Deep Learning for Classifying Creative Tasks from EEG Signal

Patrick Geitner
pgeitner@ur.rochester.edu
University of Rochester
Rochester, New York

Abstract

This work compares two approaches to classifying idea generation, idea evolution, and idea evaluation tasks from electroencephalography (EEG) signal data. One approach, inspired by a 2021 study by Zia and Jeng, uses microstate analysis and changes in upper and lower alpha band power to build a classification model to distinguish between the three creative tasks. This method is desirable because of its interpretability in terms of the activity of large scale networks associated with EEG microstates and general mental functions associated with the power of different frequency bands [1]. It was found that this method enabled reliable classification of resting behavior versus creative behavior, but was not enough to separate the three creative tasks. An alternative method is proposed that uses convolutional neural networks (CNNs) for the automated learning of features from short-period EEG spectrograms. This method, while less interpretable, was able to identify which creative mode of thinking an individual was in with 82.33% given an EEG spectrogram generated from 10-seconds of signal. Gradient-weighted class activation mapping (GradCAM) is explored to gain insight into the model's classification process.

1 Introduction

1.1 Prior Study: Idea Generation, Idea Evolution, and Idea Evaluation EEG Dataset

A recent neurocognitive study by Jia and Zeng examined EEG signal during different creative modes of thinking. The study collected raw EEG data from participants completing a modified figural Torrence Test of Creative Thinking that included three tasks and a rest period:

1. **Idea Generation**: Complete a sketch immediately triggered by a sketch stimulus
2. **Idea Evolution**: Complete a sketch that was radically distinct from what was initially triggered by the sketch stimulus
3. **Idea Evaluation**: Reflect on challenges associated with the idea generation and idea evolution tasks

The study was described as a loosely-controlled experiment as it was designed to foster the participants' creativity: allowing a flexible amount of time to complete each phase of the experiment. The study published a dataset that consisted of raw EEG signal from 29 subjects each completing the described tasks three times. EEG signal was measured using a device with 63 channels and sampled at 300Hz [1].

1.2 Prior Study: Statistical Findings

In addition to the dataset produced by the work of Jia and Zeng, significant statistical analysis was done to identify features of the EEG signal that differed between the three creative tasks: idea generation, idea evolution, and idea evaluation.

Decreases in power in the lower alpha (8-10Hz) and upper alpha (10-12Hz) frequency bands were observed to differ between some of the creative tasks. Additionally, EEG microstate analysis suggested that different large-scale networks, like the default mode network and cognitive control network, were more active during some creative tasks than others [1].

While this produced several features that differed between the EEG signal during different creative tasks, no effort was made to use these features to build predictive models. In this study, we explore the use of these features to classify what creative task a subject is completing from an EEG recording. An alternative method will also be explored that involves using convolutional neural networks to classify short-period EEG spectrograms.

2 Approach One: Feature Extraction and Classification

Given the findings listed in section 1.2 which suggested metrics that differed between EEG signal from different creativity-related tasks, this section outlines the extraction of these features from raw EEG signal and an attempt to build a predictive model that uses these features to determine whether an individual is in 1) Rest, 2) Idea Generation, 3) Idea Evolution, or 4) Idea Evaluation. The focus of this section was to be able to distinguish between these tasks given the EEG signal of that entire phase of the experiment. With this approach, the resulting dataset has 319 observations as 29 subjects each completed 3 idea generation sessions, 3 idea evolution sessions, 3 idea evaluation sessions, and 2 rest sessions.

2.1 Microstate Analysis

EEG microstates are successive short periods of stability in the scalp potential field that are thought to be attributed to the simultaneous activity of large-scale functional brain networks. Recent research has demonstrated that microstates are closely associated with resting state networks identified by fMRI [3].

To replicate the microstate analysis done by the Jia and Zeng study, the MNE Microstate python library was used to segment the raw EEG signal into 6 microstates [4]. This process involves using a modified K-Means algorithm to cluster similar activity across multiple channels. It is important to note that since K-Means involves random initialization of cluster assignments, resulting clusters may be slightly different from each run. Figure 1 shows an example of a short segment of clustered EEG signal and the microstate locations based on the electrode layout.

