Improved Deep Neural Network for Automatic Staging Applied to Assisted Sleep System

Meng Lu[1], Jiawen Chen[2], Shengpeng Shi[1] and Zhenyu Zhu[1]

[1]School of Information Engineering, Wuhan University of Technology, Wuhan, China

[2]School of Management, Wuhan University of Technology, Wuhan, China

Keywords: Sleep Stage Classification, Deep Neural Network, Assisted Sleep System, EEG Signals.

Abstract: In this paper, aiming at sleep problem, we propose an improved automatic sleep scoring model based on convolutional neural network and time-distributed Long Short-Term Memory. This model is applied to real-time assisted sleep system for the intelligent wearable device with client-server architecture in mobile application and website. In this model, we employ CNN to extract time-invariant features from raw EEG epochs, employ LSTM to learn temporal information from a sequence of EEG epochs that have been extracted features, and utilize time-distributed layers to connect the two parts. Sleep EEG signals from open Sleep-EDF database are used to evaluate the performance of our proposed model. The results show that our proposed model has good classifications for most sleep stages, especially for awake and SWS stages. Moreover, we obtain overall accuracy of 85.7%, macro F1-score of 80.5%, and Cohen's Kappa coefficient (κ) of 0.81 on 20-fold cross-validation for five sleep stages according to AASM, which outperforms other six existing methods. Due to better staging performance of our model, our assisted sleep system can have a better adaptive adjustment for people and it also has a good market promotion prospect and potential commercial value.

1. Introduction

Sleep plays an important role in human health. Nowadays, due to health problems, many people pay more attention to their sleep quality. The classification of sleep stages is an effective way to diagnose sleep disorders and monitor sleep quality, which has an important impact for medical research.

According to the Rechtschaffen and Kales (R&K) sleep scoring standard, human sleep is a dynamic process which can be divided into three main stages with awake stage (W), rapid eye movement stage (REM) and non-rapid eye movement stage (NREM). The NREM is subdivided into four sleep stages N1, N2 N3 and N4. Combining N3 and N4 into a single stage that represents slow wave sleep (SWS) is another standard submitted by the American Academy of Sleep Medicine (AASM).

With the popularity of intelligent wearable sleep monitoring devices, many research studies have been conducted to develop automatic sleep stages classification systems to reduce the cost of manual classification of sleep stages. These studies tried to develop methods for automatic sleep stage scoring based on single-channel EEG, or multiple signals such as EEG, EMG, and EOG. They also have focused on simplifying sleep staging systems into portable forms that use motion sensors, heart rate, blood oxygen, pulse oximetry and respiration levels pulse to predict sleep stages. Sleep experts are mainly based on single-channel EEG signals when classifying sleep stages, and wearable devices that record single-channel EEG signals may be more comfortable. Therefore, in this study, we aimed to develop an automatic sleep stage scoring algorithm using single-channel EEG signals. Sleep stage staging based on EEG can be roughly divided into two steps: feature extraction and classification. For feature extraction, previous methods often extracted features from EEG signals in the time domain, frequency domain, or time-frequency domain. In the step of classification, random forest classifier, decision tree, support vector machine (SVM), and neural network, etc. were employed.

  DOI: 10.25236/isaiee.2020.027

However, we believe that these methods may not be generalizable to a larger population due to the heterogeneity among subjects and the recording hardware. This is because these features are designed manually based on the characteristics of the available dataset.

In recent years, as a branch of machine learning, deep learning uses multiple layers of non-linear and linear processing units to learn features or hierarchical representations from input data and has been applied to sleep stage classification. In the works of DeepSleepNet and SeqSleepNet, Bi-directional Long-Short Term Memory (LSTM) and end-to-end Hierarchical Recurrent Neural Networks (RNN) have been developed for classifying sleep stages from raw EEG, respectively. DeepSleepNet uses representation learning with two CNNs followed by sequence residual learning with Bi-LSTM. The main drawback of this model is that since Bi-LSTM relies on larger temporal sequences to obtain better accuracy, it is necessary to input the raw EEG data of 25 epochs together to obtain 25 labels. SeqSleepNet with end-to-end Hierarchical RNN uses 3 channels, namely EEG, EMG and EOG, to give the better overall accuracy of 87.1% on the MASS dataset. Models by Sors et al. and Tsinalis et al., as well as SeqSleepNet and DeepSleepNet all use longer temporal sequences—4, 5, 10 and 25 raw EEG 30s epochs respectively. Time-Distributed CNN proposed by Koushik et al. can overcome this limitation with worse accuracy.

