Article

Vibration-Based Signal Analysis for Shearer Cutting Status Recognition Based on Local Mean Decomposition and Fuzzy C-Means Clustering

Lei Si [1,2], Zhongbin Wang [1,*], Chao Tan [1] and Xinhua Liu [1]

[1] School of Mechatronic Engineering, China University of Mining & Technology, No. 1 Daxue Road, Xuzhou 221116, China; sileicool@163.com (L.S.); tccadcumt@126.com (C.T.); l_xinhua_2006@126.com (X.L.)

[2] School of Information and Electrical Engineering, China University of Mining & Technology, No. 1 Daxue Road, Xuzhou 221116, China

* Correspondence: wangzbpaper@126.com; Tel./Fax: +86-516-8359-0758

Abstract: In order to accurately acquire shearer cutting status, this paper proposed a pattern recognition method, based on the local mean decomposition (LMD), time-frequency statistical analysis, improved Laplacian score (LS), and fuzzy C-means (FCM) clustering algorithm. The LMD was employed to preprocess the vibration signals of shear cutting coal seam, and several product functions (PFs) were obtained. Following this, 14 time-frequency statistical parameters of the original signal and optimal PF were extracted. Additionally, the improved LS algorithm was designed to ensure the accurate estimation of features, and a new feature vector could be selected. Subsequently, the obtained eigenvector matrix was fed into a FCM to be clustered, for optimal clustering performance. The experimental examples were provided to verify the effectiveness of the methodology and the results indicated that the proposed algorithm could be applied to recognize the different categories of shearer cutting status.

Keywords: pattern recognition; feature extraction; feature selection; Laplacian score; fuzzy C-means clustering

---

1. Introduction

Recently, the intellectualized mining technology of the fully-mechanized coal mining face, has increasingly attracted the attention of coal mining managers. As a key component of this fully-mechanized equipment, the shearer plays the most important role in the whole process of coal mining. The precise identification of the cutting status of the shearer is directly related to the intelligentialize level, and has a great impact on the safety of coal miners and mining efficiency [1]. Therefore, there is an urgent need to study a new pattern recognition method for obtaining an accurate recognition of shearer cutting status [2].

Vibration-based signal analysis techniques have been widely used in the field of pattern recognition and fault diagnosis. The vibration performance of a shearer's cutting part changes when the shearer cuts different geological conditions of coal seam. However, due to the extremely harsh mining environment, the vibration signals represent strong, nonlinear, non-Gaussian, and non-stationary characteristics, which need to be reprocessed in order to obtain accurate state features. Many traditional methods, such as the Fourier transformation, wavelet transform, and Wigner-Ville distribution [3–6], have been developed, in order to achieve signal analysis, but these methods have their own drawbacks, lacking the nature of a self-adaptive feature.

Taking this into consideration, Huang proposed a self-adaptive time-frequency decomposition algorithm in 1998, named the empirical mode decomposition (EMD), which has since been used in

many fields [7]. EMD can decompose the complex signal into a sum of intrinsic mode functions (IMFs), and each IMF contains the local characteristics of the original signal. However, during the process of decomposition, EMD is often faced with the end effect and mode mixing problems, etc., restricting its applications [8–10]. In 2005, Jonathan S. Smith proposed a novel signal analysis method, named the local mean decomposition (LMD), and initially used it to process the electroencephalogram [11]. LMD can adaptively decompose a complicated multi-component signal into several product functions (PFs), and each PF is the product of an amplitude envelope signal and a purely frequency-modulated signal. Because of the peculiarity of LMD, this decomposition method has attracted many scholars' attention in recent years. In [12], the LMD method was used to decompose the vibration signals of a rolling bearing, and the results were able to provide the basis for the fault diagnosis. In [13], a fault diagnosis method based on LMD and multiscale fuzzy entropy was proposed, in order to fulfill the fault pattern identifications for a rolling bearing. Wei et al. [14] proposed a fault diagnosis method for a gearbox, based on the local mean decomposition, permutation entropy, and an extreme learning machine. The research results indicate that the LMD method possesses a high suitability for analyzing non-stationary signals, which can be employed to perform the preprocessing of shearer cutting vibration signals in this paper.

Naturally, after signal reprocessing, feature extraction should be fulfilled in the pattern recognition system of shearer cutting status. Time-frequency statistical parameters can describe the waveform distribution characteristics of vibration signals, in the time and frequency domain [15,16]. Thus, the time and frequency domain feature parameters are comprehensively utilized to acquire more feature information in this paper. However, not all of the features are closely related to each kind of cutting status, and too many inputs for a classifier will consume more classification time and reduce the accuracy. Hence, in order to avoid the dimension disaster, many traditional feature selection algorithms have been proposed, such as the Fisher score (FS) [17], Laplacian score (LS) [18,19], Variant score (VS) [20], and so on. Some heuristic feature selection methods, such as the genetic algorithm, particle swarm optimization, and chaotic simulated annealing methods, have also been mentioned and applied in much of the literature [21–23].Nevertheless, these methods have their own defects and limitations. In this paper, an improved Laplacian score, based on FS, was presented, in order to select appropriate features as the new feature vectors.

In the fields of fault diagnosis and pattern recognition, various artificial intelligence techniques have been successfully developed and used in recent years, such as artificial neural networks [24–26], support vector machines [27,28], and self-organizing maps [29]. Because of the gradual changes and ambiguous boundaries between different shearer cutting statuses, the fuzzy theory is essential for identifying various cutting statuses. The fuzzy C-means (FCM) clustering algorithm owns the theoretical basis of fuzzy mathematics, and has been successfully applied in many fields [30–32]. Therefore, an FCM-based recognition algorithm was introduced in this paper, providing an effective way to fulfill the superior recognition performance for shearer cutting status, in comparison to other classification methods.

This paper is organized as follows. In Section 2, the main steps of the LMD and FCM clustering algorithm are reviewed. Section 3 introduces the flowchart and some key steps of the proposed pattern recognition method. In Section 4, the experimental results for the proposed shearer cutting status recognition method are discussed. Finally, the conclusions are summarized in Section 5.

