Gesture recognition system based on CNN-IndRNN and OpenBCI

Na Wu[1,2], *Hao* JIN [1,2,*], *Xiachuan* Pei[1,2], *Shurong* Dong[1,2], *Jikui* Luo[1,2], *Ruijian* Yan[3], and *Gang* Feng[3]

[1]Key Lab. of Advanced Micro/Nano Electronic Devices & Smart Systems of Zhejiang, College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, 310027, China
[2]ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, 310018, China
[3]Department of Orthopedic Surgery, 2nd Affiliated Hospital, School of Medicine, Zhejiang University, Zhejiang 310009, China

Abstract. Surface electromyography (sEMG), as a key technology of non-invasive muscle computer interface, is an important method of human-computer interaction. We proposed a CNN-IndRNN (Convolutional Neural Network-Independent Recurrent Neural Network) hybrid algorithm to analyse sEMG signals and classify hand gestures. Ninapro's dataset of 10 volunteers was used to develop the model, and by using only one time-domain feature (root mean square of sEMG), an average accuracy of 87.43% on 18 gestures is achieved. The proposed algorithm obtains a state-of-the-art classification performance with a significantly reduced model. In order to verify the robustness of the CNN-IndRNN model, a compact real-time recognition system was constructed. The system was based on open-source hardware (OpenBCI) and a custom Python-based software. Results show that the 10-subject rock-paper-scissors gesture recognition accuracy reaches 99.1%.

1 Introduction

Surface electromyography (sEMG) is a non-invasive technique that uses flat electrodes to record electrical signals on the surface of the skin. It is currently a popular method in muscle-computer interface, applied to arm prosthesis, human-computer interaction program, and silent speech recognition [1]. sEMG have been widely used in gesture recognition systems, especially for prosthetics. Compared with the gesture recognition based on data gloves and vision, the sEMG's gesture recognition system has advantages such as less computation and not affected by environmental changes such as lighting.

There are two common methods for measuring sEMG signals. One is based on sparse multi-channel measurement, and the other is based on high-density electrode measurement. Both multi-channel acquisition methods that require a lot of feature engineering [2] and high-density measurement methods that require lots of computation [3, 4] require deep learning to autonomously learn features. Our goal is to combine multi-channel systems with deep learning. In addition, sEMG based gesture recognition still has many problems: fuzzy

[*] Corresponding author: hjin@zju.edu.cn

marking during gesture transition, signal quality changes due to electrode position and individual differences, and training models from public databases are difficult to achieve and apply to individuals. And the accuracy of gesture recognition based multi-channel measurement needs to be improved before it can be applied to real-time applications. In order to solve these problems, we propose a CNN-IndRNN (Convolutional Neural Network-Independent Recurrent Neural Network) algorithm and train it on Ninapro dataset, which is a widely used open dataset for gesture recognition. Then we build a real-time recognition system and verify the algorithm in real-world application.

Recurrent neural networks have shown excellent classification capabilities in dealing with time series problems, especially long short-term recurrent networks (LSTM) and gated recurrent units (GRU). As for IndRNN, it is a new RNN algorithm. Compared with LSTM and GRU, IndRNN uses the ReLU activation function instead of sigmoid/tanh. Therefore, IndRNN solves the problem of gradient disappearance/explosion to a greater extent, and can form a deeper network [5]. IndRNN can learn longer time series (N>5000) and has better interpretation of individual neurons. At the same time, for simple tasks such as text classification and time series prediction, small one-dimensional (1D) CNN can replace RNNs, and are faster and cheaper. As a part of the CNN, it can also recognize the local patterns in the sequence. Here we combined the IndRNN and the 1D CNN into a novel CNN-IndRNN model, which used only one characteristic signal as input. The model evaluated the signals of 18 different gestures from 10 subjects on the Ninapro database, and the classification results were already comparable to the latest existing methods in this field.

In addition, we proposed a dual-channel real-time recognition system based on OpenBCI and self-developed software using Python to verify the model. In addition, we also introduced the concept of static recognition in the image field, and used the static data corresponding to the final gesture as our training set. Because in the end, as long as the gesture remains stable, the recorded real-time signal can be matched with the static data on the training set. This recognition process is called static recognition.

The contributions of our work are three-fold: 1) the introduction of IndRNN to better extract the spatial information between different channels and further improve the accuracy; 2) compared with the latest method, we only used a time-domain root mean square (RMS) feature and the amount of calculation was greatly reduced; 3) built a real-time recognition system for 3 kinds of gestures to verify the CNN-IndRNN network structure, which proved the robustness of the model; In the real-time system, we built a static gesture data set and let the real-time data match the training set.

