

Identification of arousals with deep neural networks (DNNs) using different physiological signals

Runnan He¹, Kuanquan Wang^{1,2}, Na zhao¹, Yang Liu¹, Yongfeng Yuan¹, Qince Li^{1*}, Henggui Zhang^{1,2,3,4*}

¹ School of Computer Science and Technology, Harbin Institute of Technology (HIT), Harbin, Heilongjiang, 150001, China

² School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK

³ Space Institute of Southern China, Shenzhen, Guangdong, 518117, China

⁴ Key Laboratory of Medical Electrophysiology, Ministry of Education, Collaborative Innovation Center for Prevention and Treatment of Cardiovascular Disease/Institute of Cardiovascular Research, Southwest Medical University, Luzhou, Sichuan, 646000, China

Abstract

Due to the increasing life pressure in modern society, more and more people are suffering from sleep disorders. The most serious case of sleep disorders called apnea is characterized by a complete breathing block, leading to awakening and subsequent sleep disturbances. In addition to apnea, sleep arousals can also be initiated, caused by partial airway obstructions, teeth grinding, partial airway obstructions, or even snoring. However, great obstacles still exist in automatic identification of arousals. In this study, a novel method was developed to detect non-apnea sources of arousals during sleep using several physiological signals. In the dataset provided, the duration of arousal regions is much less than that of non-arousal regions. In order to address this issue, a set of segments (20000 samples in length) were extracted for model training in which arousal regions take up a much larger proportion than that in the original training set. After the preprocessing, a sequence-to-sequence deep neural networks (DNNs) which consists of a series of convolutional layers with residual connections, a long short-term memory (LSTM) layer and two fully connected layers, was trained to classify samples in the segments. Results show that the area under receiver operating characteristic curve (AUROC) and area under receiver precision recall curve (AUPRC) are 0.905 and 0.453 respectively in test dataset. In this study, an effective algorithm to detect non-apnea arousals was developed, which has great potentials in the clinical diagnosis and treatment of automatic sleep disturbance in the future.

1. Introduction

Sleep disorders can make people not well recovered in energy and physical power, which is essential to maintain

the normal operation of various functions of the human body. At present, there are more than 84 defined sleep disorders including obstructive sleep apnea hypopnea syndrome (OSAS) [1]. In 1973, Guilleminauh first proposed the concept of OSAS which is a high incidence disease with serious hazard and potential danger. As hypoxia and hypercapnia often happen during sleep, OSAS may ultimately lead to progressive damages in multi-system and multi-organs. According to the survey record [2], ninety percent of the traffic accident is caused by sleep apnea disorders. However, due to the occurrence at night, these diseases are not easy to be discovered. As the disease is chronic, their symptoms are not immediately apparent [3].

The polysomnography (PSG) is the gold standard to diagnose sleep arousals. But, due to the limitation of recording duration of PSG, vast majority of patients cannot get timely treatment. Therefore, a low-cost and more convenient diagnostic method is required.

The goal of the challenge is using information from the available signals to correctly classify non-apnea arousal regions. According to the previous studies, sleep arousals were mainly investigated from two aspects, by screening the physiological parameters of breathing, blood oxygen, electrocardiograph (ECG) or analysing snoring signals.

For the first aspect, studies have proposed several simplified methods to aid the screening of sleep disturbances based on a few numbers of signals [4, 5] or even a single one, such as ECG [6, 7], pulse oximetry [8], breath sounds [8, 9], snore sounds [10, 11] or nasal airway pressure [12].

For the second aspect, in the last several decades, many works have been performed for snoring detection [13, 14]. In general, most methods of breathing signal classification used preprocessing techniques to extract finer feature information for better classification, because the discrimin-

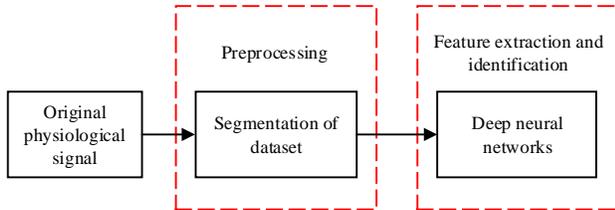


Figure 1. Flowchart diagram of the proposed method for the identification of arousals classification.

ation information is introduced by the frequency information. However, the accuracy of the existing methods can not satisfy the need of clinical diagnosis

In this study, we developed a novel effective method to detect non-apnea arousals based on deep neural networks (DNNs). Results showed that the area under receiver operating characteristic curve (AUROC) and area under receiver precision recall curve (AUPRC) are 0.905 and 0.453 in test dataset, respectively.