2. Basic Theories

2.1. Review of LMD Method

An LMD method can decompose a complicated signal into a set of product functions (PFs), each of which is the product of an amplitude envelope signal and a purely frequency-modulated signal. Specific decomposition steps can be summarized as follows:

(1) All of the local extrema n_i and the time tn_i of the original signal $x(t)$ are determined, and the mean value m_i of the two successive extrema n_i and n_{i+1} is calculated as follows:

$$m_i = \frac{n_i + n_{i+1}}{2}, \tag{1}$$

All mean values m_i are connected by straight lines between the corresponding time tn_i and tn_{i+1}, to generate the local mean segments. Then, the local mean segments are smoothed by amoving averaging method, to form a continuous local mean function $m_{11}(t)$.

(2) The local amplitude a_i can be given as follows:

$$a_i = \frac{|n_i - n_{i+1}|}{2}, \tag{2}$$

The moving averaging method is also used to smooth the local amplitude segments, in order to derive the envelope estimate function $a_{11}(t)$.

(3) The local mean function $m_{11}(t)$ is subtracted from the original signal $x(t)$, and the remnant signal, denoted by $h_{11}(t)$, can be given as follows:

$$h_{11}(t) = x(t) - m_{11}(t), \tag{3}$$

(4) The signal $h_{11}(t)$ is demodulated by the envelope estimate function $a_{11}(t)$, and the result, denoted by $s_{11}(t)$, can be calculated as follows:

$$s_{11}(t) = h_{11}(t)/a_{11}(t), \tag{4}$$

The envelope estimate function $a_{12}(t)$ of $s_{11}(t)$ is calculated. If $a_{12}(t)$ is not equal to one, $s_{11}(t)$ is not a purely frequency-modulated signal, and the above procedure for $s_{11}(t)$ should be repeated until a purely frequency-modulated signal $s_{1n}(t)$ is obtained. This is obtained when the envelope estimate function meets the condition that $a_{1(n+1)}(t)$ of $s_{1n}(t)$ is equal to one. Therefore,

$$\begin{cases} h_{11}(t) = x(t) - m_{11}(t) \\ h_{12}(t) = s_{11}(t) - m_{12}(t) \\ \quad\vdots \\ h_{1n}(t) = s_{1(n-1)}(t) - m_{1n}(t) \end{cases}, \tag{5}$$

where

$$\begin{cases} s_{11}(t) = h_{11}(t)/a_{11}(t) \\ s_{12}(t) = h_{12}(t)/a_{12}(t) \\ \quad\vdots \\ s_{1n}(t) = h_{1n}(t)/a_{1n}(t) \end{cases}, \tag{6}$$

(5) An envelope signal $a_1(t)$ can be derived by the product of all of the envelope estimate functions obtained during the iterative process described above.

$$a_1(t) = a_{11}(t)a_{12}(t)\cdots a_{1n}(t), \tag{7}$$

(6) The first product function PF_1 of the original signal can be generated by the product of the envelope signal $a_1(t)$ and the purely frequency-modulated signal $s_{1n}(t)$.

$$PF_1(t) = a_1(t)s_{1n}(t), \tag{8}$$

(7) PF_1 is then subtracted from the original signal $x(t)$, generating a new signal $u_1(t)$. The whole process is repeated k times, until $u_k(t)$ is a constant or monotonic function.

$$\begin{cases} u_1(t) = x(t) - PF_1(t) \\ u_2(t) = u_1(t) - PF_2(t) \\ \qquad\vdots \\ u_k(t) = u_{k-1}(t) - PF_k(t) \end{cases}, \tag{9}$$

Finally, the original signal $x(t)$ is decomposed into k PFs and a residual $u_k(t)$, and $x(t)$ can be reconstructed as follows:

$$x(t) = \sum_{i=1}^{k} PF_i + u_k(t), \tag{10}$$

According to above algorithm, we can realize that the LMD method is an adaptive signal decomposition method, based on the local extrema information of the signal itself.

2.2. Fuzzy C-Means Clustering Algorithm

An FCM clustering algorithm is an unsupervised dynamic clustering method which can divide a sample collection $X = \{x_1, x_2, \cdots, x_n\}$ into c categories ($2 \leq c \leq n$). The membership of the sample point x_i, belonging to the jth ($1 \leq j \leq c$) class, is defined as u_{ij}, and $\mathbf{U} = \{u_{ij}\}$ is the membership matrix, which can be used to denote the fuzzy cluster results. The membership matrix $\mathbf{U}$ possesses the following properties:

$$0 \leq u_{ij} \leq 1, \sum_{j=1}^{c} u_{ij} = 1, 0 < \sum_{i=1}^{n} u_{ij} < n, \tag{11}$$

An FCM clustering algorithm is employed to minimize the objective function J_{fcm}, under the constraint conditions (11).

$$J_{fcm}(\mathbf{U}, \mathbf{C}) = \sum_{i=1}^{n} \sum_{j}^{c} u_{ij}^m d_{ij}^2(x_i, c_j), \tag{12}$$

where m denotes the fuzzy weighting exponent, c_j is the jth clustering center, and d_{ij} denotes the Euclidean distance between the $\underline{i}$th sample and the jth clustering center, which can be described as $d_{ij} = \|x_i - c_j\|$. The main iterative process can be expressed as follows:

(1) Provide the number of clustering categories c, the fuzzy weighting exponent m, the iteration stop threshold ε, and the maximum number of iterations $T_{\max}$. Then, initialize the membership matrix $U^{(t)}$ and set the iterations number $t = 0$.

(2) The clustering center c_j can be calculated as follows:

$$c_j = \sum_{i=1}^{n} u_{ij}^m x_{ij} / \sum_{i=1}^{n} u_{ij}^m, \tag{13}$$

(3) The membership matrix $U^{(t+1)}$ can be updated as:

$$u_{ij}^{(t+1)} = \left[\sum_{l=1}^{c} \left(d_{ij}^{(t+1)} / d_{il}^{(t+1)} \right)^{2/(m-1)} \right]^{-1}, l = 1, 2, \cdots, c, \tag{14}$$

(4) If $\|U^{(t+1)} - U^{(t)}\| \leq \varepsilon$, then the iteration process terminates. Otherwise, set $t = t + 1$ and return to step (2).

