Journal of Computers Vol. 31 No. 1, 2020, pp. 274-281
doi:10.3966/199115992020023101026

Gesture Recognition System Based on Normalized Neural Network

Yangyang Guo, Hezhi Lin*, Zhibin Gao, Lianfen Huang

Department of Communication Engineering, Xiamen University, Xiamen, China
linhezhi@xmu.edu.cn

Received 13 January 2019; Revised 3 March 2019; Accepted 12 March 2019

Abstract. With the development of technology, gesture interaction has gradually become an emerging way of interaction. In this paper, the surface potential electrode is used to guide the action potential signal. After rectification and filtering, it is collected by the single-chip microcomputer NRF52832. In order to solve the problem of the inconsistent length of the active segment in the actual application, we normalize the action sequence of the active segment and process it. We use neural networks, SVM algorithms, DTW algorithms for classification, and the classification effect of neural networks is the best of the three algorithms. The network can recognize four types of gestures, fist, wrist inversion, wrist eversion, vertical The recognition rates of the four types of gesture motion waveforms are: 96%, 100%, 96%, and 100%. Using CPU on Windows, the average recognition time per gesture is about 0.0748ms.

Keywords: gesture recognition, neural network, normalized

1 Introduction

Gesture recognition is an important part of human-computer interaction, and its research and development affects the naturalness and flexibility of human-computer interaction. It can be applied to multiple scenes, such as sign language recognition, controlling prosthetics, controlling wheelchairs or applying to AR virtual scenes, etc., providing people with a lot of convenience.

Gesture recognition can use a variety of sensors for data acquisition. Common sensors are: camera, data glove, IMUS, EMG, and image sensing as a positioning of a given space may be a very good solution, but there are indeed many inconveniences for gesture recognition. Image sensing is not easy to distinguish between left and right hands, left and right hands are crossed or occluded, and there is no way to identify them, but there is no such limitation in myoelectricity. At the same time, EMG recognition is not affected by the intensity of external light and the distance of operation. For mechanical sensing, the huge size of the device itself has become a hindrance to mobilization, and even can not be used outdoors. Compared to the other two recognition methods, EMG has a slightly lower power consumption and mobility. EMG signals can be used to reflect different gestures, and EMG signals are the most stable and mature control signal source. So far, sEMG is the most used prosthetic control signal source. Signal control source. This article will use SEMG signals in gesture recognition.

1.1 Research Work of Gesture Recognition at Home and Abroad

Cai Liyu et al. used wavelet transform to analyze the EMG signal, extracted the wavelet coefficient structure feature vector, and then classified it through the neural network. The recognition rates of the four movements of the test focus, fist, forearm and forearm were respectively 90%, 100%, 90%, 80% [1]; Marie-Françoise Lucas et al. 151 explored the use of support vector machine SVM to classify the action EMG signal, the recognition rate was 95.3% [2]; Min et al. proposed to use EEMD to decompose surface signals and then use IMF to form feature vectors to reconstruct energy features. Then, the four types of

* Corresponding Author

gestures were classified by neural network, and the recognition rate was 95.0% [3]. To achieve real-time gesture recognition, Hang et al. used the moving average energy to detect the effective action segment to extract the signal features, and then implemented the gesture recognition by DTW algorithm [4]. Lian et al. used decision tree and KNN to classify, the recognition rate was 89% [5]. Aguiar et al. used MYO to collect data and extracted nine features of the gesture signal, using PCA for classification, with a recognition rate of 81% [6].

There are two commonly used gesture recognition methods: template matching and statistical methods. Template matching method such as the commonly used DTW [7] algorithm. A disadvantage of this type of method is its poor resistance to noise. Gesture recognition based on statistical methods has better dynamic adaptability, such as the commonly used hidden Markov model (HMM).

