Biomedical Signal Processing and Control 42 (2018) 107–114

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A convolutional neural network for sleep stage scoring from raw single-channel EEG

Arnaud Sors[a,b,*], Stéphane Bonnet[a,b], Sébastien Mirek[c], Laurent Vercueil[a,d], Jean-François Payen[a,e]

[a] Univ. Grenoble Alpes, F-38000 Grenoble, France
[b] CEA Leti, MINATEC Campus, 17 rue des Martyrs, F-38054 Grenoble, France
[c] Dijon University Hospital, Dpt. Anesth. and Crit. Care, 14 rue Paul Gaffarel, F-21079 Dijon, France
[d] Grenoble University Hospital, Dpt. Exploration Fonctionnelle du Système Nerveux, Avenue du Maquis du Grésivaudan, F-38700 La Tronche, France
[e] Grenoble University Hospital, Dpt. Anesth. and Crit. Care, Avenue Maquis du Grésivaudan, F-38700 La Tronche, France

ARTICLE INFO

Article history:
Received 10 February 2017
Received in revised form
14 November 2017
Accepted 2 December 2017

Keywords:
EEG
Single-channel
Sleep staging
Convolutional neural network
Classification
Sleep Heart Health Study

ABSTRACT

We present a novel method for automatic sleep scoring based on single-channel EEG. We introduce the use of a deep convolutional neural network (CNN) on raw EEG samples for supervised learning of 5-class sleep stage prediction. The network has 14 layers, takes as input the 30-s epoch to be classified as well as two preceding epochs and one following epoch for temporal context, and requires no signal preprocessing or feature extraction phase. We train and evaluate our system using data from the Sleep Heart Health Study (SHHS), a large multi-center cohort study including expert-rated polysomnographic records. Performance metrics reach the state of the art, with accuracy of 0.87 and Cohen kappa of 0.81. The use of a large cohort with multiple expert raters guarantees good generalization. Finally, we present a method for visualizing class-wise patterns learned by the network.

1. Introduction

Sleep is an essential ingredient for good human health. A number of sleep disorders exist, among which insomnias, hypersomnias, sleep-related breathing disorders, circadian rhythm sleep-wake disorders, parasomnias, sleep movement disorders. Polysomnography (PSG) is the main tool for diagnosing, following, or ruling out sleep disorders. A polysomnogram is a collection of various signals useful for monitoring the sleep of an individual. It uses physiological signals (EEG, EMG) and environmental signals (microphone, accelerometer). Sleep staging consists of dividing a polysomnographic record into short successive epochs of 20 or 30 s, and classifying each of these epochs into one sleep stage amongst a number of candidate ones, according to standardized classification rules [1,2]. Sleep staging can be carried out either on a whole polysomnogram or on a subset of its channels, and either by a trained expert or by an algorithm. In some cases the expert can use an algorithm for pre-scoring. The successive representation of sleep stages over the night is called a hypnogram. It provides a simple representation of the sleep which is useful for suspecting or diagnosing sleep disorders. Sleep staging is a tedious task which requires considerable work by human experts. Also, the quality of the rating depends on the experience and fatigue of the rater and inter-rater agreement is often less than 90% [3,4]. Hence the demand for automated sleep staging algorithms.

In this article, we consider single-channel EEG sleep staging. Whilst it constitutes a first step towards multichannel analysis systems, single-channel sleep staging is also interesting in itself because it allows light, wearable, and unobstrusive systems that can be deployed easily on mobile devices. The lightweight setup with only two or three electrodes and fewer wires also helps ensuring that sleep is not compromised by any incomfort. Most studies on single-channel EEG-based automatic sleep stage scoring adopt a two-step methodology. First, different features are extracted from the time waveforms. Second, a classifier is trained to predict sleep stages based on these extracted features. Most features belong to one of the following three categories [5]: (a) time-domain features, (b) frequency-domain features, and (c) non-linear features. For classification, the most common methods include decision trees

* Principal corresponding author at: CEA Leti, MINATEC Campus, 17 rue des Martyrs, F-38054 Grenoble, France.
E-mail addresses: arnaud.sors@m4x.org (A. Sors), stephane.bonnet@cea.fr (S. Bonnet).

https://doi.org/10.1016/j.bspc.2017.12.001
1746-8094/

Table 1
A per-class summary of the dataset.

	Wake	N1	N2	N3	REM	Total
Total epochs	1,514,280	201,431	2,169,452	719,690	779,548	5,384,401
Total equivalent days of data	525	70	753	250	271	1871

and random forests [6], support vector machines [7], and neural networks [8]. The authors of [9] use multiscale entropy and autoregressive features along with linear discriminant analysis. The authors of [10] use features from a difference visibility graph and classify using a support vector machine. In [6], time-frequency features, Renyi's entropy features and a random forest classifier are used. The authors of [11] obtain features from an Empirical Mode Decomposition and classify with bootstrap aggregating with decision trees. In [12], spectral features from a tunable *Q*-factor wavelet transform and a random forest classifier are used. [13] use iterative filtering, a discrete energy separation algorithm and various classifiers. Finally, the authors of [14] use a recurrent neural classifier on energy features.

Recently, some studies adopt the use of neural networks classifiers trained end-to-end and which serve both as feature extractors and classifiers. [8] study the use of stacked sparse autoencoders and [15] the use of convolutional neural networks. The authors of [16] use a convolutional neural network preprocessor complemented with a bi-directional long short-term memory network (LSTM). Literature results for some of these methods can be found in Table 4.

