Sleep Arousal Detection Using End-to-End Deep Learning Method Based on Multi-Physiological Signals

Haoqi Li[1], Qineng Cao[1], Yizhou Zhong[1], Yun Pan[1, 2]

[1]Hangzhou Proton Technology Co, Ltd, Hangzhou, China
[2]College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China

Abstract

We propose an end-to-end deep learning method to detect sleep arousals, especially non-apnea sleep arousals, which is the aim of Physionet/CinC Challenge 2018. We use filtered multi-physiological signals as the input of the network without any other hand-crafted features. The network automatically selects the best features to match arousal targets that we want to identify, and outputs the test result. The proposed network architecture is a 35-layer convolutional neural network (CNN) with three parts: a linear spatial filtering with 1 CNN layer, 33-layer Residual Networks (ResNets), and 1 fully connected layer. For the multi-physiological signals provided in the dataset we choose the 6-channel electroencephalography (EEG) and the 3-channel electroencephalography (EMG) signals, since these signals can better represent the characteristics of non-apnea sleep arousals. In the prediction phase, we use a sliding window method to maximize the performance of sleep arousals detection. For the training set, the result of the area under the precision-recall curve (AUPRC) is 0.3173; the area under the receiver operating characteristic curve (AUROC) is 0.8646. For the final test subset, the result of AUPRC is 0.315; AUROC is 0.858.

1. Introduction

Adequate sleep is a key point of human health. Insufficient sleep can cause serious problems, such as memory loss, obesity, decreased immunity [1], depression [2], and bad quality of life [3]. Sleep arousal is a main problem that results in sleep disorders [4]. The aim of Physionet/CinC Challenge 2018 is to detect sleep arousals and help improve the diagnosis of sleep disorders based on electroencephalography(EEG), electrooculography (EOG), electromyography(EMG), electrocardiograph (EKG), and oxygen saturation(SaO2) signals. The Obstructive Sleep Apnea Hypopnea Syndrome (apnea) consists of hypopnea, central apnea, mixed apnea, and obstructive apnea. Because of the well-studied sleep disorders of apnea [5], the apnea was not the target arousal in our experiments. That is to say, we should detect non-apnea sleep arousals and the detection of apnea will not be scored. The provided dataset includes 1,983 subjects. They are further divided into a training set and a test set, of which the number of subjects are 994 and 989, respectively.

In recent years, convolutional neural networks (CNNs) have achieved state-of-the-art results in computer vision [6]. CNNs with residual modules have better performance than ordinary CNNs. The residual network (ResNet) can solve the degradation problem caused by the increase of network layers, and thus making it easier to optimize a deep CNN [7]. In addition to computer vision, CNN has also shown good performance in signal processing recently, especially physiological signals. Schirrmeister, et al. [8] explains the decoding process of raw EEG signals using a CNN and visualizes the features that CNN learned from EEG signals in the frequency domain. Chambon, et al. [9] uses CNN to identify the sleep stages of human body based on EEG, EOG, and EMG signals. Deep learning algorithms applied to EEG signals are increasing and some of them use spatial filtering schemes in the first layer of the network [10-11]. This method is similar to Independent Component Analysis (ICA) algorithm [12]. In [13], ResNet, which was originally used in image recognition, was applied to ECG signals. It achieved remarkable results by using one-dimension convolution. CNN played an important role in the ECG classification of CinC Challenge 2017 [14] as well.

Inspired by the above, we propose a 35-layer convolutional neural network to detect non-apnea arousals based on filtered multi-physiological signals. To get the best performance of classification and keep a good trade-off between the algorithm complexity and accuracy, we choose 6-channel electroencephalography (EEG) and 3-channel electroencephalography (EMG) signals as the input signals. We regard each signal as a one-dimensional image and concatenate the extracted multi-signal features, which are used for classification. The network consists of three parts: the first part is a 1-layer linear spatial filtering,

