and diagnosis are summarized into three categories: entropy measure as a feature or health indicator, entropy criterion for wavelet parameter selection, and the usage of entropy in pattern recognition. These practices are complemented with case studies. The literature has shown that the entropy measures and their extensions are an effective and low-cost method for machine health monitoring and fault diagnosis, requiring little to none domain knowledge.

Chapter 3

The Principles of Entropy Measures

The concept of entropy has been widely used to measure the complexity of a system, be it a natural or a man-made system. Its definition encompasses numerous disciplines, ranging from logic and physics to biology and engineering. Different with traditional statistical indicators (e.g., mean and kurtosis), entropy measures do not reply on linearity. Due to its flexible applicability to the analysis of non-linear complex systems, notions of entropy are defined differently in various contexts - such as information theory and dynamical systems theory. This chapter reviews the definition of fundamental entropy measures for time series complexity analysis and clarifies the relations among them, aiming to arrive at an understanding of these approaches and clarify their relations. Some representative entropy measures are reviewed and introduced, including ShanEn and related concepts, ApEn and its variants, PerEn and its improvements, and multiple-scale entropy measures. Table 3.1 comparatively summarizes their representative characteristics in terms of merits, demerits, as well as algorithmic complexity.

3.1 Shannon Entropy and Related Concepts

ShanEn - devised by Claude Shannon - quantifies the amount of information content conveyed by messages from an information source. It is a suitable indicator that quantifies the uncertainty of a variable. When the random variable is understood as the outcome (observation or measurement) of a system, ShanEn can be interpreted as the rate of generation of new information processed by the system [49]. According to Shannon, information and uncertainty are two sides of the same coin: the reception of a certain amount of information is equivalent to a reduction in uncertainty. It explains that the more we know about what messages a process will produce, the less "surprise", the less uncertainty and the less entropy. Thus, the uncertainty quantified by ShanEn can be understood as missing information from

Table 3.1 Advantages and limitations of entropy measures in time series complexity analysis.

Year	Entropy Measures	Advantages	Limitations	Algorithmic Complexity[1]
1948	ShanEn [49]	• foundational measure to estimate the amount of information content ofa messages from probability viewpoint	• dependence on the probabilistic model of uncertainty as present in a probabilistic event space	$O(n)$
		• foundational measure to estimate the amount of information content of messages from probability viewpoint	• neglect of temporal relationship between values	
1991	ApEn [49]	• applicable to measuring the complexity change of deterministic and chaotic dynamical systems	• lack of consistency relative to SampEn	$O(n)$
		• suitable to medium-sized data	• generation of more similarity than is present	
2000	SampEn [49]	• better consistency relative to ApEn	• discontinuity and mutation at the boundary	$O(n^{\frac{3}{2}})$
		• robustness to small noisy data [131]	• sensitive to parameter selection and data length	
2007	FuzzyEn [56]	• better consistency relative to ApEn and SampEn	• sensitive to parameter selection	$O(n^{\frac{3}{2}})$
		• continuity at the boundary	• membership function needs more physical meaning	
2002	PerEn [57]	• partition naturally derived from ordinal patterns	• amplitude difference in values is neglected	$O(n^{\frac{3}{2}})$
		• invariance with respect to non-linear monotonous transformations	• cases with many equal values are not considered	
2002	Multiple-scale entropy [41]	• better classification accuracy relative to single-scale entropy measure	• efficiency differs depending on applied scale-extraction mechanism and selected single-scale entropy	$O(mn)...O(mn^{\frac{3}{2}})$
		• more robust to small degree of noise	• more time consumption because of computation of entropy measures via a range of scales	

[1] The algorithmic complexity of ApEn, SampEn, and FuzzyEn refers to optimized calculation algorithms in [132, 133]. For multiple-scale entropy measures, their computational efficiency depends on mainly selected scale-extraction mechanism and single-scale entropy method for entropy estimation. Herein, n denotes the input size in units of bits needed to represent the input, and m is the number of scales in multiple-scale entropy methods.

the information viewpoint. The larger the entropy about a system, the more uncertainty about its response, and the more information can be gained by observing the outcomes of the corresponding random variable. The ShanEn value can be obtained as follows:

Giving a random variable X whose probabilities of occurrence are

$$p(X) = \{p(x_1), p(x_2), \cdots, p(x_n)\} \tag{3.1}$$

then, ShanEn is defined as:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \cdot \log_2 p(x_i) \tag{3.2}$$

When the base for the logarithm is selected as 2, $H(X)$ coincides with the average minimum number of bits per outcome yielded by X. $H(X)$ ranges from 0 to $\log_2^n$, and $H(X)$ has the maximum value when all the outcomes are equally probable. It obtains zero when the outcome is certain, indicating that there is no information gain from the outcome.

Based on ShanEn, other related formulations were put forth in information theory. Preeminent examples are conditional entropy, mutual information, and cross-entropy. Conditional

entropy can be expressed as $H(X|Y) = H(X,Y) - H(X)$, where $H(X,Y)$ is the entropy of the joint probability distribution $P(X,Y)$. It measures the missing information and uncertainty about X upon observing another measurement of Y. Mutual information is defined as $I(X,Y) = H(X) - H(X|Y)$, and it captures the amount of information that two variables X and Y share [134]. Moreover, cross-entropy is expressed as $H(p,q) = -\sum_x p(x)\log_2 q(x)$ where $p(x)$ and $q(x)$ are typically the ground-truth and estimated probability distributions, respectively. Cross-entropy minimization has been popularly used in optimization algorithms, such as model optimization in neural networks. Also, it has been proved that ShanEn is no larger than cross-entropy[1].

In addition to information theory, ShanEn occupies center stage in complexity and chaos theory. Entropy is often linked to the degree of chaos in an observed dynamical system because uncertainty can be explained as unpredictability or irregularity in a system. In dynamical systems theory, KS entropy is a crucial concept, which is a generalization of ShanEn employed in the study of seemingly random but deterministic dynamical systems (i.e., *deterministic chaotic systems*) [135]. KS entropy analyzes how the uncertainty about a system evolves from its dynamical equations. That is, it yields the rate of generation of new information by the examined system. From an information-theoretic standpoint, chaotic behaviors are described by KS entropy through a partition of the state space [136]; thus, it is equally suitable for discrete and for continuous dynamical systems. Positive values of KS entropy are interpreted as an increase in uncertainty with respect to the system's responses [136]. Hence, systems with positive KS entropy can be regarded as chaotic systems – displaying sensitive dependence on the initial conditions [137].

In the study of non-linear dynamical systems, the Lyapunov exponent is a crucial indicator defined for quantifying the topological characteristics of the dynamics and system stability. Pesin's theorem establishes a relationship between the KS entropy and Lyapunov exponent [138]. Nevertheless, when performing numerical analysis by way of experimental data, it is usually very hard to calculate Lyapunov exponent and KS entropy directly. Added difficulty results from the fact that KS entropy relies on arbitrarily fine partitions of the state space, and from its lack of robustness to noisy measurements. Thus, typically, a large amount of measured data is required to achieve convergence [54]. In this case, numerous works concentrate on entropy development and alternative formulations attempting to estimate time-varying dynamic changes within a system, such as works by Grassberger and Procaccia [52] and Eckmann and Ruelle [139]. Thus, many entropy measure analysis methods populate the literature, which are described in this section.

[1] $H(p) = -\sum_x p(x)\log_2 p(x) \leq -\sum_x p(x)\log_2 q(x) = H(p,q)$.

Rényi entropy is a generalization of ShanEn, defined as

$$H_\alpha(p) = \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{n} p_i^\alpha\right), \tag{3.3}$$

where $\alpha \in [0,\infty)$ and $\alpha \neq 1$. Eq. 3.3 becomes ShanEn when $\alpha \to 1$.

Rényi entropy is characterized as a continuous family of entropy measures (H_α) by way of a *bias* parameter α [140]; α controls the degree of sensitivity of $H_\alpha(p)$ towards particular probability distribution functions [140] and makes $H_\alpha(p)$ non-negative for all α. Fig. 3.1 presents the probability distribution estimation of Rényi entropy for an mutually exclusive event. Other special cases of Rényi entropy include collision entropy ($\alpha = 2$) and

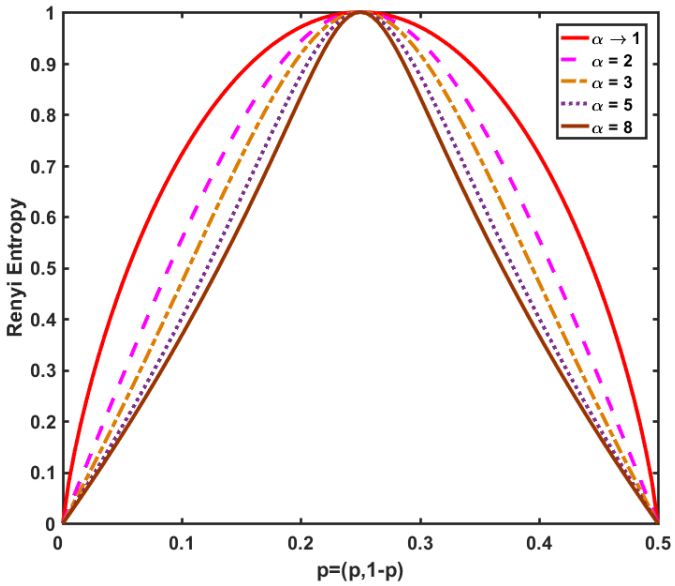


Fig. 3.1 Plot of the Rényi entropy for several positive values of parameter α. An increasing positive α value implies more sensitive to events that occur frequently.

min-entropy ($\alpha \to \infty$). Collision entropy is the negative logarithm of the probability that two independent and identically distributed random variables present the same outcome or *collide*. More likely events are more probable to collide, thus are more conspicuous under the collision entropy measure than with ShanEn.

As $\alpha \to \infty$, Rényi entropy is increasingly determined by the events of highest probability; thus, min-entropy is the negative logarithm of the probability of the most likely outcome only.

3.2 Approximate Entropy and Its Variants

ApEn, developed by Pincus, is a useful complexity indicator for the analysis of non-stationary time series [54]. It quantifies dynamic changes underlying in data and estimates its degree of irregularity, such as for the analysis of irregularity of vibration signals in the machinery [29]. ApEn is suitable for the analysis of data that has small degree of noise and short length. In [29], experimental results showed that it is applicable to discriminating between different machinery system states. It shows the prospect in identifying fault patterns in machinery fault diagnosis. The ApEn is calculated as follows:

1. Given a time series X of data length N, construct a set of m-dimensional vectors $\{\mathbf{X}_m(i), i = 1,2,\cdots,N-m+1\}$:

$$\mathbf{X}_m(i) = x_i, x_{(i+1)}, \cdots, x_{(i+m-1)} \tag{3.4}$$

2. Let $d_{i,j}^m$ represent the distance between the vector $\mathbf{X}_m(i)$ and $\mathbf{X}_m(j)$

$$d_{i,j}^m = \max_{k=1,2,\cdots,m} (|x(i+k-1) - x(j+k-1)|) \tag{3.5}$$

   where i and j range from $1,2,\cdots,N-m+1$, respectively, and N is the number of data points in time series.

3. Define $C_i^m(r)$ is the probability that any vector $\mathbf{X}_m(j)$ is within r of $\mathbf{X}_m(i)$. The vector $\mathbf{X}_m(i)$ is called the template, and an instance where a vector $\mathbf{X}_m(j)$ is within r of it is called a template match. For each vector $\mathbf{X}_m(i)$, $C_i^m(r)$ can be computed as:

$$C_i^m(r) = \frac{1}{N-m+1} \sum_{j=1}^{N-m+1} \theta\left\{r - d_{i,j}^m\right\} \tag{3.6}$$

   where m is the embedding dimension, r is the predetermined threshold, and Θ is the Heaviside function.

$$\theta\{x\} = \begin{cases} 1, \text{if } x \geq 0 \\ 0, \text{if } x < 0 \end{cases} \tag{3.7}$$

   Then, by defining

$$\Phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \log_2 C_i^m(r) \tag{3.8}$$

   where $\Phi^m(r)$ is the average of the logarithms of the functions $C_i^m(r)$.

4. For a finite time series, ApEn is defined:

$$ApEn(m,r,N) = [\Phi^m(r) - \Phi^{m+1}(r)] \tag{3.9}$$

There is a connection between ApEn and ShanEn. ApEn can be shown to be closely related to the notion of conditional entropy. According to the definition, we have

$$\begin{aligned} ApEn &= \Phi^m(r) - \Phi^{m+1}(r) \\ &= -E(\log_2(C^{m+1}(r))) - (-E(\log_2(C^m(r)))) \\ &= H(X_1,\cdots,X_{m+1}) - H(X_1,\cdots,X_m) \\ &= H(X_{m+1}|X_1,\cdots,X_m), \text{with} \, r < min(|a_j - a_k|), \\ &\text{where} \, j \neq k, \, a_j \, \text{and} \, a_k \, \text{are state space values.} \end{aligned} \tag{3.10}$$

Thus, ApEn estimates the uncertainty with respect to future observations of a time series, given the knowledge of the past observations. It is proposed that, when the behavior of the process generating the time series becomes irregular – or chaotic, ApEn increases – although a nonzero ApEn value does not certify that the dynamics are chaotic [141].

Several hyperparameters must be fine-tuned for optimal performance (such as the *embedding dimension m* and the *tolerance r*) – although empirical values are offered in the literature; when $m = 2$, values of r ranging between 0.1 to 0.25 times the standard deviation (σ) of time series can produce reasonable results [54]. For the analysis of rotating machinery, the values $m = 2$ and $r = 0.4\,\sigma$ have been suggested [29]. In the same publication, it is claimed that $N = 750 - 5000$ is sufficient for achieving consistent results. Lu et al. [142] have developed an automatic r selection approach that can reduce the computational cost while fitting the hyperparameter r, and Kaffashi et al. [143] have investigated the influence of hyperparameter selection on analyzing real-time series with ApEn.

Several modified ApEn algorithms have been proposed with alleged improved performance. One example is Cross-ApEn, also developed by Pincus, that measures the statistical independence of two concurrent time series, by capturing both spatial and temporal irregularity [144]. Another example is SampEn.

Sample Entropy

Richman and Moorman proposed SampEn as a refinement of ApEn [55]. They found the difficulties in ApEn analysis are attributed to self-matches and undefined matches resulting in undefined probabilities [145]. Therefore, SampEn refines the ApEn algorithm from two differing aspects: *i)* SampEn excludes self-matches while counting template matches

(Eq. 3.13); *ii)* in SampEn only the first $N-m$ vectors are considered (Eq. 3.14) – this ensures that for $1 \leq i \leq N-m$ both $x_k^m(i)$ and $x_k^m(j)$ are defined. SampEn is defined as follows:

1. Given a time series X of data length N, construct a set of m-dimensional vectors $\{\mathbf{X}_m(i), i = 1,2,\cdots,N-m+1\}$:

$$\mathbf{X}_m(i) = x_i, x_{(i+1)}, \cdots, x_{(i+m-1)} \tag{3.11}$$

2. Let $d_{i,j}^m$ represent the distance between the vector $\mathbf{X}_m(i)$ and $\mathbf{X}_m(j)$, and compute $d_{i,j}^m$:

$$d_{i,j}^m = \max_{k=1,2,\cdots,m} (|x(i+k-1) - x(j+k-1)|) \tag{3.12}$$

   where $i = 1,2,\cdots,N-m+1$, $j = 1,2,\cdots,N-m$. Particularly, to reduce bias, $j \neq i$ is applied to exclude self-matches.

3. Similarly, for each $\mathbf{X}_m(i)$ and a predetermined tolerance r, define $C_i^m(r)$ as

$$C_i^m(r) = \frac{1}{N-m-1} \sum_{j=1, i\neq j}^{N-m} \theta\{r - d_{i,j}^m\} \tag{3.13}$$

4. θ is the Heaviside function. Define the average of the $C_i^m(r)$ as

$$\Phi^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} C_i^m(r) \tag{3.14}$$

   where $C^m(r)$ is the probability that two sequences will match for m points.

5. For a finite time series, SampEn is defied:

$$SampEn(m,r,N) = -\log_2 \left[\frac{\Phi^{m+1}(r)}{\Phi^m(r)} \right] \tag{3.15}$$

Consequently, in the computation of SampEn, unlike that of ApEn, the logarithm is applied after Φ^m is obtained. Because the quantities $C_{ij}^m(r)$ act as surrogates of the probabilities $p(x_i)$ in (3.2), ApEn is closer to the mathematical formulation of the original entropy. Nonetheless, it has been verified that SampEn reduces bias and maintains relative consistency as compared to ApEn [131]. That is, if a time series A arising from a more ordered system than time series B, then ApEn of A has been shown to be smaller than ApEn of B for all conditions tested [141]. As an example, Yentes et al. comparatively investigated the performance of ApEn and SampEn in time series analysis. They found that SampEn is less

sensitive to the change of data length and shows better performance compared to ApEn when analyzing clinical data sets in pathological populations [146, 147].

There exist enhanced formulations of SampEn algorithm, reducing its algorithmic complexity. For example, Lu et al. [148] presented a method to accelerate the computation of ApEn and SampEn by exploiting vector dissimilarity. This method omits the computation of distances between the most dissimilar vectors, which further reduces the time complexity. Besides, Manis et al. [149] proposed three SampEn algorithms that yield identical values but are less expensive computationally speaking (by avoiding the similarity check between points in m dimensional phase space). Some works extend the applicability of SampEn estimation from one-dimentional time series analysis to two-dimensional time series analysis. For example, Silva et al. [150] applied SampEn to two-dimensional image data analysis. SampEn characterizes the irregularity of pixel patterns and texture features from image data.

A potential limitation of ApEn and SampEn resides in Eq. 3.13: the method to select template matches consists in establishing a crisp boundary. That is, the Heaviside function [2] determines two vectors are similar only when their similarity falls within a specific boundary. This vector similarity determination method is not often workable, especially when the boundary is hard to determine. Moreover, it was found that SampEn might be discontinuous and could rise or fall when the tolerance r slightly changes on account of the Heaviside function [151]. To address this shortcoming, the concept of fuzzy sets was introduced to improve the reliability of entropy analysis, which is discussed in the following.

Fuzzy Entropy

Chen et al. presented the notion of FuzzyEn for time series complexity analysis [56]. In FuzzyEn, the concept of degree of fuzzy membership, inherited from the framework of fuzzy logic, was introduced to the template matching. As a refinement of SampEn, Fuzzy is define:

1. Given a time series X of data length N, construct a set of m-dimensional vectors $\{\mathbf{X}_m(i), i = 1,2,\cdots,N-m+1\}$:

$$\mathbf{X}_m(i) = x_i, x_{(i+1)}, \cdots, x_{(i+m-1)} - u0(i) \tag{3.16}$$

   where $\mathbf{X}_m(i)$ has m consecutive data points, commencing with the ith points and generalized by removing a baseline

[2] $\theta(x) = 1, \text{if} x \geq 0; \theta(x) = 0, \text{if} x < 0.$

$$u0(i) = \frac{1}{m}\sum_{j=0}^{m-1} x_{(i+j)} \tag{3.17}$$

2. For each $\mathbf{X}_m(i)$, define the distance $d_{i,j}^m$ between $\mathbf{X}_m(i)$ and $\mathbf{X}_m(j)$ as the maximum absolute difference of the corresponding scalar components

$$\begin{aligned} d_{i,j}^m = \max_{k=1,2,\cdots,m} (|(x(i+k-1) - u0(i)) \\ - (x(j+k-1) - u0(j))|) \end{aligned} \tag{3.18}$$

3. Calculation of $C_i^m(n,r)$ as $(N-m-1)^{-1}$ times the sum of similarity degree between $\mathbf{X}_m(i)$ and $\mathbf{X}_m(j)$ where $j \neq i$.

$$C_i^m(n,r) = \frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m} \mu(d_{i,j}^m, n, r) \tag{3.19}$$

where $\mu(x)$ is the exponential function

$$\mu(d_{i,j}^m, n, r) = \exp(-(d_{i,j}^m)^n / r) \tag{3.20}$$

4. Define the average of $C_i^m(n,r)$ as

$$\Phi^m(n,r) = \frac{1}{N-m}\sum_{i=1}^{N-m} C_i^m(n,r) \tag{3.21}$$

5. For a finite time series, the FuzzyEn is then defied as follows:

$$FuzzyEn(m,n,r,N) = -\log_2\left[\frac{\Phi^{m+1}(n,r)}{\Phi^m(n,r)}\right] \tag{3.22}$$

The fuzzy boundary is the main feature of the fuzzy set, so the continuous membership value - an scalar value between $[0,1]$ - is assigned by the fuzzy membership function. The difference in similarity measurement using two different membership methods are presented in Fig. 3.2.

FuzzyEn, initially proposed in [56], employs the membership function $\exp(-d^n/r)$, where r and n control the width and gradient of the boundary respectively, and d is the maximum absolute difference of the corresponding scalar components according to Eq. 3.19. Other membership functions have been considered in the literature. For example, two exponential membership functions were developed - $\exp(-d^{\ln(\ln 2^c)/lnr}/c)$ in [87] and $\exp(-\ln 2(d/r)^n)$

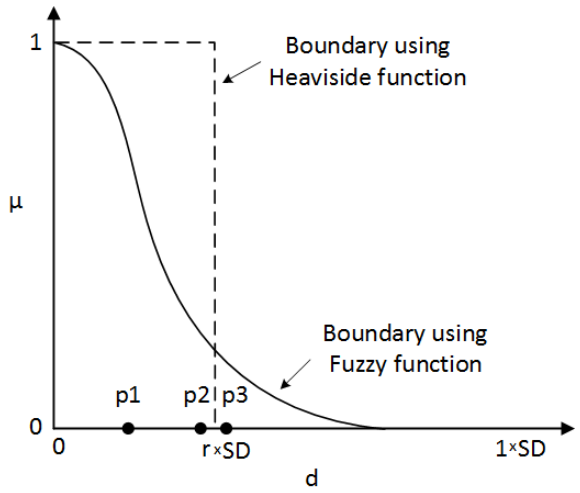


Fig. 3.2 Heaviside function (dotted line) in ApEn and SampEn estimation and fuzzy function (solid line) in FuzzyEn for similarity calculation. As can be seen, both points p1 and p2 locate within the boundary (tolerance threshold) by Heaviside function; nevertheless, the point p3 is considered dissimilar, though the p3 is very close to p2. Hence, the Heaviside function might be discontinuous due to a slight change of r using a binary decision. Comparatively, the width of the boundary in FuzzyEn is r multiply standard deviation (SD), which provides a continuous similarity estimation and greatly alleviate this issue.

in [152]. The authors proposed that assigning a value of 0.5 to similarity degree will gain more physical meaning when Heaviside boundary and fuzzy boundary intersect. Moreover, other modified FuzzyEn approaches have been developed for improved performance: a piecewise fuzzy membership function proposed in [153] and a modified FuzzyEn, which operates by increasing the number of samples during the computation of the entropy [154].

In summary, the notions of ApEn, SampEn and FuzzyEn measures are closely related to each other, and they individually contain their characteristics. Some works compared the performance of ApEn, SampEn and FuzzyEn for time series complexity analysis [155, 156], such as the consistency, dependency on parameter choice, and robustness to noise; results showed that FuzzyEn offers better consistency and has less dependence on the size of data on account of its continuous membership function, especially when analyzing real-world time-series data [156].

3.3 Permutation Entropy

PerEn, developed by Bandit and Pompe [57], measures the complexity of time series by estimating dynamic changes encoded in the ordinal pattern of a time series. Mathematically,

PerEn is ShanEn over the empirical probability distribution of the ordinal patterns naturally originated from the time series data. PerEn is defined:

1. Given a time series X of length N, the time delay λ and the embedding dimension m, the phase space of a time series can be reconstructed as:

$$\mathbf{X}_i = \{x(i), x(i+\lambda), \cdots, x(i+(m-1)\lambda)\} \tag{3.23}$$

where $1 \leq i \leq N-(m-1)\lambda$. Then, the m number of real values contained in each $\mathbf{X}_i$ can be rearranged in an increasing order as

$$x(i+(j_1-1)\lambda) \leq x(i+(j_2-1)\lambda) \leq \cdots \leq x(i+(j_m-1)\lambda) \tag{3.24}$$

2. Therefore, any vector $\mathbf{X}_i$ can be mapped onto a group of symbols as

$$\pi_n = (j_1, j_2, \cdots, j_m) \tag{3.25}$$

where π_n is one of the $m!$ symbol permutations having m distinct symbols and $n = 1, 2, \cdots, k$, $k \leq m!$ ($m!$ is the largest number of distinct symbols). Suppose $P(\pi_1), P(\pi_2), \cdots, P(\pi_k)$ denote the probability distribution of each symbol sequences respectively, and $\sum_{n=1}^{k} P(\pi_n) = 1$.

3. For each permutation π_n, the relative probability distribution can be determined by:

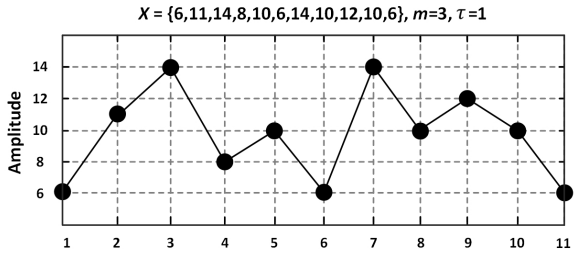
$$P(\pi_n) = \frac{\text{Number}\{\mathbf{X}_i \text{ has type } \pi_n \mid 1 \leq i \leq N-(m-1)\lambda\}}{N-(m-1)\lambda} \tag{3.26}$$

4. Then, the permutation entropy of order m is defined as [157]:

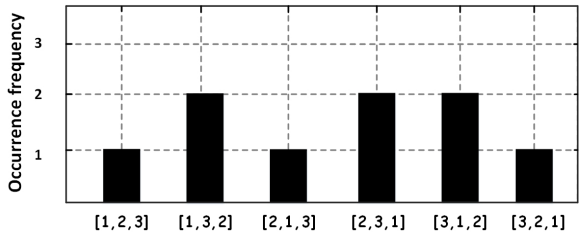
$$PE = -\sum_{j=1}^{m!} P(\pi_j) \ln(P(\pi_j)) \tag{3.27}$$

Accordingly, PerEn can be interpreted as a measure of the rate at which new permutation patterns are produced in the process of a system. The PE value ranges between $[0, \log_2 m!]$. The larger the PE value is, the more irregular the time series is. The minimum value indicates that the time series is most likely a periodic signal.

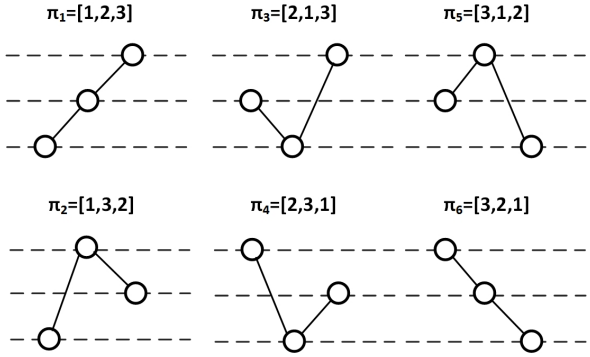
In PerEn, partitions are devised from the order structure - symbolic sequences of adjacent values in ascending order, rather than apportioning amplitudes according to different levels



(a) Plot of time series X



(b) Occurrence frequency of each ordinal pattern when $m = 3, \tau = 1$



(c) Diagram of possible ordinal patterns when $m = 3$

Fig. 3.3 A schematic illustration of counting ordinal patterns in a time series when embedding dimension $m = 3$ and time delay $\tau = 1$. The ordinal patterns are obtained by ordering neighboring values in an ascending order. For this case, the possible order permutation of a series, $\pi_n\,(1 \leq n \leq 6)$, is one of the subset in $\Omega = \{[\mathbf{123}], [\mathbf{132}], [\mathbf{213}], [\mathbf{231}], [\mathbf{312}], [\mathbf{321}]\}$.

in ApEn [158]. For instance, given two series, $\{x_\alpha, x_\beta, x_\gamma\}$ and $\{x_\gamma, x_\beta, x_\alpha\}$, their symbolic sequences, $[\alpha, \beta, \gamma]$ and $[\gamma, \beta, \alpha]$ have the possibility of revealing different temporal relationships in the symbolic dynamical system. A schematic illustration of possible ordinal patterns in a time series is presented in Fig. 3.3.

