Feature Extraction of Surface Electromyography Based on Improved Small-World Leaky Echo State Network

Xugang Xi
xixi@hdu.edu.cn
Wenjun Jiang
1301931380@qq.com
School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

Seyed M. Miran
miran@gwu.edu
Biomedical Informatics Center, George Washington University,
Washington, DC, 20052, U.S.A.

Xian Hua
huaxiang1206@126.com
Jinhua People's Hospital, Jinhua, 321000, China

Yun-Bo Zhao
ybzhao@ieee.org
Department of Automation, Zhejiang University of Technology,
Hangzhou 310023, China

Chen Yang
171060060@hdu.edu.cn
Zhizeng Luo
luo@hdu.edu.cn
School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

Surface electromyography (sEMG) is an electrophysiological reflection of skeletal muscle contractile activity that can directly reflect neuromuscular activity. It has been a matter of research to investigate feature extraction methods of sEMG signals. In this letter, we propose a feature extraction method of sEMG signals based on the improved small-world leaky echo state network (ISWLESN). The reservoir of leaky echo state network (LESN) is connected by a random network. First, we improved the reservoir of the echo state network (ESN) by these networks and used edge-added probability to improve these networks. That idea enhances the adaptability of the reservoir, the generalization ability, and the stability of ESN. Then we obtained the output weight of the network through training and used it as features. We recorded the sEMG

signals during different activities: falling, walking, sitting, squatting, going upstairs, and going downstairs. Afterward, we extracted corresponding features by ISWLESN and used principal component analysis for dimension reduction. At the end, scatter plot, the class separability index, and the Davies-Bouldin index were used to assess the performance of features. The results showed that the ISWLESN clustering performance was better than those of LESN and ESN. By support vector machine, it was also revealed that the performance of ISWLESN for classifying the activities was better than those of ESN and LESN.

1 Introduction

Surface electromyography (sEMG) is an electrophysiological reflection of skeletal muscle contractile activity. Since it can directly reflect neuromuscular activity, it has been widely used in clinical diagnosis and rehabilitation medicine (Cheng et al., 2018; Hu et al., 2018; Disselhorst-Klug, Williams, & von Werder, 2018). Multichannel sEMG can provide a safe and noninvasive control mode for artificial limb movement and other advanced man-machine interfaces (Mi, Zhou, Wei, Wang, & Zou, 2018; Gupta, Saxena, & Sazid, 2018). With the recent developments in detection technology, signal processing methods, and advances in computation, extracting useful features from original sEMG signals has become one of the hot issues in the study of sEMG (Al-Taee & Al-Jumaily, 2018).

Feature extraction, which is related to the quality of pattern recognition, is the key to the analysis and processing of sEMG. The time domain, the frequency domain, the time-frequency domain, and the nonlinear eigenstate analysis have been proposed for feature extraction (Zhang & Huang, 2015; Zhang, Huang, & Yang, 2013). Time-domain analysis (Lu, Chen, Li, Zhang, & Zhou, 2014) is based on the amplitude of the signal, and its algorithm is simple. The analysis, however, is used to deal with stationary signals, and its antijamming ability is poor (Altın & Er, 2016). Frequency domain analysis (Phinyomark, Limsakul, & Phukpattaranont, 2009) is mainly obtained by power spectral density. Time-frequency domain analysis is a combination of time and frequency that can characterize frequency changes at different moments or positions and provide a large amount of nonstationary information about the analyzed signals (Nazmi et al., 2016). Time-frequency analysis includes, for example, short-time Fourier transform, wavelet analysis, wavelet packet energy, and high-order spectrum analysis of Wigner-Ville distribution (Canal, 2010; Subasi, 2012; Farina, Févotte, Doncarli, & Merletti, 2004). The main nonlinear feature methods include fractal dimension, approximate entropy, Hurst exponent, and correlation dimension, which can extract hidden information from sEMG (Acharya, Ng, Swapna, & Michelle, 2011).

As the research in the area has advanced, many new, in-depth problems have surfaced that can hinder research progress. It is difficult to form complete support for filtering technology from the view of feature extraction alone. Statistics show that none of the features can fully reflect the mathematical characteristics of sEMG signals in the face of more, and more complex, feature extraction methods.

Peng, Peng, and Zhang (2018) used feature selection and ensemble learning to select the best feature set for gait phase detection and obtained an average accuracy of 94.1%. Bu, Guo, Ma, Xu, and Wei (2018) extracted a feature of bispectrum integration to recognize flexion and extension of continuous elbow, and the result showed that the classification accuracy of the proposed feature was higher than that of time-frequency feature. Pancholi and Joshi (2019) proposed a novel feature extraction method based on time-derivative moments to improve the performance for motion classification of the upper limb and obtained an acceptable recognition effect. Li, Li, Ju, Sun, and Kong (2018) proposed a novel feature of active muscle regions (AMR) to classify four hand motions; the result revealed that AMR had better classification performance than mean absolute value (MAV), waveform length (WL), zero crossing (ZC), and slope sign changes (SSC). Wen, Zhang, Qiu, Zeng, and Luo (2017) proposed a feature extraction method based on a two-dimensional matrix image for sEMG during finger motion, and they found that SVM can classify samples appropriately. Yu, Fan, Zhao, and Guo (2018) used zero-crossing rate, short-time energy, and linear predictive coefficient (LPC) to recognize hand gesture by backpropagation neural network and obtained an accuracy rate of 96.41% to 99.70%.