In the past, gesture recognition systems mainly used artificial neural network-based methods. The neural network method has classification characteristics and anti-interference. However, due to the different lengths of time series, its ability to process time series is not strong. The famous Fels GloveTalk system uses the neural network method as the identification technology. For the sign language signals in the analysis interval, the HMM method is usually used for modeling. HMM is a well-known and widely used statistical method. The HMM under the general topology has a very strong ability to describe the spatio-temporal variation of gesture signals. It has always dominated the field of dynamic gesture recognition, but it is due to the generality of the HMM topology. This model is too complicated to analyze sign language signals, making the HMM training and recognition calculations too large. Especially in continuous HMM, because a large number of state probability densities need to be calculated, the number of parameters to be estimated is large, which makes the training and recognition speed relatively slow.

1.2 Our Work

From a practical point of view, when people make gestures, the speed of different people's gestures may be different. When the same person does the same action at different time points, the speed of gestures may also be different. And the duration of different kinds of gestures may also be different, which leads to a problem when we apply gesture recognition to the actual situation, that is, the speed of the action will make the sequence length of the active segment different. researchers use the wavelet transform, EEMG, DTW algorithm, etc. to extract the active segment sequence features before entering the network, instead of directly inputting into the neural network for convolution and then feature extraction.

The neural network method has strong classification characteristics and strong anti-interference ability, but its ability to process time series is not strong. In this paper, we aim at classification and anti-interference. We choose neural networks. In order to solve the problem that the sequence length of the active segment is different and the time series of the neural network processing is not strong, this paper adopts the normalization process of the active segment action sequence, so that the motion sequence length of all active segments is the same. The processed active segment motion sequence is then directly input into the neural network, and the system described herein can identify four types of gesture waveforms. And we use the commonly used SVM and DTW algorithm to classify the active segment sequence, and finally compare the result with our method.

2 System Design

Fig. 1 is an overall block diagram of a gesture recognition system based on a normalized neural network.

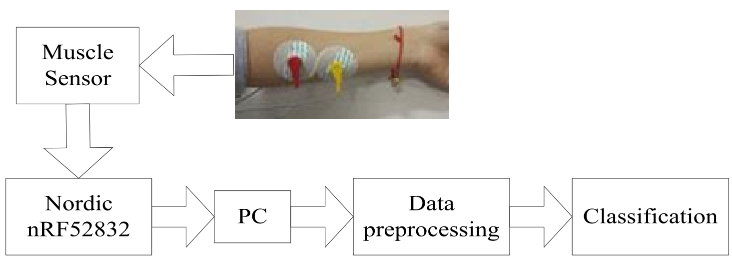


Fig. 1. Overall block diagram of gesture recognition system based on normalized neural network

2.1 Collection of Myoelectric Signals

We use Muscle Sensors to collect myoelectric signals. Muscle Sensors is shown in Fig. 2. The Muscle Sensors will amplify, correct and smooth the acquired EMG signals. The nrf52832 will acquire AD at a frequency of 500HZ, and collect the data to the computer through the serial port. And store. We use the surface electrode, the electrode placement Fig. 3.

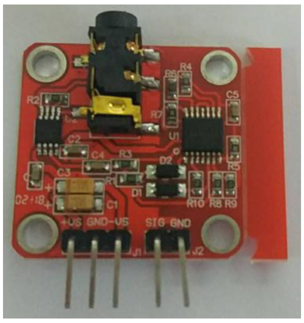


Fig. 2. Muscle Sensors

We collected gesture data for 4 subjects. There were 4 types of gestures, fist fist, wrist varus, wrist valgus, and vertical thumb. The four types of gestures are shown in Fig. 3. Among the 4 subjects, 2 males and 2 females, 50 sets of EMG signals were collected for each type of gesture, and 25 sets of gesture data were collected from each of the left and right hands of each subject. Four subjects performed four types of gestures, and collected a total of 800 sets of data. Considering the effects of muscle fatigue factors, Subjects performed a gesture every 3 seconds and took a break of 2 minutes after collecting 10 sets of data.