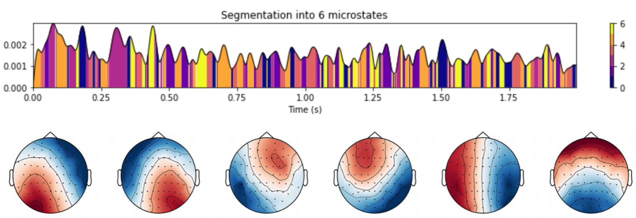

Figure 1. Microstate Segmentation and Mapping

To extract this as a meaningful feature to be used by a classification model, the relative duration of each of the six microstates was computed for each of the 319 sessions. Relative duration was used so that the feature would be invariant to the amount of time taken to complete the task. Since the experiment allowed for a flexible amount of time to complete each task, we did not want the models to be able to learn any relationship between the time taken and creative tasks.

2.2 Power in the Upper and Lower Alpha Bands

The study also suggested that differences in lower (8Hz-10Hz) and upper (10Hz-12Hz) task-related alpha power were significantly different between EEG signal from different creative modes of thinking.

To extract this as a feature for each of the sessions of the experiment, Welch's method was used to estimate the Power Spectral Density of the EEG signal and Simpson's rule was used to estimate the power in the upper and lower alpha bands for each electrode. All computations were done using the SciPy python library [5]. Figure 2 illustrates the computation of lower alpha power for a single electrode.

2.3 Modeling Creative Modes of Thinking

With the features extracted in the previous sections, we hoped to be able to train a machine learning model to be able to predict what creative mode of thinking a subject was in, given the EEG signal from a session. Support vector machine (SVM) classifiers and feed-forward neural networks were used to assess the efficacy of this approach. To evaluate the accuracy of the model, 5-fold cross-validation was used.

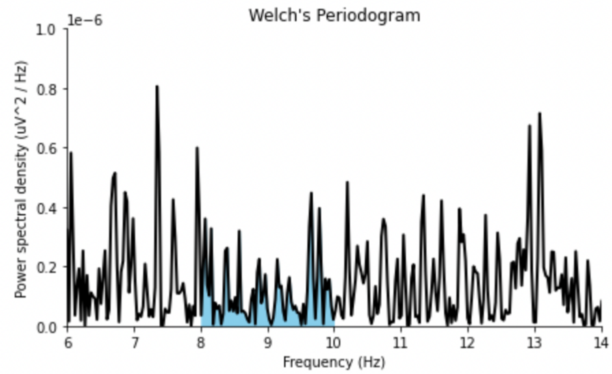

Figure 2. Using Simpson's Rule to Lower Alpha Power for a Single Electrode

2.3.1 Support Vector Machines.
Using a support vector machine with a degree-four polynomial kernel function yielded the highest cross-validation accuracy at 64%. In figure 3, we use principal component analysis to obtain a two-dimensional representation of our data and visualize a two-dimensional representation of the SVM decision boundary to gain visibility into the model's classification.

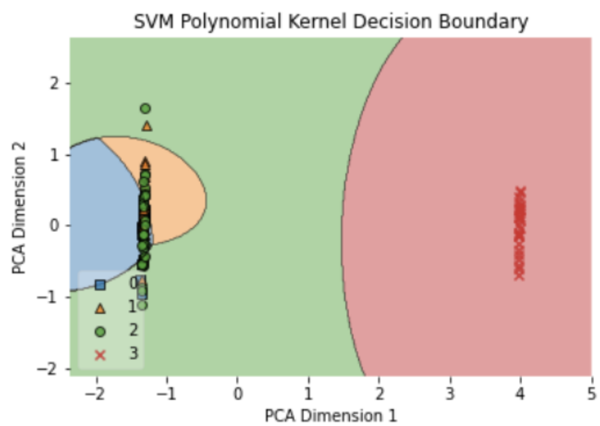

Figure 3. SVM Polynomial Decision Boundary on Principal Components

Looking at figure 3, we can see that the rest class (red) is easily separable from the three creative modes of thinking (blue, green, and orange), but idea generation, idea evolution, and idea evaluation do not seem to be easily separable using these features.

2.3.2 Neural Networks.
A feed-forward neural network was also used to attempt to distinguish between these creative modes of thinking. Because the dataset, in this format, has only 319 observations, a relatively simple model was used with only three hidden layers and rectified linear unit activation (ReLU). To combat overfitting to the relatively small training set, heavy dropout and a small learning rate were used. With those parameters, the model reached 67.6%

cross-validation accuracy, slightly outperforming the SVM model.

2.4 Results of Approach One

In this approach, features were manually extracted using EEG microstates and alpha band power analysis in an attempt to train a machine learning classification model to be able to distinguish between idea generation, idea evolution, idea evaluation, and resting tasks. Using the described features and models, the highest cross-validation accuracy reached for this four-class classification problem was 67.6%. To summarize, with these features, machine learning models were able to easily distinguish between the rest phase and creative phases, but were not able to reliably distinguish between the three creative modes of thinking. Further, the models had a particularly difficult time distinguishing between the idea generation and idea evolution tasks using these features.