In the analysis of dynamical systems, PerEn is related KS entropy, when the partition is defined based on the order of a time series. More specifically, in PerEn, permutation patterns (i.e., the partitions) result from a map, by translating into a sequence of symbols. In addition,

PerEn provides an upper bound for KS entropy when $m \to \infty$ [159] and is also related to the Lyapunov exponents of a dynamical system [57].

While PerEn is a suitable complexity indicator, it has a few limitations in time series complexity analysis. The main shortcoming is due to it only takes into account the order of neighboring elements without considering the difference in amplitudes [160, 158]. As a result, different time series may have the same PerEn value, thus lowering its performance in distinguishing between measurements that correspond to different system responses. Also, when repeated values emerge in the sensor data, PerEn assigns their sequential order according to emergence order. This results in ambiguity in the mapping from sensor data to permutations, and may introduce bias in the empirical distribution estimates. Typically repeated values are rare, but this is not the case in quasi-stationary systems or systems in an stationary operational regime [161].

To overcome these limitations, a number of improved and/or alternate formulations of PerEn have been proposed. Some of them take into account the amplitude difference – by using weighting coefficients such that the magnitudes of neighboring elements have different contribution to the relative frequencies of the permutation types. For instance, Liu and Wang presented a fine-grained PerEn [162] by adding an extra factor representing the difference in magnitudes of values in the order patterns. Fadlallah et al. introduced a modified PerEn, which weights the relative frequency of each ordinal pattern with the variance of these corresponding (phase-space) series [163]. Keller et al. proposed a robust PerEn based on counting robust ordinal patterns [164]. Azami et al. proposed an amplitude-aware PerEn which takes both the average value and differences among neighboring values into account [165]. Further, in order to tackle the problem of repeated measurements mentioned above, Bian et al. presented a solution by mapping the repeated values onto the same symbol [166].

3.4 Multiple-scale Entropy Measures

Multiple-scale entropy measures are generalized entropy methods based on scale-extraction mechanisms. By using scale-extraction framework, multiple-scale entropy estimates the complexity or irregularity over a range of temporal scales of time series. These different scales correspond to different components that contain either fine-grained or coarse-grained information, which is useful to capture the instantaneous variations in vibration analysis. For instance, the coarse-grained components usually carry useful low-frequency information in vibration signals, reflecting intrinsic frequency components. Multiple-scale entropy values can be obtained by applying single-scale entropy analysis on multiple temporal scales. In

general, multiple-scale entropy calculation consists two procedures: *i)* extract multiple-scale time series of different scales from the original time series; *ii)* calculate entropy values on obtained multiple-scale time series via a specified single-scale entropy method.

The basic idea of multiple-scale entropy methods was initially introduced by Costa [41]. A modified SampEn, named MSE, was proposed based on the coarse-graining procedure:

1. Given a time series $\{x_i, i = 1,2,3,\cdots,N\}$ of length N and a scaling factor τ, the coarse-grained time series, $\mathbf{y}^{(\tau)}$, is obtained by the relation

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad \text{for } 1 \leq j \leq \frac{N}{\tau}, \quad N > \tau. \tag{3.28}$$

2. When $\tau = 1$, $\mathbf{y}^{(\tau)}$ coincides with the original time series $\mathbf{x}$. From $\mathbf{y}^{(\tau)}$, MSE is defined as:

$$\text{MSE}(x, \tau, m, r) = \text{SampEn}(\mathbf{y}^{(\tau)}, m, r). \tag{3.29}$$

In the MSE algorithm, the raw time series is first separated into non-overlapping windows of length τ, and then a coarse-graining procedure - an averaging procedure - is used to generate time series of different scales. The diagram of coarse-graining procedure is presented in Fig. 3.4.

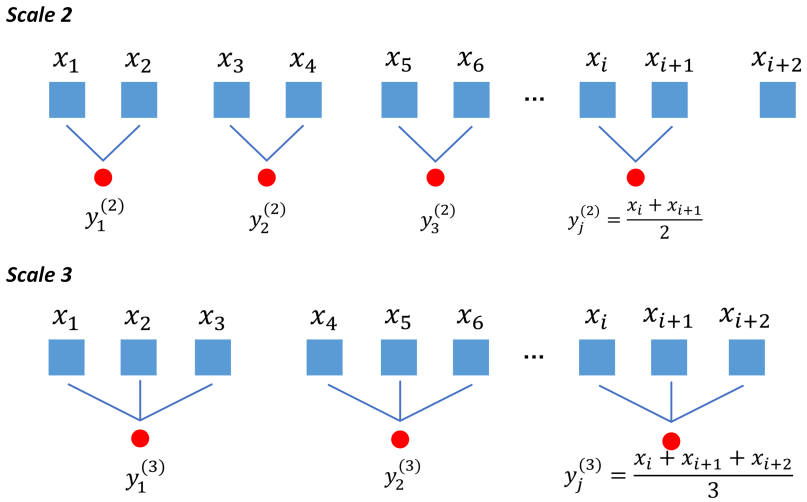


Fig. 3.4 Illustration of the coarse-graining procedure at the 2nd and 3rd scale in the MPE.

Different from SampEn, MSE can extract a set of entropy values via a range of temporal scales; thus, more information associated with the complexity of observations can be characterized [62]. Later, many variants of multiple-scale entropy algorithms were proposed

where various single-scale entropy algorithms are used for entropy estimation. It is noted that the single-scale entropy analysis provides the basis of multiple-scale entropy estimation. When calculating multiple-scale entropy values, the use of improved single-scale entropy could enhance the entropy analysis of time series. For instance, based on the coarse-graining procedure, some modified methods were developed, including multiscale approximate entropy [167], Multiscale Fuzzy Entropy (MFE) [152] and Multiscale Permutation Entropy (MPE) [168]. It reported that the MFE and MPE can achieve better entropy analysis in comparison with MSE in vibration signal complexity analysis [167]. There are many other variants of entropy methods based on the coarse-grained procedure [42].

Later, it reported that MSE has a reduced reliability of SampEn values as a time scale factor increases [65]. To improve MSE, another commonly used method, termed Composite Multiscale Entropy (CMSE) [65], is put forth:

1. Given an one-dimensional time series $\{x_i, i = 1,2,3,\cdots,N\}$ with data length N, for a scale factor τ, construct a set of coarse-grained time series $\mathbf{y}_k^{(\tau)} = \left\{y_{k,1}^{(\tau)}, y_{k,2}^{(\tau)}, \cdots, y_{k,p}^{(\tau)}\right\}$. The kth coarse-grained time series of $\mathbf{y}_k^{(\tau)}$ is defined as

$$y_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, 1 \leq j \leq \frac{N}{\tau}, 1 \leq k \leq \tau \tag{3.30}$$

2. At a scale factor the sample entropies of all coarse-grained time series are calculated and the CMSE value is defined as the means of τ entropy values.

$$\text{CMSE}(x, \tau, m, r) = \frac{1}{\tau} \sum_{k=1}^{\tau} SampEn(\mathbf{y}_k^{(\tau)}, m, r) \tag{3.31}$$

Different from MSE, given a certain scale (except the first scale), the CMSE algorithm first produces a set of coarse-grained time series based on a sliding window. Entropy values of these extracted time series are calculated and then averaged by the scale factor. By averaging entropy values at each scale, the CMSE method reduces the standard deviation of entropy values in analyzing numerical noise signals [65], thus presenting higher reliability in entropy analysis.

Inspired by the concept of the coarse-graining procedure, many modified and refined scale-extraction procedures were later developed [42, 40]. They are discussed in the following.

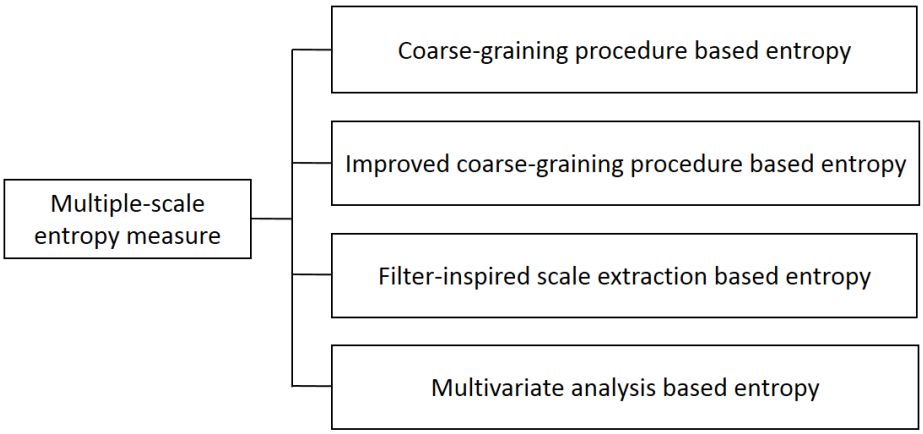


Fig. 3.5 Diagram of the main categories of multiple-scale entropy measures.

Improved coarse-graining procedure based entropy approaches

The improved multiple-scale entropy measures mainly aim at improving the efficacy of the scale-extraction procedure used for generating reliable multiple-scale time series in entropy analysis. Besides traditional coarse-graining procedure based methods, the modified methods can be broadly classified into three main categories according to the principle of scale-extraction procedures (shown in Fig. 3.5). Examples include composite MSEn [65], generalized MSEn [66], and refined composite MSEn [169]. Besides, there are many variants of improved entropy methods, where improved single-scale entropy algorithms are used in entropy estimation. The most common methods include Multiscale Permutation Entropy (MPEn) [168], refined composite MPEn [170], Multiscale Fuzzy Entropy (MFEn) [108], and modified multiscale symbolic dynamic entropy [60].

Filter-inspired scale-extraction based entropy approaches

The key idea in filter-inspired entropy measures is that the scale-extraction procedure is regarded as a filtering operation. Therefore, a set of improved multiple-scale time series are extracted. These extracted signals not only contain rich low- and high-frequency information but also have fine-grained time- and frequency-domain information. For instance, the averaging process in MSE is interpreted as filtering a time series by a filter of a piecewise constant type in [171]. In order to maintain high-frequency information in the multiple-scale time series, a hierarchical decomposition is proposed in [172].

Multivariate analysis method based entropy approaches

In order to analyze multichannel data, univariate MSE algorithm is extended to the multivariate case. The concept of multivariate sample entropy accounts for both within- and cross-channel dependencies in multiple data channels; entropy values are calculated by evaluating it over multiple temporal scales. Typical examples include multivariate MSEn [173], refined composite multivariate MFEn [107], and refined composite multivariate generalized MFEn [174].

Continuing advancements in entropy analysis have driven the emergence of more entropy measures for time series complexity analysis. Based on the notion of the scale-extraction regime, many multiple-scale entropy methods have been developed. For most of the methods, they are either based on improved single-scale entropy algorithms or improved scale-extraction mechanisms.

3.5 Motivation of Developing Improved Multiple-scale Entropy Methods

Despite the advantages and wide applications, the MSE and its variants, however, present a few limitations in time series complexity analysis in bearing diagnosis:

High-frequency information is abandoned

The basis of the MSE consists in the coarse-graining procedure which is similar to a low-pass filtering operation [42]. This procedure reduces the frequency rate of measurements to a lower value, as a result of which the high-frequency components are eliminated, thus losing the high-frequency information. However, for bearing fault diagnosis, critical fault symptoms may exist in both low-frequency and high-frequency components; thus, the elimination of high-frequency information greatly diminish the diagnosis performance of MSE for rolling bearing fault diagnosis.

Multiple-scale time series with reduced data length

The down-sampling operation by the coarse-graining procedure greatly reduces the data length of the generated coarse-grained time series. Due to the liner smoothing operation, the number of data points in the coarse-grained time series decreases with an increasing scale factor. As a result, this operation increases the variance of entropy values calculated from these extracted time series [65]. Moreover, an added difficulty may also come from potential

imprecise results with undefined entropy values (when no template vectors are matched to one another) [40]. Therefore, these factors subside the reliability of entropy measures further.

Reduced performance in analyzing non-stationary signals

The operation of the coarse-grained procedure is equivalent to the application of Finite-Impulse Response filter (FIR). This filtering is not well suitable to non-linear and non-stationary signals, especially for complex mechanical signals. The features of the frequency response of this low-pass filter are poor since it does not eliminate fast temporal scales, and thus, producing aliasing generating spurious oscillations in the frequency range from 0 to cutoff frequency. Consequently, the evaluation of the complexity of the downsampled signal is biased by the inclusion of these artifactual components [175].

3.6 Summary

Entropy measures are well suitable for the analysis of complex systems, especially in machinery fault diagnostic systems. Considering the principles of entropy measures, they can be broadly categorized into single-scale entropy measures and multiple-scale entropy measures. Single-scale entropy measures provide the basis of entropy estimation in time series complexity analysis. Comparatively, multiple-scale entropy measures enable quantifying the complexity change from a set of temporal scales. Considering its applicability in differentiating between health conditions, entropy analysis has achieved great attention and continuous improvements for machinery fault diagnosis.

Among entropy analysis, PerEn algorithm has theoretical simplicity and could extract dynamic changes in ordinal patterns from the structure of time series. Multiple-scale methods based on PerEn estimation have been used in machine early fault detection. Despite its wide range of applications, more works are still needed to further improve its reliability and robustness of analyzing non-stationary signals, in the circumstance of complex industrial settings, such as strong noise background, compound faults, and variable operating conditions.

Chapter 4

Improved Multiple-scale Entropy Measures

Multiple-scale entropy measures are suitable for complexity analysis in bearing fault diagnostic systems. They are non-linear feature indicators that can characterize underlying dynamic changes in signals collected from machinery. Different from single-scale methods, the core concept of multiple-scale entropy measures lies in the extraction of a range of temporal scales from the original signal. However, traditional multiple-scale entropy measures have several limitations on vibration analysis. For improving the reliability of entropy analysis in bearing diagnosis, this study presents an improved scale-extraction scheme and a new entropy method in bearing vibration signal analysis. Also, a new bearing diagnosis method is put forth based on the proposed entropy method and the SVM classifier. The fundamental concepts of the proposed improved multiple-scale entropy methods are introduced in the following.

4.1 Preliminary Study

In conventional multiple-scale entropy measures, the scale-extraction procedure that causes biased entropy values is due to the coarse-graining procedure. From the signal processing point of view, the coarse-graining procedure is similar to the low-pass filtering operation as well as a linear smoothing operation. In this case, advanced filtering operations are alternative to the low-pass filter to overcome its limitations.

Wavelet Packet Decomposition (WPT) is an advanced time-frequency signal analysis method. It provides various wavelet kernels that can produce appropriate transformed signals with rich time and frequency information. For instance, the Haar wavelet can be regarded as

one kind of extended coarse-graining procedures; it can generate a set of decomposed signals containing low- and high-frequency information from the original signal. Based on the WPT analysis, a new scale-extraction mechanism, named Fine-to-Coarse (F2C) procedure, is put forth, aiming to produce reliable multiple-scale time series from the original signal and yield appropriate entropy values. The principles of F2C procedure and a preliminary study is introduced in the following.

4.1.1 Wavelet Packet Decomposition

Wavelet Transform (WT) a fast-evolving mathematical and signal processing tool which is suitable for the analysis of non-stationary signals [114]. It can transform a signal in time domain into time-scale representations - wavelet coefficients. For spectrum analysis, Fourier transform is a widely used signal processing method; however, it cannot provide local information in the frequency domain and its correlations in the time domain. Short time Fourier transform is treated as an improved alternative to Fourier transform. Nevertheless, its disadvantage in the resolution of the frequency limits its applicability in the fault diagnosis system. In contrast, WT provides good time and frequency resolution. More specifically, it allows a high frequency resolution at low frequencies and high time resolution at high frequencies, as desired. Traditionally, WT can be categorized as Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), and Wavelet Packet Transform (WPT). Among these, WPT is a multi-resolution signal processing method based on digital filters, which is an extension of DWT method. The principle of WPT is introduced as follows [176], and an example of a two-level WPT tree is presented in Fig. 4.1. In each wavelet transformation, the

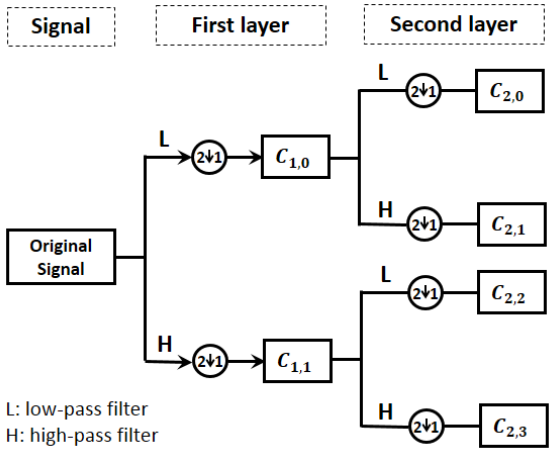


Fig. 4.1 Illustration of a two-level WPT tree.

raw signal will be decomposed into approximate and detail coefficients, respectively. The raw

signal is transformed by going through wavelet filters (i.e., low-pass and high-pass filters) and taking sub-sampling operation. A wavelet tree can thus be constructed by repeating this procedure until the desired decomposition resolution is achieved. Each node in the obtained wavelet tree represents wavelet coefficients with a specific range of frequency.

WPT can be implemented by means of a pair of low-pass and high-pass wavelet filters, denoted as $h(k)$ and $g(k) = (-1)^k h(1-k)$. Given wavelet function $\psi(t)$ and scaling function $\phi(t)$, they have the following relationships:

$$\begin{cases} \phi(t) = \sqrt{2}\sum_k h(k)\phi(2t-k) \\ \psi(t) = \sqrt{2}\sum_k g(k)\phi(2t-k) \end{cases} \tag{4.1}$$

To further extend the two-scale equation, recursive relationships are defined as follows:

$$\begin{cases} w_{2n}(t) \quad = \sqrt{2}\sum_k h(k)w_n(2t-k) \\ w_{2n+1}(t) = \sqrt{2}\sum_k g(k)w_n(2t-k) \end{cases} \tag{4.2}$$

where $n = 0$, $w_0(t) = \phi(t)$, $w_1(t) = \psi(t)$. Then, the input signal is decomposed to a set of wavelet packet coefficients which have approximation coefficients with low-frequency information and detail coefficients with high-frequency information. The decomposition of a time-domain signal $x(t)$ is described as [114]:

$$\begin{cases} C_{j+1,2n} \quad = \sum_l h(l-2k)C_{j,n} \\ C_{j+1,2n+1} = \sum_l g(l-2k)C_{j,n} \end{cases} \tag{4.3}$$

where $C_{j,n}$ denotes wavelet coefficients on the j-th decomposition level, the n-th sub-band, and l is the number of the wavelet coefficients. To prove the superiority of WPT in generating appropriate transformed signals, an example is given in Fig. 4.2. The Fig. 4.2 shows the approximation and detail coefficients decomposed on the first level in the wavelet tree based on the Haar wavelet.

Further, for comparison, Fig. 4.3 shows the modified coarse-graining procedure in the CMSE method [65]. These two figures highlight the fundamental difference in frequency information extraction between wavelet transform and the modified coarse-graining procedure. The modified coarse-graining procedure, however, neglects the high-frequency components and only considers the low-frequency components and low-frequency information. Comparatively, WPT makes full use of information hidden in both low- and high-frequency components, thereby extracting more fault information from measurements. Its advantage earns more reliability in bearing vibration signal analysis.

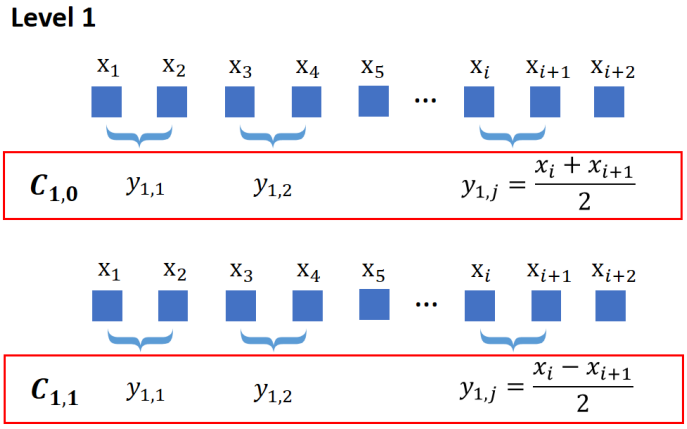


Fig. 4.2 Illustration of WPT on the first level based on Haar wavelet ($C_{1,0}$ and $C_{1,1}$ are the approximation coefficient and detail coefficient using low-pass and high-pass filters respectively).

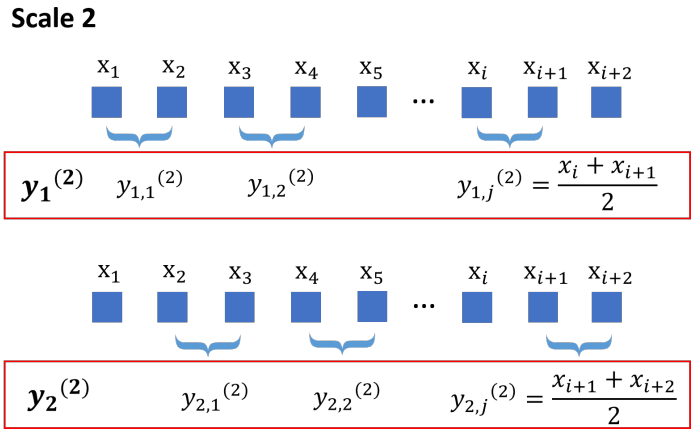


Fig. 4.3 Illustration of the coarse-graining procedure at the 2nd scale factor in the CMSE algorithm where only low-frequency components are considered.

In addition to the wavelet decomposition, the reconstruction procedure of the wavelet transform can be implemented using wavelet packet coefficients [177]:

$$C_{j,n} = \sum_{l} \left[h(k-2l) C_{j+1,2n} \right] + \sum_{l} \left[g(k-2l) C_{j+1,2n+1} \right] \tag{4.4}$$

where $h(k-2l)$ and $g(k-2l)$ denote the low-pass and high-pass wavelet reconstruction filters respectively. h is related to the scaling function and g is related to the wavelet function.

Correspondingly, given a wavelet packet tree at the j-th decomposition level, in total a set of 2^j wavelet packet coefficients, $\{C_{j,n}, 1 \leq n \leq 2^j\}$, can be obtained where n is the order of the coefficient in the j-th decomposition level. Then, based on each coefficient vector

$C_{j,n}$, a reconstructed signal $R_{j,n}$, with the same length of the original signal, can be produced by setting the all the other decomposition coefficients on level j to zero and recursively implementing the wavelet reconstruction transform in the inverse procedure until j decreases to zero [178]. Among each reconstruction procedure, the wavelet decomposition coefficient has nearly 1/2 data points by comparing with the upper level. Finally, for each $R_{j,n}$, it has an approximative frequency range with that of $C_{j,n}$ and remains the same length as the original signal. Therefore, given the sampling frequency F_s, the frequency intervals of each $R_{j,n}$ can be approximately computed by:

$$\left((n-1)*2^{-j}F_s,\ n*2^{-j}F_s\right],\ n=1,2,\cdots,2^j \tag{4.5}$$

Reconstructed signals equally partition the whole frequency spectrum of the signal and contain frequency information ranging from low to high. Furthermore, the reconstructed signals have the same data length as that of the original signal, which avoids the large variance caused by the decreased data length in calculating the MPE and CMPE values [40]. Therefore, the F2C procedure benefits from the advantages of WPT analysis, which makes the improved entropy method more suitable for entropy analysis under an improved multiple-scale framework.

4.1.2 Fine-to-Coarse Scale-extraction Procedure

The superiority of WPT allows decomposing non-stationary signals into wavelet coefficients with good time and frequency resolution. Also, the reconstruction procedure enables the inversion of each wavelet decomposition coefficient to a reconstructed sub-signal that remains the same length with the original signal. Owing to the advantages of WPT analysis, the F2C signals are produced by constructing reconstructed sub-signals with a fine-grained to coarse-grained approach [7]. To be more specific, in the F2C procedure, when the scale factor increases, the high-frequency information is consecutively removed from previously acquired F2C signals at lower scales. Hence, given the F2C signals with increasing scales, high-frequency and low-frequency information is consecutively refined and obtained from the original signal through the F2C procedure, which can contribute to appropriately characterize the dynamic changes associated with fault symptoms in vibration signals. Fig. 4.4 shows the flowchart of the proposed F2C procedure. The detail calculation procedure is described as follows:

1. Apply WPT to decompose an original signal to the j-th decomposition level where only wavelet decomposition coefficients produced from the branch of $C_{1,0}$ on the 1-th

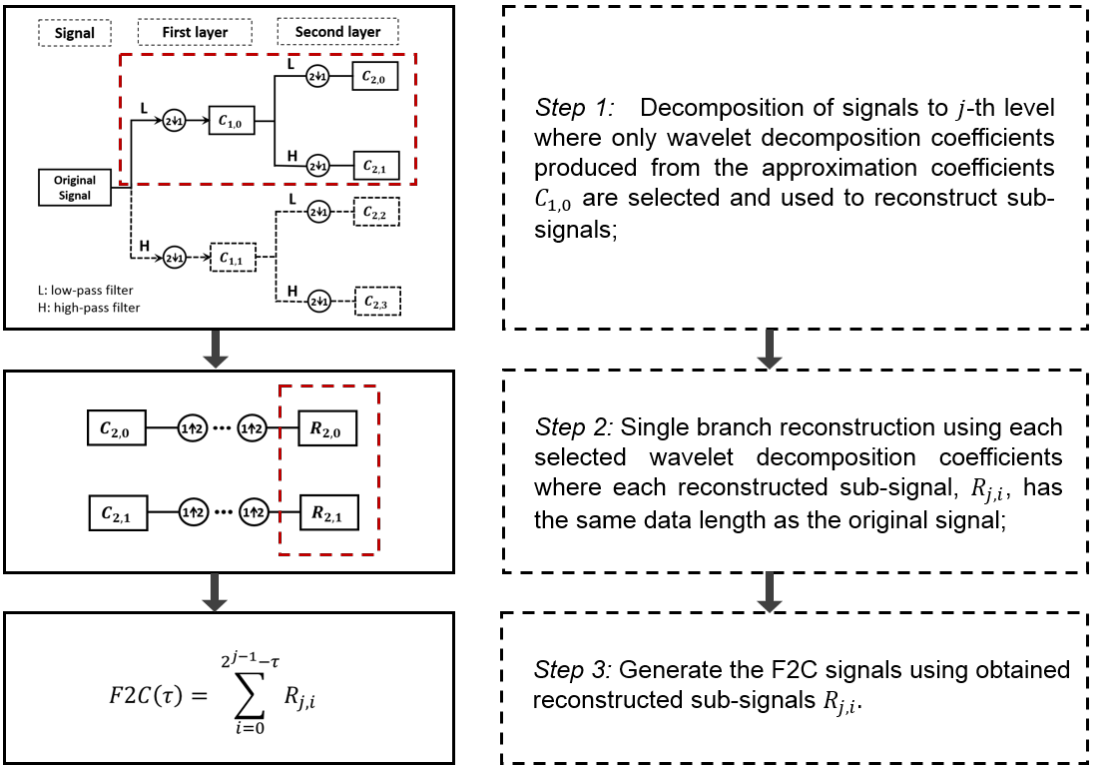


Fig. 4.4 Flowchart of the proposed F2C procedure ($C_{j,i}$ is wavelet coefficients, and $R_{j,i}$ is reconstructed sub-signal from each branch of selected wavelet coefficients $C_{j,i}$).

level are selected and used. Thereby, there are 2^{j-1} sets of wavelet decomposition coefficients $\{C_{j,n}, (0 \leq n \leq 2^{j-1}-1)\}$ are obtained and used in the next step;

2. Reconstruct single branch, $R_{j,n}$, using each acquired wavelet decomposition coefficients $C_{j,n}$ on the j-th level, by setting the coefficients of all the other vectors on level j to zero and recursively implementing the wavelet reconstruction transform in the inverse procedure until j decreases to zero. Thus, each reconstructed sub-signal has the same data length as the original signal. Therefore, totally 2^{j-1} reconstructed signals, $\{R_{j,n}, (0 \leq n \leq 2^{j-1}-1)\}$, can be produced using the wavelet reconstruction procedure;

3. F2C procedure: construct F2C signals by consecutively removing one reconstructed signal from previously obtained F2C signals, commencing from the accumulation of all 2^{j-1} reconstructed signals. Thereby, F2C signals are produced as

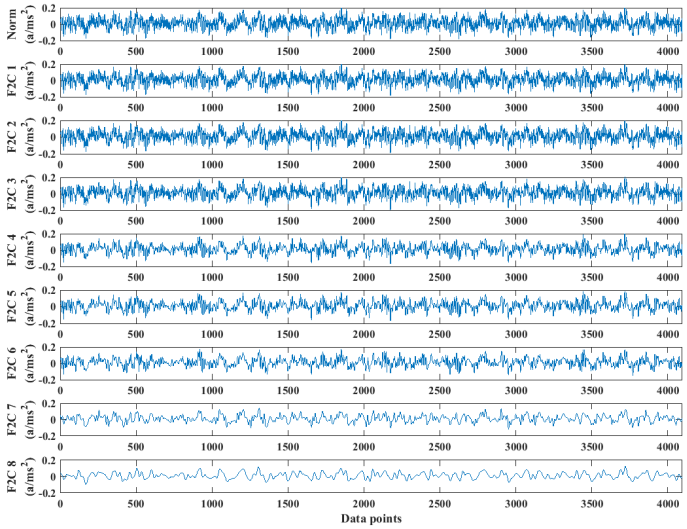
$$F2C(\tau) = \sum_{i=0}^{2^{j-1}-\tau} R_{j,i}, 0 \leq i \leq 2^{j-1}-1, 1 \leq \tau \leq 2^{j-1} \tag{4.6}$$

where j is the decomposition level, τ is the scale factor, and the maximum number of τ is equal to 2^{j-1}. Herein, the proposed F2C procedure refers to a process that produces signal branches with fine-grained to coarse-grained time-frequency information refined from the original signal.