Although the theoretical background in the previous studies is different, they have used the same approach for feature extraction of sEMG. They have tried to find some functions of sEMG and use them as features. There are, however, some problems with this approach. sEMG is a complex chaotic time series that contains a lot of information. After feature extraction, one or several descriptive quantities can be obtained, providing very limited information. As a result, in the process of feature extraction, a large amount of information is lost, which limits the improvement of automatic classification accuracy of sEMG. Hence, it is necessary to seek a new feature extraction method that can retain a large amount of original information.

The time series contains the running information of the system for awhile. If the time series is analyzed and modeled, some inherent laws contained in the system can be obtained. Through these laws, the current or future situation of the system can be predicted. The prediction of time series can provide a fundamental theoretical basis and data support for people's judgment and decision making in many fields. Therefore, the analysis and prediction of time series have become a hot research topic in scientific research and practical engineering. The early time series prediction method usually used the simple weighted average method to obtain the prediction value for the data of the past time. The autoregression moving

average model (ARMA) is the most widely used time series prediction model in the early stage (Naik et al., 2018). Jaeger and Haas (2004) proposed the echo state network (ESN) to predict the chaotic time series, and their results showed that the prediction accuracy of the ESN was 2400 times higher than that of the earlier methods. ESN has attracted extensive attention in the field of time series prediction because of its simple training and high modeling accuracy.

The neuron model is an important research direction for optimizing ESN. Related research (Holzmann & Hauser, 2010) has shown that the neuron model would affect the nonlinear approximation and memory ability of the network. Jaeger, Lukoševičius, Popovici, and Siewert (2007) proposed a leaky integrator neuron to optimize ESN, and the result showed that the performance of the ESN based on leaky-integrator neurons was better than that of the traditional ESN as long as the appropriate leaky rate was selected.

The topological structure is also one of the optimizations of the reservoir of ESN. The basic idea to improve ESN is to replace the traditional sparse connection random structure with a structure with specific characteristics. Kawai, Park, and Asada (2019) investigated the learning performance of ESN with small-world topology as a reservoir. The results showed that this performance was better than that of the traditional ESN in nonlinear time-series prediction. Cui, Xiang, and Li (2012) used small-world, scale-free topology and a mixture of small-world and scale-free topologies to improve the reservoir topologies of ESN, respectively. The results showed the proposed ESN models have better prediction capabilities compared to conventional ESN.

Recent advances in machine learning and deep learning may have made it feasible to extract features from unstructured data without depending on any prior knowledge or assumptions (Goodfellow, Bengio, & Courville, 2016). Since the original data can be completely recovered with the extracted feature vectors, the unsupervised feature extraction can effectively retain useful information.

To address the problems of sEMG feature extraction, we applied ESN to this extraction. Moreover, we improved traditional ESN and proposed a novel sEMG feature extraction method based on improved small-world echo state network (ISWLESN). This novel method can achieve the unsupervised feature extraction of sEMG signals. In the process of feature extraction of sEMG signals, the loss of information is less than that of the previous methods.

2 Methods

2.1 Echo State Network. Figure 1 shows the topology of the traditional ESN. It consists of an input layer, a hidden layer (reservoir), and an output layer, in which $u(n)$, $x(n)$, and $y(n)$ (n is time step) represent the network
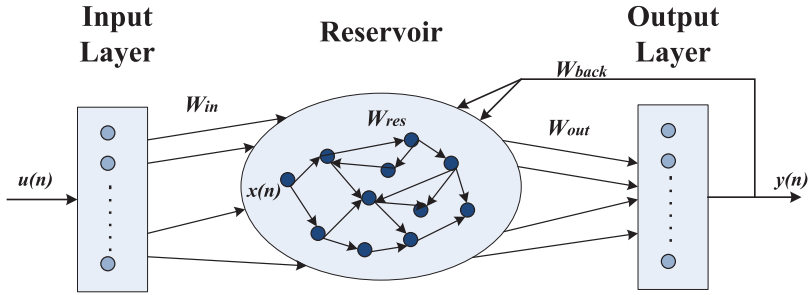
Figure 1: Topology of traditional ESN.

input signal, the state of the reservoir, and the output, respectively. Suppose that the ESN network has L input layer units, M output layer units, and N hidden layer units. Usually W_{in} is used as the connection weight matrix of the input layer, W_{res} as the connection weight matrix within the reservoir, W_{out} as the connection weight matrix of the output layer, and W_{back} as the feedback weight matrix from the output layer to the hidden layer. Among them, W_{in}, W_{res}, and W_{back} are generated randomly before the network is established, and W_{out} is calculated by training. The purpose of ESN is to determine the output weight value.

The hidden state of the ESN is updated as

$$x(n) = f(W_{in}u(n) + W_{res}x(n-1) + W_{back}y(n-1)), \tag{2.1}$$

where $f()$ represents the activation function of the internal state, usually the sigmoid function. When ESN is used for system identification and time series prediction, there is no need for output feedback (Jaeger, 2003; Li, Han, & Wang, 2012); therefore, $W_{back} = 0$.

The outputs of ESN are computed as

$$y(n) = g(W_{out}x(n)), \tag{2.2}$$

where $g()$ is the active function of the output unit (typically a linear function or a sigmoid function).