To solve these problems, we propose a novel automatic sleep scoring model with CNN and time-distributed LSTM inspired by DeepSleepNet. Additionally, this model is applied to real-time sleep staging system for the intelligent wearable device with client-server architecture in mobile application and website.

2. Methodology

2.1 Deep Neural Network

The architecture of our deep neural network model consists of two main parts as described in Fig. 1. The left part is CNN, which can be trained to learn filters to extract time-invariant features from each of raw single-channel EEG epochs. The right part is time-distributed LSTM, which can be trained to encode the temporal information such as stage transition rules from a sequence of EEG epochs in the extracted features. This architecture is designed for scoring 30s EEG epochs following the standard of AASM.

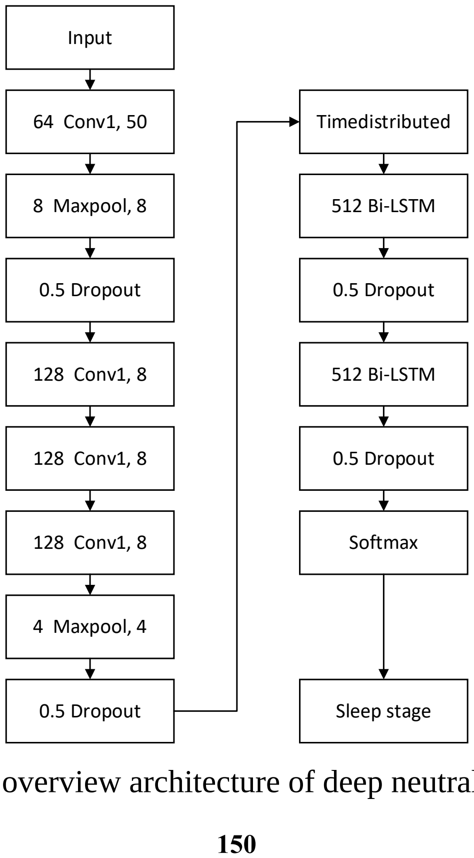

Fig 1. An overview architecture of deep neutral network.

The CNN part has four 1-D convolutional (Conv1D) layers, two 1-D max-pooling (Maxpool1D) and two dropout layers. All Conv1D layers use Rectified Linear Units (ReLU) activation (i.e., relu(x)=max(0,x)). Dropout layers with the probability of 0.5 are used throughout the model. 30s epochs of normalized EEG at 100Hz is fed into the time-distributed LTSM part. This wrapper applies a layer to every temporal sequence of an input and is followed by two repeated set of bidirectional LSTM layers with ReLU activation and dropout layers with the rate of 0.5. We finally utilize dense layers with softmax activation as output of the model.

The training uses an Adam optimizer of 0.001 with an initial learning rate of e^{-3} which is reduced each time the validation accuracy plateaus using ReduceLROnPlateau Keras Callbacks. And the CategoricalCrossentropy loss is used to computes the cross-entropy loss between the labels and predictions.

$$L_{ce} = -\frac{1}{N}\left[\sum_{i=1}^{n}(t_i \, log(p_i) + (1 - t_i) \, log(1 - p_i))\right] \quad (1)$$

For regularization parameters, we try several values for the weight decay parameters ranging from 10^{-1} to 10^{-5}. The value of 10^{-3} gives us the best performance.