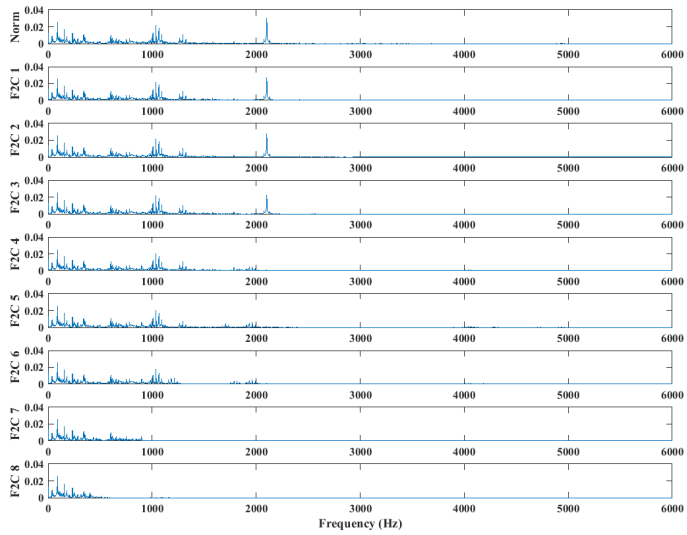
The frequency range of the F2C signals obtained from the original signal is gradually decreased since high-frequency components are consecutively removed from previously produced F2C signals. Hence, through the F2C procedure, low-frequency information finally remains in the F2C signals at high scales. With an increasing scale factor, dynamic changes hidden in lower-frequency components can thus be characterized in hierarchies.

Similarly, the F2C procedure adopts the half-frequency spectrum of the original signal in entropy analysis. Previously works [32, 40, 179, 41] have verified the effectiveness of the use of a half-frequency spectrum in bearing diagnosis, such as the MSE, MPE, and CMPE. Both of them only apply no more than half frequency spectrum of the original signal commencing from the $2nd$ scale, because the coarse-grained procedure is similar to a down-sampling operation. Besides, vibration acquisition system usually has a high sampling frequency. Therefore, very high-frequency components typically contain too much detail information that may be considered as noises to some extent, thus providing less information related to intrinsic fault symptoms.

The emergence of incipient failures in rolling bearing components typically introduces impulse waves and finally results in the occurrence of coupling frequency in both lower and higher frequency components due to periodical friction and strikes between faulty and healthy components. Thus, the F2C procedure maintains prominent low- and high-frequency information in bearing diagnosis. By using orthogonal wavelet kernels, the WPT analysis enables generating wavelet coefficients by applying low- and high-pass filters. Therefore, both low- and high-frequency information is kept in the F2C signals. Furthermore, by reconstructing coefficients to signals with the same data length as the original time series, the F2C signals can give appropriate PE values, especially when the length of the original data is already too short. For instance, a vibration signal of rolling bearing with Outer Race Fault (ORF) is analyzed using the F2C procedure, and F2C signals with 8 scales are obtained given a $4th$ decomposition level. The F2C signals and their frequency spectrums transformed using fast Fourier transform are presented in Fig. 4.6. In each wavelet transformation, the raw signal will be decomposed into approximate and detail coefficients, respectively. The raw signal is transformed by going through wavelet filters (i.e., low-pass and high-pass filters) and taking sub-sampling operation. A wavelet tree can thus be constructed by repeating this procedure until the desired decomposition resolution is achieved. Each node in the obtained wavelet tree represents wavelet coefficients with a specific range of frequency.



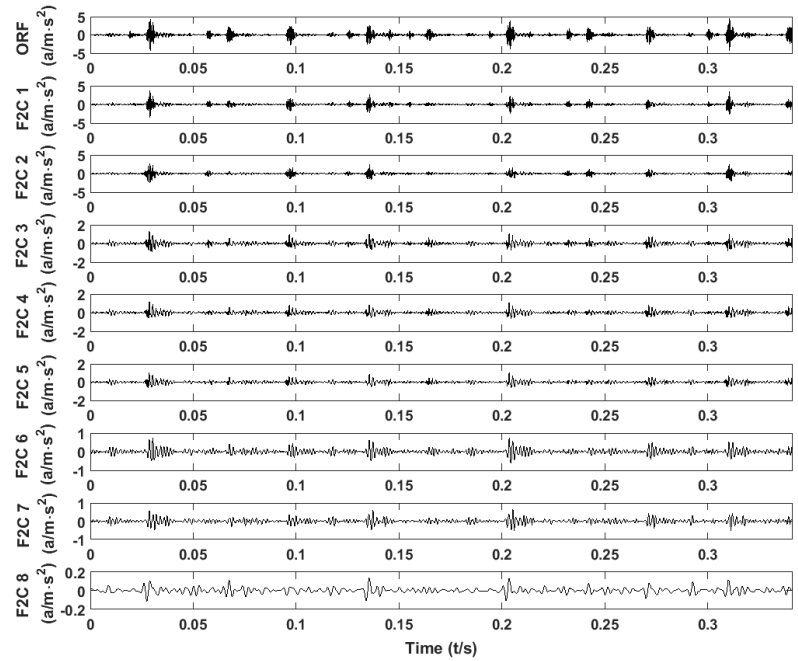
(a) Time domain waveforms of the Norm and F2C signals



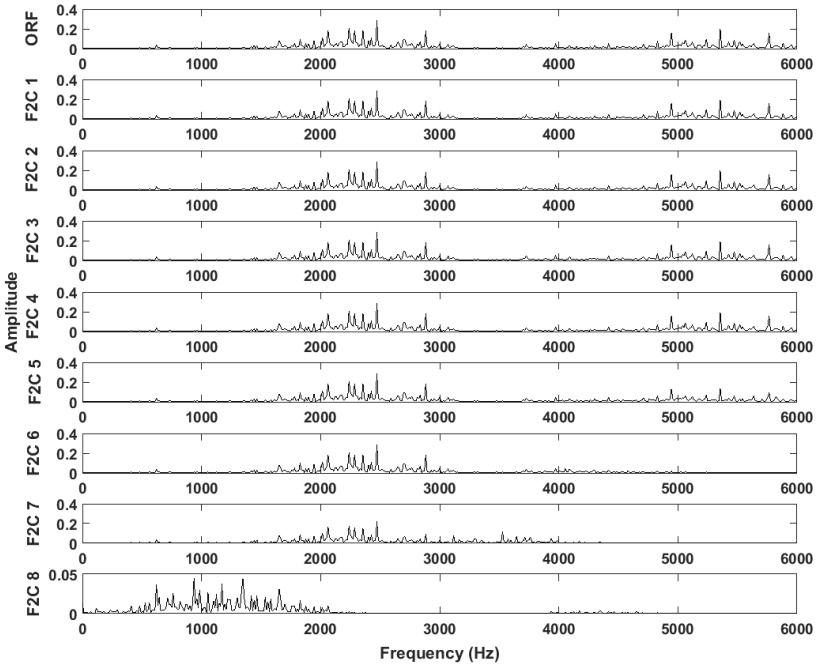
(b) Frequency spectrum of the Norm and F2C signals

Fig. 4.5 (a) Time domain waveforms and (b) frequency spectrums of the original Norm bearing signal and generated F2C signals, respectively. Herein, the Norm stands for normal bearing state.

Fig. 4.5 shows raw normal bearing signal and decomposed F2C signals as well as their spectrums. For comparison, Fig. 4.6 presents time-domain and frequency-domain signals of ORF bearing, respectively. From Fig. 4.6, it is observed that the F2C signals with increasing scale factors are becoming more and more smooth and flat because high-frequency information representing detail changes has been consecutively removed from former F2C signals according to the F2C procedure. Therefore, only low-frequency information remains



(a) Time domain waveforms of the ORF and F2C signals



(b) Frequency spectrum of the ORF and F2C signals

Fig. 4.6 (a) Time domain waveforms and (b) frequency spectrums of the original ORF bearing signal and generated F2C signals, respectively. Herein, the ORF stands for outer race fault in the bearing.

in the F2C signals that have high scales. Furthermore, Fig. 4.6 (b) indicates that the frequency range of each F2C signal is in line with the concept of the F2C procedure. In particular, both low- and high-frequency information is extracted from the original signal and exists in F2C signals at small scales. With an increasing scale, low-frequency information mainly composes the frequency spectrum because of the use of signals reconstructed from low-frequency coefficients. Besides, the frequency of the F2C signal on the first scale is very similar to that of the original ORF signal, which verifies that the F2C signals encompass rich and important frequency information refined from the original signal. For comparison, Fig. 4.5 presents raw signal and F2C signals of normal bearing.

4.1.3 Preliminarily Proposed Improved Entropy Method

Inspired by the F2C procedure, an improved entropy, named Fine-to-coarse Multiscale Permutation Entropy (F2CMPE) is proposed in the preliminary study. Its calculation relies on two procedures: 1) generation of F2C signals with a range of scales; 2) calculation of PerEn values from obtained F2C signals. The F2CMPE has advantages of considering low- and high-frequency information in entropy estimation and reduce biases in entropy calculation in comparison with traditional entropy methods [7].

4.1.4 Parameter Selection in the Generation of F2C signals

The appropriate use of mother wavelet function and decomposition level can significantly improve the performance of F2CMPE analysis based on the WPT analysis. More specifically, the desired frequency resolution in F2C signals can be achieved by using a suitable wavelet kernel and decomposition level, where the latter determines the frequency band in each F2C signal. Therefore, the selection of wavelet and decomposition level is vital for appropriately generating the F2C signals. In wavelet analysis, the performance of a mother wavelet is based on two major factors, namely the support size and the number of vanishing moments. More specifically, a mother wavelet containing a large number of vanishing moments and small support size can locate valuable information from the original signal with less redundant information [180]. Among different mother wavelets, the Daubechies and Symlet family of wavelets are well-known for their orthogonality and efficiency in filter implementation for the Mallat fast algorithm. They are considered as available wavelet kernel functions in this study.

The Relative Wavelet Energy (RWE) method has been widely applied to compare and select the appropriate mother wavelet [112, 181] in wavelet analysis. The RWE can provide information regarding relative energy distribution in transformed signals, which is

also suitable for measuring the energy ratio in the F2C signals. An appropriate mother wavelet is supposed to extract the most significant amount of energy because failures can introduce large magnitude in certain wavelet coefficients. Thus, the wavelet kernel having the highest RWE value is often considered as the optimum mother wavelet for generating F2C signals. The principle of the RWE method is described below. Given a sets of F2C signals $\left\{F2C_\tau, \tau = 1,2,\cdots,2^{j-1}\right\}$, the energy of each F2C signal can be obtained by:

$$E(\tau) = \sum_{i=1}^{N} |F2C_{i,\tau}|^2, 1 \leq i \leq N, 1 \leq \tau \leq 2^{j-1} \tag{4.7}$$

where i is the index of the data point in each F2C signal, N is the data length of the F2C signal on the scale factor τ. Then, the total energy of F2C signals obtained in the j-th decomposition level can be obtained as

$$E_{sum} = \sum_{i=\tau}^{2^j} E(\tau), 1 \leq \tau \leq 2^{j-1} \tag{4.8}$$

Finally, the normalized value represents the relative energy of each F2C signal among overall F2C signals:

$$RWE(\tau) = \frac{E(\tau)}{E_{sum}}, 1 \leq \tau \leq 2^{j-1} \tag{4.9}$$

where $\sum_{\tau=1}^{2^{j-1}} RWE(\tau) = 1$, and the energy probability distribution $RWE(\tau)$ is considered as a time-scale density. Besides, the variance of one indicator quantifies to what extent the indicator varies and fluctuates. Normally, the high variance index also means that there are extra dynamic changes and possibly additional information existed in this indicator. Hence, in this study, the variance of RWE is also applied to evaluate optional mother wavelet functions. The larger variance value of REW is, the greater possibility of extracting useful information associated with fault symptoms from non-stationary signals [182]. In this study, the RWE values and their corresponding variance are both applied to evaluate four Daubechies (“db2”,“db4”,“db6”,“db8”) and four Symlet (“sym2”, “sym4”, “sym6”, “sym8”) wavelets respectively to select the optimum one for generating F2C signals. Additionally, the wavelet decomposition level determines the range of sub-frequency band in wavelet coefficients as well as the reconstructed signals. The larger decomposition level, the higher frequency resolution in each sub-band can be obtained. Nevertheless, a very high decomposition level will require more computational time and computing resources. By taking these factors into account a five-level ($j = 5$) or six-level ($j = 6$) wavelet tree is suitable for the F2CMPE analysis.

To select the appropriate mother wavelet, the vibration signals of rolling bearing with ten conditions are randomly chosen from Case Western Reserve University (CWRU) Data

Table 4.1 Description of wavelet functions and their maximum RWE and average variance values (RWE: Relative Wavelet Energy).

Wavelet name	Maximum Relative Wavelet Energy	Variance
Daubechies2	0.110795	0.030126
Daubechies4	**0.113272**	**0.030434**
Daubechies6	**0.113340**	**0.030224**
Daubechies8	0.112933	0.030026
symlet2	0.110795	0.030126
symlet4	0.113234	0.030539
symlet6	0.113246	0.030303
symlet8	0.112933	0.030026

Center [183]. Eight number of different mother wavelet kernels are then applied to construct F2C signals. In this study, the fifth-decomposition level is used, and 32 sets of F2C signals are therefore obtained correspondingly. This experiment is operated 100 times, and the average maximum RWE values and their average variances are presented in Table. 4.1. It can be seen that "db4" and "db6" wavelet functions outperform the rest. Besides, the two indicators (namely RWE and its variance) of "db4" and "db6" wavelets are very similar, and "db4" is finally selected as the desired mother wavelet in this study.

Additionally, the calculation of PerEn also greatly affects the effectiveness of the F2CMPE feature extraction. To provide reliable PerEn measurements, the selection of the embedding dimension m and the time delay λ are necessary. Practically, when $m < 4$ it cannot detect the dynamic change of the mechanical vibration signals. Besides, when $m > 8$, not only the reconstruction of phase space will homogenize vibration signals but also the calculation of PerEn is time-consuming; hence, it cannot truly reflect the small varying range. According to literatures [32, 64], it was recommended to select $m = 4-7$. Regarding the use of time delay, when $\lambda > 5$, it cannot detect a slight change in the time series. Comparatively, the effect of time delay λ has small effects on the calculation of PerEn [179], especially when $\lambda \leq 4$. Moreover, a very short time series cannot produce prominent statistical significance on PerEn values. Therefore, in this study, $m = 4-5$ and $\lambda = 1-3$, data length of time series $N = 4,096$ are considered for calculating PerEn values from the F2C signals.

4.2 Adaptive Multiscale Weighted Permutation Entropy Measure

In the multiple-scale entropy method, entropy values are calculated from a range of multiple-scale time series based on the scale-extraction procedure. In the extraction procedure, a larger scale factor usually produces more time series that contain certain time-frequency information in the original signal. For example, the MSE method generates the coarse-grained time series with a narrower low-frequency bandwidth as the scale factor increases. Although a larger scale factor can produce more multiple-scale time series, not all extracted time series are closely related to fault information. That is, on the one hand, some extracted time series may contain unexpected redundant information that may reduce the efficiency of the data-driven diagnostic model in bearing diagnosis. On the other hand, an increasing number of high-dimensional features consume more computation resources, thus increasing computational burden. As a result, these factors may decrease the efficacy of multiple-scale entropy analysis in bearing diagnosis. Additionally, many improved single-scale entropy algorithms are available for entropy estimation. Consequently, there is a necessity to develop an improved entropy that provides both reliable and efficient entropy analysis in bearing vibration signal analysis.

For this purpose, a new Adaptive Multiscale Weighted Permutation Entropy (AMWPE) algorithm is proposed for the analysis of bearing vibration signals. It is an improvement of the F2CMPE method proposed in the preliminary study. In the AMWPE method, a refined F2C procedure is proposed, and adaptive F2C signals are produced based on correlation coefficient analysis. Also, an improved PerEn is applied to entropy estimation, which is suitable for the analysis of non-stationary vibration signals. The proposed AMWPE algorithm is introduced below.

4.2.1 Pearson Product-Moment Correlation Coefficient

The Pearson product-moment correlation coefficient is one of the association measures [184]. The correlation coefficient is widely applied to evaluate the similarity between two variables. Its value ranges between $[-1,1]$, where 1 means that two variables are completely positively correlated, 0 means they have no correlation, and -1 means they are completely negative correlated. An example of the correlation relationship is illustrated in Fig. 4.7.

Given two variables $X(n)$ and $Y(n)$, the correlation coefficient ρ is defined as:

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_x)(Y-\mu_y))}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{4.10}$$

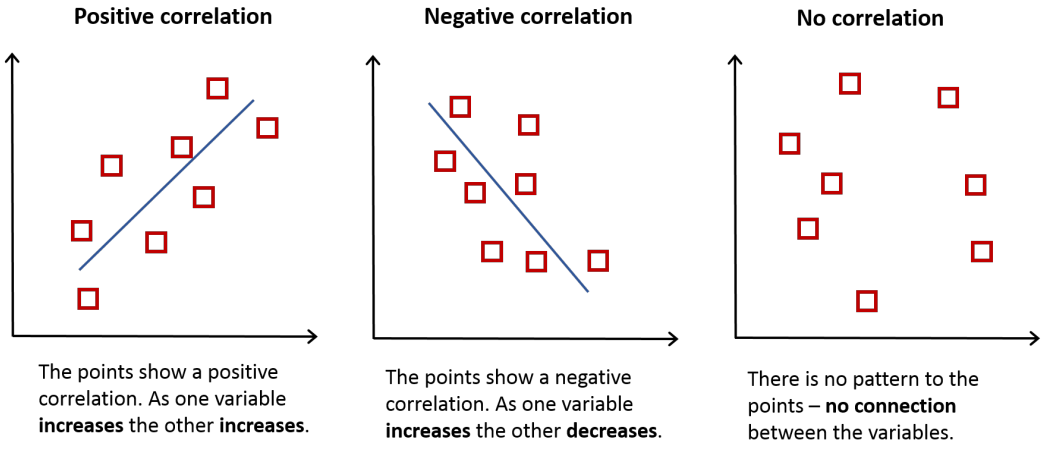


Fig. 4.7 An example of the correlation relationship.

where μ_X, σ_X, μ_Y, σ_Y denote the mean and standard deviation of X and Y, respectively. $\mathrm{Cov}(X,Y)$ is the covariance of variables X and Y.

In signal decomposition, transformed signals always have different bandwidths and maintain low-frequency or high-frequency characteristics of the original signal. Due to the filter operation, multiple-scale time series have a narrow frequency bandwidth compared to the original signal. As a result, the waveforms of extracted sub-signals (multiple-scale time series) that represent prominent fault information should have a high similarity to that of the original signal. In this case, appropriate multiple-scale time series that carry crucial fault information should have a high correlation with the original signal in the time domain. Therefore, in the AMWPE algorithm, the extracted sub-signals that have high correlation coefficients with the original signal in the time domain are applied to construct adaptive F2C signals for further entropy estimation.

4.2.2 Weighted Permutation Entropy

Although the wide application of PerEn measure, its major disadvantage lies in neglecting the amplitude difference between neighboring elements. Fig. 4.8 illustrates possible motifs that correspond to the same order permutation type when $m = 3$. The PE considers the order structure of time series merely, which inevitably results in that different amplitudes in the motif cannot differently contribute to the probability distribution of order permutations. That is, the amplitude information is neglected, and different time-series may have the same PerEn value.

In order to overcome this shortcoming, Fadlallah et al. [185] proposed an improved PerEn method - Weighted Permutation Entropy (WPerEn). The definition of WPerEn retains

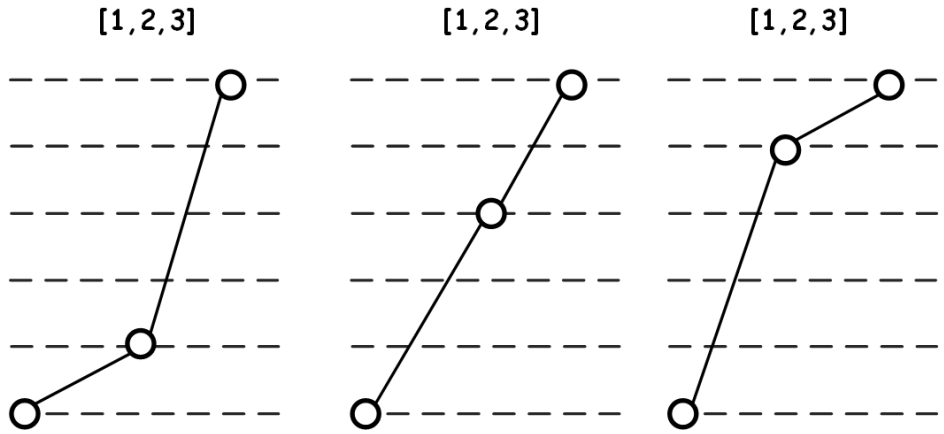


Fig. 4.8 Three possible motifs corresponding to the same permutation pattern $\pi_n = [1,2,3]$ when $m = 3$ by comparing neighboring values in an embedding vector.

most of PerEn's properties. The most significant difference consists in the definition of the relative frequency of symbol sequences. The WPerEn can distinguish vectors that have the same ordinal patterns but different amplitude elements. It takes into account the amplitude difference in different motifs. Therefore, different motifs will differently contribute to the probability distribution of permutation patterns. The WPerEn applies the concept of the variance of neighboring elements as the weighting factor w_i. The weighted relative frequency of each permutation π_n is calculated as

$$Q(\pi_n) = \frac{\sum_{i=1}^{k} 1 * w_i \mid \text{when } \mathbf{X}_i \text{ has type } \pi_n}{\sum_{n=1}^{m!} \sum_{i=1}^{k} 1 * w_i \mid \text{when } \mathbf{X}_i \text{ has type } \pi_n} \tag{4.11}$$

where k is no greater than $N-(m-1)\lambda$, and $Q(\pi_n) = 0$ only when there are no vectors $\mathbf{X}_i$ belonging to the given permutation type π_n. The weight w_i is obtained from the corresponding vector $\mathbf{X}_i$ by

$$w_i = \frac{1}{m} \sum_{k=1}^{m} \left[x_{i+(k-1)\lambda} - \bar{x_i} \right]^2 \tag{4.12}$$

where $\bar{x_i}$ is the arithmetic mean of the $\mathbf{X}_i$. Then, the WPerEn is obtained as

$$\text{WPerEn}(m,\lambda,N) = \sum_{j=1}^{m!} Q(\pi_j) \log_2(Q(\pi_j)) \tag{4.13}$$

The value of WPerEn is also in the interval of $[0, \log_2 m!]$. The WPerEn measure has been examined in many studies and shown better entropy estimation performance in bearing diagnostic systems [186–188]. In the AMWPE method, the WPerEn is applied to calculate entropy values from adaptive F2C signals.

4.2.3 Multi-class Support Vector Machine

Support Vector Machine (SVM) is a statistical machine learning technique proposed by Vapnik in 1995 [189]. The concept of SVM is intuitive, and it requires less prior knowledge and is computationally easy. Compared to traditional classifiers, SVM is robust and easy to use which is suitable for low-dimensional samples. It is a deterministic algorithm that has been extensively used for pattern recognition. The basic idea of SVM is to find the optimal linear separating hyperplane of labeled dataset. Fig. 4.9 shows a series data points for two different classes of data - red squares represent negative class and blue circles represent positive class. The SVM attempts to place a linear boundary between the two different classes. For this purpose, it minimizes the upper bound of the generalization error by maximizing the margin between the separating hyperplane and the nearest sample points. The goal of SVM is to generate a model - based on training samples - which can predict the label of testing samples based on their data features.

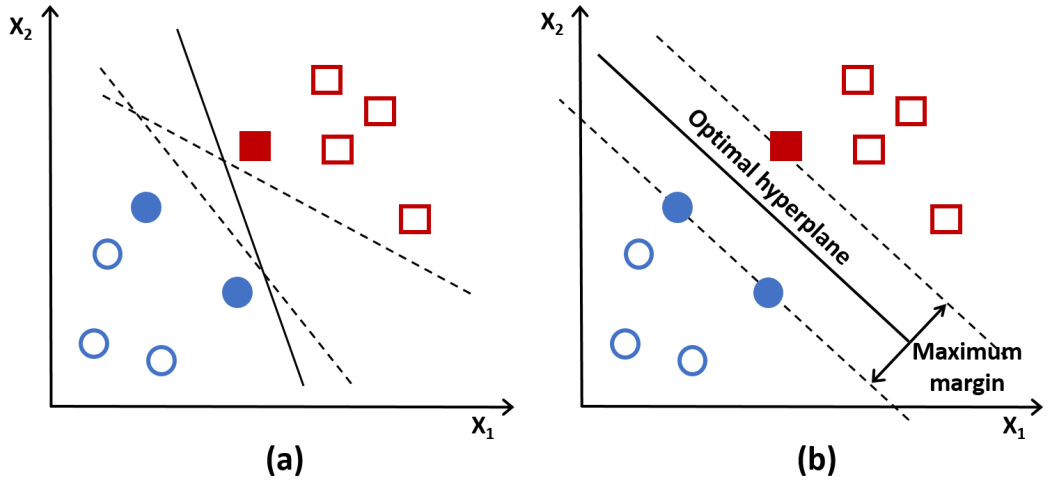


Fig. 4.9 Different separating hyperplane resulted from different algorithms: (a) the hyperplane based on liner classification algorithms; (b) the hyperplane based on the SVM algorithm.