Fig. 3. Electrode placement

3 Gesture Recognition

3.1 Data Set Preprocessing

In this paper, we use a sequence to train and test the network. Because we are doing gestures, the duration of the active segment of the gesture waveform of the same person at different time points is different, and the duration of the active segment of different human gesture waveforms is different. The duration of the active segment of the gesture waveform of the category is also different, resulting in the sequence length of all active segments may be different, but the input neurons of the neural network are fixed. So we need to preprocess the data set to normalize the sequence length of all active segments to the same.

By analyzing the waveform, the sequence length of the waveform with sequence length 784 is selected. Next, we start processing the active segment motion sequence.

Proceed as follows:

(1) Detect the length of the active segment action sequence, denoted as m;

(2) Compare m with 784;

(3) When $m > 784$, the active segment motion sequence is resampled, and the sampling frequency is m/784hz;

(4) When $m < 784$, the active segment action sequence is resampled so that the new active segment action sequence length is 784.

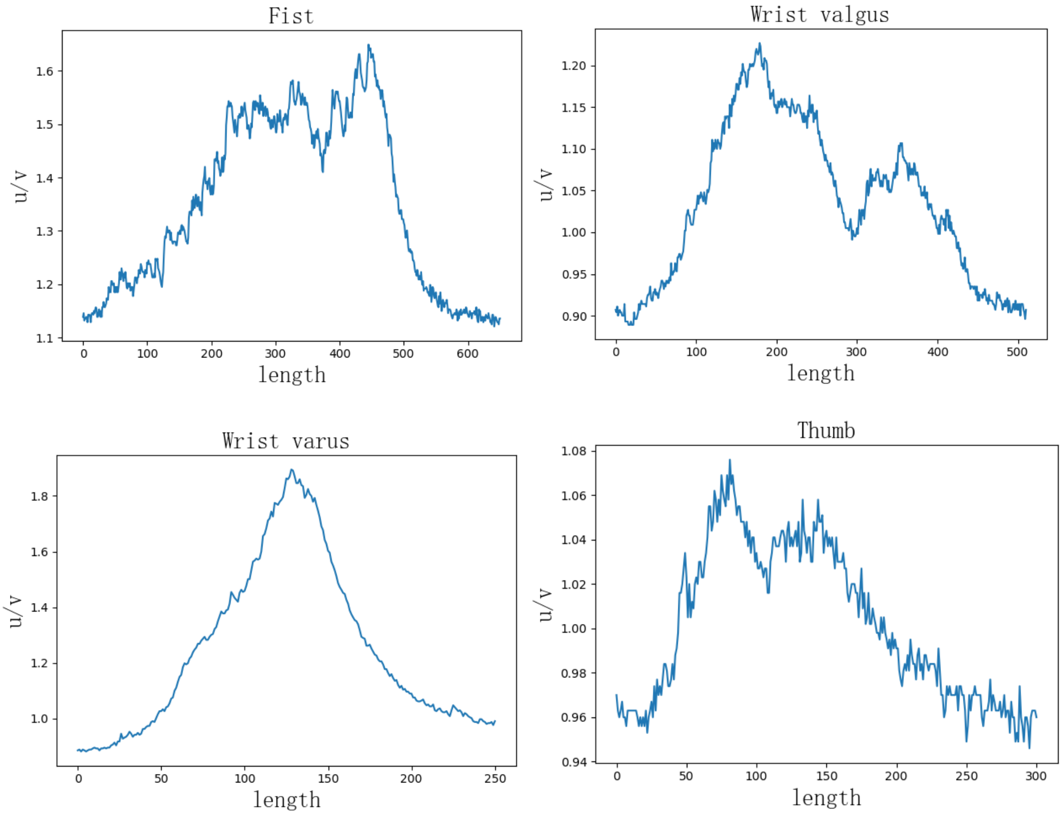


Fig. 4. Active segment motion sequence of surface EMG signals of 4 types of motion

Fig. 4 shows the active segment motion sequence of the surface EMG signal of the four types of gestures. The sequence of the active segment is processed by the nearest, linear, spline, and pchip algorithms. The effect is shown in Fig. 5. It can be seen that the smoothing effect by spline algorithm is the best, so this paper uses spline algorithm. Fig. 6 is the action sequence of the new active segment after the action sequence of the active segment of the surface EMG signal of the four types of gestures. The new active segment action sequence length is 784.