(5) Finally, an optimal membership matrix U^* and clustering center C^* can be obtained.

(6) The principle of selecting the near is adapted to recognize the unmarked object types. The Hamming near-degree H between the unmarked object A and each clustering center c_j, is used to describe the similarity of the two fuzzy subsets. Suppose the number of variables in the sample is r, and the mathematical formula of H can be given as follows:

$$H(A, c_j) = 1 - \frac{1}{r}\sum_{k=1}^{r}\left|A(k) - c_j(k)\right|, \tag{15}$$

The larger $H(A, c_j)$ signifies that the two fuzzy subsets are more similar, and the unmarked object A with the largest $H(A, c_j)$ demonstrates that it belongs to the jth category.

3. The Proposed Pattern Recognition Method

In this section, a novel pattern recognition method, based on LMD and FCM, is presented. Firstly, the LMD method is used to preprocess the measured vibration signals, producing a set of PF components. Then, the optimal PF is selected according to the Kullback-Leibler divergence values between each PF and original signal. Following this, the features of the original signal and the optimal PF are extracted, using the time-frequency statistical parameters. Finally, an improved Laplacian score algorithm is proposed, in order to rank the extracted features, and appropriate features are selected. The obtained new feature vectors are the inputs of FCM, which cluster the different cutting statuses of the shearer. The flowchart of the proposed method is shown in Figure 1.
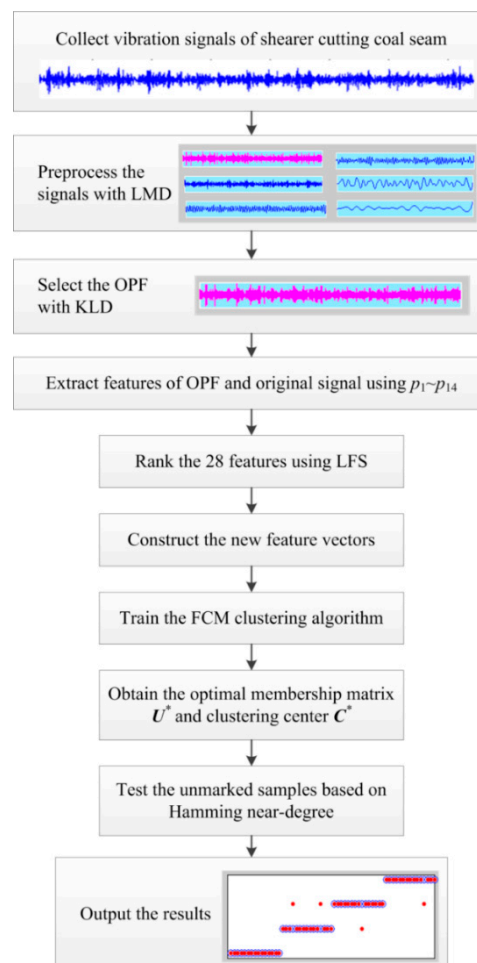

Figure 1. Flowchart of the proposed method.

3.1. The Optimal PF Component Selection Based on Kullback-Leibler Divergence

After LMD decomposition, several PF components can be obtained. However, it is unnecessary to conduct the feature extraction for all PFs. In order to choose an optimal PF component (OPF), which contains the most feature information, the Kullback-Leibler divergence (KLD) is taken as the criterion, to measure the relevancy between each PF and the original signal. The main steps of OPF selection with KLD can be described as follows:

(1) Suppose $p(x)$ and $q(x)$ are the probability density functions of original signal $x(t)$ and $PF_i(t)$, respectively. $p(x)$ can be defined as follows:

$$p(x) = \frac{1}{Nh}\sum_{i=1}^{N} K(\frac{x_i - x}{h}),\tag{16}$$

where N denotes the sampling number of $x(t)$, h is called the window width or smoothing parameter, and $K(*)$ denotes a kernel function and is commonly expressed as:

$$K(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2},\tag{17}$$

Likewise, the probability density function $q(x)$ can be obtained.

(2) The Kullback-Leibler distance between $x(t)$ and the ith *PF* can be defined as:

$$\delta_i(p,q) = \sum p(x)\log\frac{p(x)}{q(x)}, \delta_i(q,p) = \sum q(x)\log\frac{q(x)}{p(x)},\tag{18}$$

(3) The KLD between $x(t)$ and the ith *PF* can be calculated as:

$$D_i = \delta_i(p,q) + \delta_i(q,p),\tag{19}$$

The normalized KLD values can be obtained by Equation (20).

$$\hat{D}_i = D_i / \sum_{i=1}^{k} D_i,\tag{20}$$

The smaller the KLD value is, the closer the correlation between the PF and original signal is, and vice versa.

3.2. Feature Extraction Based on Time-Frequency Statistical Parameters

When the shearer drum is cutting different types of coal seams, the time domain and frequency domain properties of the collected vibration signal will change. Hence, the time domain and frequency domain feature parameters are comprehensively utilized, in order to acquire more feature information. In this paper, seven time domain feature parameters ($p_1 \sim p_7$) and seven frequency domain feature parameters ($p_8 \sim p_{14}$), are used to extract the features of the original signal. Likewise,

the 14 features ($p_{15} \sim p_{28}$) of the optimal PF component are also computed, presenting a total of 28 features. The formulas for calculating $p_1 \sim p_{14}$ are listed as follows:

$$
\begin{aligned}
&p_1 = \frac{1}{N} \sum_{i=1}^{N} x_i & &p_6 = \frac{x_{\max}}{\frac{1}{N} \sum_{i=1}^{N} |x_i|} & &p_{11} = \frac{1}{K-1} \sum_{j=1}^{K} (y_j - p_8)^2 \\
&p_2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - p_1)^2 & &p_7 = \frac{1}{N} \frac{\sum_{i=1}^{N} x_i^4}{p_3^4} & &p_{12} = \frac{1}{K} \sqrt{\sum_{j=1}^{K} \left[y_j (f_j - p_9)^2 \right]} \\
&p_3 = \sqrt{\frac{1}{N} \sum_{i=1}^{N} x_i^2} & &p_8 = \frac{1}{K} \sum_{j=1}^{K} y_j & &p_{13} = \frac{1}{K p_{12}^3} \sum_{j=1}^{K} \left[y_j (f_j - p_9)^3 \right] \\
&p_4 = \frac{1}{N} \sum_{i=1}^{N} (x_i - p_1)^3 & &p_9 = \frac{1}{\sum_{j=1}^{K} y_j} \sum_{j=1}^{K} (f_j y_j) & &p_{14} = \frac{1}{K \sqrt{p_{10}}} \sum_{j=1}^{K} \left[y_j \sqrt{|f_j - p_9|} \right] \\
&p_5 = \frac{1}{N} \sum_{i=1}^{N} x_i^4 & &p_{10} = \frac{1}{\sum_{j=1}^{K} y_j} \sqrt{\sum_{j=1}^{K} (f_j^2 y_j)}
\end{aligned}
\tag{21}
$$