Given a training dataset with l samples $(x_i, y_i,\ i = 1, 2, \cdots, l)$, where x_i is an input sample and $y_i \in \{+1, -1\}$. Through a non-linear mapping $\phi(x)$, the input data x_i is mapped into a higher dimensional feature space by the function $\phi(x)$. SVM finds a linear separating hyperplane with the maximal margin by solving the following optimization problem [190]:

$$\min \ \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{l}\xi_i \tag{4.14}$$

$$\text{subject to } \ y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \text{ and } \xi_i \geq 0. \tag{4.15}$$

where ξ_i estimates the distance between the margin and the examples $\mathbf{x}_i$ that lying on the wrong side of the margin and C is the penalty parameter. By introducing Lagrange multipliers α_i , the training procedure amounts to solving a convex quadratic problem. By projecting the original sample space into a high-dimensional with a kernel function $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$, the nonlinear separable problem becomes linearly separable in the eigenspace. Some possible kernel functions are available in the Table. 4.2.

Table 4.2 Several possible kernel functions and types.

Type of classifier	Kernel function
Gaussian RBF	$K(x_i,x_j) = \exp(-\gamma\|\|x_i - x_j\|\|^2)$
Polynomial of degree d	$K(x_i,x_j) = (x^T x_i + 1)^d$
Multi-layer perceptron	$K(x_i,x_j) = \tanh(x^T x_i + \theta)$

When the SVM is trained, the decision function is given by

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i,j=1}^{l} \alpha_i y_i \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) + b\right) \tag{4.16}$$

Taking into account different fault patterns, the rolling bearing fault pattern recognition is a multi-class classification. The multi-class classification strategy - one-against-one strategy - is applied in this study. One-against-one approach constructs $k(k-1)/2$ classifiers where each one is trained on data from two classes. For the training data from the i th and the j th classes, the following binary classification problem is solved:

$$\min \ \frac{1}{2}||\mathbf{w}^{ij}||^2 + C\sum_t \xi_t^{ij}(\mathbf{w}^{ij})^T \tag{4.17}$$

$$\begin{aligned} \text{subject to } & (\mathbf{w}^{ij})^T \phi(\mathbf{x}_t) + b^{ij} \geq 1 - \xi_t^{ij}, \ \text{if } y_t = i \\ & (\mathbf{w}^{ij})^T \phi(\mathbf{x}_t) + b^{ij} \leq -1 + \xi_t^{ij}, \ \text{if } y_t = j \\ & \xi_t^{ij} \geq 0, \quad j = 1,2,...,l \end{aligned} \tag{4.18}$$

The classification decision is made using the following strategy: if sign $(\mathbf{w}^{ij})^T \phi(\mathbf{x}_t) + b^{ij}$ says x is in the i th class, then the vote for the i th class is added by one. Otherwise, the j th is increased by one. Then x is predicted in the class using the largest vote. The voting approach described above is also called as Max Win strategy. In this study, the LIBSVM Matlab Toolbox [191] is used for bearing fault pattern recognition.

4.2.4 The Proposed AMWPE Algorithm

In the AMWPE algorithm, an improved F2C procedure is developed to construct adaptive F2C signals [192]. The advent of failures in the bearing will introduce coupling frequencies and change amplitude magnitudes in bearing vibration signals. Crucial components extracted from raw signals should maintain characteristic symptoms in the waveforms and thus have a high similarity to raw signals in the time domain. Considering this, the adaptive F2C procedure in the AMWPE algorithm selects salient reconstructed sub-signals based on correlation coefficient analysis. These selected sub-signals are closely related to the raw signals and have a high correlation in the time domain. Then, adaptive F2C signals are constructed based on these selected sub-signals, and entropy values are calculated from obtained F2C signals. The improved F2C procedure has two merits. On the one hand, these adaptive F2C signals could incorporate more crucial fault information and less redundancy. On the other hand, the improved F2C procedure can achieve higher computational efficiency compared to the F2CMPE in time series complexity analysis.

In this study, correlation coefficients are used to evaluate the similarity between reconstructed signals $\mathbf{R}_{j,n}$ and the raw signal $\mathbf{x}$ in the time domain. Fig. 4.10 presents the diagram of the AMWPE algorithm, and its detailed calculation steps are described below:

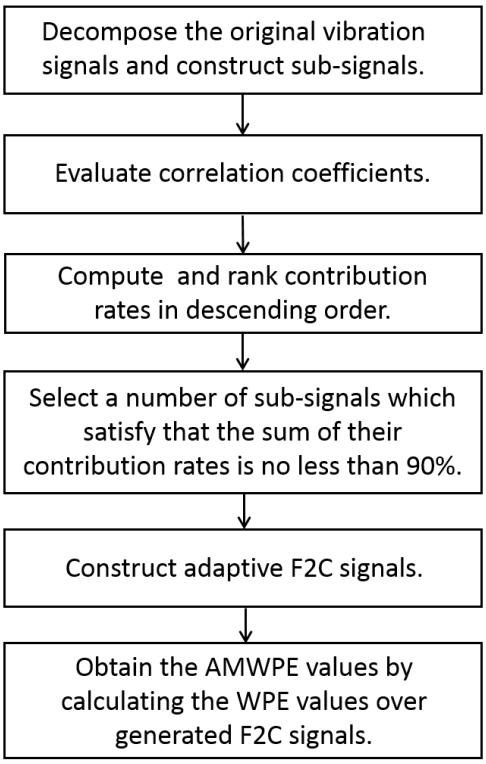


Fig. 4.10 Procedure of the AMWPE algorithm.

1. Decompose a vibration signal $\mathbf{x}$ into the j-th decomposition level using WPD. Select the wavelet coefficients $\{\mathbf{C}_{j,i}, (0 \leq i \leq 2^{j-1}-1)\}$ that are decomposed from the approximate coefficients at the first decomposition level in the wavelet tree. Reconstruct these selected wavelet coefficients to sub-signals that have the same data length to $\mathbf{x}$. Thus, totally 2^{j-1} number of reconstructed signals $\{\mathbf{R}_{j,i}, (0 \leq i \leq 2^{j-1}-1)\}$ are obtained correspondingly.

2. Compute correlation coefficients between the reconstructed sub-signal and raw signal in the time domain $\rho(\mathbf{R}_{k,i}, \mathbf{x})$ as

$$\rho(\mathbf{R}_{j,i}, \mathbf{x}) = \frac{E\left[(\mathbf{R}_{j,i} - \mu(\mathbf{R}_{j,i}))(\mathbf{x} - \mu(\mathbf{x}))\right]}{\sigma(\mathbf{R}_{j,i})\sigma(\mathbf{x})} \tag{4.19}$$

   where $\mu(\mathbf{R}_{j,i}), \mu(\mathbf{x}), \sigma(\mathbf{R}_{j,i}), \sigma(\mathbf{x})$ denote the mean and standard deviation of the reconstructed sub-signal and the original signal, respectively.

3. Contribution rates are calculated based on the correlation coefficients by

$$S_i = \frac{\rho(\mathbf{R}_{j,i}, \mathbf{x})}{\sum_{i=0}^{2^{j-1}-1} \rho(\mathbf{R}_{j,i}, \mathbf{x})} * 100\% \tag{4.20}$$

   where $0 \leq i \leq 2^{j-1}-1$, and a larger S_i indicates that the corresponding sub-signal has higher correlation with the original signal in the time domain.

4. Rank the contribution rates in descending order. For each signal, refer to n as the maximum number of its reconstructed sub-signals, which satisfies that the sum of the first n largest contribution rates is no less than 90%, namely $\sum_{i=1}^{n} S_i \geq 90\%, (n \leq 2^{j-1})$. Record the index of the selected n number of sub-signals and denote them as $\{\mathbf{U}_i, (1 \leq i \leq n)\}$.

5. Apply obtained sub-signals $\mathbf{U}_i$ to construct adaptive F2C signals accordingly, commencing from the accumulation of all n number of selected sub-signals

$$\text{F2C}^{(\tau)} = \sum_{i=1}^{n-\tau} \mathbf{U}_i \tag{4.21}$$

   where $1 \leq i \leq n$, and $1 \leq \tau \leq n$.

6. Calculate the WPerEn value over each F2C signal, the AMWPE values are finally obtained by

$$\text{AMWPE}(\mathbf{x}, \tau, m, \lambda) = \text{WPerEn}(\text{F2C}^{(\tau)}, m, \lambda) \tag{4.22}$$

The AMWPE analysis consists in wavelet analysis and WPerEn estimation. In wavelet analysis, appropriate parameters - mother wavelet and resolution of decomposition scale - can produce time-frequency components containing crucial fault information. Given a j-level wavelet tree, there are 2^{j-1} number of wavelet coefficients are totally obtained according to the Step 1) in the AMWPE algorithm. Also, a "db4" wavelet is applied as the Daubechies family of wavelets is well-known for their orthogonality and efficiency in filter implementation [193]. Besides, regarding entropy parameter configuration in the WPerEn measure, many studies have examined the performance of embedding dimension m and time delay λ in the calculation of PE values [160]. Researchers recommended that parameters, $m =$ 4-7 and $\lambda =$ 1-3, apply for bearing health monitoring [179].

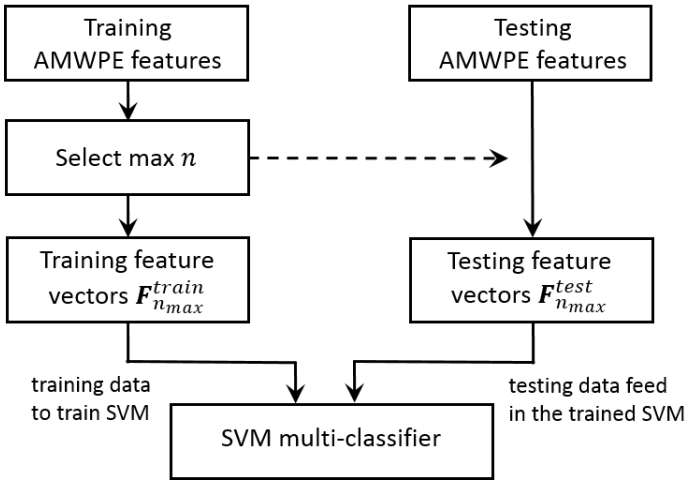


Fig. 4.11 The procedure of transferring n_{max} obtained from the training stage to testing stage.

4.2.5 A New Bearing Diagnosis Method based on the AMWPE and SVM

Based on the AMWPE and SVM, the proposed fault diagnosis method for rolling bearing is presented as follows:

1. Collect vibration signals from rolling bearings with various health conditions. For each condition, split raw data sets into training and testing data sets, respectively;

2. Calculate the AMWPE values from the training data samples. In this study, a j-level decomposition tree is used and thus $\tau = 2^{j-1}$. For each training sample, calculate the value of n; thus, a vector of n values can be obtained from all training samples. Then, specify the maximum n, denoted as n_{max}, as the number of features for constructing training feature vectors as $\mathbf{F}_{n_{max}}^{train}$;

3. Calculate the AMWPE values from the testing data samples and construct testing feature vectors $\mathbf{F}_{n_{max}}^{test}$ where n_{max} is acquired from the training data samples;

4. Apply training feature vectors $\mathbf{F}_{n_{max}}^{train}$ to train the SVM-based multi-class model for classifying bearing fault types. This procedure of transferring n_{max} parameter to testing procedure is shown in the Fig. 4.11;

5. Input testing feature vectors $\mathbf{F}_{n_{max}}^{test}$ into the obtained model to predict the health label. Thus, the fault pattern of the testing sample can be recognized. The flowchart of the proposed method based on the data collection, data analysis, and fault pattern classification is described in Fig. 4.12.

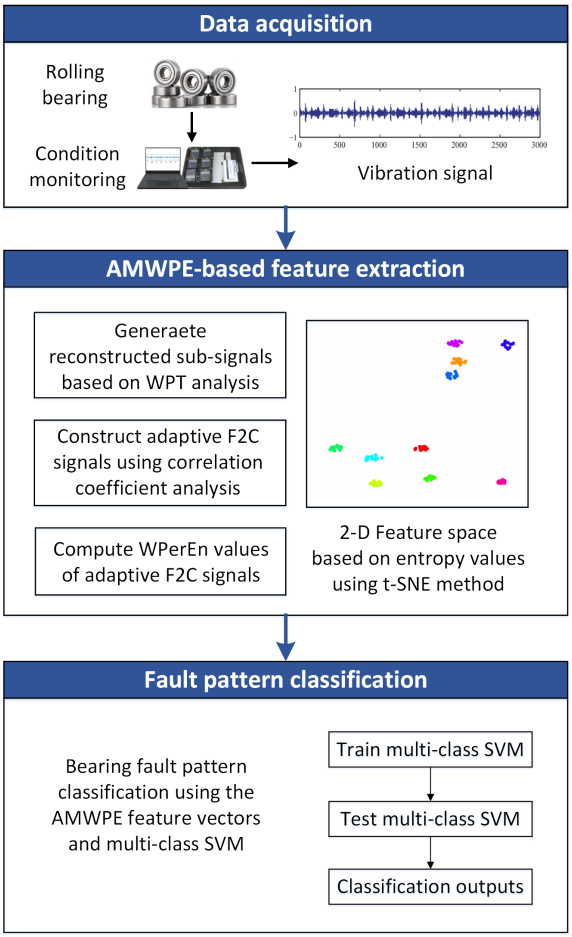
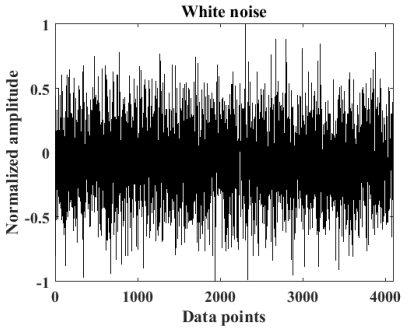


Fig. 4.12 Flowchart of the proposed bearing fault diagnosis method based on the AMWPE and SVM.

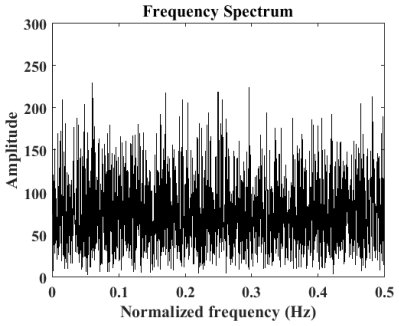
4.3 Numerical Evaluation

4.3.1 Analysis of Gaussian White Noise and 1/f Noise

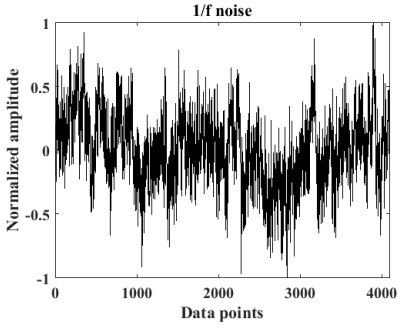
For investigating the performance of the AMWPE in entropy analysis, Gaussian white noises and 1/f noises with different data lengths (N = 512, 1024, 2048, 4096, and 8192) are analyzed in this study. Gaussian white noise is a random signal having equal intensity at different frequencies. 1/f noise, also termed pink noise, is a signal that its power spectral density is inversely proportional to the frequency of the signal. 1/f noise commonly exists in biological systems. The waveforms of white noise signals and 1/f noise signals are shown in Fig. 4.13. For comparison, the CMPE is also adopted for analyzing these signals, in which an improved coarse-graining procedure is applied. The scale factor $\tau = 32$, $m = 5$ and $\lambda = 1$ are set to calculate entropy values in these two entropy measures. In the calculation of AMWPE values, n_{max} is set to 32 as there is no training dataset for each type of noisy signal in the numerical study. Fig. 4.14 and Fig. 4.15 present calculated entropy values over 32 scales in the analysis of Gaussian white noise signals and 1/f signals, respectively.



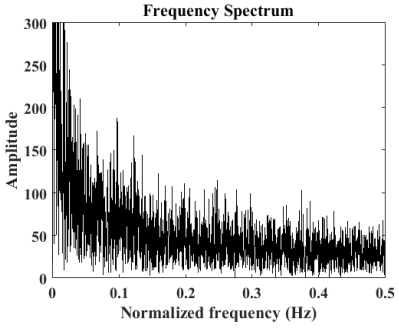
(a) Gaussian white noise signal



(b) Spectrum of Gaussian white noise signal



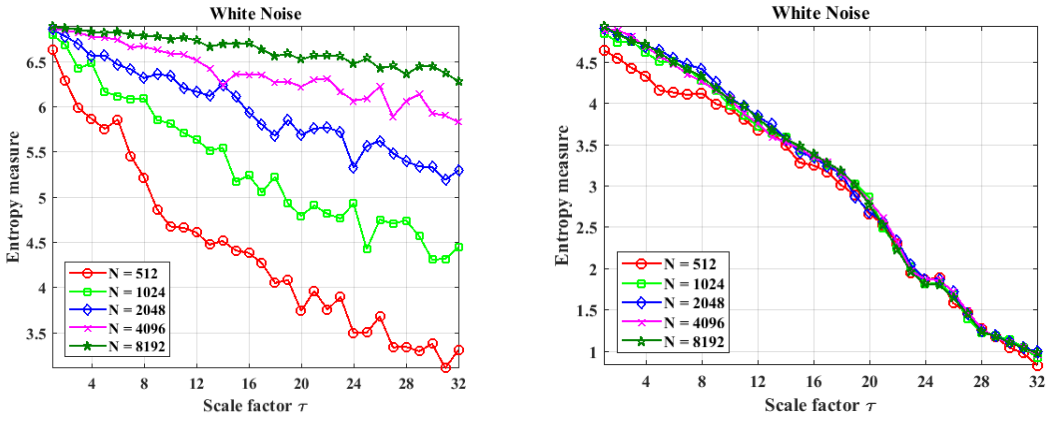
(c) 1/f noise signal



(d) Spectrum of 1/f noise signal

Fig. 4.13 Waveforms and spectrums of Gaussian white noise and 1/f noise signals.

From Fig. 4.14, it can be seen that the CMPE values sharply decrease over 32 scales with an increasing data length. The five curves representing different data length have a very large variance, which is due to the down-sampling operation in the coarse-graining procedure. In contrast, the AMWPE values exhibit stable entropy estimation under different data length, and the five feature curves have a very small deviation. This proves the advantage of the AMWPE in reducing bias in entropy values and the flexibility of choosing data length of time series according to the application scenario. Additionally, it is observed that the complexity of feature curves in two methods both decrease with an increasing scale, which can be explained that the extracted time series are becoming more and more smooth and regular, so presenting decreased irregularity and complexity degree. As a result, the larger the scale factor is, the smaller the entropy value is.



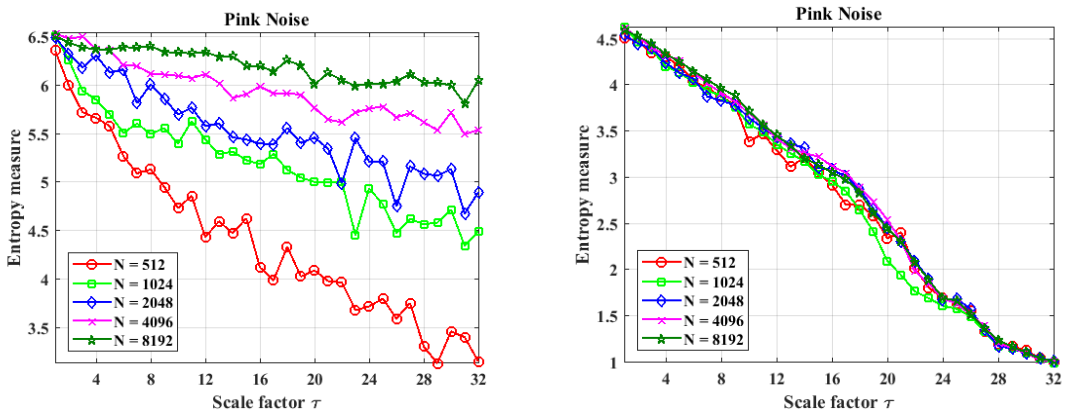
(a) CMPE values (b) AMWPE values

Fig. 4.14 Comparison of CMPE and AMWPE values of Gaussian white noise signals over 32 scales.

Similarly, from Fig. 4.15, one can observe that in the analysis of 1/f noise signals, the CMPE and F2CMPE feature curves present the same tendency. In the CMPE analysis, if a time series has less data length, its feature values will decrease to a low value more quickly with an increasing scale. However, the simulation results reveal that the AMWPE could provide consistent and stable values in entropy estimation under different data lengths.

4.3.2 Entropy Analysis of Signals with Different SNRs

Numerical mixed signals are applied to study the complexity analysis of the AMWPE and CMPE in feature representation. Gaussian white noise is added in a sinusoidal signal with a frequency 80 Hz, and the signals with SNRs: -5, 0, 5, 10, and 15 are considered in this analysis. The waveforms of noisy signals with different SNRs are illustrated in Fig. 4.16.



(a) CMPE values (b) AMWPE values

Fig. 4.15 Comparison of CMPE and AMWPE values of Gaussian 1/f signals over 32 scales.

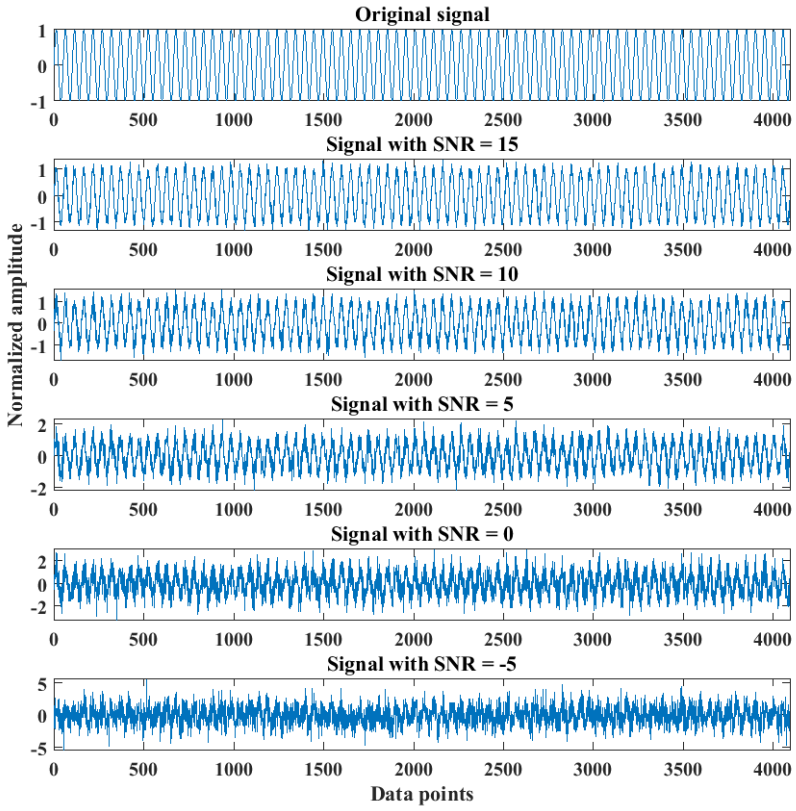
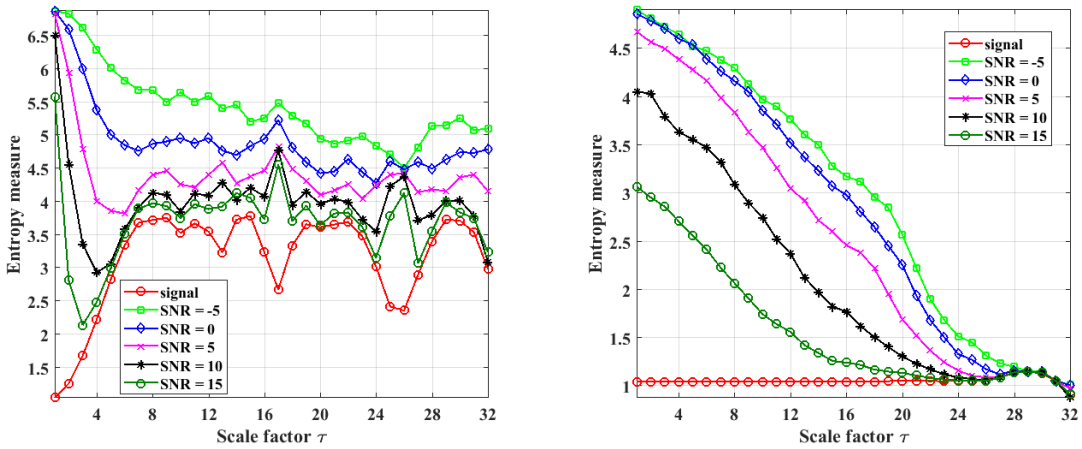


Fig. 4.16 Waveforms of the original sinusoidal signal and its noisy signals with different SNRs.

The AMWPE and CMPE values of mixed signals are then calculated, the results of which are presented in Fig. 4.17. For both AMWPE and CMPE feature curves, entropy values of signals with small SNRs are greater than that of signals with large SNRs, because the

former always has strong noise. The AMWPE and CMPE values of mixed signals are then calculated, the results of which are presented in Fig. 4.17. In regards to SNR signals, the smaller SNR value, the more noise in the signal. Thus, signals with small SNR values exhibit severe irregularity, and their waveforms have more unpredictability. As a result, for both AMWPE and CMPE feature curves, in the first several scales, entropy values of signals with small SNRs are greater compared to signals with large SNRs. However, the CMPE feature curves are very close to each other and are hard to identify. Also, their values fluctuate up and down over 32 scales and lack consistency in adjacent scales. That is, a signal with a larger scale often presents present less complexity degree and a smaller entropy value as their signal is becoming more regular. In contrast to the CMPE, the AMWPE exhibits abetter ability to distinguish between various noisy signals - their feature curves are separated from each other. It presents higher consistency over 32 scales - feature values of six signals stably decrease with an increasing scale. Therefore, the numerical analysis results demonstrate that the proposed AMWPE method has the stability of analyzing time series with different data lengths. It also has advantages of distinguishing between signals with different noise levels and producing consistent and stable entropy values as scale factor increases. Thus, the AMWPE measure has improved entropy estimation and is thus suitable for time series complexity analysis.



(a) CMPE values over 32 scales

(b) AMWPE values over 32 scales

Fig. 4.17 Comparison of CMPE and AMWPE values of synthetic noisy signals with different SNRs.

4.4 Summary

To provide reliable entropy estimation in bearing signal analysis, an improved multiple-scale entropy measure, named AMWPE, is proposed in the study. The AMWPE algorithm is aimed at providing reliable and stable entropy estimation based on the improved F2C procedure. It extracts not only low- and high-frequency information from the signal but also provides consistent entropy values with small variance as scale factor increases. Also, a new bearing fault diagnosis method is proposed based on the AMWPE and SVM classifier. Numerical studies are carried out to investigate the performance of entropy estimation in analyzing synthetic signals. Results demonstrate that the proposed AMWPE measure is robust to noise and can present satisfactory entropy values with high consistency and stability in differentiating between noisy signals.

Chapter 5

Case Studies for Fault Diagnosis of Rolling Bearing

To experimentally evaluate the proposed multiple-scale entropy measures, bearing vibration signals collected from three different machinery test rigs are analyzed, and the results are presented and discussed below. In three bearing datasets, two were measured from laboratory test rigs, and another one was acquired from an industrial-scale multistage centrifugal fan equipment. Three case studies are carried out, and their experimental setup and results are introduced for each case study, respectively. Moreover, to study the robustness of entropy analysis against noise, each bearing dataset is added with Gaussian white noise to construct noisy signals with different Signal-to-Noise Ratios (SNRs). Further, comparative experiments are carried out to compare the bearing diagnosis performance using traditional and improved entropy measures, respectively. Each case study starts with the description of the test rig and experiments, followed by diagnosis analysis using original and noisy bearing signals. Finally, experimental results and discussions are presented.