2.2 Assisted Sleep System

The entire framework of assisted sleep system is composed of wearable device, such as our assisted sleep goggles in Fig. 2, and cloud platform. The wearable device collects EEG signals and transmits them to the cloud platform via WIFI. The cloud platform uses the EEG signal processing algorithm based on wavelet transform to preprocess the EEG signal data, and then inputs them to the personalized sleep staging system based on deep neural network model of the AASM standard. Real time sleep stage outputted by the system will be used as a feedback command to make an adaptive adjustment to the working status of the whole assisted sleep system. In our sleep googles, we use white noise to help people sleep better and use light to wake people up through feedback commands. This whole process forms a closed loop feedback in Fig. 3. At the same time, our system processes the data in real-time with client-server architecture. The server (cloud platform) is responsible for receiving EEG signal at 100 Hz collected from the wearable device and applying our model for sleep staging. The client (e.g. mobile app, web page) is responsible for visualizing EEG signal and the corresponding hypnogram scored by our system.

Fig 2. Assisted sleep goggles designed by us.

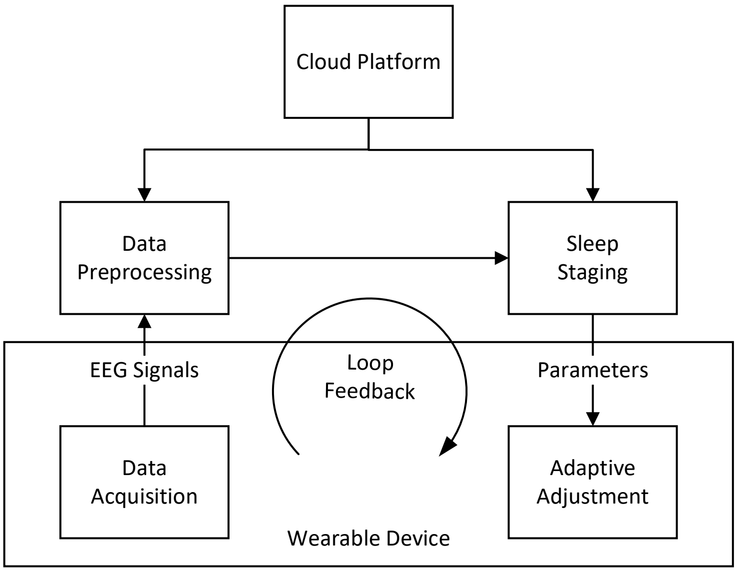

Fig 3. Overview of assisted sleep system.

3. Experiment

3.1 Dataset

The Sleep-EDF Expanded Database from Physionet is used in this study. It contains 197 whole-night Polysomnographic sleep recordings, containing EEG, EOG, chin EMG, and event markers. Some records also contain respiration and body temperature. Corresponding hypnograms (sleep patterns) were manually scored by well-trained technicians according to the R&K manual and are also available. The data comes from two studies. One named SC* is the study of age effects on sleep in healthy Caucasians aged 25-101, without any sleep-related medication. Another named ST* is the study of temazepam effects on sleep in 22 Caucasian males and females without other medication. Subjects had mild difficulty falling asleep but were otherwise healthy.

In this paper, EEG signals recorded from ten Caucasian males and ten Caucasian females (25-34 years old at the time of the recordings) without any sleep-related medication for the whole night sleep sampled at 100Hz were selected. Moreover, only a single EEG channel (Fpz-Cz) was selected to perform automatic sleep stage scoring task as the Fpz-Cz channel gave a better performance than Pz-Oz channel. Each whole EEG signals of each subject was divided into epochs of 30s (contained 3000 points of data). Each epoch was labelled as sleep stages W, R, 1, 2, 3, 4, M (Movement time) and not scored, which were scored by sleep experts according to R&K manual. In this study, very few M and not scored epochs have been removed from our experiment. Furthermore, we have also merged the N3 and N4 sleep stages as SWS sleep stage, as is currently recommended by AASM. Table I shows the number of epochs for each sleep stage that are used in this study.