2. Method

Figure 1 outlines the architecture of our proposed algorithm which included segmentation of dataset and a sequence-to-sequence deep neural networks (DNNs) trained to classify samples in the segments. Each major step was explained in more detail in the two following subsections.

2.1. The segmentation of dataset

In this challenge, there are a variety of physiological signals including electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), electrocardiology (ECG), and oxygen saturation (SaO2). Excluding SaO2, the other signals were sampled to 200 Hz. For analytic convenience, SaO2 was resampled to 200 Hz and is measured as a percentage [15].

The signals come from multiple channels, each of which has a duration of about 7 to 10 hours and a data sampling rate of 200 Hz. If each piece of data in the original data set was used as a sample for DNNs, the training process would require exhausting memory space (hundreds of GB) and have a long training period (days or even weeks). In addition, in the initial dataset, arousal interval usually accounts for a small proportion (<5 percent) of the entire sleep process. The imbalance problem of dataset will make the training model tend to mistakenly predict the arousal regions as the non-arousal regions. To address this issue, this paper provides a reconstruction method for a training set to accelerate the DNNs training process. The specific steps are as follows:

Step1: Read the sequence of labels for each piece of data in turn, and re-label them with the binary method: if the original label of a certain position is the arousal regions,

re-mark it as 1, otherwise re-mark it as 0.

Step2: Divide each piece of data and its corresponding label into small segments of fixed length (20,000 sample points) and group the intercepted segments into a matrix (each row represents a segment, and the sequence of segments is based on the time series of the original signal). After the segment is truncated, if the remaining part of the segment is less than the length of one segment, it is discarded directly.

Step3: P_a represents the proportion of segments having more than 5% points labeling 1. P_n represents the proportion of segments having all 0 labeling in the segment set. As the non-arousal regions are much bigger than the arousal regions, P_n is much larger than P_a . In order to improve the dataset imbalance, P_d proportion of segments ($P_d=P_n-P_a$) with all 0 labeling are randomly discarded.

Step4: Save the retained data segments and their corresponding annotations for the subsequent model training.

The above steps can effectively solve two problems. First, the serious imbalance of training data is improved. The gap between the non-arousal and arousal regions in the training data is narrowed, which is beneficial for the decision-making of the classification model without excessive bias. The second one is to solve the issue of training difficulty for the big dataset. This method cut the signals into small data segments (20000 sampling points) to reserve the required context information to train the model. It has less memory usage and time consumption for training the DNNs.

2.2. Sequence-to-sequence deep neural networks (DNNs)

DNNs is effective to learn features from the raw data and generally can achieve a better performance than the traditional hand-crafted features [16-18]. In this challenge, we employed a novel sequence-to-sequence DNNs to learn features from the small data segments. The structure of our network is illustrated in Figure 2. The network is a sequence-to-sequence structure, which includes a feature extraction part composed of a one-dimensional convolutional layer and a bidirectional LSTM layer, as well as a feature classification part composed of a fully connected layer, which is described in Figure 2.

(1) In the feature extraction part, as shown in Figure 3, the input signal first enters an one-dimensional convolutional layer (the number of convolution kernels is 32 and the length of the convolution kernel is 16), and then undergoes batch normalization and activation of the ReLU function, then enters the first residual module. The residual module includes an one-dimensional convolution layer, a batch normalization layer, a ReLU activation layer, a Dropout layer, a one-dimensional convolution layer, and a maximum pooling layer (the reduction factor is 2). The in-

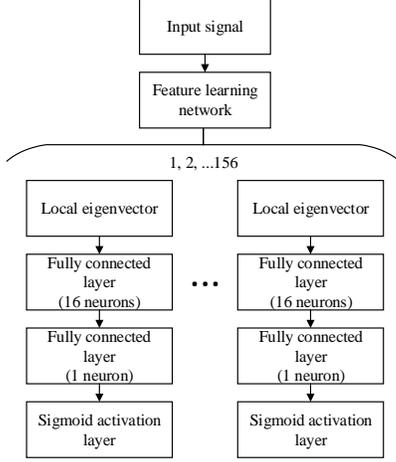


Figure 2. The structure of the proposed DNNs. There are 156 local eigenvectors each of which has a length of 128.

