Learning Sleep Stages

Ria Verma, MS[1], Ayush Baid, MS[1], Yu Jia Shai Xie, MS[1]
[1]Georgia Institute of Technology, Atlanta, GA

Motivation

Sleep quality and sleep disorder have shown high correlations with various health problems such as cardiovascular disease [1], homeostasis imbalance [2], and even Alzheimer [3]. Attention has been drawn to the Centers for Disease Control and Prevention (CDC) as well as its partners regarding sleep and sleep disorders in order to provide preventative health services to the public [4]. Since sleep quality is a major concern for sleep-related disorders, and sleep stages can be representative to the sleep behaviors, our group has determined to analyze data of electroencephalogram (EEG) collected during sleep to provide insightful information about an individual's sleep status such as polysomnography scoring.

Sleep Data and Machine Learning-Various approaches have been applied for studying sleep stages and sleep cycles: non-linear dynamical analysis [5], time-frequency analysis and entropy measures [6], and nonetheless the machine learning approach. Machine learning has been applied in the field of sleep data analysis to facilitate the automated sleep stage identification, sleep quality assessment [7], etc. One rising field of study incorporates the classification techniques in machine learning to automate the analysis of sleep stages and sleep qualities. After surveying existing literatures regarding machine learning models applied in the field of sleep analysis, our team sees potential improvements to the current pipeline and machine learning architecture to provide better automated polysomnography scoring and sleep stage classification.

Literature Survey

The approach in [7] applies CNNs on one dimensional bio-signal such as electroencephalography (EEG) and electrocardiography (ECG). The sleep stages considered for classification are Wake, Rem, N1, N2 and N3 sleep stages. Since these classes have varying frequency, class balancing approaches with random sampling has also been implemented to achieve the F1 score 81% and overall accuracy of 74% over all subjects [7].

SLEEPNET is a deep recurrent neural network (RNN) that uses expert-defined features to represent each 30-sec interval and learns to annotate EEG [8]. It is trained on the largest physiology database assembled to date, consisting of PSGs from over 10,000 patients from the Massachusetts General Hospital (MGH) Sleep Laboratory. The optimal model had 5 layers of LSTM cells with tanh activation function and dropout keep probability of 0.9. They achieved an average accuracy of 86%, which is comparable to human-level scoring performance [8].

Another approach is to use recurrent-convolutional neural networks (RCNN), where the CNN extracts sleep-specific subject-invariant features from RF signals and the RNN captures the temporal progression of sleep [9]. In [9], an adversarial training regime is adopted that discards extraneous individual or measurement condition specific information. This approach is done on a radio signals dataset and achieves an accuracy of 79.8%.

To capture higher order data from the input dataset, transfer learning was applied in [10]. Spectrograms are generated from electroencephalography, electrooculography, and electromyography. In the optimal model, the output from the spectrograms is fed as the input to the convolutional neural layers and then recurrent neural network (LSTM) layers. This achieved a weighted F1 score of 87%.

Another approach is to combine deep learning models with expert defined rules using a prototype learning framework to generate simple interpretable models [11]. This consists of the signal embedding module, where a CNN is used to get feature representation from raw polysomnogram (PSG) data. Next, there is an expert rule module that encodes each epoch into a multi-hot vector. Finally, there is the prototype learning module, where the output from the previous two modules are combined [11].

Ensemble learning with stacked sparse autoencoders can also be used to classify sleep stages. In [12], class-balanced random sampling is also used to achieve a F1 score of 84%.

Data

We will be using the PhysioNet dataset [13] in this paper. This is a publicly available dataset that contains sleep polysomnogram (PSG) data i.e. data that is consolidated from multiple sensors including electroencephalogram (EEG), an electrooculogram (EOG), an electromyogram (EMG), and an electrocardiogram (ECG).

PSG data is divided into 30 second intervals called epochs. These are then classified into the stages of sleep. According to the American Academy of Sleep Medicine (AASM) manual [14], sleep is categorized into the Rem (R) stage, 3 non-Rem stages (N1, N2, N3) and the Wake (W) stage.

Approach

Data Transformation

We will be using Spectrogram (FS) approach to average spectrogram from multiple channels of EEG is provided as the input feature to different models. We will also be comparing this against using Raw waveform (FR) as input. This is the raw waveform from 6 channels of 30 seconds of EEG data segments each.

Feature Generation

We will be using Convolutional Neural Networks (CNN) to generate features for each epoch. Thus, we will utilize CNNs to learn task specific filters without prior knowledge. The output from the spectrogram would have transformed our data into two dimensional stacks of frequency specific activity data over time that can be fed in as input to the CNN.

Loss Function

We use a triplet network to help the model distinguish between different items that belong in the class (ex: “N1”) i.e. to help with intra-class discrimination to promote better finer level retrieval. This network structure employs triplet loss, which takes an anchor (query) epoch, a positive epoch i.e. a different epoch that has the same class label, and a negative epoch i.e. an epoch that has a different class label. This should help during classification.

Using RNNs

RNNs [15] have been widely explored in automated-sleep scoring, and has been commonly used in state-of-the-art work in the last few years. Recurrent networks have the ability to preserve useful information across time and hence are quite adept at handling temporal sequences. Long-Short-Term-Memory [16] cells are the most common recurrent networks in this application. However, in the NLP domain, transformers and attention [17] have overtaken LSTMs and established themselves as the new cutting edge. We will experiment with transformers in the problem. One advantage we can think is that transformers provide a bidirectional view of the data: i.e. when we want to predict the sleep cycle label at a particular time step, we can take a look at both the data before and after this particular timestamp. This can be very helpful in avoiding predictions which are unrealistic. For example: a timestamp of label ‘awake’ between two labels of ‘REM’ is not possible in the real world. A bidirectional view of the data helps us in these cases.