where y_j denotes the frequency spectrum of the signal $x(i)$, K is the number of spectrum lines, and f_j denotes the frequency value of the jth spectrum line.

In these 14 features, the time domain feature parameters are mainly comprised of the mean, variance, root mean square, skewness, kurtosis, impulsion index, and kurtosis index. For the frequency domain feature parameters, p_8 reflects the vibration energy in the frequency domain. p_9 and p_{10} represent the change in the position of the dominant frequency. p_{11} and p_{14} characterize the degree of the spectrum's dispersion or concentration.

3.3. Feature Ranking and Selection Based on Improved Laplacian Score Algorithm

Although the extracted features can be used to identify the working status of equipment from a variety of aspects, they have different sensitivities to different working conditions. In addition, too many inputs for the FCM will consume more clustering time and reduce the clustering accuracy. Therefore, it is necessary to remove the insensitive features and refine the feature vectors. In this paper, an improved Laplacian score (LS) algorithm is introduced, in order to rank the extracted features. Following this, some optimal features are selected, according to their importance and locality preserving.

Suppose that the experiment is comprised of M samples, and that each sample has T features. In this study, L_r denotes the Laplacian score of the rth feature, and f_{ir} denotes the rth feature of the ith sample, where $i = 1, 2, \ldots, M$, $r = 1, 2, \ldots, T$. The feature vector of M samples can be used to construct a nearest neighbor graph G, and the weighting matrix S can then be calculated as the similarity matrix of graph G. The Laplacian score of the rth feature can be computed as follows [33]:

$$
L_r = \frac{\sum_{i,j=1}^{M} \left(\widetilde{f}_{ir} - \widetilde{f}_{jr} \right)^2 S_{ij}}{\sum_{i=1}^{M} \left(\widetilde{f}_{ir} - \frac{1}{M} \sum_{j=1}^{M} \widetilde{f}_{jr} \right)^2 D_{ii}} = \frac{\widetilde{\mathbf{f}}_r^T \mathbf{L} \widetilde{\mathbf{f}}_r}{\widetilde{\mathbf{f}}_r^T \mathbf{D} \widetilde{\mathbf{f}}_r},
\tag{22}
$$

where $\mathbf{f}_r = (f_{1r}, f_{2r}, \cdots, f_{Mr})^T$, $\mathbf{D} = \mathbf{SI}$, $\mathbf{I} = (1, 1, \cdots, 1)^T$, and $\mathbf{L}$ is the Laplacian matrix of graph G, and can be calculated as $\mathbf{L} = \mathbf{D} - \mathbf{S}$, $\widetilde{\mathbf{f}}_r = \mathbf{f}_r - (\mathbf{f}_r^T \mathbf{DI}) \mathbf{I} / (\mathbf{I}^T \mathbf{DI})$.

After calculating the LS value of each feature, the features can be ranked from a low score to a high score. However, the Laplacian score algorithm depends too much on the local structure information of adjacent samples. The selected features, based on LS, possess a stronger representational capacity for the similarity of adjacent samples, and a weaker distinguishing capability for global samples. To the best of our knowledge, the Fisher score (FS) algorithm is usually employed to measure the

separation degree between samples. In this work, FS is coupled with the LS algorithm, named the Laplacian-Fisher score (LFS), to select the optimal features from the entire 28 features.

The Fisher score of the *r*th feature can be written as follows:

$$F_r = \left[\sum_{ti=1}^{T} M_{ti}(\mu_{rti} - \mu)^2 \right] / \left[\sum_{ti=1}^{T} M_{ti}\sigma_{rti}^2 \right], \tag{23}$$

where *ti* denotes the category of the *i*th sample, M_{ti} denotes the number of samples of the *ti*th category, and μ_{rti} and σ_{rti} represent the mean and standard deviation of the *r*th features, respectively.

The feature with the largest Fisher score has a stronger distinguishing capability for global samples. In order to guarantee the selected features with a smaller LS and larger FS, a weighting factor α is introduced into the improved LS algorithm, and the LFS can be calculated as:

$$LF_r = |\alpha L_r - (1 - \alpha)F_r| \tag{24}$$

where $\alpha \in [0, 1]$. The improved LS method is degenerated to the basic LS method for $\alpha = 1$, while the improved LS becomes the basic FS method for $\alpha = 0$.

In order to illustrate the superiority of the improved LS method, the comparisons of LFS, LS, and FS are conducted. Two data sets: Glass and Wall-Following Robot Navigation (WFRN) [34], are selected from the UCI database for the simulations, to estimate the ranking performance of the three methods. The specific sample information is provided in Table 1.

Table 1. Statistical information of the two data sets.

