Learning from Sleep Data

Hong Yu, Ke Li, Junyi Jiao
Georgia Technological University, Atlanta, GA, USA

Motivation

Sleep is vital to human health. The quality of sleep can be used as an indicator or precursor of certain diseases, such as Alzheimer [1] and Parkinson's disease [2]. According to American Academy of Sleep Medicine (AASM) manual [3], sleep can be divided into five stages: Wake (W), Non-Rapid Eye Movement stages N1 (for drowsiness or transitional sleep), N2 (for light sleep), and N3 (for deep sleep), and Rapid Eye Movement (REM), which can be determined by analyzing polysomnograms (PSG) data of patients during sleep. PSG data is a collection of data collected from various sensors, including electroencephalography (EEG), Electromyogram (EMG) and Electrocardiogram (ECG) recordings.

Identification of sleep stages enables diagnosis of many sleep disorders. Currently, sleep stage classification relies heavily on manual inspection by well-trained physicians or technicians, which is expensive and tedious [4]. The growing need for accurate and fast sleep stage classification and limitations of manual labeling have led to an emerging research in automatic sleep stage classification (ASSC) systems. This project aims to develop an end-to-end automatic ASSC system that takes a sequence of single channel EEG recordings and outputs a sequence of sleep stage labels.

Literature Survey

There are numerous methods reported in the literature that classify sleep stages either specifically or as part of a multiple-step pipeline for sleep disorder detection. In general, an ASSC system consists of three steps: data pre-processing, feature extraction and selection, and classification. Depending on specific implementations, some of the steps can be eliminated or combined.

Hand-crafted features from PSG recordings can be roughly divided into four categories: time domain, frequency domain, time-frequency domain, and other domain specific features. Common time domain features utilized include standard statistics [5–7] and mutual information [8]. Frequency and time-frequency domain feature extraction involves Fourier transformation and wavelet analysis or their variants [8-9]. A recent study [10] incorporates meaningful phenotypes summarized by domain experts into the feature sets. These phenotype features are specific to sleep stages such as sleep spindles, which is a discriminatory feature of N2, and slow wave sleep (SWS), which is defining characteristics of N3. Incorporating these meaningful domain specific features not only improves classification performance, but also enhances explainability of the models. It is rare to use only one type of feature. Reported researches commonly used more than two types of features.

Convolutional neural networks (CNN) have been applied for both automated feature construction and classification. CNN is powerful in exploring feature sets by convolving multiple filters with small segments of input data to extract time-invariant features. Authors of [11] experimented with multiple classification models including CNN and conventional machine learning models such as logistic regression and tree boosting, and reported better performance from CNN. Study in [12] uses CNN for both automated feature extraction and classification. In such neural network based classification systems, it is common to use a softmax layer as the output layer to give probability prediction for multiple sleep stages.

While CNN is powerful in extracting time-invariant features, it fails to capture temporal features. Temporal dependence between sleep stages is important since a sleep stage at one timestamp may well depend on the previous stage. To capture the temporal feature of sleep stage transition, recurrent neural networks (RNN) are employed by multiple researchers. Long short term memory (LSTM) network, a special type of RNN gains popularity for being capable of mitigating the gradient vanishing problem in traditional RNNs. Supratak et al. [13] proposed a model architecture that consists of CNN for extracting time-invariant features and bi-directional LSTM to capture temporal dependence of sleep stage transitions. Mousavi et al. [14] introduced an attention mechanism to the CNN+LSTM architecture, trying to learn the most relevant parts of the input sequence in the decoding phase.

While the aforementioned systems utilize single channel data, authors of [15] proposed a sequence to sequence deep learning model named SeqSleepNet that takes multiple channel recordings as input. Incremental work based on [15] has been reported recently. The authors [16] use pre-trained SeqSleepNet and finetune it with single-night data from individuals outside of the training dataset to accomplish personalization.

While the overall model performance improves over iterations, the complexity of models also grows, which potentially limits their deployment on wearable devices. Currently there is no single model that performs best in all classes and for all datasets. What we aim at in this project is to develop an ASSC system with moderate complexity and capable of generalizing to multiple datasets.

It is reported that EEG signal is among the most important signals for analyzing sleep staging and thus widely used [17] . In addition, the simple configuration of single EEG channel data enables continuous monitoring at home using wearable EEG acquiring devices. In this project, single channel EEG data will be utilized in determining sleep stages. We formulate the detection of sleep stages as a sequence-to-sequence multi-class classification problem, where the input is a sequence of single channel EEG epochs, and output is a sequence of sleep stage labels. The **project goal** is to develop an end to end ASSC system. If time permits, we will attempt to predict sleep disorders using the developed ASSC system.