Experimental Setup

Platform

Our team decides to use AWS (free education credit) as the central cloud computing platform for the majority of our activities due to the large data size.

Data Storage and Processing

We plan to use AWS S3 bucket as our raw data storage location and AWS Spark as our data processing engine. We will also use our local standalone spark cluster to finetune the data processing procedures before sending it to the cloud computing engine.

Machine Learning Training

We choose Python (3.7+) as our main programming language for machine learning models. We have decided to include Scikit-learn, PyTorch, Numpy, Scipy, TensorFlow, and other popular machine learning and data manipulating packages.

Presentation, Analysis, and Evaluation

We will present our final results in the report format, and we will assess our proposed solution with standard statistical analysis and evaluations (i.e. mean, standard deviation, hypothesis testing).

Timeline

Table 1. Weekly breakdown of timeline and deliverables.

Weeks	**Milestones of Deliverables**
Week 8 (Feb 25)	Ayush/Ria/Yu Jia: Read Literature Survey; Write the Proposal; Familiarize ourselves with the dataset features; Articulate experimental setup Deadline: Project Proposal deadline
Week 9 (Mar 3)	Ayush/Ria/Yu Jia: Design and implement baseline model
Week 10 (Mar 10)	Ayush/Ria/Yu Jia: Get results on baseline models. Create approaches to improve results. Get results on the improved model.
Week 11 (Mar 17)	Ayush/Ria/Yu Jia: Discuss analysis of the results from the model. Setup the analytic infrastructure for your project
Week 12 (Mar 24)	Ayush: Write Results Ria: Describe Approach Yu Jia: Write Analysis
Week 13 (Mar 31)	Ayush/Ria/Yu Jia: Ensure Draft is completed and submitted Deadline: Project Draft Due
Week 14 (April 7)	Ayush/Ria/Yu Jia: Make the final Presentation
Week 15 (April 14)	Ayush/Ria/Yu Jia: Write the final paper
Week 16 (April 21)	Ayush/Ria/Yu Jia: Submit Final Project with code, presentation and final paper

References

1. Newman AB. Relation of Sleep-disordered Breathing to Cardiovascular Disease Risk Factors: The Sleep Heart Health Study. *American Journal of Epidemiology*. 154(1): 50–9, 2001.
2. Punjabi NM. Sleep-Disordered Breathing, Glucose Intolerance, and Insulin Resistance: The Sleep Heart Health Study. *American Journal of Epidemiology*. 160(6): 521–30, 2004.
3. Ju ES, Lucey BP, Holtzman DM. Sleep and Alzheimer disease pathology—a bidirectional relationship. *Nature Reviews Neurology*. 10(2): 115–9, 2013.
4. Liu Y, Croft JB, Wheaton AG, et al. Clustering of Five Health-Related Behaviors for Chronic Disease Prevention Among Adults, United States, 2013. *Preventing Chronic Disease*. 2016;13.
5. Acharya UR, Faust O, Kannathal N, Chua T, Laxminarayan S. Non-linear analysis of EEG signals at various sleep stages. *Computer Methods and Programs in Biomedicine*. 80(1): 37–45, 2005.
6. Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H. Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. *Computer Methods and Programs in Biomedicine*. 108(1): 10–9, 2012.
7. Tsinalis O, Matthews PM, Guo Y, and Zafeiriou S. Automatic sleep stage scoring with single-channel eeg using convolutional neural networks. *arXiv preprint arXiv:1610.01683*, 2016.
8. Biswal S, Kulas J, Sun H, Goparaju B, Westover BM, Bianchi MT, and Sun J. SLEEPNET: Automated sleep staging system via deep learning. 26 July 2017.
9. Zhao M, Yue S, Katabi D, Jaakkola TS, Bianchi MT. Learning sleep stages from radio signals: A conditional adversarial architecture. *In International Conference on Machine Learning*, pages 4100–4109, 2017.
10. Zhang L, Fabbri D, Upender R, Kent D. Automated sleep stage scoring of the sleep heart health study using deep neural networks. *Sleep*. 42(2): 1—10, 2019.
11. Al-Hussaini I, Xiao C, Westover MB, and Sun J. Sleeper: interpretable sleep staging via prototypes from expert rules. *In Machine Learning for Healthcare Conference*, pages 721–739, 2019.
12. Tsinalis O, Matthews PM, and Guo Y. Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Annals of Biomedical Engineering*. 44(5): 14 October. 2015.
13. PhysioNet: The Sleep-EDF database [Expanded]: http://www.physionet.org/physiobank/ database/sleep-edfx/ (Accessed January 2015).
14. Iber C, Ancoli-Israel S, Chesson A, and Quan SF. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Westchester, IL: *American Academy of Sleep Medicine*, 2007.
15. Mikolov T, Karafiat M, Lukas B, Cernocky J, Khudanpur S. Recurrent neural network-based language model. *Eleventh annual conference of the international speech communication association*, pages 1045—1048, 2010.
16. Hochreiter S, and Schmidhuber J. Long short-term memory. *Neural computation*. 9(8): 1735-1780, 1997
17. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in neural information processing systems*. Pages 6000-6010, 2017.