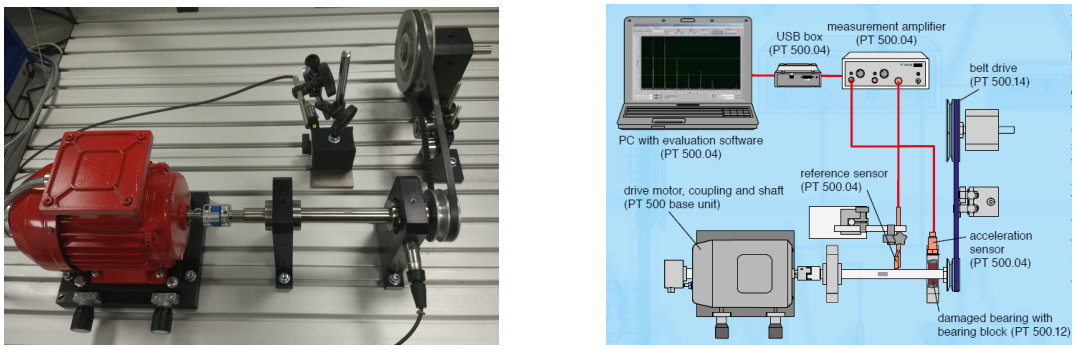
5.1 Bearing Health Diagnosis using Lincoln Dataset

5.1.1 Test Rig and Data Acquisition

In the first case study, bearing vibration signals were measured from a laboratory test rig, named PT 500 machinery diagnostic system [2], provided by the University of Lincoln, UK. The layout graph of this diagnostic system is shown in Fig. 5.1. This PT 500 test rig provides a roller bearing fault kit, which allows simulating various bearing health conditions and collecting vibration signals from the accelerometer sensor. This bearing fault kit is composed of motor assembly, motor control unit, shaft, four types of bearings, belt drive kit,

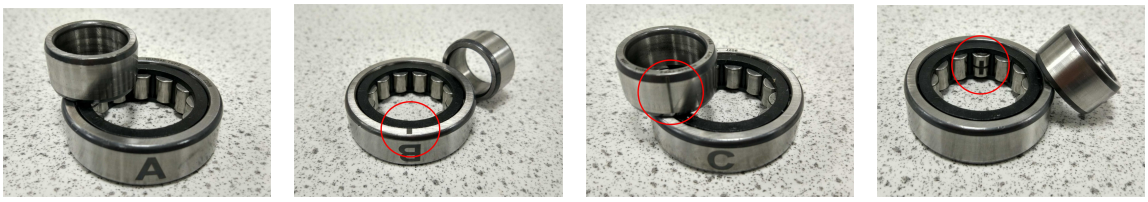
and computerised vibration analyser. The control unit is used to collect speed and horsepower data. The piezo-electric sensor and measuring amplifier are used for vibration measurement.

In this research, four states of roller bearing are considered, which are bearing A with the normal condition, bearing B with outer race damage, bearing C with inner race damage, and bearing D with rolling element damage, respectively. The type of roller bearing used in this study is NU204-E-TVP2. The inside diameter is 20 mm, the outside diameter is 47 mm, the width is 14 mm, and the number of rollers is 12. During the experiment, bearing vibration signals were collected under a sampling frequency of 8 kHz and a speed of 1500 r.p.m. Fig. 5.2 shows the original bearing signals with four health states. For each bearing state, there are 180 samples, and each sample contains 4,096 data points. Therefore, this bearing dataset contains 720 samples. The entire dataset is then split into two categories, namely 360 samples for training and 360 samples for testing, respectively. Fig. 5.3 shows the waveforms of original bearing signals with four states as well as their spectrums after FFT analysis.



(a) PT 500 test rig (b) Layout graph of the test rig

Fig. 5.1 PT 500 experimental test rig and its layout graph.



(a) Bearing A (b) Bearing B (c) Bearing C (d) Bearing D

Fig. 5.2 Four bearing health states with normal condition (A) and damages on the outer race (B), inner race (C), and roller element (D).

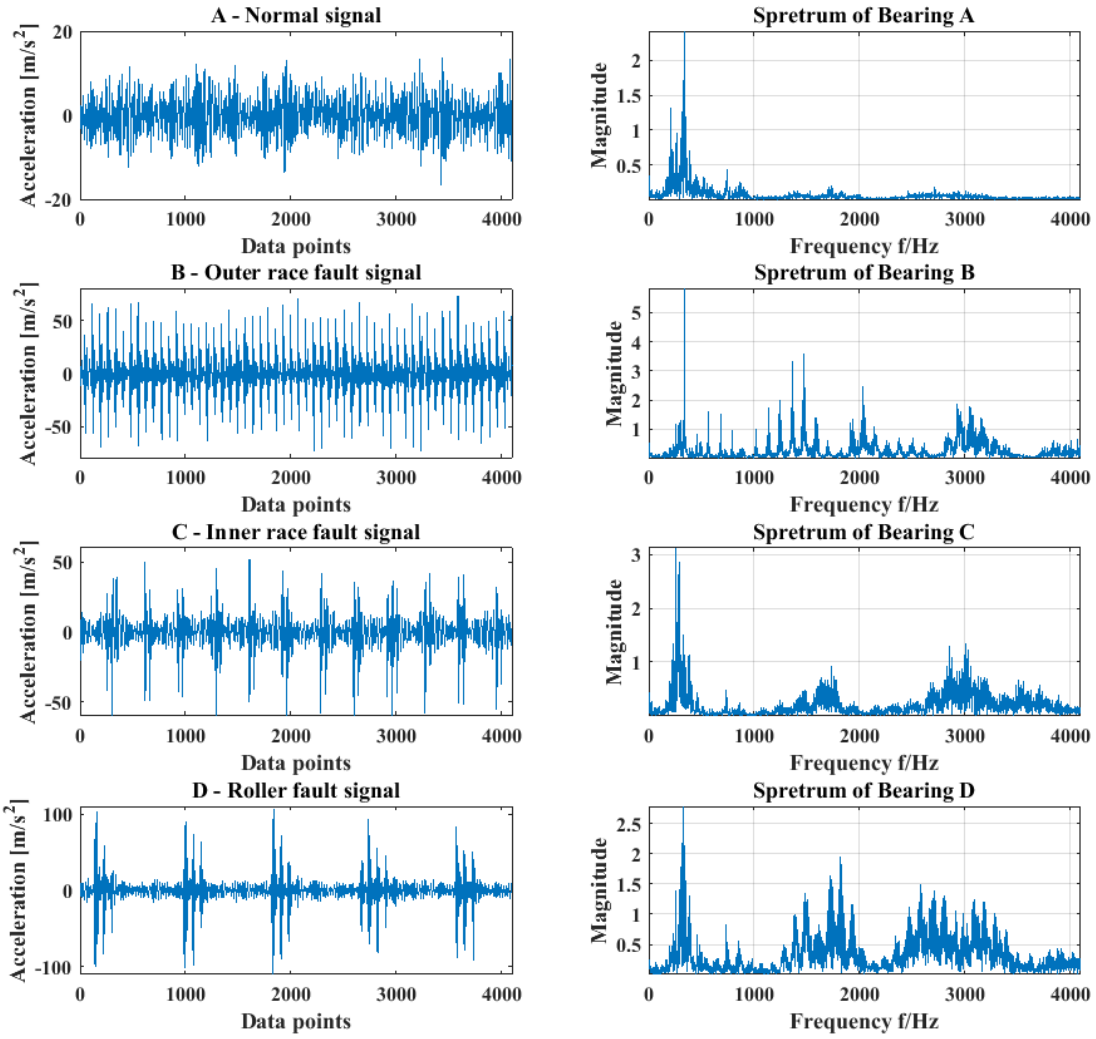


Fig. 5.3 Waveforms and spectrums of original bearing signals.

5.1.2 Experimental Analysis and Results

This experiment investigates the effectiveness of bearing diagnosis methods based on multiple-scale entropy measures. For comparison, the diagnosis performance of improved entropy measures is compared with conventional methods. First, the computation time of each entropy measure in the procedure of feature extraction is calculated and compared. A PC is used with the configuration (Intel Core i7-3770 Quad 3.40 GHz with 8GN of RAM on a Windows 7 operating system platform). Table 5.1 shows the average cost time of computing entropy values. In each entropy algorithm, PerEn or WPerEn values are calculated with specified parameters - embedding dimension $m = 5$, time delay $\lambda = 1$, scale factor and $\tau = 32$. From Table 5.1, it is observed that as the data length increases, the cost time of

Table 5.1 Cost time (s) of different entropy measures for feature extraction with $m = 5$, $\lambda = 1$, and $\tau = 32$ under different data length.

Data Length	$T_{\mathbf{MPE}}$	$T_{\mathbf{CMPE}}$	$T_{\mathbf{F2CMPE}}$	$T_{\mathbf{AMWPE}}$
512	0.1764	0.9606	1.0951	0.7673
1,024	0.2682	1.9614	1.8006	1.4398
2,048	0.4622	3.4939	3.1046	2.3976
4,096	1.0103	6.7844	5.7357	5.1682

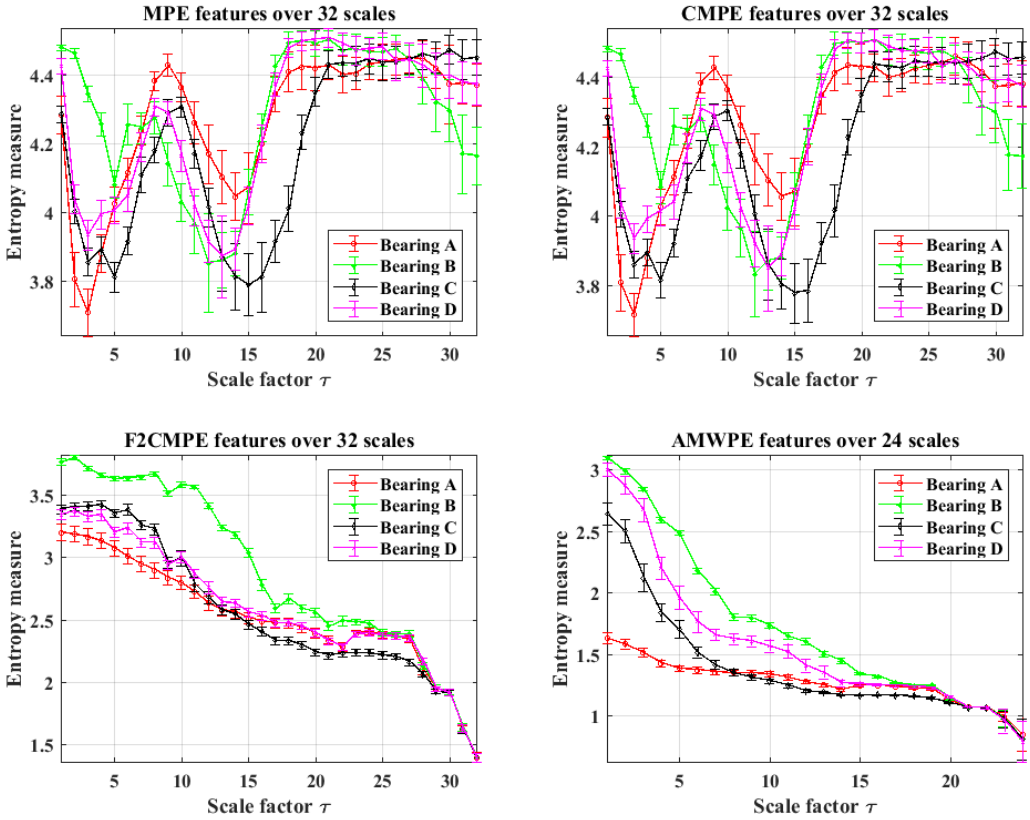


Fig. 5.4 Mean and standard deviation values of entropy features using the MPE, CMPE, F2CMPE, and AMWPE methods, respectively.

all entropy measures increases. On the one hand, when the data length is large, it will take more time to sort the adjacent values when calculating the PerEn value. On the other hand, when the m and λ are specified, the calculation time of the PerEn value is almost the same; thus, the calculation efficiency mainly depends on the scale-extraction procedure. For the traditional MPE method, the PerEn value is calculated based on a linear smoothing operation

so that the shortest calculation time can be obtained. The CMPE consumes more time than MPE because the PerEn value is an average value obtained from multiple time series on one scale - except the first scale; therefore, the larger the scale, the more time it takes. Also, since the low- and high-frequency information is extracted, the F2CMPE value consumes a little more time than that of the MPE and CMPE. In contrast, calculating AMWPE values save time compared to the CMPE and F2CEMP algorithms because a smaller number of F2C signals are applied to calculate PerEn values.

For evaluating the capability of distinguishing between different bearing health states, vibration signals are analyzed using entropy measures, and entropy feature values are obtained ($m = 5$, $\lambda = 1$, and $\tau = 32$). The mean and standard deviation values of four entropy features are shown in Fig. 5.4. From the figure, it can be seen that the MPE feature curves - representing four bearing conditions - are mixed with each other, which is difficult to distinguish. As the scale factor increases, the four curves change up and down and intersect each other. Moreover, the CMPE feature curves are similar to the MPE feature curve over 32 scales. In comparison, the F2CMPE feature curves can be distinguished from each other when the scale factor is small. The number of AMWPE features in this case study is 24 ($n_{max} = 24$) because the improved F2C procedure selects the F2C signals that are most relevant to the original signal from the training datasets. As the scale factor increases, the F2CMPE and AMWPE entropy values gradually decrease, which can be explained as the F2C signal becomes more and more smooth with a larger scale factor, so its complexity degree decreases. It is obvious that when the scale factor is small, the AMWPE feature curves are separated from each other, so the health condition of rolling bearing can be easily identified by observing these waveforms directly. After feature extraction, the feature vectors obtained using different entropy measures are then input into a multi-class SVM for fault pattern recognition. Table 5.2 lists accuracy results for bearing fault diagnosis using original vibration signals.

Table 5.2 Comparison of diagnosis performance on Lincoln bearing dataset using entropy measures and the SVM classifier.

Method	MPE	CMPE	F2CMPE	AMWPE	parameters
Accuracy (%)	98.3	99.3	99.7	100	$m = 5, \lambda = 1$
Accuracy (%)	96	98.3	99.3	100	$m = 5, \lambda = 2$
Accuracy (%)	99.3	100	100	100	$m = 5, \lambda = 3$

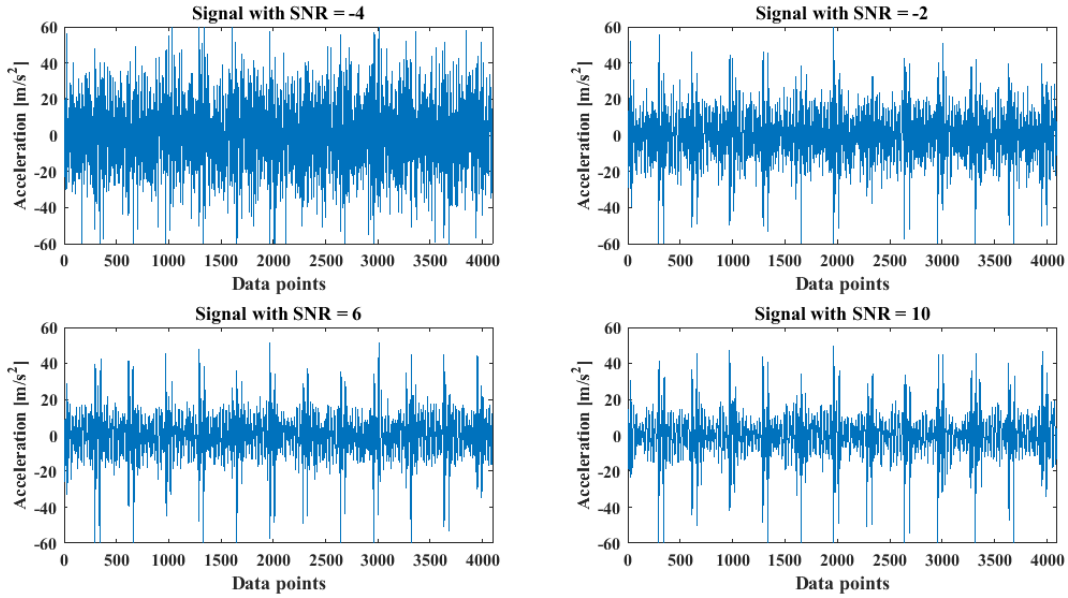


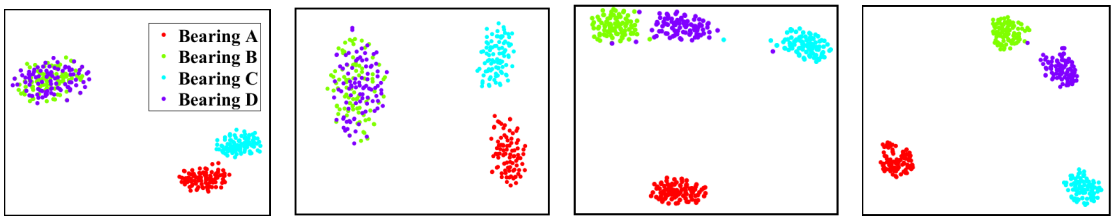
Fig. 5.5 Gaussian white noise added bearing signals with inner race fault under different SNRs: -4, -2, 6, 10.

In practical applications, rotating machinery usually operates in a noisy working environment. It is necessary to study the robustness of the diagnosis model to external disturbance and noise using different entropy measures. To further compare the robustness of various entropy measures to noise, the original bearing vibration signals are added with Gaussian white noise. The SNR is defined as the ratio of the power of a signal to the power of background noise in decibels (dB):

$$\mathrm{SNR} = 10\log_{10}(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}) \tag{5.1}$$

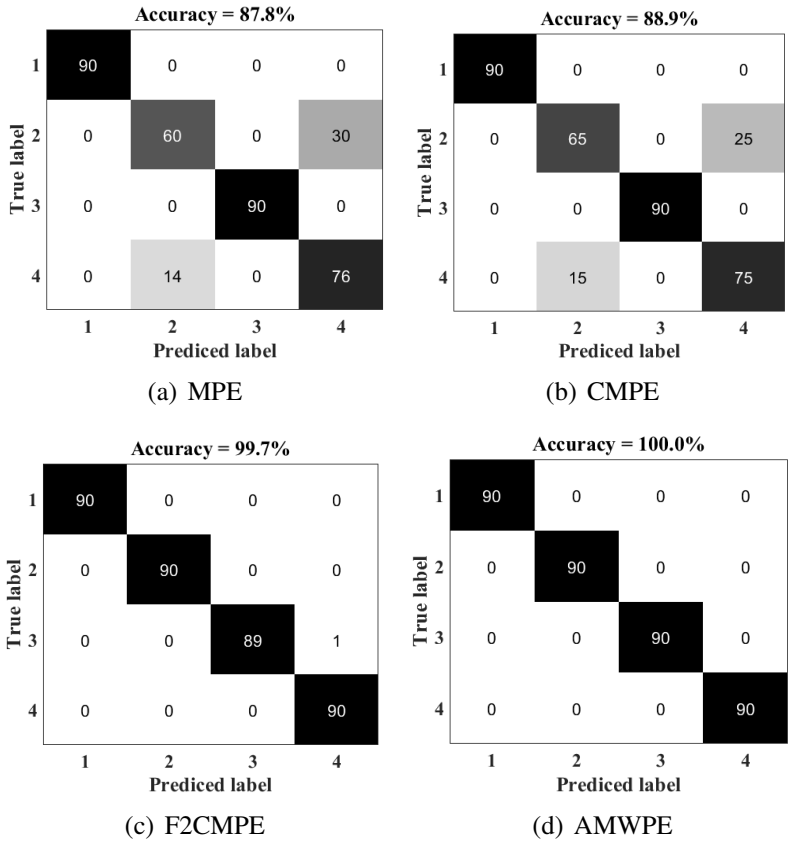
The constructed noisy bearing signals are with different SNRs from -4 to -14 dB. Fig. 5.5 presents waveforms of bearing signals with various SNRs. Obviously, when the SNE decreases (the noise level is high), the waveform of the signal will be more complicated, and it will be more difficult to extract fault information accurately. For example, the important characteristics of the original signal are contaminated servery by the noise in the signals with SNR = -4 compared to signals with SNR = 10.

For comparison, an example is given here where bearing signals with SNR = -2 are considered. First, the MPE, CMPE, F2CMPE, and AMWPE algorithms are used to calculate the entropy features over 32 scales from the noisy signals, respectively. Fig. 5.6 shows the visualized features in a two-dimensional feature space using the t-SNE method. The t-SNE technique visualizes high-dimensional data by mapping it to a two-dimensional feature space while still preserving the high dimensional clustering relationship. [194]. From Fig. 5.6,



(a) MPE (b) CMPE (c) F2CMPE (d) AMWPE

Fig. 5.6 Reduced 2-D feature space of bearing signals with SNR = 2 using t-SNE.



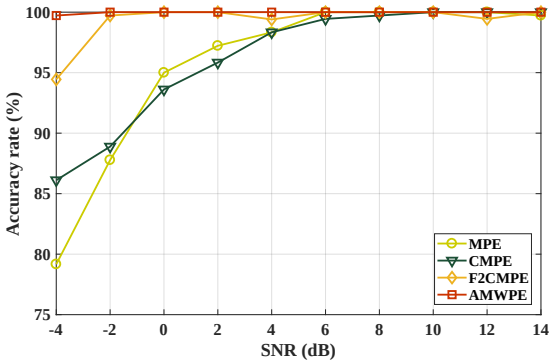
(a) MPE (b) CMPE

(c) F2CMPE (d) AMWPE

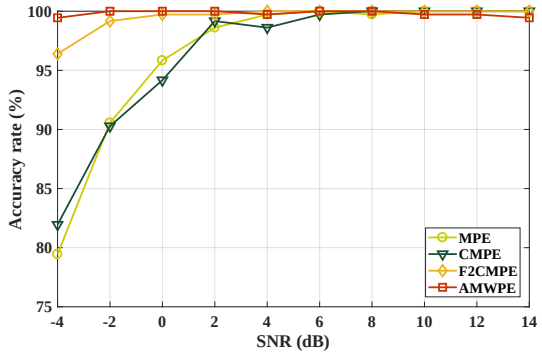
Fig. 5.7 Confusion matrix results based on different entropy measures.

it can be seen that the samples of bearing A and C are easy to distinguish, but bearing B and D are mixed together in the feature space of MPE and CMPE. In contrast, the feature space of F2CMPE and AMWPE can clearly differentiate between different bearing states, thereby providing higher accuracy of bearing fault diagnosis. Further, after feature extraction, the entropy feature vectors are input into the SVM classifier for fault pattern identification. The accuracy results of diagnosis methods using different entropy measures are presented in Fig. 5.7 based on the confusion matrix. The confusion matrix indicates the number of

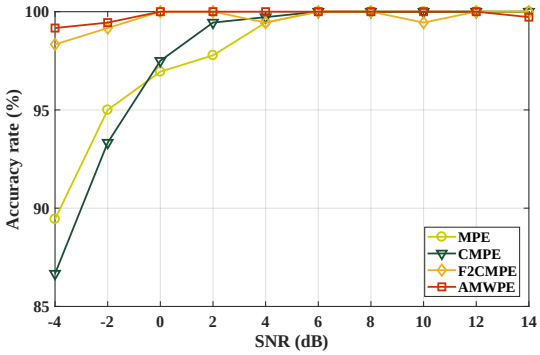
correct and incorrect predictions when classifying bearing conditions. From Fig. 5.7, it is apparent that the accuracy of the diagnosis method using different entropy measures is in accordance with the performance of its feature space. That is, if a method can more easily separate different bearing health states in the t-SNE feature space, its diagnosis accuracy is higher.



(a) Results with $m = 5, \lambda = 1$



(b) Results with $m = 5, \lambda = 2$



(c) Results with $m = 5, \lambda = 3$

Fig. 5.8 Fault diagnosis performance on noisy signals with various SNRs using entropy methods under $m = 5$ and $\lambda = 1 - 3$.

To further verify the robustness and stability of improved entropy measures in the analysis of noisy signals, three more experiments were performed, where λ ranges from 1 to 3, and $m = 5$. Fig. 5.8 shows the accuracy results. When the noise level is very high, the bearing vibration signals are severely contaminated and their complexity level increases, resulting in a reduced diagnosis performance. From the figure, when SNR = -4, the diagnosis accuracy rates of diagnosis methods using traditional entropy algorithms are about 85%. In contrast, the improved entropy methods, F2CMPE and AMWPE, can achieve more than 95% accuracy. It is obvious that the diagnosis method using the improved entropy measures is robust to signals with strong noise.

In summary, the improved entropy measures apply to the fault diagnosis of rolling bearing. In this case study, the improved entropy measures exhibit high diagnostic performance and are suitable for distinguishing between four fundamental bearing health conditions. Through the validation of noisy signal analysis further, experimental results demonstrated that the improved entropy measures have high robustness against noise. Moreover, as the noise level in the bearing signal increases, the diagnosis performance using the improved entropy measure is still satisfactory and relatively stable.

5.2 Bearing Health Diagnosis using CWRU Dataset

5.2.1 Test Rig and Data Acquisition

In the second case study, the experimental rolling bearing dataset is provided by Case Western Reserve University (CWRU) [195] bearing data centre. The schematic of the CWRU test rig is shown in Fig.5.9. This test rig comprises a 0 to 3 horsepower electric motor, a torque transducer, a dynamometer, control electronics, and SKF deep-groove ball bearings with the type of 6205-2RS JEM. In data acquisition, bearing vibration data was collected from the drive end channel. Ten health states of the rolling bearing are considered, including four conditions of bearing fault and three different fault severity levels. Single point failures were introduced into SKF bearings using electro-discharge machining with local fault diameters of 0.1778 mm, 0.3556 mm, and 0.5334 mm and fault depth of 0.2794 mm.

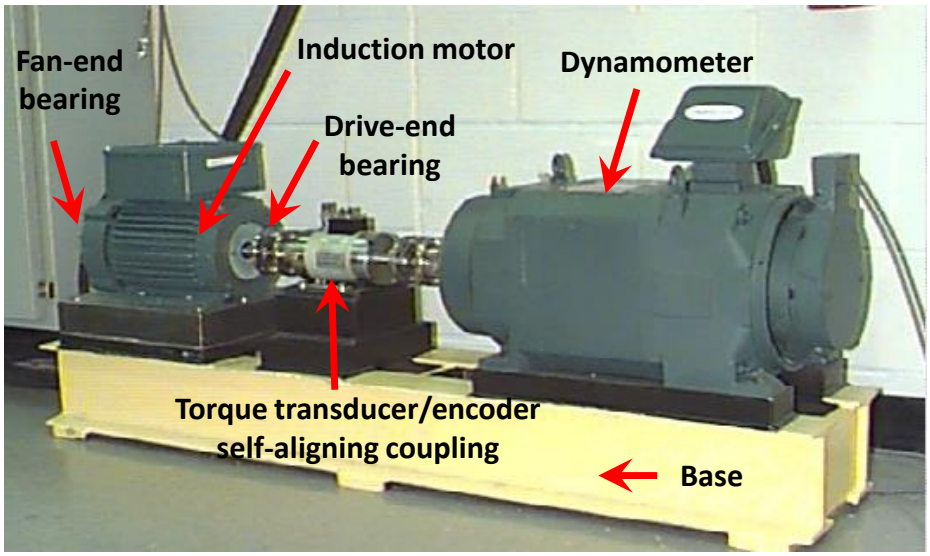


Fig. 5.9 CWRU bearing test rig

Table 5.3 Description of the bearing state and its class label in the CWRU bearing dataset.