The reservoir, the core part of ESN, is a randomly generated, large-scale, sparse-join recursive structure. The parameter setting of reservoirs is a critical step in the training process of ESN, which has a significant influence on performance. The parameters of reservoir include reservoir size, internal connection power spectral radius, sparsity degree, and input scaling.

Here, reservoir size is the number of neurons in the reservoir, and the size selection of the reservoir is related to the number of samples, which has a high impact on network performance. As the reservoir size becomes

larger, the description of the given dynamic system by ESN becomes more accurate. Increasing the size, however, raises the overfitting problem.

In the experiments, W_{in} is initialized with random values from a uniform distribution. Then W_{res} is scaled by spectral radius λ_{max}

$$W_{res} = \alpha \frac{W_0}{\lambda_{max}}, \tag{2.3}$$

where W_0 is an initial connection weight matrix drawn from a uniform distribution, λ_{max} is the maximum absolute value of the eigenvalue of W_0 in the reservoir, and α is a scaling constant expressing the eigenvalue of W_{res}. Only when spectral radius satisfies $\alpha < 1$ does ESN have the property of echo state.

Not all neurons have connections. The sparsity degree indicates the percentage of processing unit connections in reservoir (usually 1% to 5% connections). This parameter can measure the richness of the vectors contained in the reservoir and affect the nonlinear approximation ability of the network. Higher richness of the vector of the network leads to its stronger linear approximation ability.

Input scaling is a scale factor needed to multiply the input of the reservoir before connecting to the internal neuron of the reservoir, that is, a scaling of the input signal. The stronger linearity of the target results in larger input scaling. Here the input scaling is set to 1.

The ESN training process determines the output weight matrix W_{out} by a given training sample. The first m values of the training sample are used to eliminate the effect of the initial states, as these internal states $x(1), x(2), \ldots, x(m)$ may be affected by the initial states. Therefore, the internal state $x(n)$ needs to be collected from the certain time $m + 1$ to calculate W_{out}. Then W_{out} can be calculated according to the pseudo-inverse operation,

$$W_{out} = YM^+, \tag{2.4}$$

where Y is the target output, $M = [\, x(m+1) \;\; x(m+2), \;\; \ldots, \;\; x(n)\,]$ is the internal state matrix, and M^+ is the pseudo-inverse matrix of M.

2.2 Leaky ESN. A leaky ESN (LESN) network is an improved model of the ESN network, with its reservoir composed of leaky integrating neurons. This type of neuron has independent state dynamics information and can be adapted to the timing characteristics of network learning tasks in various ways. LESN has the same topological structure as ESN, but modifies only the state update equation of neurons in the reservoir of ESN,

$$\begin{aligned}
x(n) = (1 - a)x(n - 1) \\
+ f(W_{in}u(n) + W_{res}x(n - 1) + W_{back}y(n - 1)),
\end{aligned} \tag{2.5}$$

where a is leaky rate.

It can be seen from equation 2.5 that ESN is a particular case of LESN at $a = 1$. Another parameter can neutralize the effect of changing one parameter on the result. As a result, the echo state characteristics of the network can remain unaffected. The state update of reservoir of traditional ESN is less affected by the previous state that will affect the memory function of the network to a certain extent. LESN, however, makes the state update have continuity, and it has certain advantages in slow learning and a continuous change system. The low leaky rate also leads to the gradual change of $x(n)$ in the internal neuron state that significantly enhances ESN's short-term memory ability. Therefore, when the parameter selection is appropriate, the reservoir performance of LESN is superior to that of traditional ESN.

The LESN mathematical model can be regarded as a low-pass filter acting on the state neurons of reservoir. The leaky rate controls the degree of preservation of the neuron state at the previous moment, and its cutoff frequency is determined by parameter a. The small value will result in the slow change of internal neuron state $x(n)$, which significantly enhances the short-term memory ability of ESN. LESN changes only the output connection weights in the training phase; the other weights are fixed.

2.3 Improved Small-World Leaky ESN. A small-world network is a kind of network with a short, characteristic path length and a high clustering coefficient. It is a new network structure that combines the advantages of regular and random networks. Through the connection weights of nodes, the adjacency matrix of the small-world network is formed. The value of the elements in the adjacency matrix is 1 or 0. This kind of connection is a deterministic one that makes the small-world network unable to reflect a complex network with ambiguity.

To address the problems we have identified, we propose an improved small-world network that directly represents the connection weights of the two nodes by using the edge-adding probability p. By establishing the function between the edge-adding probability and the distance between nodes, the weight value of a sparse connection between nodes is obtained. The range of values, which is between zero and one, indicates the degree of connection between nodes. A shorter distance between nodes predicts a higher probability of adding edges p, and, hence, a larger value of elements in the corresponding adjacency matrix. The range of elements in the adjacency matrix of the improved small-world network is between zero and one. We structured a regular network in a flat sheet, and each connection was rewired to a randomly selected node with probability p. In this letter, the value of the edge-adding probability p decreases exponentially as the distance between the nodes increases, as follows:

$$p = \alpha \times e^{-\beta \times d}, \tag{2.6}$$

where the range of p, which is between zero and one, represents the connection weights between nodes, and d denotes the Euclidean distance between

nodes, α is used to adjust the distance sensitivity, and β is used to adjust the overall density of the network. In this letter, we chose $\alpha = 0.2$, $\beta = 10$.