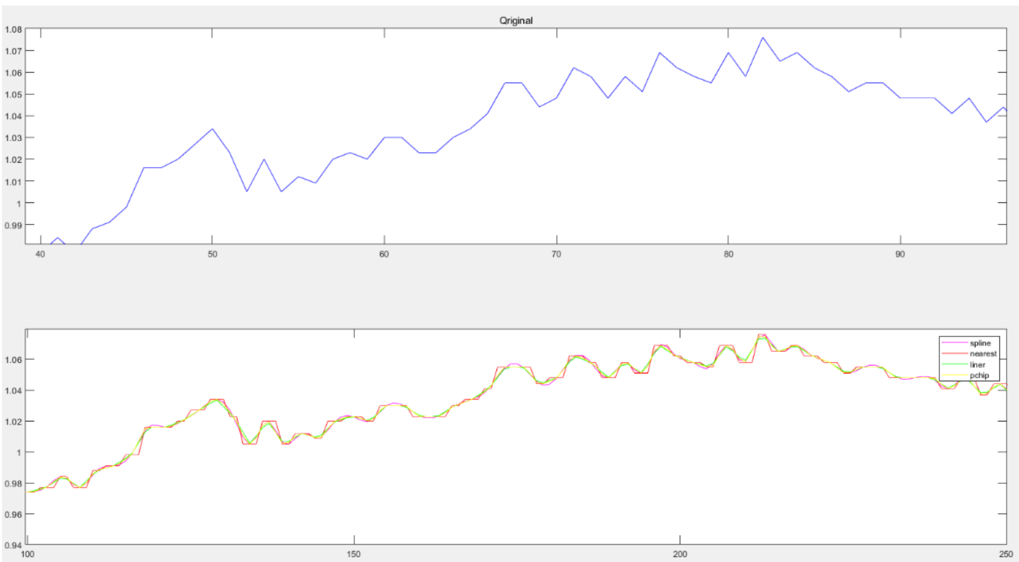


Fig. 5. Nearest, linear, spline, pchip algorithm on the active segment sequence processing effect