put of the module is subjected to bitwise addition operation producing the output of its main operation after the maximum pooling layer down sampling, which is fed as the input to the subsequent part of the network. Then there are six residual modules of the same structure, each of which includes a batch normalization layer, a ReLU activation layer, a Dropout layer, a one-dimensional convolution layer, a batch normalization layer, a ReLU activation layer, a Dropout layer, an one-dimensional convolution layer and a maximum pooling layer. The convolutional layer in the first module contains 32 convolution kernels, each of which has a length of 16. After each of previous two modules, the number of convolution kernels is increased by 32, and the convolution kernel length is reduced to the half of previous one. The same as the first residual module, the input of subsequent each module is subjected to a bitwise addition operation producing the output of its main operation after the maximum pooling layer down sampling, and the result is input to the subsequent part of the network. The dropping rate of all the above Dropout layers is 0.5. After all the residual modules, the data is then passed through a normalization layer, a ReLU activation layer, and a bidirectional LSTM layer. The LSTM layer performs sequence-to-sequence learning on the input data that means each output corresponding to each input sequence element, and the length of input and output sequence is the same. The LSTM layer contains 128 units, which means that each element of the output sequence is a vector of the length of 128. Each vector in the sequence represents the feature of its corresponding location. The input Dropout rate and the cyclic Dropout rate of the LSTM layer are both 0.2.

(2) The feature classification section processes each of the feature vectors learned as described above to determine whether the corresponding position belongs to the arousal

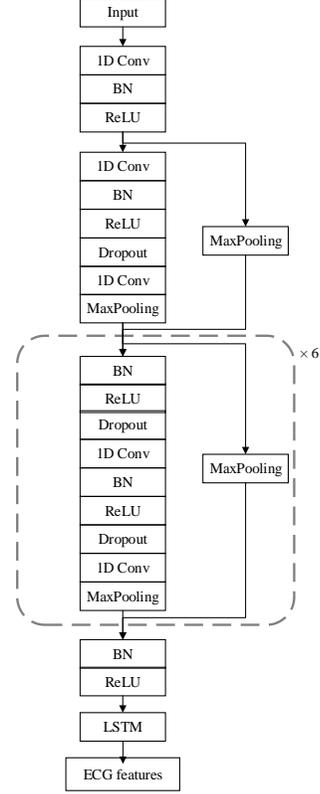


Figure 3. The structure of feature learning network

regions. This part contains two fully connected layers, the first and second layers contain 16 and 1 neuron. Since the network performs the binary classification problem, the activation function of the second layer is Sigmoid, and its formula is as follows:

$$S(t) = \frac{1}{1 + e^{-x}} \quad (1)$$

(3) For each training sample, the objective function to be optimized during the model training process is the cross entropy function, and its formula is as follows:

$$L(X, r) = -\frac{1}{m} \sum_{i=1}^m \log p(R = r_i | X) \quad (2)$$

where x is the input data of the training sample, r is the marking sequence of the training sample, m is the length of the marking sequence in the training sample, and p is the probability of the i^{th} output flagged as r_i . The DNNs apply the Adam optimization method and uses Keras based on the TensorFlow engine to train.

Based on the method described above, the classification in test dataset of PhysioNet database were estimated.

3. Results and discussions

Results showed that the area under receiver operating

characteristic curve (AUROC) and area under receiver precision recall curve (AUPRC) are 0.905 and 0.453 respectively in test dataset, which is acceptable in spite of the existence of various physiological signals and the influences of different noises in the test dataset.

It is notable that the advantages of the proposed method include the novel methods of data segmentation and automatic feature extraction. The good accuracy shows it is a potential choice for clinical automatic arousal identification in the future. In the further work, new network construction should be added to improve our method for arousal analysis.

4. Conclusion

In this study, using different physiological signals, DNNs have been proposed for identification of arousals. A novel method of segmenting datasets was proposed in preprocessing stage. the method reorganized the training set to accelerate the DNNs training and narrowed the gap between the non-arousal and arousal regions. For the feature learning and classification, the DNNs incorporate residual model and recurrent neural network (RNN) to adapt to sequence learning problems with long time series dependence, which can effectively identify the arousal regions.