2 Algorithm developed on Ninapro

2.1 Description of algorithm architecture and data flow

The data flow on Ninapro dataset are: (1) 19 channels (16 for sEMG and 3 for acceleration), firstly passed the data cleansing and 1 Hz high-pass filtering part. (2) Then it was divided by a continuous sliding window with a size of 150 samples (750 ms) and a step length of 5 samples. The window matrix obtained the RMS value through the feature extraction stage. The window length and step length values of the sliding window were the result of repeated adjustment. (3) Data feed into CNN-IndRNN network. After the batch normalization (BN) layer, 1D convolutional layer (ConvNet) was used to extract the features in the channel. Then, the features of each channel were integrated into the feature map matrix as the input of IndRNN, followed by a BN layer, an IndRNN layer, and finally a fully connected layer in Figure 1. We divided the overall process into three parts: data preprocessing, feature extraction, and CNN-IndRNN model classification. We will explain them in detail.

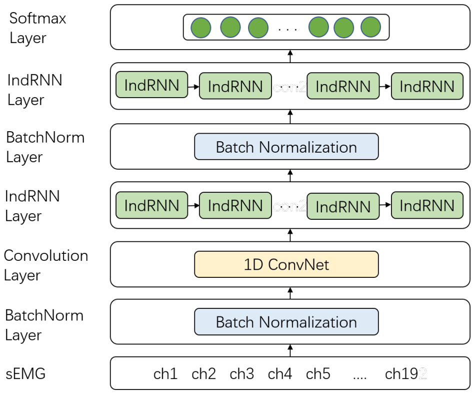


Fig. 1. Structure of the CNN-IndRNN model.

Data cleansing was to replace the missing values and outlaw values of 19 electrodes with 0. Then, 1 Hz high-pass filtering was performed on the data, and the data with DC bias was filtered (the second-order Butterworth filter). After the data was processed, the training set and test set were divided. 2 and 5 sets of data in DB5 [1] were used as test sets, and the remaining 4 sets were used as training sets.

The window length was 150 samples (making the backtracking time less than 0.8 s), and the step length was 5 samples (Myo's sampling rate is 200 Hz). The sliding window extracted RMS values from the pre-processed data. Because RMS is characteristic of the time domain, it is simpler and faster to compute than the frequency domain. This feature is the variation characteristic of sEMG signal amplitude and the sum of muscle activity.

The RMS values obtained for each row were fed to the CNN-IndRNN network for training in the shape of 1x19. After our test, the same eigenmatrix was inputted to SVM (It is usually a preferred method for gesture recognition based on sEMG signals), and the effect was far less than our model. Moreover, feature input was better than original data input. The hyperparameter settings in the CNN-IndRNN network are listed below. For the 1D CNN network, we used 64 1x1 convolution kernels to fully extract the information contained in the channel by deepening the depth of the network. The activation function is ReLU. The multi-channel feature graphs extracted from 1D CNN were successively passed into 32 IndRNN layers, BN layer and 32-unit IndRNN layers. Finally, 18 probability scores were obtained from the evaluation of the cross-entropy loss function. Each IndRNN layer was followed by a Dropout layer, with dropout rates of 0.1.

2.2 Experimental results based on Ninapro

To evaluate our proposed CNN-IndRNN model, we selected 17 gestures from Ninapro DB5, which are the same as [4]. This dataset has provided sEMG signals and accelerometer signals from 10 complete subjects. The subjects repeated each of the 17 gestures 6 times (5 s) with their right hands, with rest gestures (3 s) in between.

The CNN-IndRNN model was implemented using Keras. Other super-parameters were set as follows: the batch size of training set was 100, the epoch was 10, and the initial learning rate was 0.01. The optimizer adopted stochastic gradient descent method and used momentum (0.9) to correct the optimization direction.

Under the above conditions, the model accuracy of the 18 gestures in Ninapro DB5 was 87.43%. The results were comparable to the effects of 17 gestures, 4 features, and 500 epochs [4]. But we only used RMS feature and 10 epochs, and the computational amount was far lower than them. The total hyperparameters of our model was only 5278, and we did not use the GPU during the whole process. The processor is Intel i5-10210U, so it is suitable for the use of ordinary computers. The resulting confusion matrix for the 10 subjects was shown in the Figure 2.

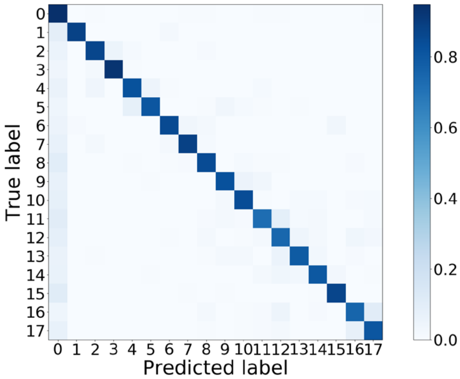
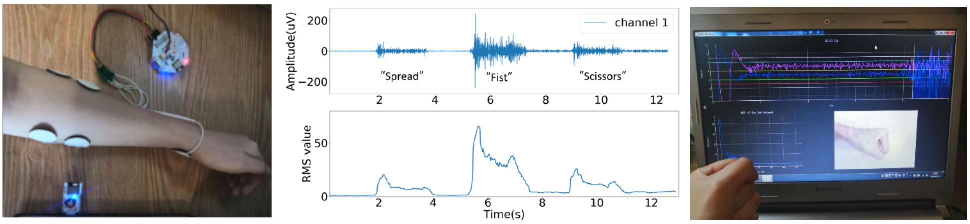


Fig. 2. 18 Gesture recognition confusion matrix in Ninapro DB5 data set.

