DEEP LEARNING METHOD FOR ENHANCING SLEEP STAGING CLASSIFICATION

By

Zhengzhong Zhu

Senior Thesis in Computer Engineering

University of Illinois at Urbana-Champaign

Advisor: Jimeng Sun, Mark Hasegawa-Johnson, Zhen Lin

May 2022

Abstract

Traditionally, doctors classify the sleep staging of patients by manually examining the spectrogram of electroencephalogram (EEG) data at different time intervals. We propose a two-stage pipeline that automatically predicts the staging of sleep given enough confidence and requests for manual labeling from doctors given enough uncertainty. In the first stage, multiple deep learning models, such as VGG-16 and ResNet-50, take in EEG data preprocessed with fast Fourier transform (FFT) and output a prediction and some confidence scores. In the second stage, multiple models with trainable parameters take in the confidence scores and output the decision of whether accept or reject the prediction. Among different combinations of models, RESNET-50 and SCRIB achieve the lowest average class-specific risk.

Subject Keywords: Deep Learning; Sleep Staging Classification

Contents

1. Introduction

In recent years, deep learning models have been dramatically advanced with the emergence of large public datasets such as ImageNet [1] and the immense amount of computational power supported by the GPUs. Over the past decades, spectrograms of electroencephalogram (EEG) data were manually inspected by doctors to determine the sleep staging of their patients. However, the process of manual inspection is labor-intensive and time-consuming. Deep learning models can be applied to expedite such a process.

Spectrograms of EEG data are images that can be processed by convolutional neural networks. Convolutional neural network models, such as VGG-16 [2] and ResNet-50 [3], have powerful representation learning capacities and perform well on real datasets.

Recently, rejection models such as SGR [4] and SCRIB [5] have been proposed to reject neural network model predictions based on the confidence scores output from the neural network model. Models such as SGR control the overall risk, and models such as LABEL [6] and SCRIB control class-specific risks.

This thesis proposes a two-stage pipeline that automatically predicts the sleep staging of patients given spectrograms of EEG data and rejects with confidence. In the first stage, a convolutional neural network takes in spectrograms of EEG data of patients and generates predictions and confidence scores. In the second stage, a rejection model takes in the confidence scores generated in the first stage and determines whether to reject the prediction or not. The doctors would receive a list of stages with confidence scores if the rejection model rejected the prediction.

The main contributions of this thesis can be summarized as follows:

- Experiments of two-stage pipeline with different convolutional neural network models and different rejection models

- Comparisons on model performance

The rest of the paper is organized as follows. In Section 2, we describe the different convolutional neural network models and rejection models used in the two-stage pipeline. In Section 3, we describe the methodology and discuss the experimental results. Section 4 provides a conclusion to the thesis.

2. Literature Review

Three convolutional neural network models, AlexNet [7], VGG-16, and ResNet-50, are used in the thesis.

In 2012, AlexNet was proposed to solve the image classification problem on the ImageNet dataset. The key contribution of this paper was the demonstration of the parallel uses of the GPUs, which makes the training of deep convolutional neural network models feasible. AlexNet takes in 224 x 224 images, passes through five convolution layers along with max-pooling and three fully connected layers, and outputs confidence scores of different classes.

In 2014, VGG-16 was proposed to solve the same problem as AlexNet. Different from AlexNet, VGG-16 has 13 convolution layers and twice the number of parameters. The main contribution of this paper was the constant use of very small (3 x 3) convolution filters, which dramatically reduce the number of parameters per layer.

In 2015, ResNet-50 was proposed to solve the same problem as AlexNet and VGG-16. ResNet introduced the concept of residual blocks and skip connections. Residual blocks use even smaller convolution filters to reduce the number of parameters, while skip connections provide additional paths for gradients to mitigate the vanishing gradient problem. Although ResNet-50 has 48 convolution layers, it only has a third of the number of parameters of AlexNet.

Three rejection models, SGR, LABEL, and SCRIB, are used in the thesis.

In 2017, SGR was proposed to control the overall risk of the base classifier. The main contribution of this paper was the application of the predicted Maximum Class Probability on the selection function, which theoretically guaranteed the overall risk given accurate predictions.

In 2018, LABEL was proposed to control the class-specific risks of the base classifier. The key contribution was an analytic solution to control the unconditional class-specific risks and some theoretical guarantees.

In 2021, SCRIB was proposed to solve the same problem as LABEL. However, the difference between SCRIB and LABEL is that SCRIB controls both the unconditional class-specific risks and conditional class-specific risks.

3. Description of Research Results

3.1 Datasets

Two sleep staging datasets are used in the thesis: Sleep-EDF Expanded [8] and ISRUC-SLEEP [9].

The Sleep-EDF Expanded dataset is a publicly available sleep staging dataset open-sourced in 2013. It contains 197 whole-night Polysomnographic sleep recordings in the EEG form. We split the dataset into a training set (413,456 samples), a validation set (51,682 samples), and a test set (51,682 samples). The sleep staging classes 0-4 are corresponding to W/N1/N2/N3/REM.