Acknowledgements

The work is supported by the National Natural Science Foundation of China (NSFC) under Grant nos.61572152, 61601143, 61571165 and 81770328, the Science Technology and Innovation Commission of Shenzhen Municipality under Grant nos. JSGG20160229125049615 and JCYJ20151029173639477, and Heilongjiang and China Postdoctoral Science Foundation under Grant nos.2015M581448.

References

[1] Gislason, T., Lindholm, M. D., Almqvist, M., Birring, D. D. S., & Boman, M. D. (1988). Sleep apnea syndrome. *Arch Otolaryngol Head Neck Surg*, 114(1), 45-51.

[2] Caples, S. M., Gami, A. S., & Somers, V. K. (2005). Obstructive sleep apnea. *Annals of Internal Medicine*, 142(3), 187-97.

[3] van Maanen, J. P., & De, V. N. (2014). Long-term effectiveness and compliance of positional therapy with the sleep position trainer in the treatment of positional obstructive sleep apnea syndrome. *Sleep*, 37(7), 1209-15.

[4] Chervin, R. D., & Burns, J. W. (2011). Engineering better sleep. *Med Biol Eng Comput* 49:623-625.

[5] Flemons, W. W., Littner, M. R., Rowley, J. A., Gay, P., Anderson, W. M., Hudgel, D. W., ... & Loube, D. I. (2003). Home diagnosis of sleep apnea: a systematic review of the literature: an evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the

American Thoracic Society. *Chest*, 124(4), 1543-1579.

[6] Abdullah, H., Maddage, N. C., Cosic, I., & Cvetkovic, D. (2010). Cross-correlation of EEG frequency bands and heart rate variability for sleep apnoea classification. *Medical & biological engineering & computing*, 48(12), 1261-1269.

[7] Kesper, K., Canisius, S., Penzel, T., Ploch, T., & Cassel, W. (2012). ECG signal analysis for the assessment of sleep-disordered breathing and sleep pattern. *Medical & biological engineering & computing*, 50(2), 135-144.

[8] Yadollahi, A., Giannouli, E., & Moussavi, Z. (2010). Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals. *Medical & biological engineering & computing*, 48(11), 1087-1097.

[9] Kulkas, A., Huupponen, E., Virkkala, J., Tenhunen, M., Saastamoinen, A., Rauhala, E., & Himanen, S. L. (2009). New tracheal sound feature for apnoea analysis. *Medical & biological engineering & computing*, 47(4), 405-412.

[10] Fiz, J. A., Jane, R., Solà-Soler, J., Abad, J., Garcia, M. A., & Morera, J. (2010). Continuous analysis and monitoring of snores and their relationship to the apnea-hypopnea index. *The Laryngoscope*, 120(4), 854-862.

[11] Solà-Soler, J., Fiz, J. A., Morera, J., & Jané, R. (2012). Multiclass classification of subjects with sleep apnoea-hypopnoea syndrome through snoring analysis. *Medical Engineering and Physics*, 34(9), 1213-1220.

[12] van Houdt, P. J., Ossenblok, P. P., Van Erp, M. G., Schreuder, K. E., Krijn, R. J. J., Boon, P. A., & Cluitmans, P. J. (2011). Automatic breath-to-breath analysis of nocturnal polysomnographic recordings. *Medical & biological engineering & computing*, 49(7), 819-830.

[13] Cavusoglu, M., Kamasak, M., Eroglu, O., Ciloglu, T., Serinagaoglu, Y., & Akcam, T. (2007). An efficient method for snore/nonsnore classification of sleep sounds. *Physiological measurement*, 28(8), 1-13.

[14] Dafna, E., Tarasiuk, A., & Zigel, Y. (2013). Automatic detection of whole night snoring events using non-contact microphone. *PloS one*, 8(12), 1-14.

[15] Mohammad M Ghassemi, Benjamin E Moody, Li-wei H Lehman, Christopher Song, Qiao Li, Haoqi Sun, Roger G. Mark, M Brandon Westover, Gari D Clifford, You Snooze, You Win: the PhysioNet/Computing in Cardiology Challenge 2018, *Computing in Cardiology Volume 45*. Maastricht, Netherlands, 2018, 1-4

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

[17] He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Identity mappings in deep residual networks. In *European conference on computer vision*. Springer, Cham, 630-645

[18] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., and Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*.

Address for correspondence.

Henggui Zhang
Room 3.07, Shuster building
The School of Physics and Astronomy
The University of Manchester, Oxford Road
Manchester, M13 9PL, UK
E-mail:henggui.zhang@manchester.ac.uk