Database	Number of Training Samples	Number of Testing Samples	Number of Categories	Number of Features
Glass	150	50	6	9
WFRN	500	100	4	24

The weighting factors α are set as 0.5, 1, and 0, and the corresponding feature selection methods are LFS, LS, and FS, which are used to rank the features of the two databases. The FCM clustering algorithm is used as the classifier and the classification results with different numbers of features are shown in Figure 2. In Figure 2, it can be clearly seen that the classification accuracies are significantly influenced by the feature ranking techniques. Compared with FS and LS, the LFS method has a superior ranking ability for the highest number of selected features, which may have a better purpose in terms of shear cutting status recognition. The reason for this is that LFS can comprehensively consider the power of locality preserving and global distinguishing capability, in order to evaluate the importance of a feature. Hence, the LFS method is utilized to rank the 28 features of vibration signals, and select an important feature vector for training the FCM clustering model.
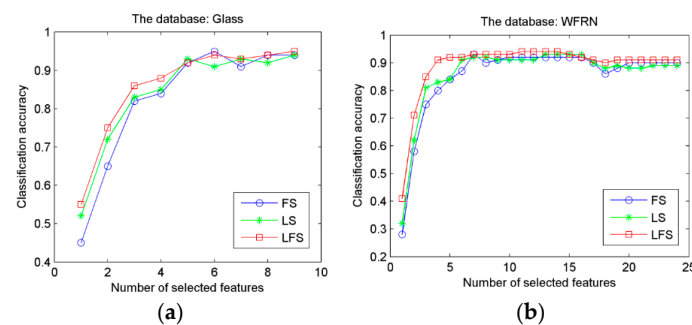

Figure 2. The classification results of different feature selection methods using the UCI database: (**a**) Glass; (**b**) WFRN.

4. Experimental Validation

In order to verify the effectiveness of the proposed method, the vibration signals under different shearer cutting statuses were analyzed. The experimental data were collected from a self-designed experimental system for shearer cutting coal, as shown in Figure 3. The coal seam was composed of three parts, according to the Protodikonov's hardness coefficient f. The first part was the coal seam with $f = 2$, the second was the coal seam with $f = 3$, and the last part was the coal seam with gangues. Different cutting statuses mean that the shearer was cutting different types of coal seam. Thus, the shearer mainly contained four cutting statuses, including the idling pattern, the coal seams with $f = 2$ and 3, and the coal seam with gangues, which were represented by the symbols of F1, F2, F3, and F4, respectively.
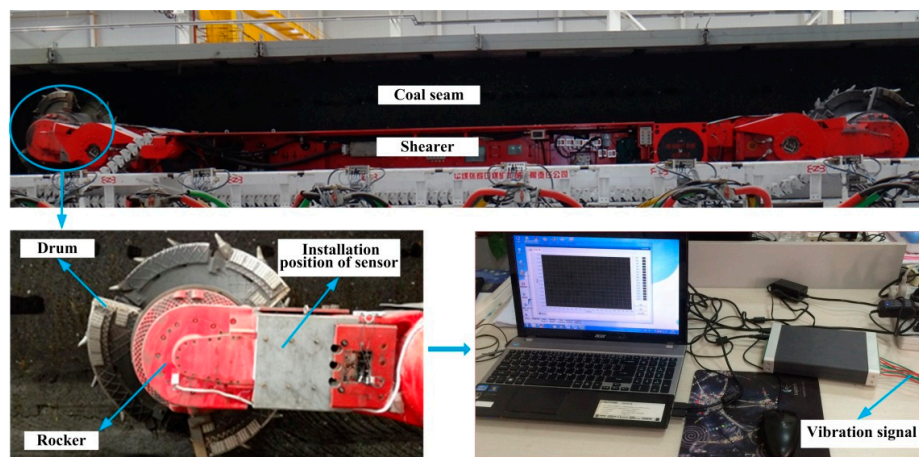
Figure 3. The self-designed experimental system for shearer cutting coal.

From the experimental system, 200 samples, with 50 data samples under each cutting status, were collected to verify the proposed algorithm, and each sample contained 5000 data points in it. The time domain waveforms of vibration acceleration signals under the four cutting statuses, were plotted in Figure 4.
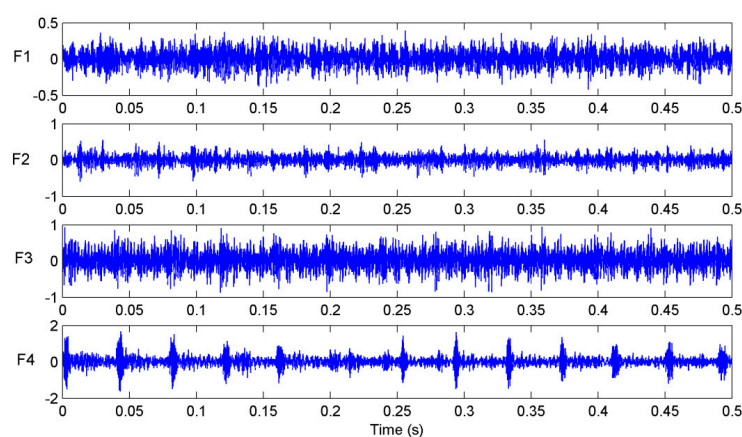

Figure 4. The vibration acceleration signals of each shearer cutting status.

In the proposed method, the vibration signals should be decomposed by initially using the LMD method. The vibration signals under the different cutting conditions could be decomposed into a sum of PF components. To save space, only the decomposition results of the vibration signal of shearer cutting coal seam with $f = 2$ were shown in Figure 5, as a representative.
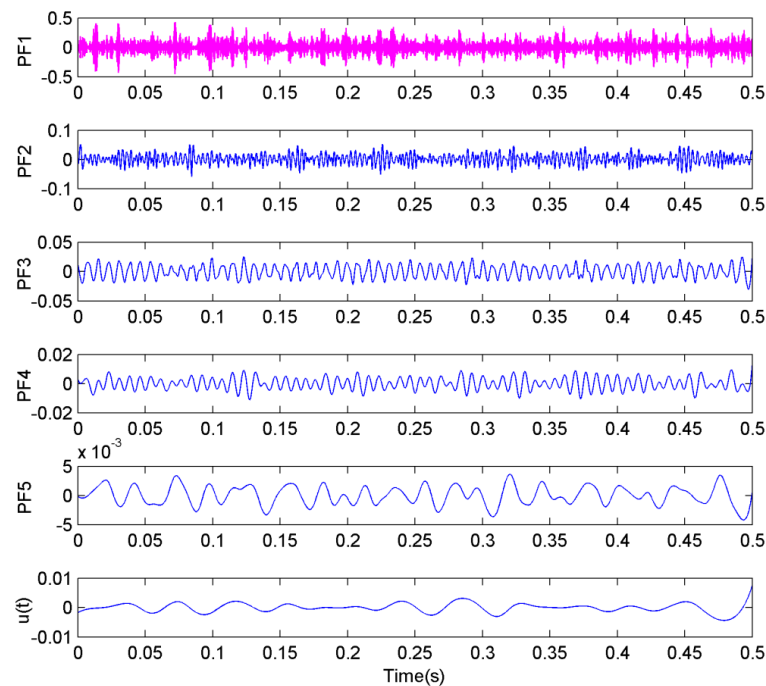
Figure 5. The LMD decomposition results of vibration signal of shearer cutting coal seam with $f = 2$.