The ISRUC-SLEEP dataset is another publicly available sleep staging dataset open-sourced in 2016. It contains 100 subjects (89,283 samples) recorded in the EEG form. We split the dataset into a training set (71,431 samples), a validation set (8,926 samples), and a test set (8,926 samples). Similarly, the sleep staging classes 0-4 are corresponding to W/N1/N2/N3/REM.

3.2 Methodology

The two-stage pipeline consists of two models: a convolutional neural network model and a rejection model. The input of the two-stage pipeline is the spectrogram of EEG data of patients, and the output of the two-stage pipeline is the candidate(s) of the sleep stage and whether the candidates should be trusted or not.

The first stage of the two-stage pipeline is a classifier modeled by a convolutional neural network chosen from AlexNet, VGG-16, and ResNet-50. The input of the first stage is the spectrogram rescaled to the model input size (224 x 224). The model of the first stage is composed of two parts: pre-trained convolution layers trained on the ImageNet and two fully connected layers fine-tuned on the sleep staging datasets. The outputs of the first stage are the confidence score of each sleep stage and a prediction generated by argmax.

The second stage of the two-stage pipeline is a rejector chosen from SGR, LABEL, and SCRIB. The input of the second stage is the confidence scores generated from the first stage. The model of the second stage has some trainable parameters, either an overall threshold of the confidence scores or a set of thresholds of the confidence scores for each sleep staging class. The model is trained on the sleep staging datasets. The outputs of the second stage are the candidate(s) of the sleep stage and whether to trust the candidate(s) or not.

3.3 Experiments

A total of 18 sets of experiments are conducted. For each sleep staging dataset, the two-stage pipeline chooses one convolutional neural network model from AlexNet, VGG-16, and ResNet-50 and chooses one rejection model from SGR, LABEL, and SCRIB.

Table 1   Experiments on the Sleep-EDF Expanded dataset

First-Stage Model	Second-Stage Model	Average Class-Specific Risk
AlexNet	SGR	26.20
AlexNet	LABEL	13.74
AlexNet	SCRIB	4.92
VGG-16	SGR	18.43
VGG-16	LABEL	6.87
VGG-16	SCRIB	0.94
ResNet-50	SGR	15.29

ResNet-50	LABEL	6.44
ResNet-50	SCRIB	0.76

Comparing the first-stage models, ResNet-50 learns representations more accurately than VGG-16 than AlexNet. Hence, given the same second-stage model, the average class-specific risk of ResNet-50 is lower than that of VGG-16 than that of AlexNet.

Comparing the second-stage models, SGR controls the overall risk, LABEL controls the unconditioned class-specific risks, and SCRIB controls both the unconditioned class-specific risks and the conditioned class-specific risks. Consequently, the average class-specific risk of SGR is higher than that of LABEL than that of SCRIB.

Table 2 Experiments on the ISRUC-SLEEP dataset

First-Stage Model	Second-Stage Model	Average Class-Specific Risk
AlexNet	SGR	18.45
AlexNet	LABEL	12.01
AlexNet	SCRIB	6.96
VGG-16	SGR	10.14
VGG-16	LABEL	5.95
VGG-16	SCRIB	2.28
ResNet-50	SGR	8.73
ResNet-50	LABEL	4.02
ResNet-50	SCRIB	1.61

4. Conclusion

In this thesis, we evaluate different deep learning methods for sleep staging classification. We propose a two-stage pipeline, where the first stage contains a convolutional neural network model, and the second stage contains a rejection model. From the experiments performed on two sleep staging datasets, we demonstrated that SCRIB is more effective in controlling the class-specific risks. On both datasets, the combination of ResNet-50 and SCRIB achieves the lowest average class-specific risks.

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: a large-scale hierarchical image database,” *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *International Conference on Learning Representations (ICLR)*, 2015.

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[4] Y. Geifman and R. El-Yaniv, “Selective classification for deep neural networks,” *Advances in Neural Information Processing Systems*.

[5] Z. Lin, C. Xiao, L. Glass, M. Westover, and J. Sun, “SCRIB: set-classifier with class-specific risk bounds for blackbox models,” *Association for the Advancement of Artificial Intelligence Conference (AAAI)*, 2022.

[6] M. Sadinle, J. Lei, and L. Wasserman, “Least ambiguous set-valued classifiers with bounded error levels,” *Journal of the American Statistical Association*, 114(525).

[7] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” *Conference on Neural Information Processing Systems (NeurIPS)*, 2012.

[8] B. Kemp, A. Zwinderman, B. Tuk, H. Kamphuisen, and J. Oberye, “Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG,” *IEEE Transactions on Biomedical Engineering (IEEE-BME)*, 2000.

[9] S. Khalighi, T. Sousa, J. Santos, and U. Nunes, “ISRUC-Sleep: a comprehensive public dataset for sleep researchers,” *Computer Methods and Programs in Biomedicine*, 2016.