3 Approach Two: Time-Frequency Domain and Automated Learning of Features

Given the usefulness of using changes in upper and lower alpha band power in distinguishing between idea evaluation and the other two creative modes of thinking, we were curious to see if using information about the power of different frequency bands would also help to train a model that can distinguish between all three creative modes of thinking. However, rather than just using the power of each frequency band, we opted to use EEG spectrograms. Spectrograms use a short-time Fourier transform to retain the frequency components of the signal and also the temporal information. The hope with this method was that there would be subtle patterns in the power of different frequencies over time between the creative tasks that a model could exploit to be able to distinguish between them.

3.1 Generating EEG Spectrograms

To evaluate the efficacy of this approach, all EEG signal was segmented into 10-second fragments. The pipeline for generating the final spectrograms is depicted in figure 4.

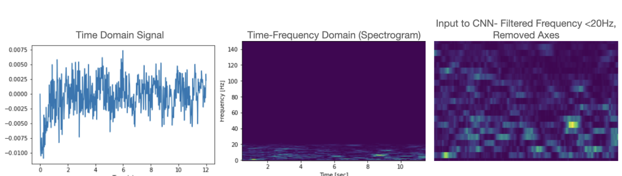

Figure 4. Spectrogram Generation Pipeline

As observed in figure 4, for each 10-second segment of raw EEG signal, a short-time Fourier transform is used to generate a spectrogram using the SciPy signal processing library [6]. Additionally, the spectrograms were filtered to only show activity under 20Hz, as that was where a majority of the activity was observed. Lastly, the axes and title were removed and the spectrograms were saved as a three-channel (red-green-blue) image to be passed into a convolutional neural network.

Since the spectrograms were passed to the models as three-channel images, the color palette used to generate the spectrograms and the resolution of the spectrograms will have an effect on a model's ability to extract features from the spectrograms. Higher classification accuracy was reached when spectrograms were generated with a more divergent color palette. Additionally, higher classification accuracy was observed with a lower resolution spectrogram. Figure 5 shows an example of a spectrogram generated with the parameters that yielded optimal classification accuracy.

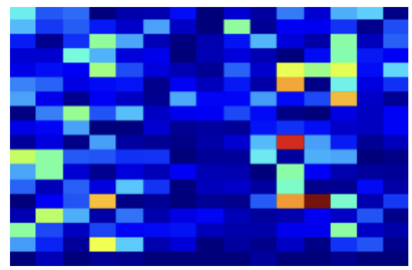

Figure 5. Spectrogram with Optimal Color Palette and Resolution

3.2 Convolutional Neural Networks (CNNs) for Classifying Creative Tasks

This approach frames the problem as an image classification problem for which convolutional neural networks are well-suited. Convolutional neural networks on EEG spectrograms have seen success in emotion recognition, motor imagery, mental workload, and seizure detection tasks [7]. However, this is typically done with spectrograms generated from much longer time periods of EEG signal, with some using hours of EEG signal.

Initially, it was hypothesized that it would be essential for the CNN models to learn a relationship between the activity in different frequency ranges and the different creative modes of thinking. For that reason, 1-Dimensional convolution and pooling layers were thought to be optimal as they would allow the frequency information, the y-axis of the spectrograms, to be retained. Although, it was quickly realized that this was not the case. With 2-dimensional convolution and pooling layers, the models experienced much more stable learning and reached far greater performance.

The CNN architecture that has reached the highest accuracy on unseen data has been a simplified model based off of the original VGG model [7]. The model consists of four convolutional layers with rectified linear unit (ReLU) activation, two max-pooling layers, a flatten layer, two fully-connected layers with ReLU activation, and an output layer with softmax activation. It is a relatively small model with less than two million trainable parameters. Figure 6 contains a diagram of the current model architecture.

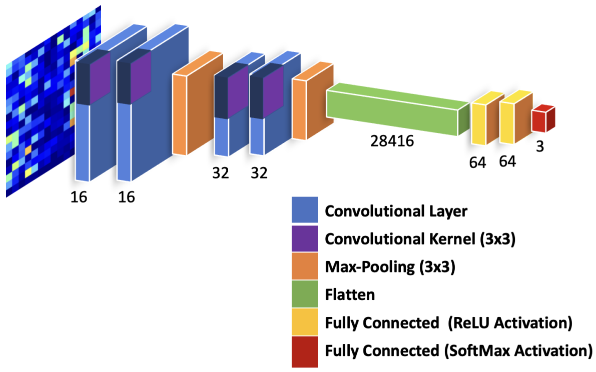

Figure 6. CNN Architecture

Heavy data augmentation, including vertical and horizontal flip and shift, were also found to be very useful in training a model that generalized well to unseen data. The learning curves of the described architecture with and without data augmentation are shown in figure 7.