 ISSN: 2325-887X DOI: 10.22489/CinC.2018.083

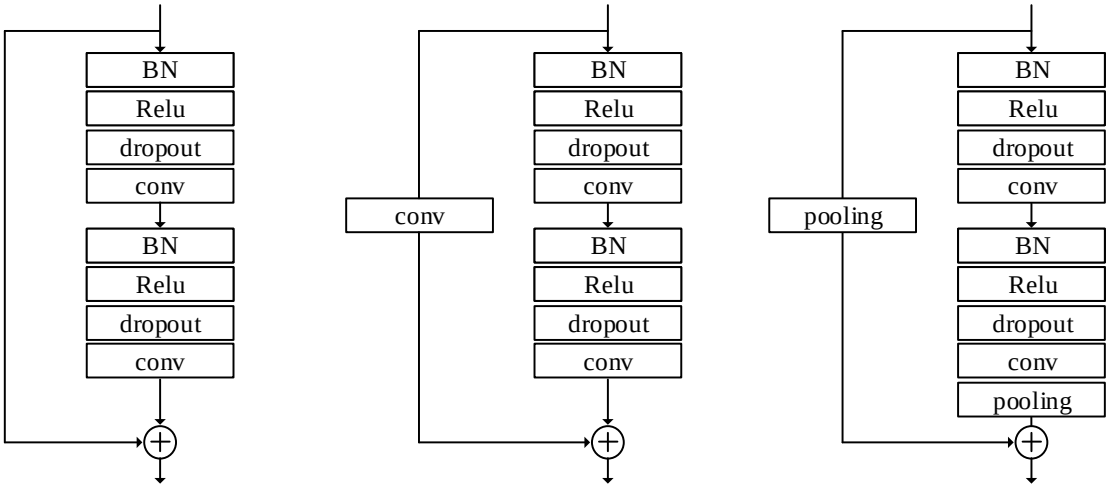

Figure 2: Three structures of ResNet. Structure (a) is the normal ResNet; structure (b) is used to increase feature dimensions with increased convolution filters; structure (c) is used to reduce dimensions with max pooling.

which increases the signal-to-noise ratio. The second part is a 33-layer ResNet used to extract features in signals. The last part is a fully connected layer with softmax to classify the extracted features.

2. Network Architecture

The proposed network architecture is shown in Fig. 1. Due to the spatial filtering method we used, the EMG and EEG signals should be processed respectively. Otherwise the characteristics between different signals will be cut down. After the spatial filtering, we use the ResNet to extract features, and finally they are concatenated together to be classified with softmax.

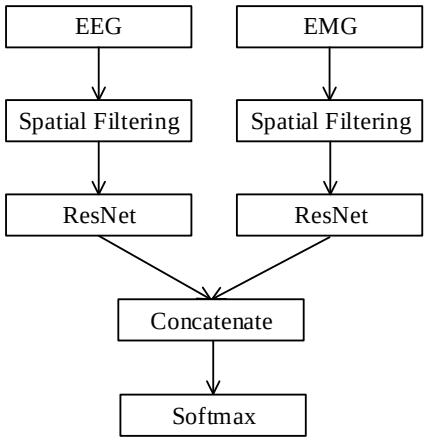

Figure 1: The proposed CNN architecture with 3 modules: Spatial Filtering, ResNet and a fully connected layer with Softmax.

2.1. Spatial Filtering

The first layer of the network is a spatial filtering, which is used to increase the signal-to-noise ratio of the data. Spatial filtering doesn't change the shape of input data and the process of it is equivalent to the unmixing process of ICA algorithm. Table 1 shows more details about the spatial filtering. For M channels of EEG signals with time series of length T, the dimension of (M, T) is firstly reshaped to (M, T, 1). Then, the data with dimension of (M, T, 1) is input to a 2-dimensional convolutional layer, and we get the output with dimension of (1, T, M). Finally, we permute the dimension of output data from (1, T, M) to (M, T, 1), which is the same as the reshaped input's dimension. EMG signals with M' channels are the same as EEG signals.

Table 1: Structure of spatial filtering for EEG signals. The size of convolution kernel and stride is (M, 1) and (1, 1), respectively. The activation function is linear.