TABLE I. The number of epochs for each sleep stage that are used in this study

Sleep Stage	Number of Epochs
W	8285
N1	2804
N2	17799
SWS	5703
REM	7717
Total	42308

3.2 Implementation

To evaluate the performance of our model, our experiments used the K-fold cross-validation method and the K is set to 20. We divided all types of sleep stage EEG signals epoch data into training sets, validation sets and test sets. Specifically, we randomly selected 80% of all data sets as train sets, 10% as validation sets and the rest as test sets during each training session. In this experiment, the code of this paper used Python 3.7 and deep learning framework TensorFlow 2.0 to complete our experiment. To speed up the training of the model, we have a Linux server with TITAN X (Pascal) GPU as our experimental platform.

3.3 Metrics

In our study, we use the recall (RE), precision (PR) and F1-score (F1) to evaluate the performance of classification of each sleep stage (called as Per-stage Metrics). In addition, we used widely used indexes such as macro-averaging F1-score (MF1), overall accuracy (ACC) and Cohen's Kappa coefficient (κ) to evaluate the performance of our automatic sleep stage scoring model. The index MF1 refers to the average of all F1 of each sleep stage. The index ACC represents the proportion of the correct number of sleep stage epochs classification to all sleep stage epochs. The index Cohen's kappa coefficient (κ) is used to measure inter-rater reliability for sleep stage epochs. The MF1 and ACC are calculated as follows:

$$\begin{aligned} \mathrm{ACC} &= \frac{\sum_1^n \mathrm{TP}}{\mathrm{n}} \\ \mathrm{MF1} &= \frac{\sum_1^m \mathrm{F1}}{\mathrm{m}} \end{aligned} \tag{2}$$

Where TP is the true positives, and n is the total number of test epochs, F1 is per-class F1-score, and m is the number of sleep stages.

4. Result

4.1 Performance

Table II show confusion matrices obtained from the 20-fold cross-validation on the Fpz-Cz channel from Sleep-EDF dataset. Each row and column represent the number of 30s EEG epochs of each sleep stage classified by the sleep expert and our model, respectively. The numbers in bold indicate the number of epochs that were correctly classified by our model. The last three columns in each row indicate per-class performance metrics computed from the confusion matrix. The most noticeable results in the experiment were that the sleep stages of W, SWS and REM have excellent classification performance. For example, the accuracy of sleep stage W, SWS and REM classification is close to 89%. We observe that the sleep stage N1 has the worst classification performance as the sleep stage N1 is mainly mistakenly classified as REM and N2 stages. The reason may be that the background EEG waves are quite similar for the three sleep stages. Another reason might be the lack of epochs for N1 stage, which is 2804 in total, significantly fewer than other four sleep stages, referring to Table I. In one person's whole-night sleep, the N1 stage has only about 50 sleep stage epochs of EEG signals of 30s. The inequality of distribution of the five sleep stages may result in a poor classification results for N1 stage using CNN model.

TABLE II. Confision matrix obtained from 20-fold cross-validation on fpz-cz channel from the sleep-edf dataset

	Predicted					**Pre-class Metrics**		
	W	N1	N2	SWS	REM	PR	RE	F1
W	**7411**	5457	96	46	187	89.3	89.5	89.4
N1	287	**1451**	599	35	432	51.5	51.7	51.6
N2	230	449	**15427**	1096	5977	91.8	86.7	89.2
SWS	143	26	349	**5176**	9	81.4	90.8	85.8
REM	226	345	336	7	**6803**	84.7	88.2	86.4
ACC = 85.7%; MF1 = 80.5%; κ = 0.81								

4.2 Comparison

Table III shows a comparison between our method and other sleep stage scoring methods across ACC, MF1, κ They all use the 20-fold cross-validation on the Fpz-Cz channel from Sleep-EDF dataset except Hsu et al., which used the 10-fold cross-validation. Hsu et al. utilize hand-engineered energy features and recurrent neural classifier. We believe the practical evaluation scheme should not include any epochs from the test subjects. Also, it has been shown that the non-independent scheme resulted in an improvement of the performance. For automatic sleep staging, our method achieves better results than other methods, with the overall accuracy of 85.7%, macro F1-score of 80.5%, and Cohen's Kappa coefficient (κ) of 0.81. However, all these methods have bad performance on the stage N1.