States	Fault diameter	Class label	States	Fault diameter	Class label
IR1	0.1778	1	OR3	0.5334	6
IR2	0.3556	2	BE1	0.1778	7
IR3	0.5334	3	BE2	0.3556	8
OR1	0.1778	4	BE3	0.5334	9
OR2	0.3556	5	Norm	0	10

In this experiment, vibration data of rolling bearing include ten conditions, i.e., the normal condition (Norm), and the damages on the inner race (IR), the outer race (OR) at 6 o'clock, and the ball element (BE). In this study, three defect sizes (namely, 0.1778 mm, 0.3556 mm, and 0.5334 mm) of single-point fault were considered as fault states. The rotating speed is 1730 r/min with Load 3 HP, and the sampling rate is 12 kHz. These ten conditions are labelled as Norm, IR1, IR2, IR3, OR1, OR2, OR3, BE1, BE2, and BE3, respectively. Then, these vibration signals were split into a set of non-overlapping segments with a specified data length ($N = 4,096$). The detail specification of each rolling bearing state is presented in Table 5.3. Fig. 5.10 shows the original vibration signals of rolling bearing under ten health conditions.

5.2.2 Experimental Analysis and Results

The experiment first compares the computation time of four entropy measures in analyzing the CWRU bearing vibration signal in the second case study. The average cost time results are shown in Table 5.4. As with the first case study, the MPE algorithm consumes the least time. Compared with CMPE and F2CMPE, the AMWPE entropy measure has higher computational efficiency.

For detecting and identifying bearing health conditions, vibration signals were analyzed using four entropy measures for feature extraction. Further, to compare the performance of fault feature extraction, entropy feature values are visualized in Fig. 5.11. It can be seen that most of the MPE and CMPE features are mixed with each other on 32 scales, and their features have large standard deviation values. In comparison, both F2CMPE and AMWPE feature values have relatively small standard deviation values. Moreover, the feature curves of the ten bearing states over the first 10 scales can be distinguished by observing

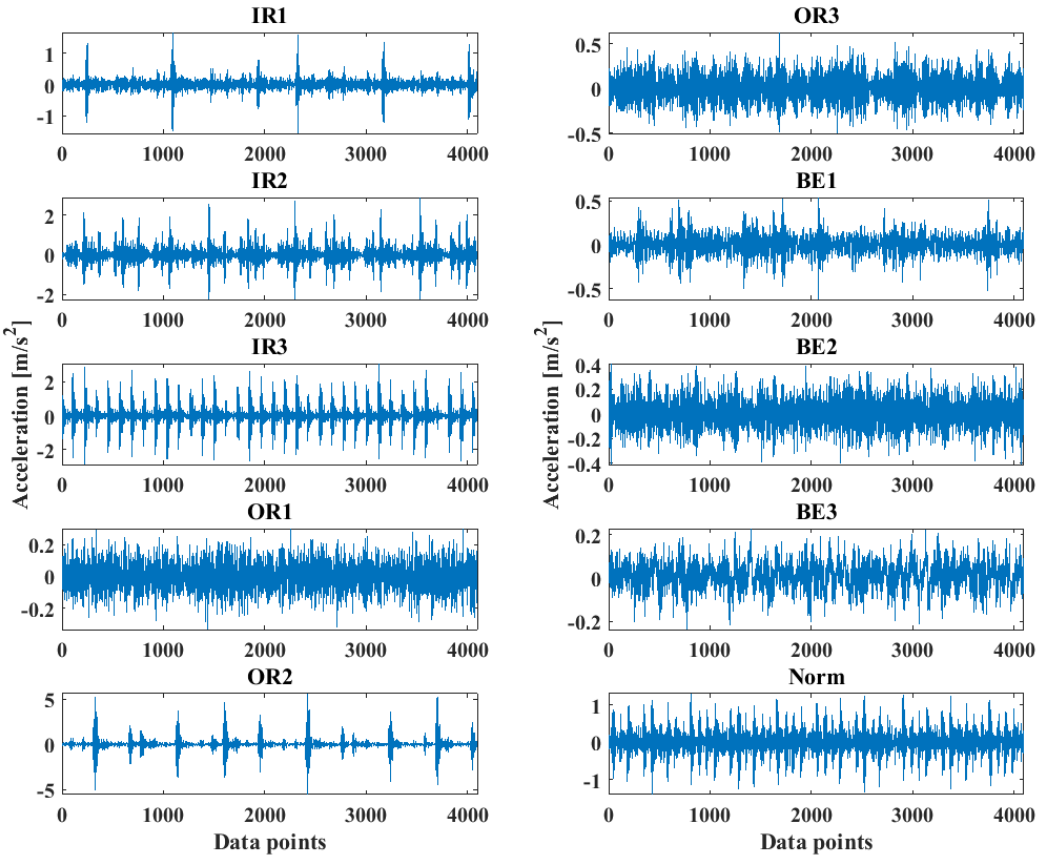


Fig. 5.10 Time-domain waveforms of bearing vibration signals with ten conditions of rolling bearing.

the AMWPE features. Later, feature vectors were input into the SVM for fault pattern recognition. Diagnosis results using the four entropy measures based on the CWRU bearing dataset are shown in Table 5.5.

The robustness of the improved entropy measures to noise is further investigated in the second case study (where ten bearing states are considered). Noise signals are produced where their SNR ranges from -4 to -14 dB. A comparison of the original and noisy signals is given in Fig 5.12. From the figure, the overall trend of the noisy signal (with SNR = -2) and its several peak values can still be identified by observing the waveform. However, when the noise level increases to SNR = -4, the IR signal will be seriously contaminated by high-frequency components, which makes it difficult to correctly diagnose the health condition of rolling bearings.

For comparison, an experiment was carried out where bearing signals with SNR = 2 are considered in analyzing CWRU bearing data. Entropy features were extracted from noisy signals using entropy measures with parameters of $m = 5, \lambda = 1, \text{scale} = 32$. The obtained

Table 5.4 Cost time (s) of different entropy measures for feature extraction with $m = 5$, $\lambda = 1$, and $\tau = 32$ under different data length.

Data Length	$T_{\mathbf{MPE}}$	$T_{\mathbf{CMPE}}$	$T_{\mathbf{F2CMPE}}$	$T_{\mathbf{AMWPE}}$
512	0.1665	0.8352	1.0103	0.5272
1,024	0.2520	1.7130	1.6416	1.0647
2,048	0.4251	3.1223	2.9379	2.0959
4,096	0.7965	6.2057	5.5404	4.0092

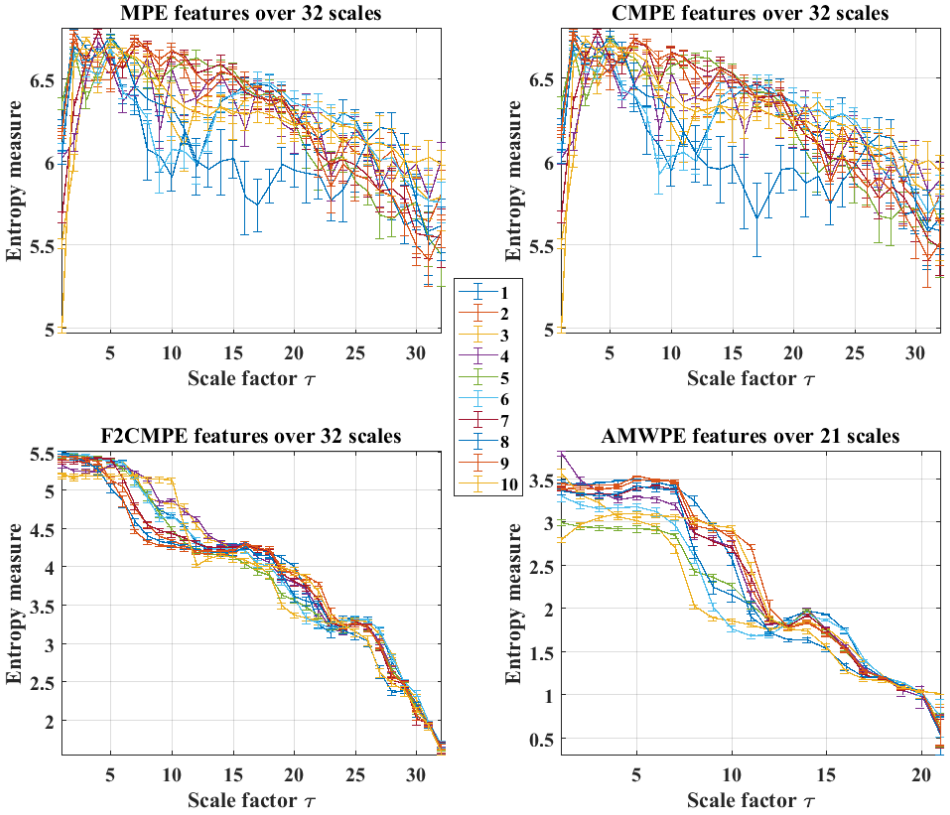


Fig. 5.11 Mean and standard deviation values of entropy features over 32 scales.

feature values are displayed in reduced 2-D feature space, as shown in Fig. 5.13. From the figure, most of the bearing samples represented by the MPE and CMPE feature values are mixed and cannot be easily distinguished. In comparison, the samples in the feature space of F2CMPE and AWMPE are relatively separated and are easy to identify.

Besides, several more experiments were performed to study the robustness of various entropy measures to different noise levels. To be more specific, a set of bearing signals

Table 5.5 Comparison of diagnosis performance on CWRU bearing dataset using entropy measures and the SVM classifier.

Method	MPE	CMPE	F2CMPE	AMWPE	parameters
Accuracy (%)	98.3	99.3	100	100	$m = 5, \lambda = 1$
Accuracy (%)	96	98.3	99.3	100	$m = 5, \lambda = 2$
Accuracy (%)	99.3	100	100	100	$m = 5, \lambda = 3$

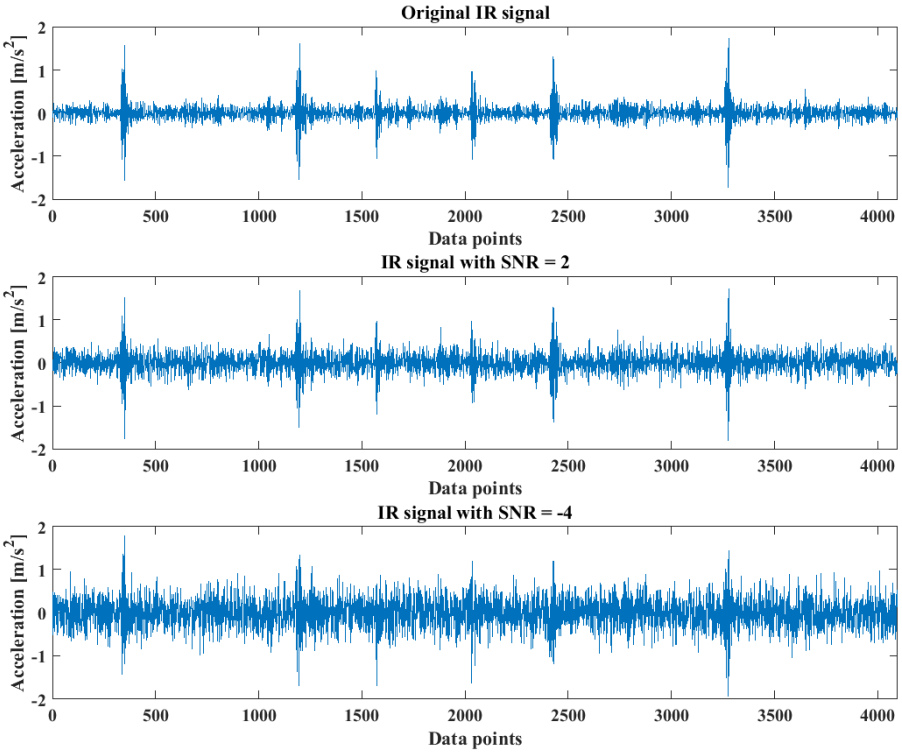
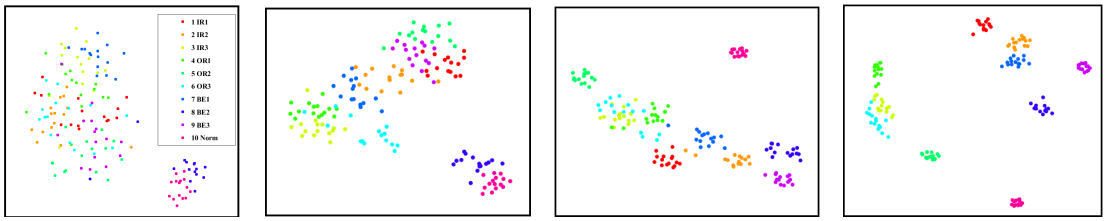


Fig. 5.12 Original and noisy bearing vibration signals with inner race fault in the CWRU bearing dataset.

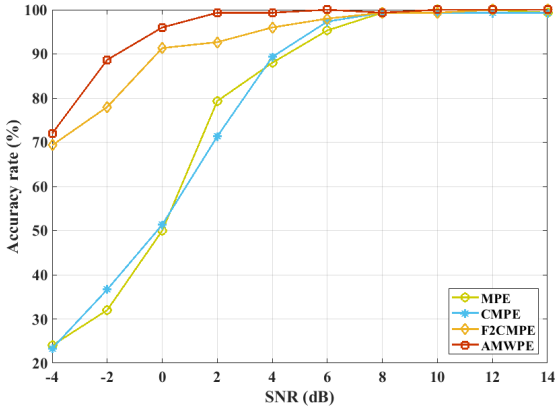
were tested where their SNR levels range between -4 and 14. Moreover, the diagnosis performance using entropy measures is compared where entropy feature values are calculated using different time delay parameters - $m = 5, \lambda = 1, 2, 3$, respectively. The diagnosis results are presented in Fig. 5.14.

In summary, the improved entropy measures apply to detect and diagnose ten bearing health conditions. Feature values calculated using the improved entropy measures not only have small standard deviation values but also can differentiate between different bearing states. In addition, diagnosis results with high accuracy rates demonstrated the effectiveness

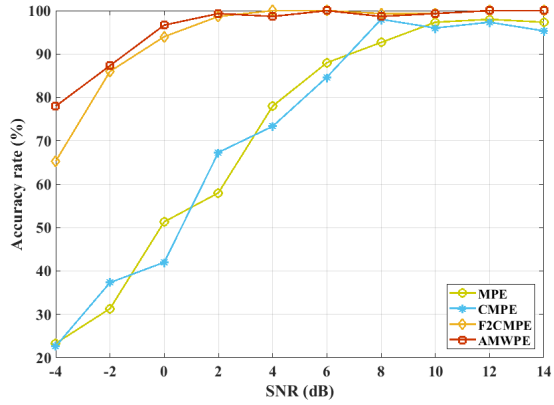


(a) MPE (b) CMPE (c) F2CMPE (d) AMWPE

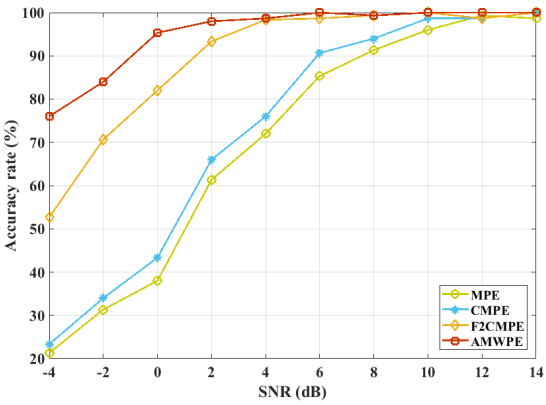
Fig. 5.13 Reduced 2-D feature space of bearing signals with SNR = 2 based on t-SNE.



(a) Results with $m = 5, \lambda = 1$



(b) Results with $m = 5, \lambda = 2$



(c) Results with $m = 5, \lambda = 3$

Fig. 5.14 Fault diagnosis performance on noisy signals with various SNRs using entropy features under $m = 5$ and $\lambda = 1 - 3$.

of the proposed method for bearing diagnosis, where features are extracted from the improved entropy measures and the SVM is used for fault pattern recognition, respectively.

5.3 Bearing Health Diagnosis using GDUPT Dataset

5.3.1 Test Rig and Data Acquisition

In the third case study, the bearing dataset [196] is collected from an industrial-scale multi-stage centrifugal air pump unit at the Guang Dong University of Petrochemical Technology (GDUPT), China. The schematic of this equipment is shown in Fig. 5.15. This unit comprises a base, electrical motor, torque converter, gearbox, rolling bearing, and air compressor. The specification of the multistage centrifugal air compressor unit is described in Table. 5.6.

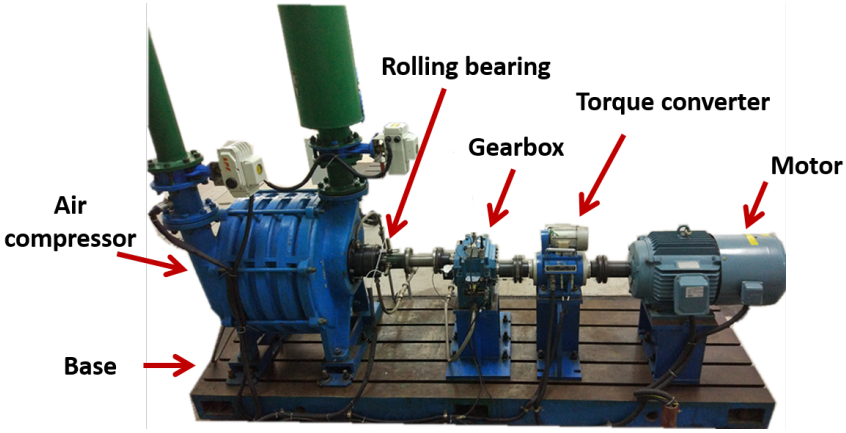
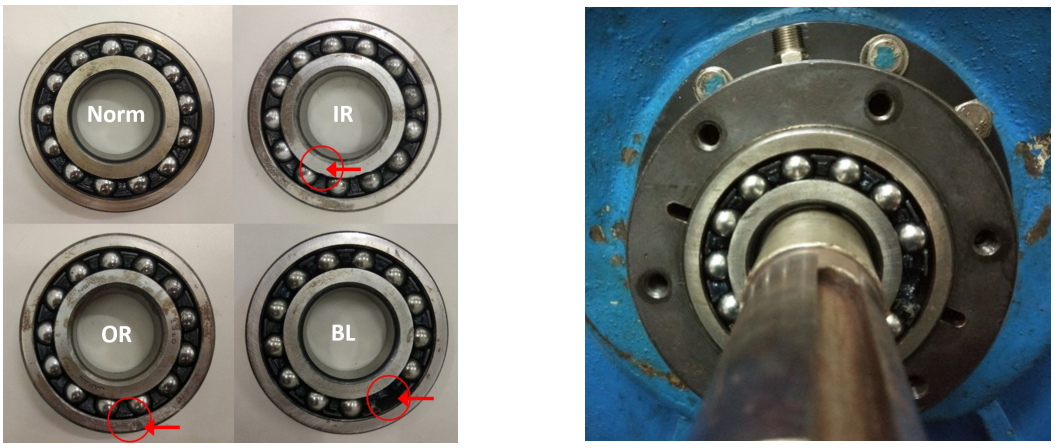


Fig. 5.15 Multistage centrifugal air pump unit in the GDUPT.

Table 5.6 Detailed specification of the tested multistage centrifugal air compressor unit.

Device	Model	Parameters
Inverter motor	YP-50-112	Triangular junction Circuit 380V, rated voltage 24.8 A, motor rated power 11 kw
Air compressor	C8-2000	Rated power = 11 kw, maximum speed 2970 r/min, blowing rate 8 m^3/min
Torque converter	CYT-302	Rated torque 100N·m, speed range 0 3000 r/min, temperature coefficient -0.027% \ °C, precision ± 0.2% FS
Data collector	EMT 390	Acceleration 0.1 199.9 m/s^2, acquisition frequency 10 Hz-10 kHz, number of groups 1 100, sampling rate: 512,1024,2048Hz

In this case study, two experiments are carried out to study the diagnosis performance of the proposed method for single-fault bearing fault diagnosis and multi-fault bearing fault diagnosis, respectively. In the single-fault bearing diagnosis test, four bearing health states are considered, including the normal bearing (Norm), bearing with inner race (IR) fault, bearing with outer race (OR) fault, and bearing with a rolling ball lacked (BL). Fig. 5.16



(a) Four types of bearings (b) Bearing house

Fig. 5.16 Four types of rolling bearings and bearing house in the multistage centrifugal air pump equipment.

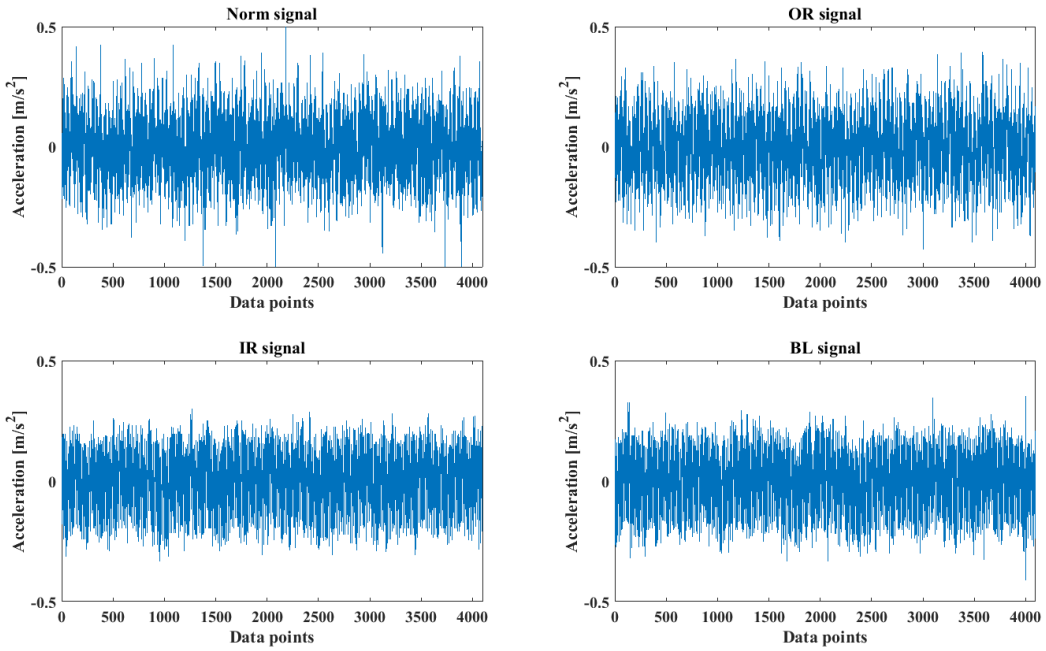


Fig. 5.17 Waveforms of bearing and gearbox vibration signals.

shows four types of bearings and bearing house. In the multi-fault bearing diagnosis, six machinery health conditions are considered including three bearing faults, one gearbox fault, and two compound faults. Specifically, three bearing health states include normal bearing (Norm), bearing with inner race (IR) fault, bearing with outer race (OR) fault. Gearbox with one missing teeth (MT) fault is considered. There are two compound faults that combine bearing fault and gearbox fault together - IR fault with MT fault, and OR fault with MT fault,

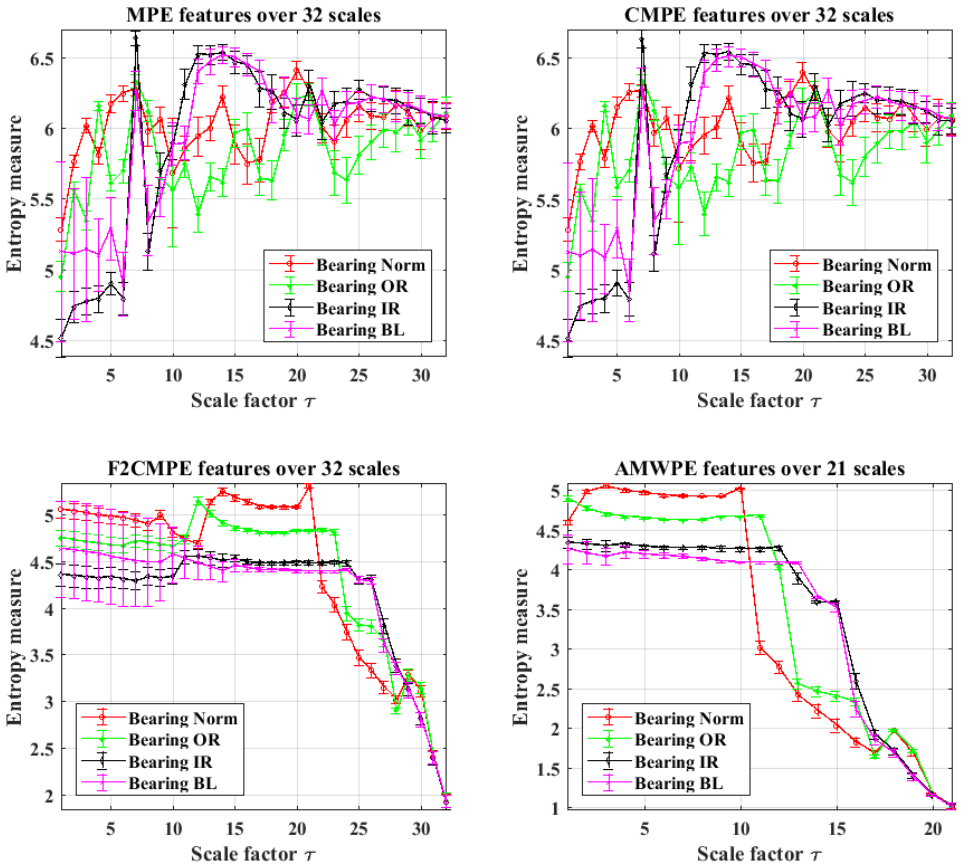


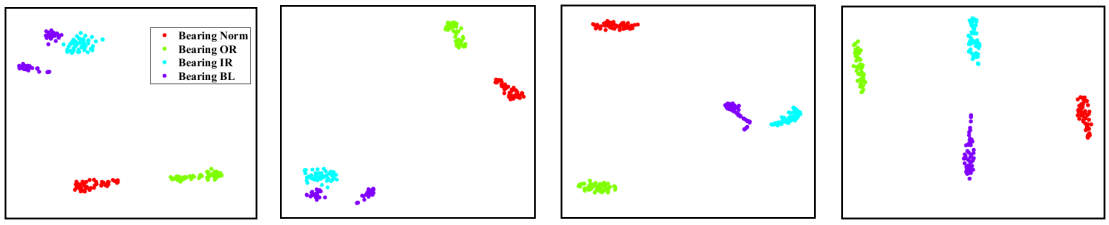
Fig. 5.18 Mean and standard deviation values of entropy features over 32 scales.

respectively. In the two experiments, vibration signals were collected from the multistage centrifugal air pump equipment using an EMT390 data collector with a sampling rate of 1,024Hz. The speed of rotation is 1000 r.p.m, and the motor rated power is 11kW. Measured signals are analyzed using the proposed diagnosis methods.

5.3.2 Analysis and Results of Single-fault Bearing Diagnosis

The waveforms of single-fault bearing vibration signals are shown in Fig. 5.17. As can be seen, signals collected from real industrial-scale equipment are far more complicated compared to the small-scale test rig (Lincoln and CWRU test rig). Measured signals have a high-level noise caused by background noise and additional vibrations generated by coupling components.

In this case study, bearing vibration signals are first analyzed using entropy measures with specific parameters ($m = 5, \lambda = 1$, and scale $= 32$). Extracted entropy values of four types of bearing signals are presented in Fig. 5.18. It is obvious that the MPE and CMPE



(a) MPE (b) CMPE (c) F2CMPE (d) AMWPE

Fig. 5.19 Reduced 2-D feature space of original vibration signals using different entropy measures based on t-SNE.