Approach

The general workflow or pipeline of the proposed approach mainly comprises four essential steps, including feature engineering, training, evaluation, and deployment. A brief description of each step is given below.

Feature engineering

Extraction of highly relevant features from the raw input data is the very first critical step in the whole workflow. EEG signals are time series data of local electrical potentials, which are gathered from electrodes placed on multiple regions of a subject's scalp, each representing a data channel. The Rechtschaffen and Kales standard (R&K) rules and the American Academy of Sleep Medicine (AASM) define the criteria for stage scoring or staging of sleep for adults. Both R&K and AASM recommend the use of 30-second epochs of PSG signals for sleep staging. An expert scorer assigns one of the five stage names to each 30 seconds of the EEG data using the standard scoring rules.

There are three different feature representations pertaining to EEG signals, including raw EEG data, spectrogram, and expert-defined features. The raw EEG data can be considered as a three-dimensional tensor of *n* epochs, *m* channels in each epoch, and *k* (i.e., 30 multiplied with the sampling rate) data points in each epoch. The time series of EEG data can be converted into the frequency domain through Fourier transformation to obtain a spectrogram, which can be described with another three-dimensional tensor of *n* epochs, *m* (e.g., 29) sub-epochs of a 2-second duration with a 1-second overlap, and *k* (e.g., 257) frequency bins. Features can also be manually defined by experts who examine the both time series and spectrogram of the EEG data and consult the AASM rule sets. The expert-defined features will be used as the ground truth to assess the predictive accuracy of the machine learning models.

Training

The training step is to select and train an appropriate classification model which can autonomously annotate the EEG data epochs using the standard stage names and the constructed features. We propose to start with several machine learning algorithms, such as logistic regression and random forests, and then apply deep learning methods, including Concurrent Neural Networks (CNN), Recurrent Neural Networks (RNN), and a combination of both (CNN-RNN). We will be using a number of filters to convolve the feature matrix to produce preactivation feature maps and then be invoking the Rectifier Linear Unit (ReLu) as non-linear activation functions before passing the features through a max-pooling layer to reduce the spatial size of the representation. We plan to implement the RNN formulation using the Long Short Term Memory (LSTM) method in Tensorflow and to incorporate dropout regularization to avoid overfitting. For the hybrid CNN-RNN model, we are going to use CNN to extract the spatial features from EEG, which is time invariant and independent in each step, and pass them to a RNN model, which learns the temporal dependency present of the spatial feature already extracted by CNN.

Evaluation

In the evaluation step, the data will be split into training, validation, and test sets. We propose to use 90% of the considered EEG data as the training data and the remaining 10% as the test data. For each classification method evaluated, we will be using random search and then grid search, along with cross-validation, to tune hyperparameters. The metrics to be used include precision, ROC-AUC, F1 score, and Cohen's kappa.

Deployment

The deployment of the selected model will be implemented in local machines and Google Colab. Sci-kit learn will be used for the machine learning models (e.g., logistic learning and random forests), while TensorFlow will be employed for the deep learning methods (e.g., CNN and RNN).

Data

There are a few public datasets available of different sizes. Our overall strategy is to start with one smaller data set so that it will be easier to set up the environment on our local machines and to establish our data processing and learning pipeline. We will then apply the pipeline to a larger dataset for a more comprehensive study.

For the smaller dataset, we plan to use Sleep-EDF Database Expanded available at PhysioNet [18], which has been adopted in a number of studies. It contains 197 whole night PolySomnoGraphic sleep recordings with multiple channel data, containing EEG, EOG, chin EMG, and event markers. The data is in edf or edf+ file format, and different sleep stages are annotated. The average recording duration was around 9 hours, with subjects' age ranging from 25 to 101 years old. The sampling rate for both EEG and EOG signals was 100 Hz.

We will also consider a larger data set that contains PSG recordings of 2000 subjects from Massachusetts General Hospital (MGH). The data was randomly selected from a mixture of diagnostic and split night recordings collected from patients whose ages range from 42 years old to 64 years old, with an average age of 53. The EEG signals were sampled at 200 Hz.

Experimental Setup

In the early exploration/setup stage, we plan to start with the smaller data set (PhysioNet, 8GB of data) to get ourselves familiar with the data and establish the basic structure of the data processing and learning pipeline. In this case, we are going to run on local clusters with Spark. In the later stage of our study, we will apply our methods to a larger dataset in Google Colab to achieve our "stretch goal".