The performance of the LESN is related to its global parameters, such as leaky rate and spectral radius. However, these parameters of LESN are time invariant. To obtain the best prediction performance, it is necessary to modify the global parameters of LESN dynamically according to the characteristics of the time series. Since the small-world network can modify the topology of the network by changing the edge-adding probability, it has time-varying characteristics. Therefore, we propose an improved small-world LESN (ISWLESN), which can dynamically improve the reservoir topology of LESN by using the improved small-world network. That is, the improved small-world network is used to replace LESN's reservoir, which makes the LESN translate into ISWLESN. The dynamic adjacency matrix of the improved small-world network is treated as the sparse connection weight matrix of the reservoir of LESN. It not only guarantees the sparse connection between the neurons of the reservoir, but also reduces the blindness of random connections. The adaptability of the reservoir of LESN is improved.

ISWLESN uses the improved small-world network to dynamically direct the sparse connection of the reservoir, which makes ISWLESN become a time-varying and complex network. ISWLESN's equation of the state can be written as

$$x(n) = Ax(n-1)$$
$$+ f(W_{in}u(n) + W_{res}x(n-1) + W_{back}y(n-1)), \tag{2.7}$$

where A represents the connection weight matrix of the neuron of the reservoir.

Equation 2.7 is equivalent to the state update equation of the LESN reservoir in the ISWLESN model, and the leaky rate becomes a matrix compared to the traditional LESN. Therefore, A can also be regarded as the leaky rate parameter matrix of the ISWLESN model. When $A = (1-a)I$ (I is an identity matrix), equation 2.7 is converted to equation 2.5, which indicates that LESN is a special case of ISWLESN.

2.4 Feature Extraction of sEMG Based on ESN. Since the input weight W_{in} and internal weight W_{res} of ESN are fixed and the output weight is only weight adjusted, W_{out} of the time series realized reflects the inherent difference in the internal dynamics of our time series.

In this way, we propose a new feature extraction method of sEMG that can automatically learn sEMG features and adequately reflect the state of activities.

1. According to the specified parameters, the corresponding reservoir network is generated randomly, and the reservoir network is no longer changed.

2. The collected and processed sEMG signal $x_i(n)$ is used as the input signal of the network (i represents ith channel sEMG signal, and n is the sampling number of sEMG signal), and $x_i(n + 1)$ is used as the target of the network to train the ESN network. The output weight W_{out}^i after training is regarded as the feature of sEMG signal.

2.5 Performance Measurement. To evaluate the features extracted by three different methods, we transform the features into a class separability index (CSI) through Fisher discriminant function (Fisher, 1936; Huang, Zhu, Zhou, & Peng, 2018).

Suppose that the sample vector is $X^i = \{x_1^i, x_2^i, \ldots, x_N^i\}$ $i = 1, 2, \ldots, C$, with C and N the number of classes and sample of class i, respectively. Then the between-class scatter matrix S_b is computed as

$$S_b = \sum_{i=1}^{C} (m_i - m_m)(m_i - m_m)^T, \tag{2.8}$$

and the within-class scatter matrix S_W is computed as

$$S_W = \sum_{i=1}^{C} \sum_{j=1}^{N} (x_j^i - m_i)(x_j^i - m_i)^T, \tag{2.9}$$

where m_m is the mean value of all classes and m_i is the mean value of class i.

The CSI is computed as

$$CSI = \frac{\text{tr}(S_b)}{\text{tr}(S_W)}, \tag{2.10}$$

where $tr(S_b)$ is the trace of a matrix S_b, that is, the sum of the diagonal elements of the matrix. Therefore, a bigger CSI shows that the feature is better.

The Davies-Bouldin index (DBI) has been introduced to verify the reliability of the feature (Coelho et al., 2012). DBI is a validity index for unsupervised fuzzy clustering, which is used to determine whether the test set can be divided into several categories. In unsupervised clustering, it is often not clear how many types of sample data need to be classified. The principle is rooted in the degree of separation between different clusters and the degree of dispersion in individual cluster (Davies & Bouldin, 1979). Mathematically, DBI is computed as

$$DBI = \frac{1}{K} \sum_{i=1}^{K} \max_{j \neq i} \left\{ \frac{C_i + C_j}{d_{ij}} \right\} \tag{2.11}$$

where C_i and C_j are the degree of dispersion in class i and class j, d_{ij} represents the distance between the center of clusters i and j, and K is the number of clusters. Therefore, a smaller DBI shows that the feature is better.

Since DBI must be used under the unsupervised clustering method, we introduce the fuzzy C-means clustering (FCM) method (Wu & Yang, 2002). FCM is a clustering algorithm that determines the degree of each data point belonging to a certain cluster by membership degree. The main idea is to divide n vectors ($i = 1, 2, \ldots, n$) into c fuzzy groups, find the clustering centers of each group, and minimize the nonsimilar indexes(such as Euclidean distance).

3 Experimentation

Six healthy men (age range, 25 ± 2 years; weight range, 65 ± 5 kg; height range, 170 ± 5 cm) and six healthy women (age range, 23 ± 2 years; weight range, 46 ± 2 kg; height range, 162 ± 2 cm) were selected to investigate the recognition effect of lower limb movement and study the feature extraction method based on ESN. All the subjects read and signed an informed consent form approved by an institutional review board.

Trigno Wireless EMG (Delsys Inc., Natick, MA, USA) was used to record sEMG signals. Trigno sensors provide a 16-bit resolution, a 20 Hz to 450 Hz bandwidth, and baseline noise of less than 1.25 μV.