3 Real-time recognition system

In order to verify our model, a dual differential channel sEMG gesture recognition system was constructed. We used a commercial OpenBCI 32-bit system, through which surface electrodes can be connected to obtain sEMG signals. The system is able to record the raw data, process and display the results at a personal computer with a self-developed graphic user interface (GUI). Two differential channels and a reference electrode are utilized to record the sEMG signals, and the differential electrodes are placed on the flexor and extensor muscles of the left arm as shown in Figure 3 (a) for measurements.



(a) (b) (c)

Fig. 3. (a) Physical map of electrodes position and gesture, (b) Real-time raw sEMG signals and RMS curve of three gestures and (c) The software interface in real-time recognition.

For this experiment, we collected three widely used rock, paper, scissors gestures from 10 volunteers. Each volunteer was asked to do 6 sessions for each gesture, with four sets for training, one for validation and one for testing. Static gesture data was collected for 20 seconds in each group. Figure 3 (b) shows a typical one channel real-time acquired raw sEMG signal and corresponding RMS curve of the rock-paper-scissors gestures (gesture conversion was separated by a period of rest). It can be seen directly that the amplitude spectrum and RMS corresponding to the three gestures are different.

Preprocessing includes signal acquisition, data cleansing and signal processing. Only two pairs of differential electrodes were used for signal acquisition. Data cleansing part was simply done by padding some missing values of the two-channel data. After that, a 50 Hz 3rd-order Butterworth notch filtering was applied, followed by a 2-45 Hz band-pass filtering process to remove noise, and finally, a full-wave rectification was carried out.

The vector values of the two channels were first taken for the above data preprocessing, and then signal was divided into sliding window with a length of 200 samples (0.8s) and step length of 3 samples. Assign a label to each window, and the label was determined by majority voting. Finally, the RMS characteristics of each data segment were extracted.

And then, the constructed CNN-IndRNN network was used to verify RMS. In addition, we verified ten people and only took two pairs of differential electrodes and single RMS feature, so the CNN-IndRNN network input became 1x2. In order to avoid insufficient feature learning, we increased the complexity of the model here to improve the accuracy. The IndRNN layer in both the two layers was replaced with 64 cells, and the overall accuracy of the three gestures reaches 99.1%. The robustness of our constructed model applies not only to open source datasets, but also to our systems. Finally, the recognition result and the latest 1250 sets of data waveform were displayed on the GUI as shown in Figure 3 (c). Accuracy and robustness of the designed system were evaluated, and under static condition, the original algorithm can meet the accuracy requirement perfectly.

4 Conclusions

We developed a CNN-IndRNN network model, which used one layer of 1D CNN and two layers of IndRNN to extract the features in and between channels respectively, and combine the spatial features and temporal features of time series signals. This model was tested on the Ninapro DB5 dataset of 18 gestures, and achieved 87.43% recognition accuracy with only RMS feature and 10 epochs, which was similar to the latest technology. Most importantly, our calculation amount was far lower than them. We also developed a real-time recognition system for three gestures using OpenBCI and Python, and proposed the concept of the static recognition. The model has been tested by 10 people, and the recognition accuracy is 99.1%, which proved the robustness of the model. In the future, we hope to be able to test more categories of gestures and further improve our accuracy.

This work was funded by National Key R&D Programs of China (Nos.2018YFA0701400, 2018YFC0809200), Zhejiang Province Key R&D program (No.2020C03039), Joint Construction Project of Zhejiang Province and Health Ministry (WKJ-ZJ-2029), and Zhejiang Provincial Public Technology Research and Social Development (No. LGF19F010007).

References

1. S. Pizzolato, L. Tagliapietra, *et al*. Comparison of six electromyography acquisition setups on hand movement classification tasks, PLoS One., 12 (2017).
2. M. Tavakoli, C. Benussi, P. Alhais Lopes, *et al*. Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classfier, Biomed. Signal Process. Control, **46**, 121–130 (2018).
3. Geng, W. *et al*. Gesture recognition by instantaneous surface EMG images, Sci. Rep., **6**, 36571 (2016).
4. B. Xie, J. Meng, B. Li and A. Harland, Gesture Recognition from Bio-signals Using Hybrid Deep Neural Networks, ICAICA, 493-499 (2020).
5. S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, Independently Recurrent Neural Network (IndRNN): Building a Longer and Deeper Rnn, CVPR, 5457–5466 (2018).