Attention decoding using Deep learning & Conditional variational autoencoder for data augmentation

Introduction

We employ deep convolutional neural networks to detect the direction of incoming speech for hearing-impaired individuals, additionally generating artificial data for improving model performance.

Main text

The human brain is a highly complex and sophisticated system, characterized by its ability to perform a wide range of cognitive functions, including attention, perception and memory.

This study focuses on the auditory attention task, which is a fundamental cognitive process that allows us to selectively attend to relevant sounds while filtering out irrelevant ones, commonly referred to as the "cocktail-party effect". In everyday situations, when we engage in conversation in a noisy environment, our brain can effectively suppress background noise and focus on the attended speech. However, individuals with hearing impairment (HI) face significant challenges in this context. Therefore, developing Deep Learning-based models that can decode the attended speech in the presence of multiple talkers and background noise can help HI individuals to focus on and understand the speech of interest.

For this purpose, two deep convolutional neural network (DCNN) models were created to accomplish two distinct tasks. The first model was able to classify incoming sound segments as noise or speech-in-noise, while the second model was able to identify the direction (left vs. right) of incoming attended speech. The DCNN models achieved accuracy (ACC) scores of 69.9% and 84.9%, and area under the curve (AUC) scores of 77.5% and 92.3%, respectively, for the two objectives.

In addition, two conditional variational autoencoders (CVAEs) were trained to generate artificial data for data augmentation, with the goal of improving the performance of the final models by learning from a latent space of training data to generate unique data for each respective class.

The performance of the DCNN models was further improved with the use of augmented data from the CVAE models, resulting in ACC scores of 70.5% and 86.6%, and AUC scores of 78.3% and 93.6%, respectively. The EEG data was processed using a time window of 1 second, allowing the models to function in real-time scenarios.

The proposed methods were tested on a data set of 32 participants who performed an auditory attention task. Participants were instructed to attend to one of two talkers in the front and ignore the talker on the other side and background noise behind them, while high-density EEG was recorded. The EEG data consists of 66 channels in total, and all channels were used in this study.

The findings of this study demonstrate the high capability of the proposed DCNN models in accurately detecting the direction of incoming speech and differentiating between babble noise and speech of interest, even with a small time window of just 1 second using multi-channel EEG data making the models useable in real-time applications. Moreover, the results highlight the success of the CVAE models as a valuable tool for data augmentation, generating synthetic data that closely approximates the latent space information of the training data. This suggests the potential of CVAE for improving the performance of deep learning models in EEG-based attention decoding tasks.