In this article, we introduce a method for single-channel EEG-based sleep staging using a deep supervised convolutional neural network(CNN) on raw signal samples. CNNs have been used in other domains on raw continuous signal with great results, starting with image recognition [17,18], followed by many other domains such as natural language processing [19], recommender systems [20], and other supervised pattern recognition tasks. Since recently, CNNs have also been used on short EEG time series for various applications such as Brain Computer Interfaces [21,22] including motor imagery [23] and Steady State Visually Evoked Potentials (SSVEP) [24], as well as seizure detection [25], driver's cognitive performance [26], and eye tracking [27]. Since recently, CNNs are also used for sleep scoring [15,16]. The goal of our work is to show that CNNs are suitable and offer competitive sleep scoring performance on a large multi-center sleep scoring dataset. Such systems may then be applied in various conditions such as critically ill patients where continuous EEG recording after brain injury is showing a growing interest. The advantage of using an end-to-end approach is that no feature engineering phase is required. The network, described in Section 2, is trained to learn feature detectors that are suited to the classification task at hand and are likely to perform better than hand-engineered features. As discussed in Sections 3 and 4, the method has state-of-the-art performance when applied on a large sleep scoring dataset.

2. Materials and methods

2.1. Dataset

The Sleep Heart Health Study (SHHS) [28] is a multi-center cohort study, initiated by the American National Heart Lung and Blood institute to determine whether sleep-disordered breathing is associated with a higher risk of various cardiovascular diseases. The study includes two rounds of polysomnographic recordings. We use only the first round (SHHS-1) because it includes almost all patients and because all records have the same sampling rate (125 Hz), contrary to the second round where records can be sampled at 125 or 128 Hz. Dataset SHHS-1 contains 5793 polysomnographic records. Recorded channels include two bipolar EEG channels (C4-A1 and C3-A2), two EOG channels, one EMG channel, one ECG channel, two inductance plethysmography channels (thoracic and abdominal), a position sensor, a light sensor, a pulse oxymeter, and an airflow sensor. Each record was manually scored for sleep stages by a single technician on 30-s epochs according to Rechtschaffen and Kales scoring rules [1], resulting in several sleep stages: Wake, N1, N2, N3, N4 (non-REM), and REM. The total number of technicians involved is not reported. More details about montages and scoring modalities are provided in [29].

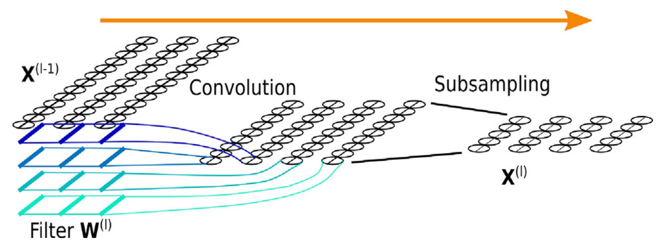

Fig. 1. Composition of a 1D convolution layer, including convolution and subsampling. The nonlinearity is not represented.

2.2. Preprocessing

In such polysomnographic records, for most subjects a long 'wake' period before the patient goes to sleep and another after he or she wakes up is observed. These wake periods are trimmed so that the number of pre-and-post-sleep wake epochs is not larger than the most represented other class. Since available EEG channels are symmetrical, they yield comparable performance. In the following we use C4-A1. As suggested in recent recommendations [2], stages N3 and N4 are merged into a single stage N3. The very few patients containing no epoch for a given sleep stage are excluded because they might be outliers. The resulting number of epochs (resp. relative importance) per stage and total number of epochs are shown in Table 1. As in any PSG study, classes are very unbalanced. Stage N1 is particularly under-represented. No preprocessing is done on the EEG signals themselves.

2.3. CNN classifier

2.3.1. Architecture

A complete CNN is usually composed of a number of convolutional layers, followed by one or two fully-connected layers, and a softmax regression layer that outputs class probabilities. The structure of a convolutional layer for one dimensional signals is shown in Fig. 1. Each layer l convolves the set $\boldsymbol{X}^{(l-1)}$ of its input feature maps with a set of learnable kernels (also called filters) $\boldsymbol{W}^{(l)}$ and adds biases $\boldsymbol{b}^{(l)}$. With $n^{(l-1)}$ the number of input feature maps and $n^{(l)}$ the number of output feature maps, and $k^{(l)}$ the width of the kernel, $\boldsymbol{W}^{(l)}$ has shape $(k^{(l)}, n^{(l-1)}, n^{(l)})$. Since inputs have one channel only, $n^{(0)}$ equals 1. Let $\boldsymbol{x}_j^{(l)}$ denote the jth feature map in $\boldsymbol{X}^{(l)}$, and $\boldsymbol{w}_{ij}^{(l)}$ the slice of $\boldsymbol{W}^{(l)}$ that applies from input feature map i to output feature map j. With such notation, we have:

$$\boldsymbol{x}_j^{(l)} = \sigma \circ g_{p^{(l)}} \left(\sum_{i=1}^{n^{(l-1)}} \boldsymbol{x}_i^{(l-1)} * \boldsymbol{w}_{ij}^{(l)} + \mathrm{b}_j^{(l)} \right) \tag{1}$$