During human lower limbs motor, different muscles are involved in different activities; therefore, there may be obvious difference in the signals from different activities. These signals also have good periodicity. Therefore, we mainly collect these muscle signals for research and analysis. Each subject wore a Trigno wireless EMG sensor on his or her left leg during the test. The sampling frequency of the sensor was set to 1000 Hz, and each group of experiments consisted of falling, walking, sitting down, squatting, going upstairs, and going downstairs, as shown in Figure 2. The participants were instructed that walking, going upstairs, and going downstairs are controlled at a speed of about 1 m/s. We told them, "During sitting down, keep your feet as wide as you stand, and keep your upper body straight. During squatting, keep your feet as wide as you stand, and your upper body in a straight state. For falling, fall on the ground when the experimenter trips over an obstacle on the ground." Every activity was repeated 20 times and completed in 2 seconds. There was enough time before every activity to avoid muscle fatigue.

The collected signals were processed by the wavelet threshold denoising algorithm (Azzalini, Farge, & Schneider, 2005). The number of layers in wavelet decomposition was selected to compare scale coefficients with the threshold. Then these wavelet coefficients were reconstructed to obtain the denoised signal. Falling on the ground was done when the experimenter tripped over an obstacle. Figure 3 shows the four-channel sEMG of falling.

(A) falling   (B) walking   (C) sitting down

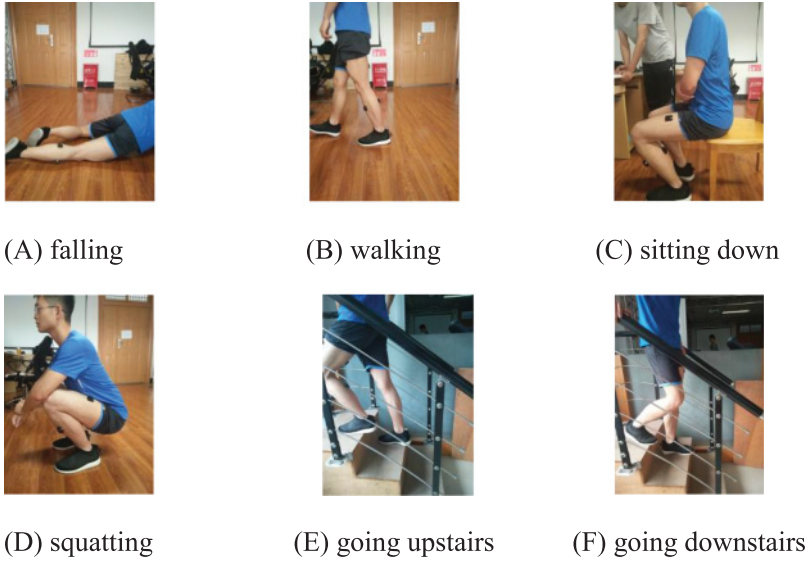(D) squatting   (E) going upstairs   (F) going downstairs
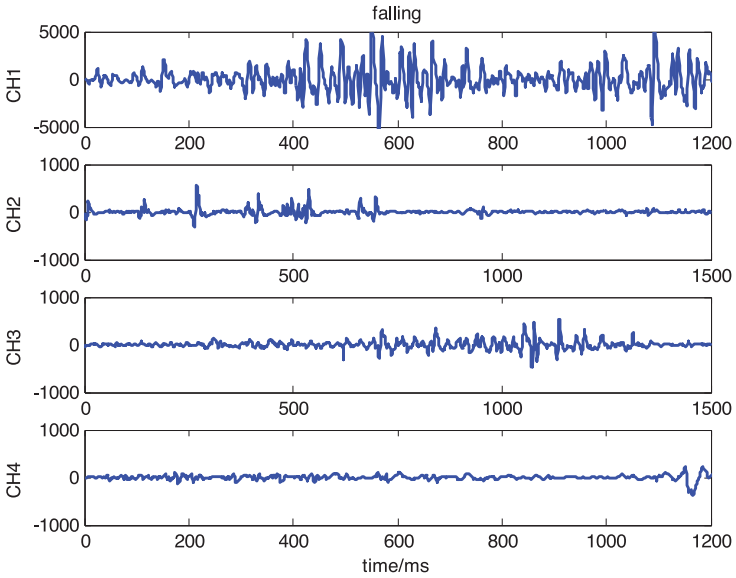
Figure 2: Experimental activities.

Figure 3: sEMG signals of falling.

In this figure, CH1, CH2, CH3, and CH4 represent gastrocnemius, tibialis anterior, vastus medialis, and vastus lateralis, respectively.

Since the network has 50 neurons and the dimension of the feature is 50, the large dimension reduces the quality of the feature space, recognition speed, and accuracy. Therefore, we need to reduce the dimension of the features. Principal component analysis (PCA) is the most commonly used linear dimension-reduction method (Jolliffe, 2011). Its goal is to map high-dimensional data to low-dimensional spatial representations through some linear projection. It is expected that the variance of the data in the projected dimension is the largest. This method helps to use fewer data dimensions while retaining more of the characteristics of the original data. Therefore, PCA reduces the feature dimensions to two dimensions.