According to the flowchart of the proposed recognition system, the Kullback-Leibler divergence (KLD) was then used to measure the relevancy between each PF and the original signal. The PF component with the lowest KLD value could be selected as the OPF component. Considering the LMD decomposition results in Figure 5 as an example, the normalized KLD value of each PF was calculated as 0.0018, 0.0125, 0.1454, 0.3578, and 0.4825, respectively. Therefore, the first PF component possessed the lowest KLD value (0.0018), and could be selected as the OPF component for further analysis.

The 14 time-frequency statistical parameters were utilized to extract the features of the original signal and the OPF component, and it was possible to obtain a total of 28 features. In the experiment, 140 samples with 35 data samples under each cutting status were randomly selected for training the FCM clustering model, and the rest of the samples were used to test its clustering performance. Furthermore, in order to determine a more appropriate weighting factor α, the improved LS algorithm was then used to rank the extracted 28 features, under different weighting factors. The weighting factor was set to 0, 0.2, 0.4, 0.6, 0.8, and 1, respectively, and the ranking results for each α, based on the LFS values, were shown in Table 2. From this table, it can be observed that the features have different sequences with various weighting factors.

Table 2. The feature sequences under different weighting factors.

Weighting Factor	New Order of the 28 Features
0	23, 26, 12, 7, 1, 9, 13, 6, 20, 18, 22, 11, 14, 2, 17, 10, 24, 3, 15, 8, 21, 5, 27, 25, 4, 28, 19, 16
0.2	23, 5, 6, 3, 15, 13, 10, 14, 28, 9, 24, 7, 1, 18, 19, 21, 20, 11, 27, 26, 12, 22, 2, 17, 8, 25, 4, 16
0.4	23, 9, 13, 10, 24, 19, 21, 5, 6, 20, 7, 1, 3, 15, 18, 22, 4, 8, 11, 14, 28, 27, 25, 2, 17, 26, 12, 16
0.6	23, 4, 8, 22, 11, 28, 27, 2, 17, 10, 3, 24, 5, 6, 15, 7, 1, 12, 16, 19, 21, 14, 13, 20, 18, 25, 26, 9
0.8	23, 15, 18, 5, 6, 20, 2, 17, 21, 7, 24, 19, 9, 13, 26, 12, 22, 4, 10, 1, 11, 14, 3, 8, 27, 25, 28, 16
1	23, 26, 12, 2, 17, 9, 13, 1, 3, 15, 21, 5, 6, 14, 28, 22, 4, 24, 19, 25, 16, 11, 27, 7, 18, 20, 10, 8

For each feature sequence in Table 2, the FCM clustering algorithm was used to cluster the training samples with different feature subsets, and the testing samples were classified according to the principle of selecting the near. Finally, we were able to obtain the highest classification accuracy, and the corresponding dimension of feature subsets under each weighting factor, as shown in Figure 6.
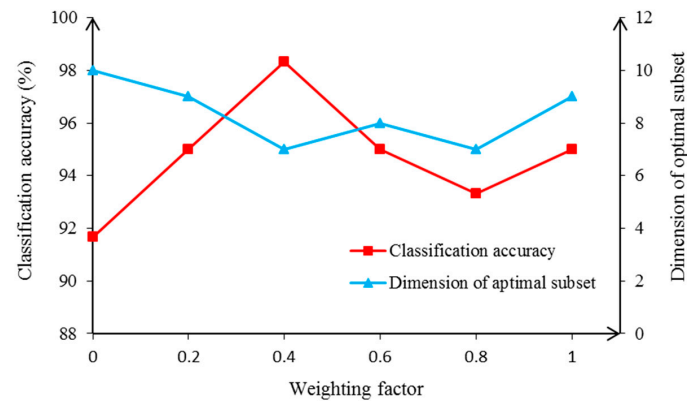
Figure 6. The classification results of FCM under each weighting factor.

It can be observed in Figure 6 that the dimension of the feature vectors for FCM can affect the clustering centers and the classification results of the FCM clustering model. The FCM clustering model was able to produce the highest classification accuracy (98.33 percent), and the lowest dimension of the optimal feature subset (7), for a weighting factor of $\alpha = 0.4$. The simulation results indicated that the LFS method possessed a superior ranking ability, when compared to other two methods, and the proposed pattern recognition method was applicable to shearer cutting status.

To verify the necessity of preprocessing the shearer cutting vibration signals with the LMD method, two simulations were performed. During the first, the original signal was decomposed by an EMD method, and the optimal IMF component was determined, according to the KLD values. Then, the 14 statistical parameters were used to extract the features of the original signal and the optimal IMF. In another simulation, the original signal was not preprocessed using any decomposition methods, and the 14 statistical parameters were directly used to extract the features of the original signal. The LFS method with $\alpha = 0.4$ was employed to select the optimal feature subset for the two simulations. Through the same process that was mentioned in the proposed methodology section, the classification results of the testing samples were produced, and can be seen in Figures 7 and 8. From Figure 7, three samples were misclassified by FCM clustering using the EMD method, and the classification accuracy only reached 95 percent, which was lower than the LMD method. As shown in Figure 8, when the original signal was not preprocessed by any decomposing methods, four samples were misclassified by FCM clustering, and the classification accuracy only reached 93.33 percent, which is lower than the other two methods. Therefore, the comparison results demonstrate the necessity to preprocess the original vibration signals with the LMD method, before extracting the characteristics. The reason for this is that the LMD method can restrain the interference noise and highlight the feature information hidden in the original signals.
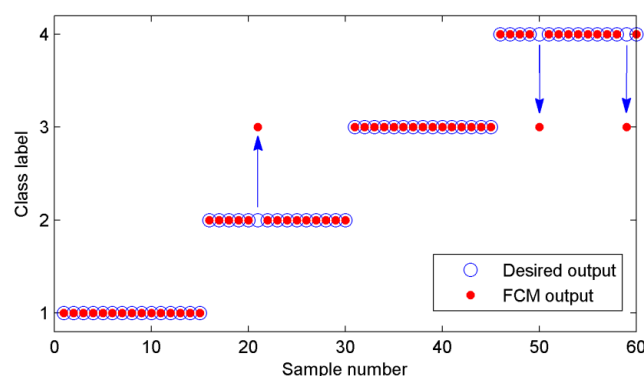