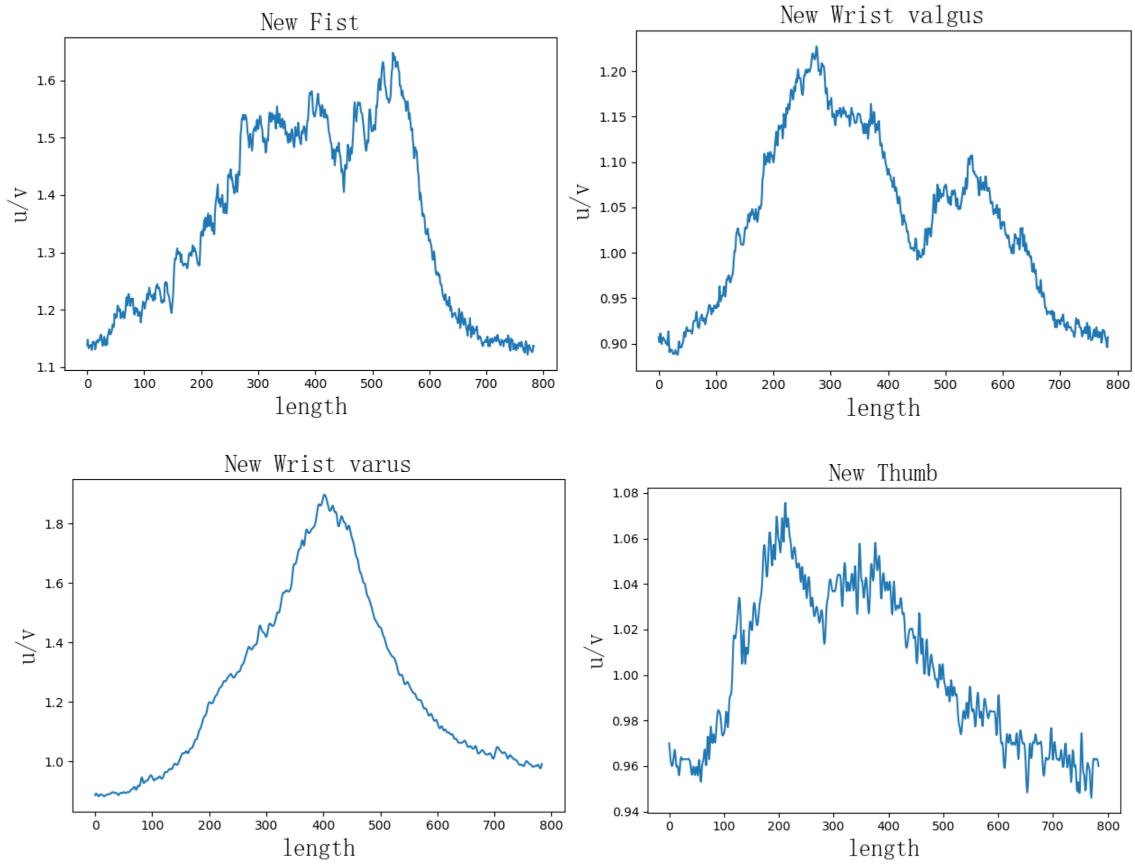


Fig. 6. New active segment action sequence after 4 types of preprocessing

In this paper, 800 gesture motion waveform sequences are generated into vectors, each waveform sequence is 1x784 row vector, and the vector of 600 gesture motion waveform sequences is used as the input data of the training set. In the training set, there are 150 groups of waveform data for each type of gesture action. The vector of 50 gesture motion waveform sequences is used as the input data of the verification set. There are 20 groups of data for each type of gesture motion waveform in the verification set. The vector of 150 gesture motion waveform sequences is used as the input data of the test set. There are 50 groups of waveform data for each type of gesture action in the test set.

Since there are 4 types of output, we use one-hot code to encode the 4 types of output tags, As shown in Table 1.

Table 1. Output label coding

Types of gestures	One-hot code
Fist	0001
Wrist valgus	0010
Wrist varus	0100
Thumbs up	1000

3.2 Identification and Classification

We use neural networks and SVMs commonly used in one-dimensional sequence classification to identify and classify active segment motion sequences, and compare the classification results of the two.

Neural Networks. In this paper, a simple BP neural network is used to classify the active segments (see Fig. 7). The entire network structure consists of 4 layers. Layer 1 and Layer 2 are hidden layers. There are 784 neurons in the input layer and 1500 neurons in Layer 1. Layer 2 has 100 neurons, and each layer has a sigmoid activation function. The BP algorithm is used for iterative training. In order to prevent the algorithm from falling into an infinite loop, the training data is scrambled in each iteration, and the training data is divided into 100 small batches for the process of speeding up the learning, and the gradient is divided by each batch. Calculated separately, not on the entire training set. We use the logistic cost function.

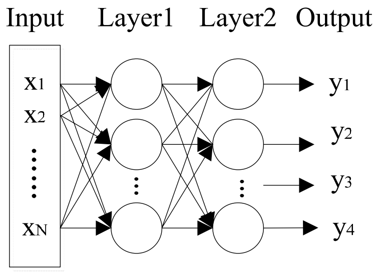


Fig. 7. Network structure

$$J(w) = -\sum_{i=1}^{n} y^{(i)} \log(a^{(i)}) + (1 - y^{(i)}) \log(1 - a^{(i)}). \quad \textbf{(1)}$$

Where $a^{(i)}$ is the sigmoid excitation function used to calculate the i-th element in a layer during forward propgation.

