MethodsX 3 (2016) 144–155

Contents lists available at ScienceDirect

MethodsX

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

SegWay: A simple framework for unsupervised sleep segmentation in experimental EEG recordings

Farid Yaghouby, Sridhar Sunderam *

Department of Biomedical Engineering, University of Kentucky, Lexington, KY, USA

A B S T R A C T

Sleep analysis in animal models typically involves recording an electroencephalogram (EEG) and electromyogram (EMG) and scoring vigilance state in brief epochs of data as Wake, REM (rapid eye movement sleep) or NREM (non-REM) either manually or using a computer algorithm. Computerized methods usually estimate features from each epoch like the spectral power associated with distinctive cortical rhythms and dissect the feature space into regions associated with different states by applying thresholds, or by using supervised/unsupervised statistical classifiers; but there are some factors to consider when using them:

- Most classifiers require scored sample data, elaborate heuristics or computational steps not easily reproduced by the average sleep researcher, who is the targeted end user.
- Even when prediction is reasonably accurate, small errors can lead to large discrepancies in estimates of important sleep metrics such as the number of bouts or their duration.
- As we show here, besides partitioning the feature space by vigilance state, modeling transitions between the states can give more accurate scores and metrics.

An unsupervised sleep segmentation framework, "*SegWay*", is demonstrated by applying the algorithm step-by-step to unlabeled EEG recordings in mice. The accuracy of sleep scoring and estimation of sleep metrics is validated against manual scores.

A R T I C L E I N F O

Method name: SegWay: simple sleep segmentation

Keywords: Sleep scoring, EEG, EMG, Clustering, Hidden Markov model, Bout duration, Vigilance state, Sleep–wake, Behavior, Classifier

Article history: Received 27 August 2015; Accepted 17 February 2016

* Corresponding author at: Department of Biomedical Engineering, University of Kentucky, Lexington, KY 40506-0108, USA.
E-mail address: ssu223@uky.edu (S. Sunderam).

http://dx.doi.org/10.1016/j.mex.2016.02.003

Methods

Approach

In our simple framework for unsupervised sleep segmentation, dubbed "*SegWay*", unlabeled data are first separated into clusters that correspond to each of the vigilance states based on their location in the feature space. This clustering is seen to give reasonably accurate predictions of vigilance state from sample feature measurements. But correctly predicting scores for a large proportion of epochs of data alone does not guarantee that state transitions are accurately identified. Therefore, in addition to clustering the data to determine the vigilance states, dynamical transitions between the states are modeled using a Markov chain. The two aspects in combination—i.e., observations conditioned on latent states and probabilistic state transitions—is known as a hidden Markov model [1]. Sleep scores are predicted from the HMM, which can be used to infer the sequence of vigilance states underlying a time series of sample points in the feature space. The use of the HMM leads to improved prediction of state and estimation of dynamical attributes like the proportion, number of bouts, and mean bout duration of each vigilance state, which are of interest in sleep research [2]. The HMM, when fitted to baseline data, can then be used to score future recordings in the same animal under other conditions. Analysis steps within this framework are demonstrated on the Matlab environment with sample data and visual aids where appropriate. The associated script and sample data are as supplementary material with this article.

Step 1. Animal procedures, EEG/EMG measurements, and manual sleep scoring

EEG and EMG measurements were recorded for 24 h each (7 a.m.–9 p.m. Light, 9 p.m.–7 a.m. Dark) in C57BL/6J mice ($n = 18$, 8–10 weeks old, 24–29 g weight) from Jackson Labs (Bar Harbor, Maine) using procedures approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Kentucky. Each animal was stereotaxically implanted under isoflurane anesthesia with a headmount (Pinnacle Tech., Lawrence, Kansas) that was affixed to the skull by four screws, of which two served as epidural EEG electrodes (one frontal and one parietal) and the other two as common reference and ground respectively. The frontal screws were located 3.0 mm anterior to bregma. Two silver wires extending from the rear of the headmount were inserted into the nuchal muscle to record bipolar EMG. The reader is referred to previous work [3] for further details of the implantation technique, which is standard practice for such behavioral recordings in rodents.

After allowing the animal to recover for about two weeks, it was placed for monitoring in a $7'' \times 7''$ plexiglass cage with bedding and free access to food and water. A $100\times$ preamplifier was clipped to the headmount and conveyed the EEG and EMG signals via a slip-ring commutator to a biosignal amplifier (Pinnacle), which further amplified ($50\times$), bandpass-filtered (0.5–100 Hz for EEG, 10–100 Hz for EMG) and sampled the data at 400 Hz and 16-bit resolution for storage and analysis on a computer. Time-stamped digital video of the animal was recorded using a webcam with infrared illumination to assist in scoring of behavior in both Light and Dark conditions.

A human rater scored each 24-h recording in 4-s epochs as Wake, NREM, and REM vigilance states using standard criteria. Sleep and Wake were differentiated based on muscle tone measured by the EMG. During sleep, when the EMG is low in amplitude, epochs were labeled as NREM and REM depending on whether EEG spectral power was concentrated in the delta (0.5–4 Hz) or theta (6–9 Hz) band respectively and quasirhythmic oscillations in these frequency ranges were observed. The human rater's scores were stored for later use in the validation of the *SegWay* algorithm. The EEG/EMG recordings and manual scores were split into Light (7 a.m. to 9 p.m.) and Dark (9 p.m. to 7 a.m.) data sets so that the Light data could be used for developing the classifier and the Dark data for validation as out-of-sample data.

Step 2. Selection and extraction of EEG/EMG signal features

All analysis from this step onward was performed on Matlab™ (Mathworks, Natick, Massachusetts). Three features were extracted in each recording: 1. The r.m.s. power of the EMG after bandpass-filtering from 80 to 100 Hz, which expresses muscle tone and differentiates sleep from