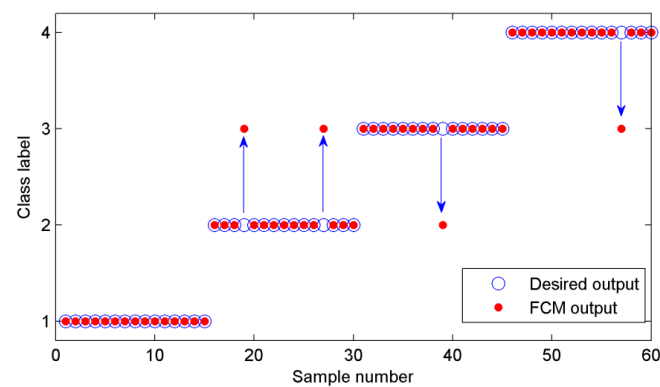**Figure 7.** The classification results of FCM clustering with EMD method.

Figure 8. The classification results of FCM clustering without reprocessing method.

Furthermore, in order to illustrate the significance of ranking the features based on the LFS method, seven features were randomly selected from the 28 features of the original and OPF components, and were then fed into the FCM, in order to distinguish the various shearer cutting conditions. The simulation conditions remained the same as those mentioned above, and the clustering results of the 60 testing samples can be seen in Figure 9. It can be observed from Figure 9 that the FCM clustering model can distinguish between the idling pattern and the other cutting conditions, but possesses a weaker clustering ability for identifying various cutting statuses. This is because the random selection of features did not contain enough feature information to form appropriate clustering centers for the FCM clustering model.
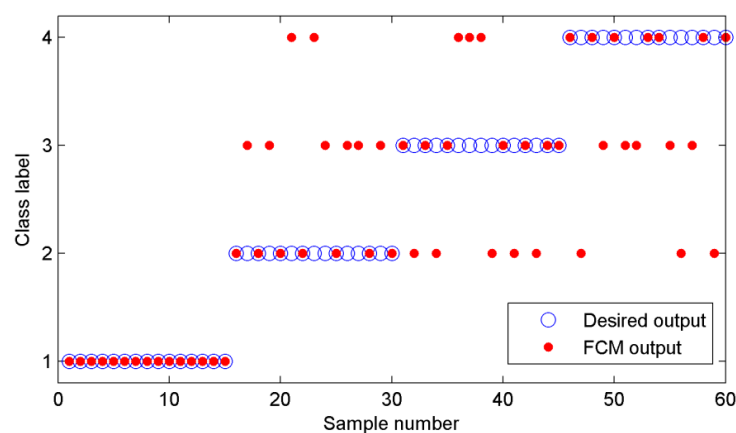

Figure 9. The classification results of FCM clustering without LFS method.

In order to illustrate the potential application of FCM in shearer cutting status recognition, a comparative study between the present work and other common methods, is shown in Table 3. The common methods include two clustering methods of K-means and a self-organizing map (SOM), and two artificial intelligence techniques of a BP neural network (BPNN) and support vector machine (SVM). Obviously, BPNN, SVM, and FCM classifiers, had a better recognition ability than the other two clustering methods, and the FCM clustering classifier obtained the highest recognition accuracy out of the five methods. The boundaries of the different shearer cutting statuses were usually ambiguous and could not be strictly distinguished. An FCM clustering algorithm has the theoretical basis of fuzzy mathematics and can reveal better classification performance in the recognition of shearer cutting statuses.

Table 3. Comparisons between the current work and some common methods.

Classification Methods	Training Samples (%)	Testing Samples (%)
K-means	95.71	86.67
SOM	94.82	83.33
BPNN	97.86	93.33
SVM	98.57	96.67
FCM	99.29	98.33

5. Conclusions

This paper presents a novel pattern recognition algorithm for shearer cutting status, based on LMD and FCM. In the proposed method, LMD is employed to preprocess the vibration signal of the shearer, in order to obtain a sum of PF components, and the KLD of each PF is calculated to select the OPF component. Then, 14 time-frequency statistical parameters are used to extract features from the original signal and the OPF. Furthermore, to solve the dimension disaster of the obtained 28 features, an improved LS method is presented, to rank the features and automatically select the optimal features by analyzing their importance and locality preserving. In addition, the FCM clustering model is introduced, to fulfill the classification of shearer cutting statuses. Finally, the experiments and some comparisons are presented, and the proposed method is proved feasible, illustrating a superior performance for the identification of the different cutting categories and working conditions of a shearer.

Author Contributions: Lei Si and Zhongbin Wang conceived and designed the experiments; Xinhua Liu and Lei Si performed the experiments; Chao Tan analyzed the data; Lei Si wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Si, L.; Wang, Z.; Tan, C.; Liu, X. A novel approach for coal seam terrain prediction through information fusion of improved D-S evidence theory and neural network. *Measurement* **2014**, *54*, 140–151. [CrossRef]
2. Si, L.; Wang, Z.; Liu, Z.; Liu, X.; Tan, C.; Xu, R. Health condition evaluation for a shearer through the integration of a fuzzy neural network and improved particle swarm optimization algorithm. *Appl. Sci.* **2016**, *6*, 171. [CrossRef]
3. Lee, J.; Kim, J.; Kim, H. Development of enhanced Wigner-Ville distribution function. *Mech. Syst. Signal Process.* **2001**, *13*, 367–398. [CrossRef]
4. Lin, J.; Qu, L. Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *J. Sound Vib.* **2000**, *234*, 135–148. [CrossRef]
5. Sanz, J.; Perera, R.; Huerta, C. Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms. *J. Sound Vib.* **2007**, *302*, 981–999. [CrossRef]
6. Fu, Y.; Chui, C.; Teo, C. Accurate two-dimensional cardiac strain calculation using adaptive windowed Fourier transform and Gabor wavelet transform. *Int. J. Comput. Assist. Radiol.* **2013**, *8*, 135–144. [CrossRef] [PubMed]
7. Huang, N.; Shen, Z.; Long, S.; Wu, M.; Shih, H.; Zheng, Q.; Yen, N.; Tung, C.; Liu, H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys.* **1998**, *454*, 903–995. [CrossRef]
8. Cheng, G.; Chen, X.; Li, H.; Li, P.; Liu, H. Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition. *Measurement* **2016**, *91*, 140–154. [CrossRef]
9. Ren, Y.; Suganthan, P.; Srikanth, N. A novel empirical mode decomposition with support vector regression for wind speed forecasting. *IEEE Trans. Neural Netw. Learn.* **2016**, *27*, 1793–1798. [CrossRef] [PubMed]
10. Yang, Y. Empirical mode decomposition as a time-varying multirate signal processing system. *Mech. Syst. Signal Process.* **2016**, *76–77*, 759–770. [CrossRef]