4 Results

To verify the prediction effect of ISWLESN for the sEMG series, we trained the network to obtain output weight W_{out} by sEMG signals. Then we used a trained network to predict the track of the sEMG signal during the falling, with the prediction performance evaluated by root mean squared error (RMSE). The prediction result of sEMG can be seen in Figure 4. The result shows that ISWLESN can predict the track of sEMG appropriately, and the prediction error is very small ($RMES = 3.7463e^{-6}$V). Since the original data can be completely recovered with the extracted features by ISWLESN, we can assume that all the key information of the original data is likely to be stored in the extracted feature. Therefore, the output weight as a feature may have a better classification effect.

Scatter plots are used to show the clustering degree of the features, with a better clustering effect denoting a better feature. Figure 5 is the scatter plot of the sEMG of the vastus lateralis after three different feature extraction methods. Here, the extracted two-dimensional feature is defined as feature 1 and feature 2.

These three feature extraction methods can be well distinguished in Figure 5 for squatting and going downstairs. For falling, the ESN feature extraction method is not very effective, but LESN and ISWLESN can be easily distinguished. For the other three types of activities, the ESN feature distribution has a point crossing, and the effect is unsatisfactory. For walking and going upstairs, the feature points of the LESN are crossed and, unlike that of ISWLESN, cannot be easily distinguished. In short, ISWLESN clustering is better than that of LESN, and the LESN's is better than the ESN's.

To evaluate the significant difference in the six activities, a one-way analysis of variance (ANOVA) was performed for these activities. ANOVA was calculated using the *anova1()* function of Matlab. Table 1 shows the results. Because the *p*-value of ESN, LESN, and ISWLESN is less than 0.05, there are significant differences in these activities. Moreover, the *p*-value of ISWLESN
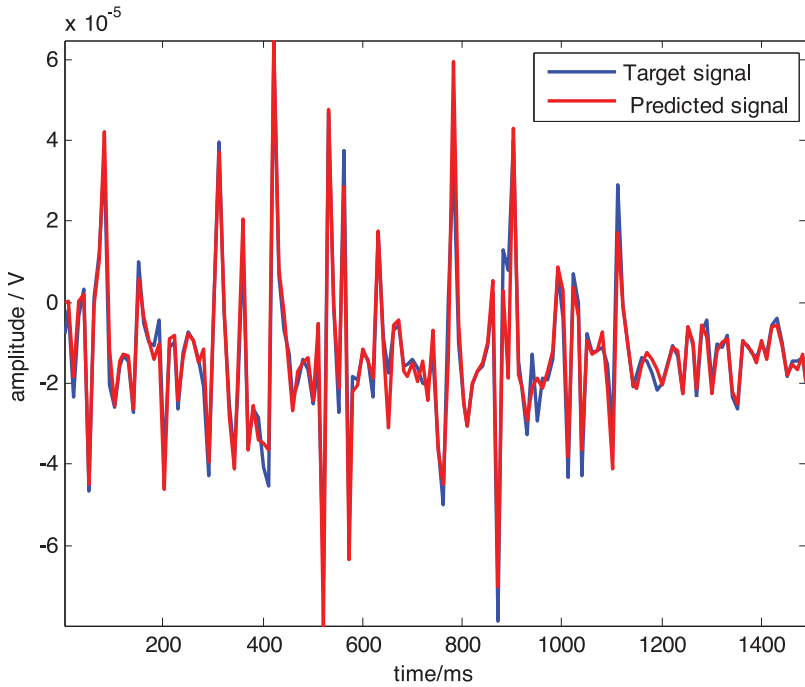
Figure 4: Prediction result of sEMG.

is the least one among the three features that indicates ISWLESN outperforms ESN and LESN significantly.
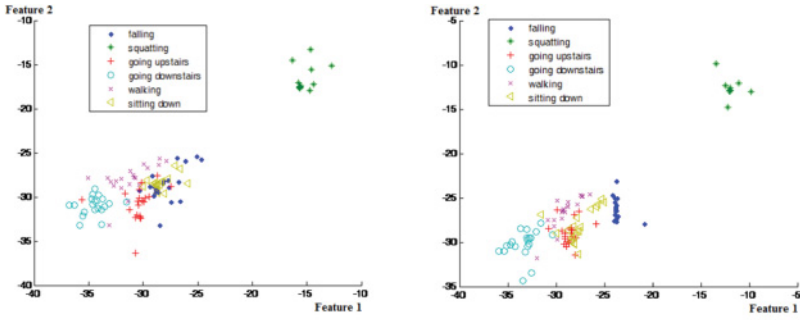
To compare the feature proposed in this letter with traditional features, we introduce the integral of absolute value (IAV) (Cheng, Chen, & Shen, 2013) and mean frequency (MF; Zardoshti-Kermani, Wheeler, Badie, & Hashemi, 1995).

The CSI of ISWLESN is bigger than the other features (see Table 2). The CSI of ESN is less than that of IAV and MF, which indicates that ESN clustering is less than the traditional features IAV and MF. But LESN and ISWLESN clustering is greater than that of IAV and MF.
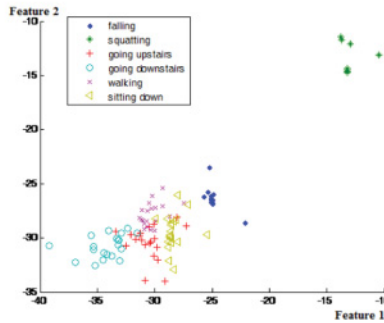
Also, we calculate the DBI of IAV, MPF, ESN, LESN, and ISWLESN, with the result recorded in Table 3. The table shows that it is possible to notice that ISWLESN is better able to separate these activities according to DBI, as the DBI of ISWLESN is smaller than those achieved by the others.