TABLE III. Comparison between our method and other sleep stage scoring methods across acc, mf1, κ

Authors	ACC	MF1	κ
Supratak et al.	82.0	76.9	0.76
Mousavi et al.	84.3	79.7	0.79
Koushik et al.	83.5	76.2	0.68-0.76
Tsinalis et al.	78.9	73.7	-
Tsinalis et al.	74.8	69.8	-
Hsu et al.	90.3	76.5	-
Our method	85.7	80.5	0.81

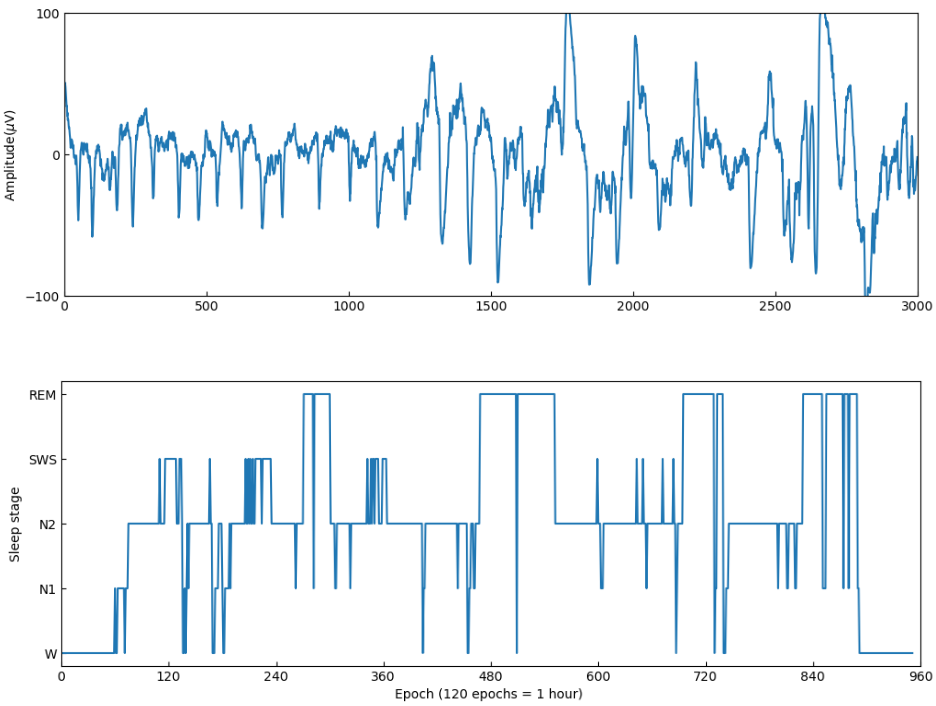

Fig 4. Experimental data chart

5. Conclusion

In this paper, we have proposed a real-time automatic sleep scoring system using single-channel EEG signals without utilizing any hand-engineered features. Our model utilizes CNNs to extract time-invariant features, and bidirectional-LSTMs to learn stage transition rules among sleep stages from EEG epochs. Our results showed that, without changing the model architecture and the training algorithm, our model was able to automatically learn features for sleep stage scoring from different raw single-channel EEGs from dataset and our model analysis results also demonstrated that our model learned several features that are consistent with the AASM manual. We apply our model to the wearable device and build a real-time automatic sleep scoring system. As our system automatically learn features from raw EEG, we believe that it achieves a remote sleep monitoring compared to the hand-engineering ones.

In the future, we plan to improve our model to be able to better classify EEG signals on the stage N1 and improve our real-time system to reduce delays for better performance.

References

[1] Li Yanyan. Saving sleep--This article is dedicated to the developing mental health business in China [J]. Beijing Literature (wonderful reading), 2019(04):52-75.

[2] Wu Dejun, Zhang Zhonghan, Qing Shangyun. The right sleep to ensure health [J]. Chinese and Foreign Health Abstracts, 2012, 9 (49):406-407.

[3] Gu Wenjia. Visual thinking of user interaction experience in UI design [D]. Hangzhou: China Academy of Art, 2012.