Layer	Layer Type	EEG Output Dimension
1	Input	(M, T)
2	Reshape	(M, T, 1)
3	Convolution 2D	(1, T, M)
4	Permute	(M, T, 1)

2.2. ResNet

After spatial filtering, we use the ResNets to extract features of EEG and EMG signals, respectively. The ResNet consists of 33-layer CNN with 16 residual modules. As shown in Fig. 2, the ResNet contains three

structures, and each residual module contains 2 convolutional layers. Structure (a) is the prototype of the ResNet, and (b) and (c) are the variants of (a). For structure (a), the output dimension is the same as the input dimension; for structure (b), as the number of convolution filters increases, the feature dimension of the output also increases; for structure (c), the pooling layer is added to (a) to reduce the dimensions of the data. In the ResNet, (a) and (c) are arranged alternately, and (b) appears once every four modules.

After extracting the features of EEG signals and EMG signals, we concatenate them in series and classify them by a softmax classifier for the result.

3. Experiments

3.1. Preprocessing

6-channel EEG signals are all filtered by a low-pass FIR filter with a cut-off frequency of 40 Hz. For 3-channel EMG signals, only one signal named Chin1-Chin2 passes through a high-pass FIR filter with a cut-off frequency of 15 Hz and the other two-channel signals are not preprocessed. After filtering, all signals are standardized individually with zero mean and unit variance.

3.2. Training

For training, we divide signals into 30s segments. Because regions marked by "-1" are not scored, we only do the two-class training with "1" and "0" tags, which represent non-apnea arousals and normal states respectively, and keep the proportion of the number of two classes in balance. One problem we need to solve during the training is about the arousal duration in segments. Since non-apnea arousals in one segment only account for a certain proportion in most cases, we need to find a good trade-off between arousals and non-arousals to define an arousal label, i.e., determining an optimal ratio of arousals in segments with arousal labels to maximize the performance of arousals detection.

Table 2 shows the results of different arousals duration in 30s segments of "1" tags. We choose tr03-series subjects in this experiment, since the training data set is large, and the result is with 5-fold cross validation. Table 2 illustrates that when the duration of arousals is at least 15s, the result of AUPRC is optimal as 0.197, and the result of AUROC is 0.792. According to this result, we choose 15s duration as the standard of arousal segments. In the training process, we use Adam [15] as the optimization algorithm and the cross entropy as the loss function; the batch-size is set to 16 and the dropout probability is set to 0.5.

Table 2: Results of different arousals duration in arousals segments under tr03-series training set.

Arousals Duration	ACC Ave	AUROC	AUPRC
30s	0.686	0.742	0.165
15s	0.723	0.792	0.197
10s	0.721	0.795	0.184
5s	0.722	0.792	0.175
2s	0.723	0.796	0.191

3.3. Evaluation

Since the distribution of non-apnea arousals is random , while the segmentation of signals is regular, we use a sliding window scheme to predict arousal regions more accurately. As shown in Fig. 3, for a signal with N segments, only the first $N-1$ segments are slided to be predicted. If the length of the sliding window is T, it's necessary for $N-1$ segments to slide N/T times. For the segments of one signal, the second one to $(N-1)th$ segment will be predicted $N/T+1$ times, and the number of prediction for adjacent left and right segments decreases by one every T time.

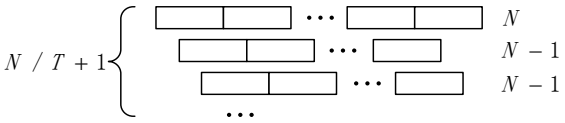

Figure 3: Sliding window prediction for evaluation

Table 3 shows the test results with the sliding window. In this experiment, we use all data for training. The result is better than that using only tr03-series. As the time of sliding window decreases, the result of AUPRC improves. Considering computation complexity, we finally adopted a sliding window with 3 seconds. Under 5-fold cross validation of the training set, the result of AUPRC is 0.317 and the result of AUROC is 0.865. In the final test subset, our results are: AUPRC, 0.315; AUROC, 0.858.