The features extracted by the three methods are respectively reduced by PCA, and then input into a support vector machine (SVM). Table 3 shows the recognition result of IAV, MF, ESN, LESN, and ISWLESN.

Table 4 demonstrates that ISWLESN has considerable success in the activity classification by comparing it with others. The accuracy of ESN is

Table 2: The CSI of IAV, MF, ESN, LESN, and ISWLESN.

Feature	IAV	MF	ESN	LESN	ISWLESN
CSI	0.0138	0.0154	0.0121	0.0180	0.0212

Table 3: DBI of Feature Based on IAV, MF, ESN, LESN, and ISWLESN.

Feature	IAV	MF	ESN	LESN	ISWLESN
DBI	4.8632	3.2546	5.0331	2.4779	1.8319

Table 4: Recognition Result of IAV, MF, ESN, LESN, and ISWLESN.

Feature	Falling	Squatting	Going Upstairs	Going Downstairs	Walking	Sitting Down
IAV	80.17	90.16	84.57	92.72	80.82	82.73
MF	90.36	95.54	86.92	95.28	89.25	84.63
ESN	74.36	97.46	80.46	88.36	78.36	75.59
LESN	96.49	97.78	85.78	97.64	92.36	87.27
ISWLESN	96.58	98.83	96.83	98.35	97.45	98.57

is highly dependent on the extracted features. This study aims to propose a feature extraction method of sEMG based on an improved small-world leaky echo state network. In this letter, we collected the sEMG during the activities of falling, walking, sitting, squatting, going upstairs, and going downstairs. Then we extracted corresponding features by ESN, LESN, and ISWLESN and used PCA to reduce feature dimension. To compare the features proposed in the letter with traditional features, we introduce IAV and MF. Results show that the clustering performance of ISWLESN is better than that of other features.

sEMG is a dynamic chaotic time series, and its traditional feature extraction is usually complex and challenging. First, a lot of useful information is lost in the process of feature extraction, which makes it difficult for classification with the traditional feature extraction methods to achieve higher accuracy. Second, there may be greater redundancy among feature extraction of sEMG from different angles. Third, feature sets of sEMG apply only to specific classification tasks. If different kinds of sEMG can be correctly distinguished in accordance with the descriptions of sEMG, the description of sEMG can be used as a feature. Therefore, the selection of features depends on the prior sample set, and it is worth considering whether the prior sample set can represent all the cases.

The feature extraction based on ISWLESN is approximately reversible— that is, the original sEMG can be restored almost from the extracted features

of sEMG. Thus, the loss of information in the feature extraction process is less, which ensures that the classification of sEMG based on ISWLESN can achieve higher classification accuracy. Moreover, ISWLESN can describe nonlinear chaotic time series appropriately. Therefore, the feature extraction of sEMG based on ISWLESN can well reflect the nonlinear dynamic characteristics of the sEMG time series.

Since feature extraction of sEMG based on ISWLESN is an unsupervised method, more useful information is retained in the feature extraction process rather than selecting useful information for a particular classification task. Therefore, this method has a broad application prospect in multitask classification.

Acknowledgments

References

Acharya, U. R., Ng, E. Y. K., Swapna, G., & Michelle, Y. S. L. (2011). Classification of normal, neuropathic, and myopathic electromyograph signals using nonlinear dynamics method. *Journal of Medical Imaging and Health Informatics*, *1*(4), 375–380.

Al-Taee, A. A., & Al-Jumaily, A. (2018). Optimal feature set for finger movement classification based on sEMG. In *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5228–5231). Piscataway, NJ: IEEE.

Altın, C., & Er, O. (2016). Comparison of different time and frequency domain feature extraction methods on elbow gesture's EMG. *European Journal of Interdisciplinary Studies*, *2*(3), 35–44.

Azzalini, A., Farge, M., & Schneider, K. (2005). Nonlinear wavelet thresholding: A recursive method to determine the optimal denoising threshold. *Applied and Computational Harmonic Analysis*, *18*(2), 177–185.

Bu, D., Guo, S., Ma, H., Xu, H., & Wei, C. (2018). Pattern recognition of continuous elbow joint movements using bispectrum-based sEMG. In *Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation* (pp. 551–556). Piscataway, NJ: IEEE.

Canal, M. R. (2010). Comparison of wavelet and short time Fourier transform methods in the analysis of EMG signals. *Journal of Medical Systems*, *34*(1), 91–94.

Cheng, J., Chen, X., & Shen, M. (2013). A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals. *IEEE Journal of Biomedical and Health Informatics*, *17*(1), 38–45.

Cheng, J., Wei, F., Li, C., Liu, Y., Liu, A., & Chen, X. (2018). Position-independent gesture recognition using sEMG signals via canonical correlation analysis. *Computers in Biology and Medicine*, *103*, 44–54.

Coelho, G. P., Barbante, C. C., Boccato, L., Attux, R. R., Oliveira, J. R., & Von Zuben, F. J. (2012). Automatic feature selection for BCI: An analysis using the Davies-Bouldin index and extreme learning machines. In *Proceedings of the 2012 International Joint Conference on Neural Networks* (pp. 1–8). Piscataway, NJ: IEEE.

Cui, H., Xiang, L., & Li, L. (2012). The architecture of dynamic reservoir in the echo state network. *Chaos*, *22*(3), 455.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*, 224–227.

