Computers in Biology and Medicine 89 (2017) 441–453

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Identifying sleep spindles with multichannel EEG and classification optimization

Ning Mei [a], Michael D. Grossberg [b], Kenneth Ng [a], Karen T. Navarro [a], Timothy M. Ellmore [a,*]

[a] *Department of Psychology, The City College of the City University of New York, USA*
[b] *Department of Computer Science, The City College of the City University of New York, USA*

ARTICLE INFO

Keywords:
Thresholding
Machine learning
Optimization
Sleep spindle
Memory consolidation

ABSTRACT

Researchers classify critical neural events during sleep called spindles that are related to memory consolidation using the method of scalp electroencephalography (EEG). Manual classification is time consuming and is susceptible to low inter-rater agreement. This could be improved using an automated approach. This study presents an optimized filter based and thresholding (FBT) model to set up a baseline for comparison to evaluate machine learning models using naïve features, such as raw signals, peak frequency, and dominant power. The FBT model allows us to formally define sleep spindles using signal processing but may miss examples most human scorers would agree are spindles. Machine learning methods in theory should be able to approach performance of human raters but they require a large quantity of scored data, proper feature representation, intensive feature engineering, and model selection. We evaluate both the FBT model and machine learning models with naïve features. We show that the machine learning models derived from the FBT model improve classification performance. An automated approach designed for the current data was applied to the DREAMS dataset [1]. With one of the expert's annotation as a gold standard, our pipeline yields an excellent sensitivity that is close to a second expert's scores and with the advantage that it can classify spindles based on multiple channels if more channels are available. More importantly, our pipeline could be modified as a guide to aid manual annotation of sleep spindles based on multiple channels quickly (6–10 s for processing a 40-min EEG recording), making spindle detection faster and more objective.

1. Introduction

1.1. Background

The functional role of sleep in mammals remains a matter of debate [2–6]. One of the theories is that occurrences of particular neural events during sleep reflect the processes associated with memory consolidation [7,8]. It has become a challenge to identify these neural events by simply viewing the data because it is time consuming and prone to different interpretations by different viewers, especially in high definition neural recordings, which contain thousands of data points in just a few seconds of data. One of the neural recording techniques is to record and monitor using scalp electroencephalography (EEG).

Macro and micro structures are typically found in segmented EEG recordings. Macro-structured neural events refer to segments that are usually 20–30 s long and represent different sleep stages, or levels of sleep compared to the awake condition [9–11]. On the other hand, micro-structured neural events refer to local and short segments, such as sleep spindles. Sleep spindles typically occur during sleep stage 2, and they are believed to be generated from the thalamus [6,12]. Based on the dominant frequency of a segment around a spindle, each is classified as a slow spindle (9–10 Hz [13 14]; 10–12 Hz [13,15,16]) or a fast spindle (13–15 Hz [15,17]; 12–14 Hz [16]), which are believed to occur during different phases of slow oscillations (<1 Hz) [18]. Measuring sleep spindles and analyzing their relationship to behavior and cognition may provide insight into how these neural events influence memory performance, as well as provide diagnostic measurements for various sleep disorders. It is not completely understood how the brain integrates past information to generate new memories. Thus, the recording of sleep spindles provides common and quantifiable measurements of sleep so that we can connect sleep spindles with memory and describe how the brain processes information during sleep.

* Corresponding author. The City College of New York, 160 Convent Avenue, Department of Psychology, NAC 7/120, New York, NY, 10031, USA.
E-mail address: tellmore@ccny.cuny.edu (T.M. Ellmore).

http://dx.doi.org/10.1016/j.compbiomed.2017.08.030
Received 17 May 2017; Received in revised form 28 August 2017; Accepted 29 August 2017

1.2. Related work

Given that identifying these neural events may provide a powerful tool to study the relationship between sleep and memory, it is critical that we have a standard to define these events [45,46]. Unfortunately, the definition of spindle features varies across studies. This complicates the aim to classify spindles automatically [19–23]. Studies that incorporate machine learning algorithms to classify these neural events usually make classifications based on a single EEG channel (i.e. Cz) and long period (>7 h) of recordings. Among all the automated algorithms, filtering based and thresholding (FBT) approaches have shown some promising results for classifying sleep stages, spindles, and k-complexes [24,25]. Methods like template-based filtering and continuous wavelet transforms (CWTs) [12], Support Vector Machine classifiers (SVM) [26], decision-tree classifiers [27], and artificial neural networks [28] have also been investigated. However, there are few studies that classify spindles and other neural events (k-complexes) simultaneously using a unified framework [29–33]. Visually, a typical spindle (11–16 Hz) has a unique symmetrical shape along the temporal axis, looking like a football, while other neural events, such as k-complexes, usually have an asymmetrical shape. This difference limits regular approaches that rely on an explicitly characterizing both events using signal processing. While distinguishing different patterns is easy, it remains a challenge to recognize the different patterns with the same system. The current study focused mainly on classifying sleep spindles and investigates how the current results might guide us to classify many other neural events, such as sleep stages and k-complexes.