Regarding the programming framework and packages, we will use pySpark for data processing, and then switch to python (sklearn and tensorflow) for model training and tuning. We will also explore the use of Flint for time series data analysis and feature extraction.

Timeline

The proposed milestones of the project are tabulated below.

Table 1. Proposed project timeline

Milestone	Description	Duration	Responsibility
1	Understand the data and set up a simple data processing pipeline. Goal: Able to extract **basic** usable features for traditional classification algorithms (e.g., Logistic regression) and deep learning models.	10/10 -10/18	All
2	Feature engineering and model exploration. Goal: Extract useful features for model training and train at least one baseline model (e.g., logistic regression).	10/19 - 10/25	All
3	Data processing/feature extraction and deep learning models Goal: train a deep learning model and compare its performance with the baseline model.	10/25 - 11/01	All
4	Work on a project draft. Apply the data processing and modeling pipeline to a different data set if time allows	11/01 - 11/08	All
5	Interpret the results from the model and compare them to literature.	11/09 - 11/22	All
6	Finalize analysis and work on report and final presentation.	11/22 - 12/08	All

References

1. Lim MM, Gerstner JR, Holtzman DM. The sleep-wake cycle and Alzheimer's disease: what do we know? Neurodegener Dis Manag 2014;4:351–62. doi:10.2217/nmt.14.33
2. Cooray N, Andreotti F, Lo C, et al. Detection of REM sleep behaviour disorder by automated polysomnography analysis. Clin Neurophysiol 2019;130:505–14. doi:10.1016/j.clinph.2019.01.011
3. Berry RB, Brooks R, Gamaldo CE, et al. The AASM manual for the scoring of sleep and associated events. … Academy of Sleep … 2012.
4. Aboalayon K, Faezipour M, Almuhammadi W, et al. Sleep stage classification using EEG signal analysis: A comprehensive survey and new investigation. Entropy 2016;18:272. doi:10.3390/e18090272
5. Hassan AR, Bashar SK, Bhuiyan MIH. On the classification of sleep states by means of statistical and spectral features from single channel Electroencephalogram. In: 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE 2015. 2238–43. doi:10.1109/ICACCI.2015.7275950
6. Hassan AR, Hassan Bhuiyan MI. Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. Biocybernetics and Biomedical Engineering 2016;36:248–55. doi:10.1016/j.bbe.2015.11.001
7. Khalighi S, Sousa T, Pires G, et al. Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels. Expert Syst Appl 2013;40:7046–59. doi:10.1016/j.eswa.2013.06.023
8. Herrera LJ, Mora AM, Fernandes C, et al. Symbolic representation of the EEG for sleep stage classification. In: 2011 11th International Conference on Intelligent Systems Design and Applications. IEEE 2011. 253–8. doi:10.1109/ISDA.2011.6121664
9. Sanders TH, McCurry M, Clements MA. Sleep stage classification with cross frequency coupling. Conf Proc IEEE Eng Med Biol Soc 2014;2014:4579–82. doi:10.1109/EMBC.2014.6944643
10. Al-Hussaini I, Xiao C, Westover MB, et al. SLEEPER: interpretable Sleep staging via Prototypes from Expert Rules. arXiv preprint arXiv … 2019.
11. Biswal S, Kulas J, Sun H, et al. SLEEPNET: automated sleep staging system via deep learning. arXiv preprint arXiv … 2017.
12. Tsinalis O, Matthews PM, Guo Y, et al. Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. arXiv preprint arXiv … 2016.
13. Supratak A, Dong H, Wu C, et al. DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG. IEEE Trans Neural Syst Rehabil Eng 2017;25:1998–2008. doi:10.1109/TNSRE.2017.2721116
14. Mousavi S, Afghah F, Acharya UR. SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach. PLoS ONE 2019;14:e0216456. doi:10.1371/journal.pone.0216456
15. Phan H, Andreotti F, Cooray N, et al. SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging. IEEE Trans Neural Syst Rehabil Eng 2019;27:400–10. doi:10.1109/TNSRE.2019.2896659
16. Phan H, Mikkelsen K, Chén OY, et al. Personalized automatic sleep staging with single-night data: a pilot study with Kullback–Leibler divergence regularization. Physiol Meas 2020;41:064004. doi:10.1088/1361-6579/ab921e
17. Kayikcioglu T, Maleki M, Eroglu K. Fast and accurate PLS-based classification of EEG sleep using single channel data. Expert Syst Appl 2015;42:7825–30. doi:10.1016/j.eswa.2015.06.010
18. Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000;101:E215-20. doi:10.1161/01.cir.101.23.e215