Table 5.7 Comparison of diagnosis performance on GDUPT single-fault bearing dataset based on entropy measures and the SVM classifier.

Method	MPE	CMPE	F2CMPE	AMWPE	parameters
Accuracy (%)	98	98.3	100	100	$m = 5, \lambda = 1$
Accuracy (%)	95.6	96	98	99.3	$m = 5, \lambda = 2$
Accuracy (%)	96	98	99.3	100	$m = 5, \lambda = 3$

feature curves representing four types of bearings are mixed together and are difficult to identify. With an increasing scale factor, these four curves gradually overlap, making it difficult to provide valuable feature information. Comparatively, the proposed F2CMPE and AMWPE feature curves can better distinguish between four bearings when the scale factor ranges in the middle value. When the scale factor increases from 1, the high-frequency information is gradually stripped from the signal, and the low-frequency signal containing the fault information can gradually distinguish different bearing fault characteristics. Fig. 5.19 plots the reduced 2-D feature space of four entropy feature values. One can see that the way data sample distributes in the feature space is in line with feature curves presented in Fig. 5.18. The proposed F2CMPE and AMWPE feature values properly separate four health states of bearings in the feature space. Diagnosis accuracy results using four entropy measures (under $m = 5$ and $\lambda = 1 - 3$, and scale $= 32$) and SVM are presented in Table. 5.7.

5.3.3 Analysis and Results of Multi-fault Bearing Diagnosis

In the multi-fault bearing diagnosis, an additional gearbox fault states is considered, and it is combined with bearing inner race fault and outer race fault, respectively. Compared to single-fault diagnosis, multi-fault diagnosis has the difficulty in classifying single faults and

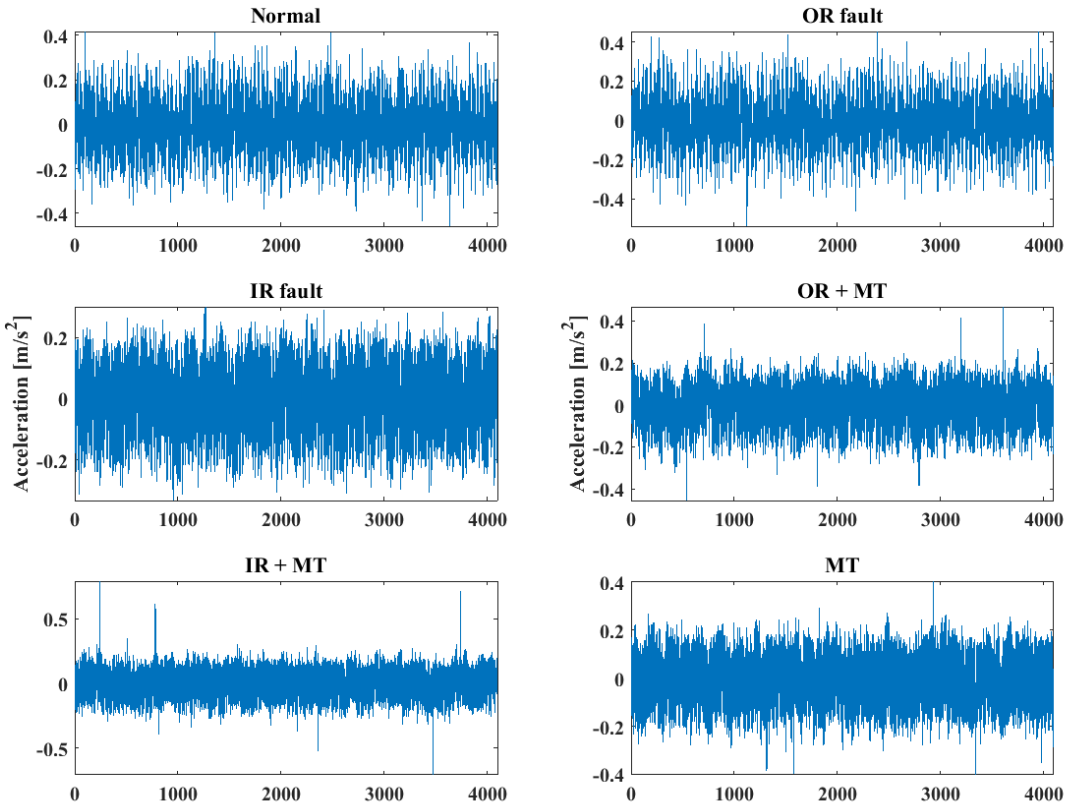


Fig. 5.20 Waveforms of bearing and gearbox vibration signals.

compound faults. The waveforms of multi-fault vibration signals are shown in Fig. 5.20. With the gearbox fault, waveforms of compound faults are always more complicated than single-fault bearing signals. Also, the coupling components will exhibit more non-linear characteristics, thus making it more difficult to appropriately extract fault information and identify fault patterns correctly.

In this case research, vibration signals are analyzed using four different entropy measures ($m = 5, \lambda = 1$, and scale $= 32$). The mean and standard deviation values of entropy feature values, respectively, extracted from four entropy algorithms are shown in Fig. 5.21. It can be observed that the MPE and CMPE feature values still present irregular and chaotic waveforms over 32 scales. However, the F2CMPE and AMWPE feature values exhibit more separated feature curves that can distinguish six bearing health states. Compared with single-fault bearing diagnosis, the feature values of bearings with compound faults are mixed with the feature values of single faults. For example, in the F2CMPE features, IR feature curve and "IR + MT" curve are intertwined and thus are difficult to distinguish. In contrast, with the AMWPE algorithm, feature values of bearing signals under six conditions have relatively lower standard deviation values. Also, these six machinery states can be separated from the

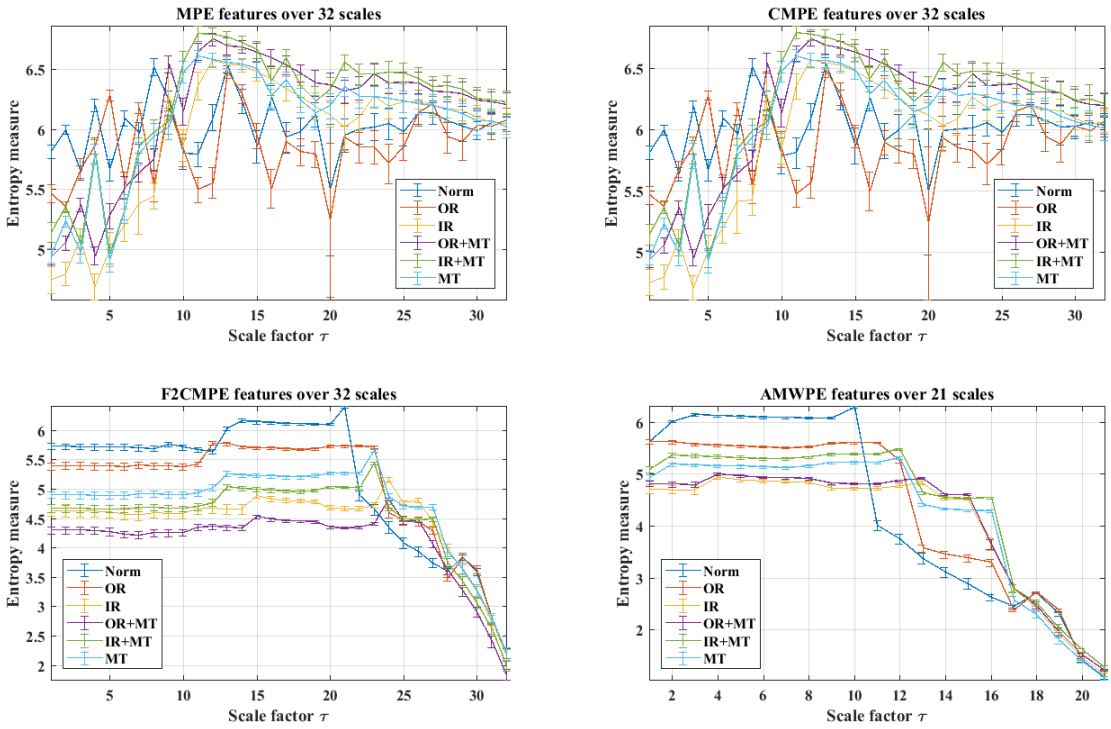
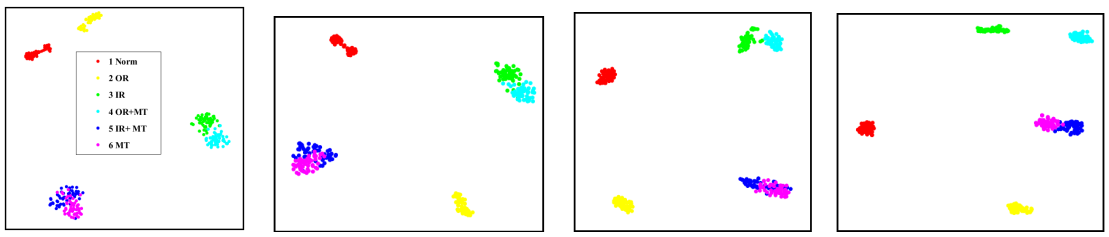


Fig. 5.21 Mean and standard deviation values of four different entropy measures.

feature curves, thus it can give a higher fault classification accuracy in multi-fault machinery diagnosis. The feature spaces of four types of entropy features are displayed in Fig 5.22.

To verify the effectiveness of entropy measures for bearing diagnosis, feature vectors were then input into the SVM for fault pattern classification. Table 5.8 shows the diagnosis performance results using four different entropy measures for comparison. It is observed that the improved entropy measures gain higher accuracy rate compared to the MPE and CMPE. Their diagnosis performance is stable when the time delay parameter λ ranges from 1 to 3.



(a) MPE (b) CMPE (c) F2CMPE (d) AMWPE

Fig. 5.22 Reduced 2-D feature space of original vibration signals using different entropy measures based on t-SNE.

Table 5.8 Comparative performance on GDUPT multi-fault bearing diagnosis using different entropy measures and SVM classifier.

Method	MPE	CMPE	F2CMPE	AMWPE	parameters
Accuracy (%)	94.6	95.3	96.6	98.3	$m = 5, \lambda = 1$
Accuracy (%)	86.3	91.3	93.3	98	$m = 5, \lambda = 2$
Accuracy (%)	83.6	86.3	92.6	96.6	$m = 5, \lambda = 3$

5.4 Summary

In this chapter, three case studies are carried out to investigate the effectiveness of the proposed multiple-scale entropy measures for bearing fault diagnosis. The experiments examined the performance of the proposed entropy measures in analyzing bearing vibration signals in terms of cost time of calculating entropy values, mean and standard deviation values, and the robustness against noise, respectively. The advantage of the proposed multiple-scale entropy algorithms was seen by the fact that they save calculation time than traditional methods, but they can give more reliable and robust diagnosis performance. On the one hand, the feature values of the proposed entropy measures on a given scale have a small variation, and their waveforms over a set of scale factor can clearly differentiate between various bearing health states. Also, the reduced 2-D feature space demonstrated the advantage of the proposed methods in classifying bearing states. However, the feature values of traditional methods are mixed and are difficult to identify a certain bearing state by observing feature curves. On the other hand, the analysis of noise signals with various SNRs verified that the proposed entropy measures are robust to noise and can accurately identify bearing states, even with a certain strong noise level. In the study of bearing faults in the industrial-scale multistage centrifugal air compressor equipment, the advantage of the proposed entropy algorithms was seen by the fact that they provide satisfactory diagnosis accuracy in detecting both single-fault and multi-fault bearing fault in the unit. The developed entropy measures and diagnosis methods will contribute to developing improved and reliable bearing diagnosis techniques, thus preventing costly breakdowns and saving financial losses.

Chapter 6

Conclusions and Future Works

This chapter summarizes the research presented in this dissertation and outlines potential future work for intelligent data-driven methods in machinery fault diagnosis using entropy measures.

6.1 Summary

Entropy measures have exhibited practical effectiveness in time series complexity analysis and wide application to fault detection and diagnosis in rotating machinery. The goal of this dissertation is to develop new multiple-scale entropy measures and diagnostic techniques for fault diagnosis of rolling bearings. The research particularly aims to bridge an existing gap between the state-of-the-art entropy analysis in the academia and current practice on the factory floor.

Chapter 1 introduced the research background of this dissertation and discussed the motivations and contributions of this research work on bearing fault diagnosis. Bearing fault diagnosis methodologies were systematically reviewed, including model-based, signal-based, and data-driven-based fault diagnosis techniques. Limitation of traditional statistical features on analyzing bearing vibration signals measured from non-linear machinery systems was discussed. The research on multiple-scale entropy measures for analyzing vibration signals was reviewed, aiming at improving the efficiency and reliability of the scale-extraction mechanism in entropy estimation. The lack of systematic research in entropy analysis using reliable and improved scale-extraction mechanisms was identified.

While the wide application of entropy measures to the analysis of complex systems, their notions are defined differently in various contexts. To locate the gap between entropy analysis in the academia and fault diagnostic application in industrial-level machinery, chapter 2 systematically reviewed the theoretical development of some fundamental entropy

measures. Then, typical usages and applications of entropy analysis for machine fault diagnosis are summarized, including entropy as a feature indicator, entropy criterion for parameter selection, and entropy usage in pattern recognition. Further, insights into potential applications of entropy measures are explained, as to where and how these measures are useful for data-driven fault diagnosis methodologies. Based on the systematical survey, limitation of traditional entropy measures on estimating complexity change of bearing signal analysis was identified.

Chapter 3 presented the principles of typical single-scale and multiple-scale entropy measures and clarified the relations between different notions of entropy algorithms, respectively. This helps researchers arrive at an understanding of some of the most significant principles of entropy measures. The research summarizes representative characteristics of entropy measures in time series complexity analysis in terms of their merits, demerits, and algorithmic complexity. Through systematic survey work, it was found that despite the wide application of traditional multi-scale entropy measures, they have some limitations in the analysis of non-stationary time series from the non-linear system, particularly for bearing vibration signal analysis. On the one hand, high-frequency information is abandoned in traditional scale-extraction procedures. On the other hand, extracted multiple-scale time series with greatly reduced data length lead to inconsistent and biased entropy values with large variance, consequently resulting in reduced diagnosis performance in bearing diagnosis. These critical motivations drive the presented research in this dissertation in continually developing new entropy measures with improved and reliable diagnosis performance in bearing diagnosis, especially towards real industrial diagnostic applications.

Chapter 4 proposed a novel Fine-to-Coarse (F2C) mechanism for generating improved multiple-scale signals in the scale-extraction stage. The selection of appropriate wavelet kernels used to generate the F2C signals was evaluated using the relative wavelet energy criterion. Based on the F2C procedure, a preliminary study was carried out and a novel Fine-to-Coarse Multiscale Permutation Entropy (F2CMPE) was put forth. Further, an improved entropy measure - Adaptive Multiscale Weighted Permutation Entropy (AMWPE) - was proposed for bearing vibration signal analysis. Also, a new bearing diagnosis method is developed based on the AMWPE and SVM classifier for bearing fault pattern identification. Numerical evaluation study demonstrated that the proposed AMWPE measure offers coherent and stable entropy values in entropy analysis in comparison with traditional entropy methods.

In chapter 5, three case studies were carried out to verify the effectiveness of the proposed entropy algorithms in bearing fault diagnosis. Original signals and noisy signals were analyzed using entropy measures, respectively. Results demonstrated that the proposed algorithms have higher computational efficiency in comparison with traditional modified

entropy methods. Feature values of the improved entropy measures have smaller variance and higher consistency over a range of scale factors. Also, their feature waveforms enable distinguishing between various bearing health states in the feature space, corresponding to higher diagnostic accuracy compared to traditional methods. Particularly, with respect to the CWRU dataset with ten bearing states, the proposed methods can detect and identify different fault types and fault severity levels in the bearing. Also, results demonstrated that the proposed diagnosis methods have high robustness to noise. Even if there is strong noise interference, they can accurately identify fault patterns and fault severity levels. These advantages make the proposed entropy measures and their applications suitable for bearing diagnosis in industrial-scale machinery equipment.

In addition, bearing signals measured from an industrial-level multistage centrifugal air pump unit were analyzed, where studies of single-fault and multi-fault bearing diagnosis are considered, respectively. Experimental results demonstrated that for both single-fault and multi-fault bearing diagnosis, the proposed methods exhibit satisfactory diagnostic performance, such as in feature representation and bearing classification accuracy rate. Improved abilities of entropy estimation and robustness against noise verified the effectiveness and reliability of the proposed entropy measures in bearing signal analysis for machinery fault diagnosis. It is worth mentioning that, in this thesis, a few fault patterns and severity levels are considered and simulated. Data sets may vary due to different operating conditions, experimental environments and applications. The proposed diagnostic models are suitable for detecting known fault types and severity levels. Once new fault patterns added, the fault detection model need updates by training new data sets.

As a summary, this dissertation systematically reviews state-of-art entropy measures in machinery fault diagnosis and summarizes potential usages and roles of entropy measures in diagnostic applications. Limitation of traditional multiple-scale entropy measures on bearing signal analysis is identified. For providing improved and reliable entropy analysis in bearing diagnosis, this research proposes a new F2C scale-extraction procedure and an improved entropy measure - AMWPE, as well as another improved method proposed in the preliminary study. A new bearing diagnosis method is developed based on the AMWPE and SVM classifier. Considering three case studies as benchmarks, the presented work has three main advantages:

1. The proposed entropy measures can characterize fine-grained fault information incorporating both low- and high-frequency information from the original signal. This merit earns the proposed entropy measures better ability in detecting incipient faults in the bearing and distinguishing between various bearing health conditions.

2. The improved F2C scale-extraction procedure enables yielding coherent entropy values with small variance and slowly reduced values with an increasing scale factor, thus avoiding incoherent and biased entropy values and improving entropy analysis in bearing diagnosis.

3. Through the analysis of three case studies, experimental results verified the effectiveness of the improved entropy methods in bearing diagnosis. For single-fault bearing diagnosis, they are capable of identifying various bearing conditions, defect sizes, and severity levels with satisfactory performance, such as less computational cost, higher fault classification accuracy and higher robustness to noise, in comparison with traditional entropy methods. With respect to multi-fault bearing diagnosis, the proposed methods also exhibit the applicability to identify compound faults with satisfactory diagnostic performance, where interferences due to coupling components and strong background noise exist.

6.2 Future works

In addition to the work presented in this dissertation, some other researches need to be carried out in the future, which are listed in the following.

1. The presented work aims to analyze bearing vibration signals using permutation entropy estimation under the F2C scale-extraction scheme for bearing diagnosis. Since different single-scale entropy algorithms estimate different types of dynamic changes in the time series, it is possible to design various multiple-scale entropy methods under the framework of the F2C procedure. They may produce more comprehensive fault information using improved single-scale entropy algorithms for bearing signal analysis.

2. In three experimental case studies, a single wired acceleration sensor was used to collect bearing vibration signals from the machine. Therefore, accurate diagnostic performance entirely depends on the normal operation of a single sensor, but the premise is that the sensor node works without malfunction, so it can collect sensor data without abnormal values. Once the sensor fails, relying on the sole sensor node may lead to greatly reduced reliability of fault diagnostic systems and erroneous diagnostic decisions. To overcome this shortcoming, a multivariate entropy algorithm is a suitable tool to analyze the complexity of multichannel data using multiple sensor nodes. Therefore, the proposed entropy measures are expected to generalize to the multivariate entropy algorithm further in the future.

3. Feature representation is one of the critical procedures in the data-driven bearing diagnosis methodology. Considering the advantage of entropy measures in complexity analysis, various entropy feature indicators can be applied to extract fault information from signals. After that, extracted feature values can be fused, on the feature level, using information fusion techniques, after which advanced fault pattern classifiers are available for diagnostic decision-making. On the other hand, a variety of classifiers can be applied to make decisions, and their predicted results can be fused on the decision-level further; thus, more robust and accurate diagnostic decision-making is achieved by incorporating the advantages of different artificial intelligence-based fault pattern classifiers.

With these potential future works, more advanced techniques can further enhance the robustness of data-driven fault diagnostic systems, thus improving the reliability and safety of mechanical systems in industrial settings.

References

[1] PJ Tavner. Review of condition monitoring of rotating electrical machines. *IET Electric Power Applications*, 2(4):215–247, 2008.

[2] Zhiqiang Huo, Yu Zhang, Pierre Francq, Lei Shu, and Jianfeng Huang. Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures. *IEEE Access*, 5:19442–19456, 2017.

[3] Andrew KS Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510, 2006.

[4] Report of large motor reliability survey of industrial and commercial installations, Part I. *IEEE Trans. Ind. Appl.*, IA-21(4):853–864, Jul./Aug. 1985.

[5] Report of large motor reliability survey of industrial and commercial installations, Part II. *IEEE Trans. Ind. Appl.*, IA-21(4):865–872, Jul./Aug. 1985.

[6] Report of large motor reliability survey of industrial and commercial installations, Part III. *Part III, IEEE Trans. Ind. Appl.*, IA-23(1):153–158, Jan./Feb. 1987.

[7] Zhiqiang Huo, Yu Zhang, Lei Shu, and Michael Gallimore. A new bearing fault diagnosis method based on fine-to-coarse multiscale permutation entropy, Laplacian score and SVM. *IEEE Access*, 7:17050–17066, 2019.

[8] Analyzing gearbox failure and preventing it. *Renewable Energy World Magazine*, March/April; 2015.

[9] Glinsky C. Sheng S, Keller J. Gearbox reliability collaborative update. *Sandia Reliability Workshop Albuquerque*, NREL/PR-5000-60141, August 13-14; 2013.

[10] Managing the wind: reducing kilowatt-hour costs with condition monitoring. *Refocus*, page 48–51, May/June 2005.

[11] Ian Howard. A review of rolling element bearing vibration'detection, diagnosis and prognosis'. Technical report, Defence Science and Technology Organization Canberra(Australia), 1994.

[12] Robert B Randall and Jerome Antoni. Rolling element bearing diagnostics - A tutorial. *Mechanical systems and signal processing*, 25(2):485–520, 2011.

[13] Henrique Dias Machado de Azevedo, Alex Maurício Araújo, and Nadège Bouchonneau. A review of wind turbine bearing condition monitoring: State of the art and challenges. *Renewable and Sustainable Energy Reviews*, 56:368–379, 2016.

[14] Zhiqiang Huo, Yu Zhang, Richard Sath, and Lei Shu. Self-adaptive fault diagnosis of roller bearings using infrared thermal images. In *IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society*, pages 6113–6118. IEEE, 2017.

[15] Zhiqiang Huo, Yu Zhang, Zhangbing Zhou, and Jianfeng Huang. Crack detection in rotating shafts using wavelet analysis, shannon entropy and multi-class SVM. In *International Conference on Industrial Networks and Intelligent Systems*, pages 332–346. Springer, 2017.

[16] Naresh Tandon and Achintya Choudhury. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology international*, 32(8):469–480, 1999.

[17] Y Han and YH Song. Condition monitoring techniques for electrical equipment - A literature survey. *IEEE Transactions on Power delivery*, 18(1):4–13, 2003.

[18] Zeyu Zhang, Amjad Mehmood, Lei Shu, Zhiqiang Huo, Yu Zhang, and Mithun Mukherjee. A survey on fault diagnosis in wireless sensor networks. *IEEE Access*, 6:11349–11364, 2018.

[19] Yaguo Lei, Jing Lin, Zhengjia He, and Ming J Zuo. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical systems and signal processing*, 35(1-2):108–126, 2013.

[20] Yu Wei, Yuqing Li, Minqiang Xu, and Wenhu Huang. A review of early fault diagnosis approaches and their applications in rotating machinery. *Entropy*, 21(4):409, 2019.

[21] Zhiwei Gao, Carlo Cecati, and Steven X Ding. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Trans. Ind. Electron.*, 62(6):3757–3767, 2015.

[22] Patricia Henriquez, Jesus B Alonso, Miguel A Ferrer, and Carlos M Travieso. Review of automatic fault diagnosis systems using audio and vibration signals. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(5):642–652, 2013.

[23] Zhipeng Feng, Ming Liang, and Fulei Chu. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mechanical Systems and Signal Processing*, 38(1):165–205, 2013.

[24] Zhiwei Gao, Carlo Cecati, and Steven X Ding. A survey of fault diagnosis and fault-tolerant techniques - Part II: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Trans. Ind. Electron.*, 62(6):3768—-3774, 2015.

[25] Mariela Cerrada, René-Vinicio Sánchez, Chuan Li, Fannia Pacheco, Diego Cabrera, Jose Valente de Oliveira, and Rafael E Vasquez. A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 99:169–196, 2018.

[26] Ruonan Liu, Boyuan Yang, Enrico Zio, and Xuefeng Chen. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108:33–47, 2018.

[27] Weifeng Sun, Min Tang, Lijun Zhang, Zhiqiang Huo, and Lei Shu. A survey of using swarm intelligence algorithms in IoT. *Sensors*, 20(5), 2020.

[28] Tian Ran Lin, Kun Yu, and Jiwen Tan. Condition monitoring and fault diagnosis of roller element bearing. *INTECH*, pages 39–75, 2017.

[29] Ruqiang Yan and Robert X Gao. Approximate entropy as a diagnostic tool for machine health monitoring. *Mechanical Systems and Signal Processing*, 21(2):824–839, 2007.

[30] Klaus Mainzer. *Thinking in complexity: The computational dynamics of matter, mind, and mankind*. Springer Science & Business Media, 2007.

[31] Zhiqiang Huo, Miguel Martínez-García, Yu Zhang, Ruqiang Yan, and Lei Shu. Entropy measures in machine fault diagnosis: Insights and applications. *IEEE Transactions on Instrumentation and Measurement*, 69(6):2607–2620, 2020.

[32] Ruqiang Yan, Yongbin Liu, and Robert X Gao. Permutation entropy: a nonlinear statistical measure for status characterization of rotary machines. *Mechanical Systems and Signal Processing*, 29:474–484, 2012.

[33] Yongbo Li, Xianzhi Wang, Zhenbao Liu, Xihui Liang, and Shubin Si. The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review. *IEEE Access*, 6:66723–66741, 2018.

[34] Adam L Berger, Vincent J Della Pietra, and Stephen A Della Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71, 1996.

[35] Marta Borowska. Entropy-based algorithms in the analysis of biomedical signals. *Studies in Logic, Grammar and Rhetoric*, 43(1):21–32, 2015.

[36] Rongxi Zhou, Ru Cai, and Guanqun Tong. Applications of entropy in finance: A review. *Entropy*, 15(11):4909–4931, 2013.

[37] Nicholas Georgescu-Roegen. The entropy law and the economic problem. *Valuing the earth: Economics, ecology, ethics*, pages 75–88, 1993.

[38] Daniel R Brooks, Edward O Wiley, and DR Brooks. *Evolution as entropy*. University of Chicago Press Chicago, 1988.

[39] Zhiqiang Huo, Yu Zhang, Lei Shu, Yunrong Lv, and Shuiquan Lin. Bearing fault diagnosis using multi-sensor fusion based on weighted DS evidence theory. In *2018 18th International Conference on Mechatronics-Mechatronika (ME)*, pages 1–6. IEEE, 2018.

[40] Anne Humeau-Heurtier. The multiscale entropy algorithm and its variants: A review. *Entropy*, 17(5):3110–3123, 2015.

[41] Madalena Costa, Ary L Goldberger, and C-K Peng. Multiscale entropy analysis of complex physiologic time series. *Physical review letters*, 89(6):068102, 2002.

[42] Meng Hu and Hualou Liang. Multiscale entropy: Recent advances. In *Complexity and Nonlinearity in Cardiovascular Signals*, pages 115–138. Springer, 2017.

[43] Q.W. Gao, W.Y. Liu, B.P. Tang, and G.J. Li. A novel wind turbine fault diagnosis method based on intergral extension load mean decomposition multiscale entropy and least squares support vector machine. *Renewable Energy*, 116:169 – 175, 2018.