1.3. Motivation

The studies mentioned above provide evidence that sophisticated machine learning algorithms perform better in classifying sleep spindles against "non-spindles" in segments of neural recordings. However, it is not useful for small datasets or practical use. First, machine learning algorithms, especially multi-layer neural networks, usually require a large amount of data (>1000 training samples) and features besides raw signals are extracted to improve classification performance, but neither is a common approach in many clinical evaluations of sleep. As we mentioned above, features for defining a typical spindle varies from study to study, and features in time-frequency space are usually extracted to add to the feature list so that traditional single layer machine learning models learn better about the patterns between spindles and non spindles. Second, machine learning models take a structured segment of the data (namely an epoch) and return probabilities of whether this segment of data contains spindles or not. Applying a machine learning model is challenging for practical use in localizing spindles for several reasons.

The first reason is that it is difficult to localize and recognize neural events. Neural events like spindles can occur at any moment of a recording with varying duration (~0.5–2 s), and this makes it difficult to define a segmenting window to sample representative training data for machine learning models [36]. It is difficult to construct sampling windows to sample segments containing a full cycle of spindle. With small windows, we might capture part of a spindle, while with large windows we might capture too much irrelevant signal around the spindle. To localize spindles, a model must take the varying duration of spindles into account and return the locations (time) and the durations (length) of the spindles. Such a goal could be achieved by using flexible kernels within a machine learning model and is usually easier to address in recurrent neural networks with long-short term memory (LSTM) neurons [34].

The second reason is that the sample size of spindles could be small while the sample size of non-spindles could be large, which are not optimal for machine learning models [6]. Researchers usually only are able to sample about 50–200 spindles in a 30-min short nap [7], regardless of the duration of the spindles. The total sample size of spindles is usually a small fraction of the total recording. However, a sufficient machine learning training takes more than 1000 training samples, and a LSTM neural network would take more than 5000 training samples. It is common in short nap periods to sample imbalanced sample sizes of spindles and non-spindles (e.g., 5 spindles and 95 non-spindles). Assuming a machine learning model is flexible to take the varying duration of spindle and non-spindle samples into account, the model could report at least 95% accuracy by claiming all the samples are "non-spindles", but this is not we want to see in practical settings. Therefore, the imbalanced sample size of the spindle and non-spindle classes makes preparing training data a difficult problem in sleep studies [35].

The third reason is that the classification of a spindle is usually made based on signals of a single channel. Classifying spindles based on a single channel could misrepresent the global characteristics of spindles, which we might capture better through a multi-channel approach. Studies have shown that spindles can be recorded across multiple channels [7]. The FBT approach includes particular spindle and non-spindle samples based on global signal patterns across multiple channels ($n > 2$). Making classifications based on multiple channels, we could identify spindles that consistently occur in several regions of the brain.

1.4. Objectives

The objectives of the current study include implementing the FBT approaches [19] to classify spindles using a short period of high definition EEG recordings. The FBT approach is designed to classify spindles quickly with flexible parameters that capture temporal and spectral variations of the EEG representations of spindles (e.g., frequency, duration, amplitude, etc.) and serves as a classification bench mark for further investigation of machine learning models. Furthermore, this study aims to optimize feature parameters that are used to speed up and aid the sampling of enough data to train a more accurate fully automated process. With enough training data, we hope to eventually define spindles probabilistically by intuitive features. Thus, we applied our algorithms to the publically available DREAMS project data [1] with few human inputs to classify spindles based on a single EEG channel. An additional objective is to present a state of the art multi-channel FBT approach which encodes current characterization of spindles with a flexible range of features. To make a fair comparison among the FBT approach, machine learning models, and experts' scores, a cross-validation criterion that is described in the Ray et al. study [37] is used. We applied this algorithm to our data with six channels of interest to examine how machine learning models perform better than the FBT approach.

1.5. Novelty and outline of the present study

There are two novel aspects in the current study. The first is that we propose a nested model of FBT and machine learning, using a fast processing FBT model to guide machine learning in model selection. The second is that we propose a nested machine learning model derived from the FBT model that can perform well by using simple signal information, such as raw signal values, peak frequencies, and corresponding power density values sampled across multiple recording channels. The paper is structured first by outlining the preprocessing of the EEG data. Then the implementation, optimization, and cross-validation is detailed. The development of the machine learning pipeline is then detailed followed lastly by a comparison of the FBT and machine learning models.

2. Methods

Data Acquisition. A total of 64-channels of EEG data, including 2 EOG electrodes, were continuously recorded at 1 kHz sampling with an actiCHamp active electrode system (Brain Products, GmbH) while subjects napped on a bed inside a sound-attenuated testing booth (IAC industries). Experiments were carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Each subject provided written informed consent and completed the study