Table 3: Results of different window sliding time under 15s arousals duration in arousal segments. The training result is under 5-fold cross validation with all training data. The final test result is using 3s window slide on a subset of test data.

Set	Slide Time	AUROC	AUPRC
Training	30s	0.831	0.270
	15s	0.853	0.304
	3s	0.865	0.317
Test	3s	0.858	0.315

4. Discussion

The aim of Physionet/CinC Challenge 2018 is to detect the arousals in addition to apnea. Compared to EOG and other signals, we found that the EEG and EMG signals played the most important role among all signals for this aim. The reason why we choose the 30-second duration of segments is it's the optimal length we have found through the experiments. Since the distribution of arousals is random, we decide to use a sliding window scheme to improve the performance of arousals detection. The 3-second length of the sliding window is a good trade-off between computation complexity and algorithm performance. The imbalance of the data is the major problem in our experiments. Regions of non-apnea arousals account only for very few proportion among the whole signal. In order to get results without bias from training, we discarded redundant normal segments to keep it balanced with the arousals. Further problems about how to make use of the discarded data need to be solved. Considering the size of model with deep learning, we only use a 33-layer ResNet. Deeper ResNets may improve the algorithm performance, and we will achieve this idea in the future. Recurrent Neural Network (RNN) is another important network structure in deep learning, which can extract the associated information between different input data [16]. We will also try to use RNNs on this problem in our future work.

5. Conclusion

In this paper, we propose a CNN structure that combines spatial filtering and ResNet, which is effective to detect non-apnea arousals. By selecting EEG and EMG signals, we implement an end-to-end deep learning method, i.e., the network gives the classification results using the raw filtered input data. By adjusting arousals duration in arousal segments and applying a sliding window scheme, we maximize the algorithm performance. The final result of AUPRC and AUROC in the test subset is 0.315 and 0.858, respectively.

Acknowledgements

This work has been partially supported by Zhejiang Key Research and Development Program of Zhejiang Science and Technology Bureau under Grant No.2016CSA160100.

References

[1] Trammell R A, Miller A V. Sleep and the immune system. Encyclopedia of Sleep, 2013;8:568-571.
[2] Nutt D, Wilson S, Paterson L. Sleep disorders as core symptoms of depression. Dialogues Clin Neurosci, 2008;10(3):329-336.
[3] Lee M, Choh A C, Demerath E W, et al. Sleep disturbance in relation to health-related quality of life in adults: The fels longitudinal study. Journal of Nutrition Health & Aging, 2009;13(6):576.
[4] Broughton R J. Sleep disorders: Disorders of arousal? Science, 1968;159(3819):1070-1078.
[5] Terzano M G, Parrino L, Boselli M, et al. Polysomnographic analysis of arousal responses in obstructive sleep apnea syndrome by means of the cyclic alternating pattern. Journal of Clinical Neurophysiology, 1996;13(2):145-155.
[6] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017;39(4):640-651.
[7] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv151203385 2015;7(3):171–180.
[8] Schirrmeister R T, Springenberg J T, Ldj F, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, 2017;38(11):5391-5420.
[9] Chambon S, Galtier M N, Arnal P J, et al. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. IEEE Transactions on Neural Systems & Rehabilitation Engineering, 2018;PP(99):1-1.
[10] Stober S, Cameron D J, Grahn J A. Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings. Advances in Neural Information Processing Systems, 2014;1449-1457.
[11] Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: A compact convolutional network for EEG-based brain-computer interfaces. Journal of Neural Engineering, 2018.
[12] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Networks, 2000; 13(4):411-430.
[13]Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv170701836 2017; .
[14] Andreotti F, Carr O, Pimentel M A F, et al. Comparing feature based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG. In Computing in Cardiology. 2017; .
[15] Kingma DP, Ba J. Adam: A method for stochastic optimization. In Proc. Int. Conf. on Learn. Representations (ICLR) 2015; .
[16] Sak H, Senior A, Beaufays F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. Computer Science; 2014:338-342.

Address for correspondence.

Hangzhou Proton Technology Co., Ltd, Hangzhou, China
Haoqi Li
haoqi.li@protontek.com