[44] Ruqiang Yan and Robert X Gao. Complexity as a measure for machine health evaluation. *IEEE Transactions on instrumentation and measurement*, 53(4):1327–1334, 2004.

[45] Arieh Ben-Naim. *Information, Entropy, Life and the Universe: What We Know and What We Do Not Know*. World Scientific Publishing Co., Inc., 2015.

[46] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[47] Jeremy Rifkin. Entropy: a new world view.[social and political implications of the Second Law of Thermodynamics]. 1980.

[48] James V Stone. *Information theory: a tutorial introduction*. Sebtel Press, 2015.

[49] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[50] Robert M Gray. *Entropy and information theory*. Springer Science & Business Media, 2011.

[51] Roman Frigg and Charlotte Werndl. Entropy-a guide for the perplexed. 2011.

[52] Peter Grassberger and Itamar Procaccia. Estimation of the Kolmogorov entropy from a chaotic signal. *Physical review A*, 28(4):2591, 1983.

[53] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

[54] Steven Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.

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

[56] Weiting Chen, Zhizhong Wang, Hongbo Xie, and Wangxin Yu. Characterization of surface EMG signal based on fuzzy entropy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(2):266–272, 2007.

[57] Christoph Bandt and Bernd Pompe. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.*, 88(17):174102, 2002.

[58] Peng Li, Chengyu Liu, Ke Li, Dingchang Zheng, Changchun Liu, and Yinglong Hou. Assessing the complexity of short-term heartbeat interval series by distribution entropy. *Medical & biological engineering & computing*, 53(1):77–87, 2015.

[59] Xiaofeng Liu, Aimin Jiang, Ning Xu, and Jianru Xue. Increment entropy as a measure of complexity for time series. *Entropy*, 18(1):22, 2016.

[60] Yongbo Li, Yuantao Yang, Guoyan Li, Minqiang Xu, and Wenhu Huang. A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection. *Mechanical Systems and Signal Processing*, 91:295–312, 2017.

[61] Hua Li, Tao Liu, Xing Wu, and Qing Chen. Research on bearing fault feature extraction based on singular value decomposition and optimized frequency band entropy. *Mechanical Systems and Signal Processing*, 118:477–502, 2019.

[62] Madalena Costa, Ary L Goldberger, and C-K Peng. Multiscale entropy analysis of biological signals. *Physical review E*, 71(2):021906, 2005.

[63] Yongbo Li, Minqiang Xu, Rixin Wang, and Wenhu Huang. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *Journal of Sound and Vibration*, 360:277–299, 2016.

[64] S Wu, P Wu, C Wu, J Ding, and C Wang. Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy*, 14(8):1343–1356, 2012.

[65] Shuen-De Wu, Chiu-Wen Wu, Shiou-Gwo Lin, Chun-Chieh Wang, and Kung-Yen Lee. Time series analysis using composite multiscale entropy. *Entropy*, 15(3):1069–1084, 2013.

[66] Madalena D Costa and Ary L Goldberger. Generalized multiscale entropy analysis: application to quantifying the complex volatility of human heartbeat time series. *Entropy*, 17(3):1197–1203, 2015.

[67] Naiquang Su, Xiao Li, Qinghua Zhang, and Zhiqiang Huo. Composite fault diagnosis for rotating machinery of large units based on evidence theory and multi-information fusion. *Shock and Vibration*, 2019, 2019.

[68] Zhiqiang Huo, Yu Zhang, and Lei Shu. A short survey on fault diagnosis of rotating machinery using entropy techniques. In *International Conference on Industrial Networks and Intelligent Systems*, pages 279–284. Springer, 2017.

[69] Gang Niu, Achmad Widodo, Jong Son, Bo Yang, and Dong Hwang, Donand Kang. Decision-level fusion based on wavelet decomposition for induction motor fault diagnosis using transient current signal. *Expert Systems with Applications*, 35(3):918–928, 2008.

[70] Van Tran, Bo Yang, Myung Oh, and Andy Tan. Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. *Expert Systems with Applications*, 36(2, Part 1):1840 – 1849, 2009.

[71] M Elforjani and D Mba. Accelerated natural fault diagnosis in slow speed bearings with acoustic emission. *Engineering Fracture Mechanics*, 77(1):112–127, 2010.

[72] Yu Yang, Dejie Yu, and Junsheng Cheng. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *Journal of Sound and Vibration*, 294(1):269 – 277, 2006.

[73] Yuan Xie and Tao Zhang. Fault diagnosis for rotating machinery based on convolutional neural network and empirical mode decomposition. *Shock and Vibration*, 2017, 2017.

[74] Jing Yuan, Yu Wang, Yizhen Peng, and Chenjun Wei. Weak fault detection and health degradation monitoring using customized standard multiwavelets. *Mechanical Systems and Signal Processing*, 94:384–399, 2017.

[75] Cheng Fei, Guang Bai, Wen Tang, and Shuang Ma. Quantitative diagnosis of rotor vibration fault using process power spectrum entropy and support vector machine method. *Shock and Vibration*, 2014, 2014.

[76] Quansheng Jiang, Yehu Shen, Hua Li, and Fengyu Xu. New fault recognition method for rotary machinery based on information entropy and a probabilistic neural network. *Sensors*, 18(2), 2018.

[77] Yan Ai, Jiao Guan, Cheng Fei, Jing Tian, and Feng Zhang. Fusion information entropy method of rolling bearing fault diagnosis based on n-dimensional characteristic parameter distance. *Mechanical Systems and Signal Processing*, 88:123–136, 2017.

[78] Yongyong He and Xinming Zhang. Approximate entropy analysis of the acoustic emission from defects in rolling element bearings. *Journal of Vibration and Acoustics*, 134(6):061012, 2012.

[79] Diego Sampaio and Rodrigo Nicoletti. Detection of cracks in shafts with the approximated entropy algorithm. *Mechanical Systems and Signal Processing*, 72:286–302, 2016.

[80] Fengtao Wang, Chenxi Liu, Wensheng Su, Zhigang Xue, Qingkai Han, and Hongkun Li. Combined failure diagnosis of slewing bearings based on MCKD-CEEMD-ApEn. *Shock and Vibration*, 2018, 2018.

[81] Quansheng Jiang, Yehu Shen, Hua Li, and Fengyu Xu. New fault recognition method for rotary machinery based on information entropy and a probabilistic neural network. *Sensors*, 18(2):337, 2018.

[82] Wenfeng Hu, Hong Chang, and Xingsheng Gu. A novel fault diagnosis technique for wind turbine gearbox. *Applied Soft Computing*, page 105556, 2019.

[83] Keheng Zhu, Xigeng Song, and Dongxin Xue. Fault diagnosis of rolling bearings based on IMF envelope sample entropy and support vector machine. *Journal of Information and Computational Science*, 10(16):5189–5198, 2013.

[84] Minghong Han and Jiali Pan. A fault diagnosis method combined with LMD, sample entropy and energy ratio for roller bearings. *Measurement*, 76:7–19, 2015.

[85] Mourad Kedadouche, Marc Thomas, Antoine Tahan, and Raynald Guilbault. Nonlinear parameters for monitoring gear: comparison between Lempel-Ziv, approximate entropy, and sample entropy complexity. *Shock and Vibration*, 2015, 2015.

[86] Ke Feng, Kesheng Wang, Qing Ni, Ming J Zuo, and Dongdong Wei. A phase angle based diagnostic scheme to planetary gear faults diagnostics under non-stationary operational conditions. *Journal of Sound and Vibration*, 408:190–209, 2017.

[87] Guo Xiong, Long Zhang, He Liu, Hui Zou, and Wei Guo. A comparative study on ApEn, SampEn and their fuzzy counterparts in a multiscale framework for feature extraction. *Journal of Zhejiang University SCIENCE A*, 11(4):270–279, Apr 2010.

[88] Jinde Zheng, Junsheng Cheng, and Yu Yang. A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy. *Mechanism and Machine Theory*, 70:441–453, 2013.

[89] Xihui Chen, Gang Cheng, Hongyu Li, and Min Zhang. Diagnosing planetary gear faults using the fuzzy entropy of LMD and ANFIS. *Journal of Mechanical Science and Technology*, 30(6):2453–2462, 2016.

[90] Wu Deng, Shengjie Zhang, Huimin Zhao, and Xinhua Yang. A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing. *IEEE Access*, 2018.

[91] Hu Zhang, Lei Zhao, Quan Liu, Jingjing Luo, Qin Wei, Zude Zhou, and Yongzhi Qu. An improved feature extraction method for rolling bearing fault diagnosis based on MEMD and PE. *Polish Maritime Research*, 25(s2):98–106, 2018.

[92] Zongli Shi, Wanqing Song, and Saied Taheri. Improved LMD, permutation entropy and optimized K-means to fault diagnosis for roller bearings. *Entropy*, 18(3):70, 2016.

[93] B Qin, GD Sun, LY Zhang, JG Wang, and J Hu. Fault features extraction and identification based rolling bearing fault diagnosis. In *Journal of Physics: Conference Series*, volume 842, page 012055, 2017.

[94] Liye Zhao, Wei Yu, and Ruqiang Yan. Gearbox fault diagnosis using complementary ensemble empirical mode decomposition and permutation entropy. *Shock and Vibration*, 2016, 2016.

[95] Yu Wei, Minqiang Xu, Yongbo Li, and Wenhu Huang. 1983. gearbox fault diagnosis based on local mean decomposition, permutation entropy and extreme learning machine. *Journal of Vibroengineering*, 18(3), 2016.

[96] Shenghan Zhou, Silin Qian, Wenbing Chang, Yiyong Xiao, and Yang Cheng. A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier. *Sensors*, 18(6):1934, 2018.

[97] Xiaoyuan Zhang, Yitao Liang, Jianzhong Zhou, et al. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement*, 69:164–179, 2015.

[98] Xiaoming Xue and Jianzhong Zhou. A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery. *ISA transactions*, 66:284–295, 2017.

[99] Long Zhang, Guoliang Xiong, Hesheng Liu, Huijun Zou, and Weizhong Guo. Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Systems with Applications*, 37(8):6077–6085, 2010.

[100] Tian-Yau Wu, Chang-Ling Yu, and Da-Chun Liu. On multi-scale entropy analysis of order-tracking measurement for bearing fault diagnosis under variable speed. *Entropy*, 18(8):292, 2016.

[101] Shuang Pan, Tian Han, Andy CC Tan, and Tian Ran Lin. Fault diagnosis system of induction motors based on multiscale entropy and support vector machine with mutual information algorithm. *Shock and Vibration*, 2016, 2016.

[102] Shuen-De Wu, Chiu-Wen Wu, Kung-Yen Lee, and Shiou-Gwo Lin. Modified multiscale entropy for short-term time series analysis. *Physica A: Statistical Mechanics and its Applications*, 392(23):5865–5873, 2013.

[103] Yongbo Li, Minqiang Xu, Rixin Wang Wang, and Wenhu Huang. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy.

[104] Jinde Zheng, Haiyang Pan, and Junsheng Cheng. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mechanical Systems and Signal Processing*, 85:746–759, 2017.

[105] Anne Humeau-Heurtier, Chiu-Wen Wu, and Shuen-De Wu. Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence. *IEEE Signal Processing Letters*, 22(12):2364–2367, 2015.

[106] Yongbo Li, Minqiang Xu, Yu Wei, and Wenhu Huang. 1585. bearing fault diagnosis based on adaptive mutiscale fuzzy entropy and support vector machine. *Journal of Vibroengineering*, 17(3), 2015.

[107] Jinde Zheng, Deyu Tu, Haiyang Pan, Xiaolei Hu, Tao Liu, and Qingyun Liu. A refined composite multivariate multiscale fuzzy entropy and Laplacian score-based fault diagnosis method for rolling bearings. *Entropy*, 19(11), 2017.

[108] Yongbo Li, Xianzhi Wang, Shubin Si, and Shiqian Huang. Entropy based fault classification using the case western reserve university data: A benchmark study. *IEEE Transactions on Reliability*, 2019.

[109] Yuantao Yang, Huailiang Zheng, Jiancheng Yin, Minqiang Xu, and Yushu Chen. Refined composite multivariate multiscale symbolic dynamic entropy and its application to fault diagnosis of rotating machine. *Measurement*, 151:107233, 2020.

[110] Yunguang Ye, Yongxiang Zhang, Qingbo Wang, Zhiwei Wang, Zhenjie Teng, and Hougui Zhang. Fault diagnosis of high-speed train suspension systems using multiscale permutation entropy and linear local tangent space alignment. *Mechanical Systems and Signal Processing*, 138:106565, 2020.

[111] Ronald R Coifman and M Victor Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, March 1992.

[112] Ruqiang Yan. Base wavelet selection criteria for non-stationary vibration analysis in bearing health diagnosis. *Electronic Doctoral Dissertations for UMass Amherst*, 2007.

[113] Pavan Kankar, Satish Sharma, and Suraj Prakash Harsha. Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing*, 11(2):2300–2312, 2011.

[114] Ruqiang Yan, Robert X Gao, and Xuefeng Chen. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal processing*, 96:1–15, 2014.

[115] Sukhjeet Singh and Navin Kumar. Detection of bearing faults in mechanical systems using stator current monitoring. *IEEE Transactions on Industrial Informatics*, 13(3):1341–1349, 2016.

[116] K Deák, T Mankovits, and I Kocsis. Optimal wavelet selection for the size estimation of manufacturing defects of tapered roller bearings with vibration measurement using Shannon entropy criteria. *Strojniski Vestnik/Journal of Mechanical Engineering*, 63(1), 2017.

[117] Shuting Wan, Xiong Zhang, and Longjiang Dou. Shannon entropy of binary wavelet packet subbands and its application in bearing fault extraction. *Entropy*, 20(4):260, 2018.

[118] Rui Xiao, Qunfang Hu, and Jie Li. Leak detection of gas pipelines using acoustic signals based on wavelet transform and support vector machine. *Measurement*, 146:479–489, 2019.

[119] Ankit Darji, PH Darji, and DH Pandya. Fault diagnosis of ball bearing with WPT and supervised machine learning techniques. In *Machine Intelligence and Signal Analysis*, pages 291–301. Springer, 2019.

[120] Biao Wang, Yaguo Lei, Naipeng Li, and Ningbo Li. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, 2018.

[121] Rui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, and Robert X Gao. Deep learning and its applications to machine health monitoring: A survey. *arXiv preprint arXiv:1612.07640*, 2016.

[122] Miguel Martinez-Garcia, Yu Zhang, Jiafu Wan, and Jason Mcginty. Visually interpretable profile extraction with an autoencoder for health monitoring of industrial systems. In *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*, pages 649–654. IEEE, 2019.

[123] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[124] Douglas M Kline and Victor L Berardi. Revisiting squared-error and cross-entropy functions for training neural network classifiers. *Neural Computing & Applications*, 14(4):310–318, 2005.

[125] Siyu Shao, Ruqiang Yan, Yadong Lu, Peng Wang, and Robert Gao. DCNN-based multi-signal induction motor fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 2019.

[126] Xiang Li, Wei Zhang, Nan-Xi Xu, and Qian Ding. Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places. *IEEE Transactions on Industrial Electronics*, 2019.

[127] Minghang Zhao, Shisheng Zhong, Xuyun Fu, Baoping Tang, and Michael Pecht. Deep residual shrinkage networks for fault diagnosis. *IEEE Transactions on Industrial Informatics*, 2019.

[128] Bin Yang, Yaguo Lei, Feng Jia, and Saibo Xing. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing*, 122:692–706, 2019.

[129] Ruqiang Yan, Fei Shen, Chuang Sun, and Xuefeng Chen. Knowledge transfer for rotary machine fault diagnosis. *IEEE Sensors Journal*, 2019.

[130] Miguel Martínez-García, Yu Zhang, Kenji Suzuki, and Zhang Yudong. Measuring system entropy with a deep recurrent neural network model. In *Industrial Informatics (INDIN), 2019 IEEE International Conference on*. IEEE, 2019.

[131] Raul Alcaraz and Jose Joaquin Rieta. A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms. *Biomedical Signal Processing and Control*, 5(1):1–14, 2010.

[132] Yu-Hsiang Pan, Yung-Hung Wang, Sheng-Fu Liang, and Kuo-Tien Lee. Fast computation of sample entropy and approximate entropy in biomedicine. *Computer methods and programs in biomedicine*, 104(3):382–396, 2011.

[133] Karsten Keller, Anton Unakafov, and Valentina Unakafova. Ordinal patterns, entropy, and EEG. *Entropy*, 16(12):6212–6239, 2014.

[134] Gauthier Doquire and Michel Verleysen. Mutual information-based feature selection for multilabel classification. *Neurocomputing*, 122:148–155, 2013.

[135] G Sinai Ya. On the concept of entropy for dynamical systems (in Russian) Dokl. *Akad. Nauk SSSR*, 124:768–771, 1959.

[136] Roman Frigg. In what sense is the Kolmogorov-Sinai entropy a measure for chaotic behaviour? – bridging the gap between dynamical systems theory and communication theory. *The British journal for the philosophy of science*, 55(3):411–434, 2004.

[137] Claus Beisbart and Stephan Hartmann. *Probabilities in physics*. Oxford University Press, 2011.

[138] Yakov Borisovich Pesin. Characteristic Lyapunov exponents and smooth ergodic theory. *Uspekhi Matematicheskikh Nauk*, 32(4):55–112, 1977.

[139] J-P Eckmann and David Ruelle. Ergodic theory of chaos and strange attractors. In *The Theory of Chaotic Attractors*, pages 273–312. Springer, 1985.

[140] Alfréd Rényi. On measures of entropy and information. Technical report, in: Fourth Berkeley Symposium on Mathematics, Statistics and Probability, 1960.

[141] Steve Pincus. Approximate entropy (ApEn) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):110–117, 1995.

[142] Sheng Lu, Xinnian Chen, Jørgen K Kanters, Irene C Solomon, and Ki H Chon. Automatic selection of the threshold value *r* for approximate entropy. *IEEE Transactions on Biomedical Engineering*, 55(8):1966–1972, 2008.

[143] Farhad Kaffashi, Ryan Foglyano, Christopher Wilson, and Kenneth Loparo. The effect of time delay on approximate & sample entropy calculations. *Physica D: Nonlinear Phenomena*, 237(23):3069–3074, 2008.

[144] Steven M Pincus, Thomas Mulligan, Ali Iranmanesh, Sylvia Gheorghiu, Michael Godschalk, and Johannes D Veldhuis. Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males. *Proceedings of the National Academy of Sciences*, 93(24):14100–14105, 1996.

[145] Joshua S Richman, Douglas E Lake, and J Randall Moorman. Sample entropy. In *Methods in enzymology*, volume 384, pages 172–184. Elsevier, 2004.

[146] Jennifer Yentes, Nathaniel Hunt, Kendra Schmid, Jeffrey Kaipust, Denise McGrath, and Nicholas Stergiou. The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of biomedical engineering*, 41(2):349–365, 2013.

[147] Jennifer Yentes, William Denton, John McCamley, Peter C Raffalt, and Kendra Schmid. Effect of parameter selection on entropy calculation for long walking trials. *Gait & posture*, 60:128–134, 2018.

[148] Yun Lu, Mingjiang Wang, Rongchao Peng, and Qiquan Zhang. Accelerating the computation of entropy measures by exploiting vectors with dissimilarity. *Entropy*, 19(11):598, 2017.

[149] George Manis, Md Aktaruzzaman, and Roberto Sassi. Low computational cost for sample entropy. *Entropy*, 20(1), 2018.

[150] LEV Silva, ACS Senra Filho, VPS Fazan, JC Felipe, and LO Junior. Two-dimensional sample entropy: assessing image texture through irregularity. *Biomedical Physics & Engineering Express*, 2(4):045002, 2016.

[151] Ruhi Mahajan and Bashir Morshed. Unsupervised eye blink artifact denoising of eeg data with modified multiscale sample entropy, kurtosis, and wavelet-ICA. *IEEE journal of Biomedical and Health Informatics*, 19(1):158–165, 2015.

[152] Jinde Zheng, Junsheng Cheng, Yu Yang, and Songrong Luo. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mechanism and Machine Theory*, 78:187–200, 2014.

[153] Jean Girault and Anne Humeau-Heurtier. Centered and averaged fuzzy entropy to improve fuzzy entropy precision. *Entropy*, 20(4), 2018.

[154] Tao Zhang, Wanzhong Chen, and Mingyang Li. Fuzzy distribution entropy and its application in automated seizure detection technique. *Biomedical Signal Processing and Control*, 39:360–377, 2018.

[155] Weiting Chen, Jun Zhuang, Wangxin Yu, and Zhizhong Wang. Measuring complexity using FuzzyEn, ApEn, and SampEn. *Medical Engineering and Physics*, 31(1):61–68, 2009.

[156] Samantha Simons, Pedro Espino, and Daniel Abásolo. Fuzzy entropy analysis of the electroencephalogram in patients with Alzheimer's disease: Is the method superior to sample entropy? *Entropy*, 20(1):21, 2018.

[157] Yinhe Cao, Wen-wen Tung, JB Gao, Vladimir A Protopopescu, and Lee M Hively. Detecting dynamical changes in time series using the permutation entropy. *Phys. Rev. E*, 70(4):046217, 2004.

[158] Massimiliano Zanin, Luciano Zunino, Osvaldo A Rosso, and David Papo. Permutation entropy and its main biomedical and econophysics applications: a review. *Entropy*, 14(8):1553–1577, 2012.

[159] Karsten Keller and Mathieu Sinn. Kolmogorov–Sinai entropy from the ordinal viewpoint. *Physica D: Nonlinear Phenomena*, 239(12):997–1000, 2010.

[160] Müller Riedl, A Müller, and N Wessel. Practical considerations of permutation entropy. *The European Physical Journal Special Topics*, 222(2):249–262, 2013.

[161] Karsten Keller, Teresa Mangold, Inga Stolz, and Jenna Werner. Permutation entropy: New ideas and challenges. *Entropy*, 19(3):134, 2017.

[162] Xiao Liu and Yue Wang. Fine-grained permutation entropy as a measure of natural complexity for time series. *Chinese Physics B*, 18(7):2690, 2009.

[163] Bilal Fadlallah, Badong Chen, Andreas Keil, and José Príncipe. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Phys. Rev. E*, 87:022911, Feb 2013.

[164] Karsten Keller, Anton Unakafov, and Valentina Unakafova. Ordinal patterns, entropy, and EEG. *Entropy*, 16(12):6212–6239, 2014.

[165] Hamed Azami and Javier Escudero. Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation. *Computer methods and programs in biomedicine*, 128:40–51, 2016.

[166] Chunhua Bian, Chang Qin, Qian Ma, and Qinghong Shen. Modified permutation-entropy analysis of heartbeat dynamics. *Phys. Rev. E*, 85:021906, Feb 2012.

[167] Amritpal Singh, Barjinder Singh Saini, and Dilbag Singh. An adaptive technique for multiscale approximate entropy (MAEbin) threshold (r) selection: application to heart rate variability (HRV) and systolic blood pressure variability (SBPV) under postural stress. *Australasian physical & engineering sciences in medicine*, 39(2):557–569, 2016.

[168] Wajid Aziz and Muhammad Arif. Multiscale permutation entropy of physiological time series. In *9th International Multitopic Conference, IEEE INMIC 2005*, pages 1–6. IEEE, 2005.

[169] Shuen-De Wu, Chiu-Wen Wu, Shiou-Gwo Lin, Kung-Yen Lee, and Chung-Kang Peng. Analysis of complex time series using refined composite multiscale entropy. *Physics Letters A*, 378(20):1369–1374, 2014.

[170] Anne Humeau-Heurtier, Chiu-Wen Wu, and Shuen-De Wu. Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence. *IEEE signal processing letters*, 22(12):2364–2367, 2015.

[171] Yuesheng Xu and Liang Zhao. Filter-based multiscale entropy analysis of complex physiological time series. *Physical Review E*, 88(2):022716, 2013.

[172] Ying Jiang, C-K Peng, and Yuesheng Xu. Hierarchical entropy analysis for biological signals. *Journal of Computational and Applied Mathematics*, 236(5):728–742, 2011.

[173] Mosabber Uddin Ahmed and Danilo P Mandic. Multivariate multiscale entropy analysis. *IEEE Signal Processing Letters*, 19(2):91–94, 2011.

[174] Hamed Azami and Javier Escudero. Refined composite multivariate generalized multiscale fuzzy entropy: A tool for complexity analysis of multichannel signals. *Physica A: Statistical Mechanics and its Applications*, 465:261–276, 2017.

[175] J. F. Valencia, A. Porta, M. Vallverdu, F. Claria, R. Baranowski, E. Orlowska-Baranowska, and P. Caminal. Refined multiscale entropy: Application to 24-h holter recordings of heart period variability in healthy and aortic stenosis subjects. *IEEE Transactions on Biomedical Engineering*, 56(9):2202–2213, Sept 2009.

[176] Todd Ogden. *Essential wavelets for statistical applications and data analysis*. Springer Science & Business Media, 2012.

[177] Gang Zhang, Tian Yi, Tianqi Zhang, and Li Cao. A multiscale noise tuning stochastic resonance for fault diagnosis in rolling element bearings. *Chinese Journal of Physics*, 56(1):145 – 157, 2018.

[178] L Law, J Kim, W Liew, and S Lee. An approach based on wavelet packet decomposition and Hilbert-Huang transform (WPD-HHT) for spindle bearings condition monitoring. *Mechanical Systems and Signal Processing*, 33:197 – 211, 2012.

[179] Jinde Zheng, Haiyang Pan, Shubao Yang, and Junsheng Cheng. Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 99:229–243, 2018.

[180] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.

[181] Pavan Kankar, Satish Sharma, and Suraj Harsha. Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing*, 11(2):2300–2312, 2011.

[182] J. Rafiee and P.W. Tse. Use of autocorrelation of wavelet coefficients for fault diagnosis. *Mechanical Systems and Signal Processing*, 23(5):1554–1572, 2009.

[183] Case Western Reserve University Bearing Data Center. (Date last accessed 15-Aug.-2020).

[184] Karl Pearson and Alice Lee. On the laws of inheritance in man: I. Inheritance of physical characters. *Biometrika*, 2(4):357–462, 1903.

[185] Bilal Fadlallah, Badong Chen, Andreas Keil, and José Príncipe. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Physical Review E*, 87(2):022911, 2013.

[186] Jianan Xia, Pengjian Shang, Jing Wang, and Wenbin Shi. Permutation and weighted-permutation entropy analysis for the complexity of nonlinear time series. *Communications in Nonlinear Science and Numerical Simulation*, 31(1-3):60–68, 2016.

[187] Shenghan Zhou, Silin Qian, Wenbing Chang, Yiyong Xiao, and Yang Cheng. A novel bearing multi-fault diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier. *Sensors*, 18(6):1934, 2018.

[188] Cheng He, Tao Wu, Changchun Liu, and Tong Chen. A novel method of composite multiscale weighted permutation entropy and machine learning for fault complex system fault diagnosis. *Measurement*, page 107748, 2020.

[189] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[190] Achmad Widodo and Bo-Suk Yang. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical systems and signal processing*, 21(6):2560–2574, 2007.

[191] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:1–27, 2011.

[192] Zhiqiang Huo, Yu Zhang, Gbanaibolou Jombo, and Lei Shu. Adaptive multiscale weighted permutation entropy for rolling bearing fault diagnosis. *IEEE Access*, 8:87529–87540, 2020.

[193] Zhiqiang Huo, Yu Zhang, and Lei Shu. Fine-to-coarse multiscale permutation entropy for rolling bearing fault diagnosis. IEEE Conference on International Wireless Communications & Mobile Computing Conference (IWCMC 2018), 2018.

[194] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[195] Wade A Smith and Robert B Randall. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing*, 64:100–131, 2015.

[196] Jianbin Xiong, Qinghua Zhang, Jianfu Wan, Liang Liang, Pinghua Cheng, and Qiong Liang. Data fusion method based on mutual dimensionless. *IEEE/ASME Transactions on mechatronics*, 23(2):506–517, 2017.