To reduce the degree of overfitting, a regularization term has been added:

$$L2 = \lambda \| w \|_2^2 = \lambda \sum_{j=1}^{m} w_j^2. \quad \textbf{(2)}$$

Since we are a 4 classification, the classifier returns an output vector containing 4 elements, which is compared with the 4x1 dimensional target vector represented by one-hot code, and the cost function of all excitation units j is calculated as:

$$J(w) = -\sum_{i=1}^{n} \sum_{k=1}^{t} y_j^{(i)} \log(a_j^{(i)}) + (1 - y_j^{(i)}) \log(1 - a_j^{(i)}). \quad \textbf{(3)}$$

Add a regularization penalty:

$$J(w) = -\left[\sum_{i=1}^{n} \sum_{k=1}^{t} y_j^{(i)} \log(a_j^{(i)}) + (1 - y_j^{(i)}) \log(1 - a_j^{(i)}) \right] + \frac{\lambda}{2} \sum_{l=1}^{L-1} \sum_{i=1}^{u1} \sum_{j=1}^{ul+1} (w_j^{(l)}, t)^2. \quad \textbf{(4)}$$

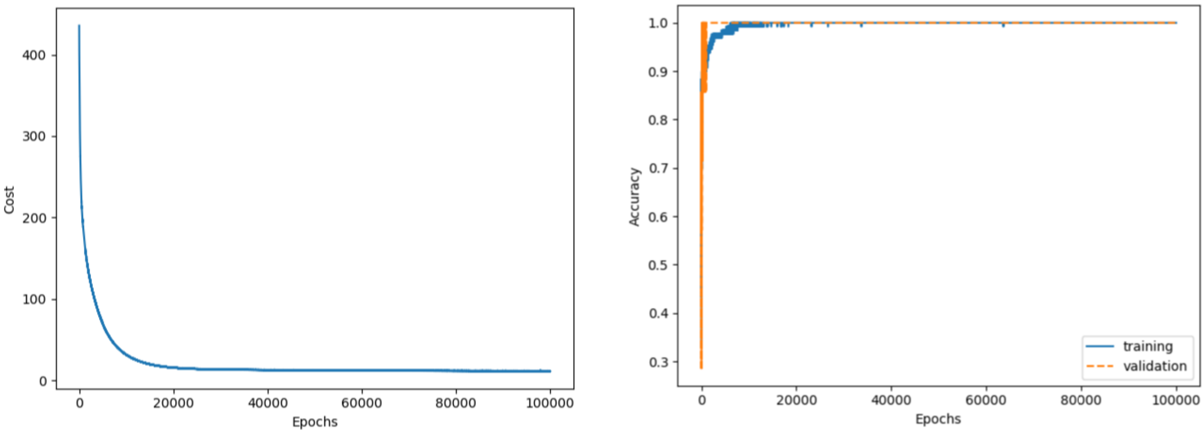
Finally, the cost is minimized by the back propagation algorithm $J(w)$.

SVM. The basic model of the Support Vector Machine (SVM) is to find the best separation hyperplane on the feature space so that the positive and negative sample intervals on the training set are the largest. SVM can easily obtain the nonlinear relationship between data and features at the time of small and medium sample size, which can avoid the use of neural network structure selection and local minimum value problem. It can be explained with strong interpretability and can solve high dimensional problems. SVM is often used in one-dimensional data classification, so we use SVM to compare with neural networks.

DTW. DTW algorithm is often used in the case of inconsistency of time series length, and DTW algorithm can automatically detect active segments. Therefore, this paper also uses DTW to classify gesture motion waveforms. The shortest distance that the same type of gesture sequence accumulates is 0.3~35, and different types of gesture sequences have a minimum distance of 45 or more.

4 Performance Analysis

When using neural networks for classification, the number of iterations basically converges around 40,000 times, the cost is minimized, and The change of the verification set and the training set with the number of iterations is shown in the Fig. 8.the recognition rates of the four types of gestures of fist, wrist inversion, wrist eversion, and thumbs up were 96%, 100%, 96%, and 100%, respectively. The average recognition time per gesture is about 0.0748ms (see Table. 2).