11. Smith, J. The local mean decomposition and its application to EEG perception data. *J. R. Soc. Interface* **2005**, *2*, 443–454. [CrossRef] [PubMed]

12. Cheng, S.; Zhang, K.; Yang, Y. Local mean decomposition method and its application to roller bearing fault diagnosis. *Chin. J. Mech. Eng.* **2009**, *20*, 2711–2717.

13. Li, B.; Xu, M.; Wang, R.; Huang, W. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *J. Sound Vib.* **2016**, *360*, 277–299. [CrossRef]

14. Wei, Y.; Xu, M.; Li, Y.; Huang, W. Gearbox fault diagnosis based on local mean decomposition, permutation entropy and extreme learning machine. *J. Vibroeng.* **2016**, *18*, 1459–1473.

15. Wodecki, J.; Stefaniak, P.; Obuchowski, J.; Wylomanska, A.; Zimroz, R. Combination of principal component analysis and time-frequency representations of multichannel vibration data for gearbox fault detection. *J. Vibroeng.* **2016**, *18*, 2167–2175.

16. Fu, S.; Liu, K.; Xu, Y.; Liu, Y. Rolling bearing diagnosing method based on time domain analysis and adaptive fuzzy C-means clustering. *Shock Vib.* **2016**, *2016*, 9412787.

17. Chen, Y.; Lin, C. Combining SVMs with various feature selection strategies. *Feature Extr.* **2006**, *207*, 315–324.

18. Ou, L.; Yu, D. Rolling Bearing fault diagnosis based on Laplacian score and fuzzy C-means clustering. *Chin. J. Mech. Eng.* **2014**, *25*, 1352–1357.

19. Feng, H.; Chang, G.; Kong, J. Fault diagnosis method of bearing based on Laplacian score and hypersphere support vector machine. *Comput. Meas. Control* **2015**, *23*, 1102–1105.

20. Xiao, Y.; Sheng, L.; Zhang, D. Supervised and unsupervised parallel subspace learning for larger scale image recognition. *IEEE Trans. Circuits Syst. Video* **2012**, *22*, 1497–1511.

21. Wang, L.P.; Wang, Y.L.; Chang, Q. Feature selection methods for big data bioinformatics: A survey from the search perspective. *Methods* **2016**, *111*, 21–31. [CrossRef] [PubMed]

22. Oh, I.-S.; Lee, J.-S.; Moon, B.-R. Hybrid genetic algorithms for feature selection. *IEEE Trans. Pattern Anal.* **2004**, *26*, 1424–1437.

23. Wang, L.P.; Zhou, N.; Chu, F. A general wrapper approach to selection of class-dependent features. *IEEE Trans. Neural Netw.* **2008**, *19*, 1267–1278. [CrossRef]

24. Kanai, R.; Desavale, R.; Chavan, S. Experimental-based fault diagnosis of rolling bearings using artificial neural network. *J. Tribol. Trans. ASME* **2016**, *138*, 031103. [CrossRef]

25. Chine, W.; Mellit, A.; Lughi, V.; Malek, A.; Sulligoi, G.; Pavan, A. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. *Renew. Energy* **2016**, *90*, 501–512. [CrossRef]

26. Zhang, L.; Suganthan, P.N. A Survey of randomized algorithms for training neural networks. *Inf. Sci.* **2016**, *364–365*, 146–155. [CrossRef]

27. Yang, C.; Hou, J. Fed-batch fermentation penicillin process fault diagnosis and detection based on support vector machine. *Neurocomputing* **2016**, *190*, 117–123. [CrossRef]

28. Wang, L.P.; Fu, X. *Data Mining with Computational Intelligence*; Springer: Berlin/Heidelberg, Germany, 2005.

29. Song, Y.; Jiang, Q.; Yan, X. Fault diagnosis and process monitoring using a statistical pattern framework based on a self-organizing map. *J. Cent. South Univ.* **2015**, *22*, 601–609. [CrossRef]

30. Liu, H.; Wang, X.; Jiao, Y. Damage identification for irregular-shaped bridge based on fuzzy C-means clustering improved by particle swarm optimization algorithm. *J. Vibroeng.* **2016**, *18*, 2149–2166.

31. Liu, L.; Sun, S.; Yu, H.; Yue, X.; Zhang, D. A modified fuzzy C-means (FCM) clustering algorithm and its application on carbonate fluid identification. *J. Appl. Geophys.* **2016**, *129*, 28–35. [CrossRef]

32. Rajaby, E.; Ahadi, S.; Aghaeinia, H. Robust color image segmentation using fuzzy C-means with weighted hue and intensity. *Digit. Signal Process.* **2016**, *51*, 170–183. [CrossRef]

33. He, X.; Cai, D.; Niyogi, P. *Laplacian Score for Feature Selection. Proceedings of Advances in Neural Information Processing System*; The Neural Information Processing Systems (NIPS) Foundation: Vancouver, BC, Canada, 2005; pp. 507–514.

34. Lichman, M. *UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]*; School of Information and Computer Science, University of California: Irvine, CA, USA, 2013.