Disselhorst-Klug, C., Williams, S., & von Werder, S. C. (2018). Surface electromyography meets biomechanics or bringing sEMG to clinical application. In *Proceedings of the International Conference on NeuroRehabilitation* (pp. 1013–1016). Cham: Springer.

Farina, D., Févotte, C., Doncarli, C., & Merletti, R. (2004). Blind separation of linear instantaneous mixtures of nonstationary surface myoelectric signals. *IEEE Transactions on Biomedical Engineering*, *51*(9), 1555–1567.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Gupta, R., Saxena, S., & Sazid, A. (2018). Channel selection in multi-channel surface electromyogram based hand activity classifier. In *Proceedings of the 2018 4th International Conference on Computational Intelligence and Communication Technology* (pp. 1–7). Piscataway, NJ: IEEE.

Holzmann, G., & Hauser, H. (2010). Echo state networks with filter neurons and a delay&sum readout. *Neural Networks*, *23*(2), 244–256.

Huang, Z., Zhu, H., Zhou, J. T., & Peng, X. (2018). Multiple marginal Fisher analysis. *IEEE Transactions on Industrial Electronics*, *66*, 9798–9807.

Hu, Y., Wong, Y., Wei, W., Du, Y., Kankanhalli, M., & Geng, W. (2018). A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. *PLOS One*, *13*(10), e0206049.

Jaeger, H. (2003). Adaptive nonlinear system identification with echo state networks. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems, 15* (pp. 609–616). Cambridge, MA: MIT Press.

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, *304*(5667), 78–80.

Jaeger, H., Lukoševičius, M., Popovici, D., & Siewert, U. (2007). Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Networks*, *20*(3), 335–352.

Jolliffe, I. (2011). *Principal component analysis*. Berlin: Springer.

Kawai, Y., Park, J., & Asada, M. (2019). A small-world topology enhances the echo state property and signal propagation in reservoir computing. *Neural Networks*, *112*, 15–23.

Li, D., Han, M., & Wang, J. (2012). Chaotic time series prediction based on a novel robust echo state network. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(5), 787–799.

Li, G., Li, J., Ju, Z., Sun, Y., & Kong, J. (2018). A novel feature extraction method for machine learning based on surface electromyography from healthy brain. *Neural Computing and Applications*, *31*, 1–10.

Lu, Z., Chen, X., Li, Q., Zhang, X., & Zhou, P. (2014). A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices. *IEEE Transactions on Human-Machine Systems*, *44*(2), 293–299.

Mi, C., Zhou, T., Wei, B., Wang, Y., & Zou, L. (2018). Design of high-accuracy eight-channel surface electromyography acquisition system and its application. *European Physical Journal Special Topics*, *227*(7–9), 933–942.

Naik, G. R., Selvan, S. E., Arjunan, S. P., Acharyya, A., Kumar, D. K., Ramanujam, A., & Nguyen, H. T. (2018). An ICA-EBM-based sEMG classifier for recognizing lower limb movements in individuals with and without knee pathology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(3), 675–686.

Nazmi, N., Abdul Rahman, M., Yamamoto, S. I., Ahmad, S., Zamzuri, H., & Mazlan, S. (2016). A review of classification techniques of EMG signals during isotonic and isometric contractions. *Sensors*, *16*(8), 1304.

Pancholi, S., & Joshi, A. M. (2019). Time derivative moments based feature extraction approach for recognition of upper limb motions using EMG. *IEEE Sensors Letters*, *3*(4), 1–4.

Peng, F., Peng, W., & Zhang, C. (2018). Evaluation of sEMG-based feature extraction and effective classification method for gait phase detection. In *Proceedings of the International Conference on Cognitive Systems and Signal Processing* (pp. 138–149). Singapore: Springer.

Phinyomark, A., Limsakul, C., & Phukpattaranont, P. (2009). A novel feature extraction for robust EMG pattern recognition. *J. Comput.*, *1*, 71–80.

Subasi, A. (2012). Classification of EMG signals using combined features and soft computing techniques. *Applied Soft Computing*, *12*(8), 2188–2198.

Wen, T., Zhang, Z., Qiu, M., Zeng, M., & Luo, W. (2017). A two-dimensional matrix image based feature extraction method for classification of sEMG: A comparative analysis based on SVM, KNN and RBF-NN. *Journal of X-Ray Science and Technology*, *25*(2), 287–300.

Wu, K. L., & Yang, M. S. (2002). Alternative c-means clustering algorithms. *Pattern Recognition*, *35*(10), 2267–2278.

Yu, H., Fan, X., Zhao, L., & Guo, X. (2018). A novel hand gesture recognition method based on 2-channel sEMG. *Technology and Health Care*, *26*(S-1), 1–10.

Zardoshti-Kermani, M., Wheeler, B. C., Badie, K., & Hashemi, R. M. (1995). EMG feature evaluation for movement control of upper extremity prostheses. *IEEE Transactions on Rehabilitation Engineering*, *3*(4), 324–333.

Zhang, X., & Huang, H. (2015). A real-time, practical sensor fault-tolerant module for robust EMG pattern recognition. *Journal of Neuroengineering and Rehabilitation*, *12*(1), 18.

Zhang, X., Huang, H., & Yang, Q. (2013). Real-time implementation of a self-recovery EMG pattern recognition interface for artificial arms. In *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5926–5929). Piscataway, NJ: IEEE.