(a) Iteration number and cost curve

(b)Verification set and training set change with the number of iterations

Fig. 8. Neural network classification result

Table 2. Classification using neural networks

Gestures	Number	Fist	Wrist valgus	Wrist varus	Thumbs up	Average time
Fist	50	48	2	0	0	0.0748ms
Wrist valgus	50	0	50	0	0	
Wrist varus	50	0	0	48	2	
Thumbs up	50	0	0	0	50	

When using SVM for classification, the recognition rates of the four types of gestures of fist, wrist inversion, wrist eversion, and thumbs up were 94%, 94%, 96%, and 100%, respectively. The average recognition time per gesture is about 0.921ms (see Table. 3).

Table 3. Classification using SVM

Gestures	Number	Fist	Wrist valgus	Wrist varus	Thumbs up	Average time
Fist	50	47	3	0	0	0.921ms
Wrist valgus	50	3	47	0	0	
Wrist varus	50	0	0	48	2	
Thumbs up	50	0	0	0	50	

When using DTW for classification, the recognition rates of the four types of gestures of fist, wrist inversion, wrist eversion, and thumbs up were 100%, 100%, 100%, and 100%, respectively. The average recognition time per gesture is about 7.95s (see Table 4). DTW calculation complexity is $O(n^2)$, it is slower to calculate.

Table 4. Classification using DTW

Gestures	Number	Fist	Wrist valgus	Wrist varus	Thumbs up	Average time
Fist	50	50	0	0	0	7.95s
Wrist valgus	50	0	50	0	0	
Wrist varus	50	0	0	50	0	
Thumbs up	50	0	0	0	50	

We can see that the classification effect of neural networks is the best of the three algorithms.

5 Conclusion

The gesture recognition system based on normalized neural network introduced in this paper uses the surface muscle electrode to guide the potential signal, and uses the NRF52832 single-chip microcomputer to collect the sEMG signal to obtain the active segment of the four types of gesture motion waveforms as sample data. In order to solve the problem that the sequence length of the active segment is different, the resampling method is used to normalize the active segment waveform sequence to obtain the active segment with the same sequence length.

We use neural networks, SVM algorithms, DTW algorithms for classification, And compare the classification effects. We can see that the classification effect of neural networks is the best of the three algorithms.

Using neural networks for classification, the recognition rates of four types of gesture motion waveforms: fist, wrist inversion, wrist eversion, and thumbs up are: 96%, 100%, 96%, and 100%.

References

[1] X. Hu, Z. Wang, X. Ren, Z. Yan, G. Wang, Classification of surface EMG signal based on wavelet transform with nonlinear scale, Journal of Biomedical Engineering 23(6)(2006) 1232-1236.

[2] M.-F. Lucas, A. Gaufriau, S. Pascual, C. Doncarli, D. Farina, Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization, Biomedical Signal Processing and Control 3(2)(2008) 169-174.

[3] X. Zji, S. Lv, L. Fan, X. Yu, The EEG signal process based on EEMD, in: Proc. International Symposium on Intelligence Information Processing & Trusted Computing, 2011.

[4] C. Hang, R. Zhang, Z. Chen, C. Li, Z. Li, Dynamic gesture recognition method based on improved DTW algorithm, in: Proc. International Conference on Industrial Informatics-computing Technology, 2017.

[5] K.-Y. Lian, C.-C. Chiu, Y.-J. Hong, W.-T. Sung, Wearable armband for real time hand gesture recognition, in: Proc. IEEE International Conference on Systems, Man and Cybernetics, 2017.

[6] L.F. Aguiar, A.P.L. Bó, Hand gestures recognition using electromyography for bilateral upper limb rehabilitation, in: Proc. Life Sciences Conference, 2018.

[7] C. Myers, L. Rabiner, Connected digit recognition using a level-building DTW algorithm, IEEE Transactions on Acoustics Speech & Signal Processing 29(3)(1981) 351-363.