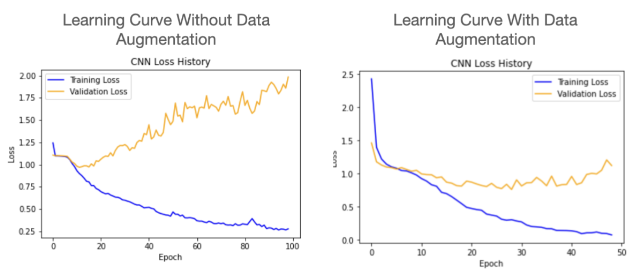

Figure 7. CNN Loss History With and Without Data Augmentation

3.3 Results of Approach Two

Using this simple CNN model architecture to classify creative mode of thinking from only 10-seconds of EEG signal in the form of a spectrogram, we were able to reach 82.33% accuracy on a completely unseen test set with 300 observations. Figure 8 shows the confusion matrix for model predictions on the test set.

Figure 8 demonstrates that using this method, a model can easily distinguish between all three creative modes of thinking from only 10-seconds of EEG signal. Additionally, it seems the model is extremely effective at classifying the idea evaluation state. This is reasonable as the idea evaluation phase of the experiment involved a more analytical task, while the idea evolution and idea generation phases consisted of very similar tasks involving drawing an image from a stimulus.

With that said, given the similarity between the tasks in the idea evolution and idea generation phases, it is surprising that a model is able to distinguish between them at this level of accuracy with only 10-seconds of signal.

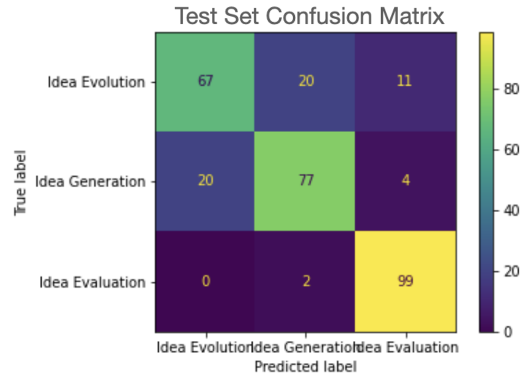

Figure 8. CNN Model Confusion Matrix on Unseen Test Set

3.3.1 Gradient-weighted Class Activation Mapping (GradCAM).

Given the success of this method and the general lack of interpretability associated with CNN models, a method was desired to understand how the model was making these classifications. Gradient-weighted class activation mapping is a tool that uses the gradient of each target class flowing into the final convolutional layer of a model to give a coarse localization of what areas of the image are causing the activations [8]. This can be done without altering the architecture of the model. Figure 9 shows examples of a class activation map for a spectrogram generated from EEG signal from each class. Each pairing shows the input spectrogram and the activation map of the correctly predicted class.

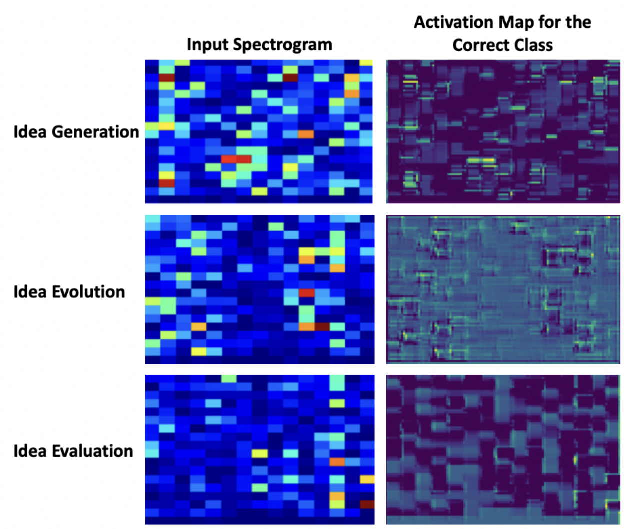

Figure 9. Example Input Spectrogram/Class Activation Maps Pairings From Each Class

Looking at the activation maps for spectrograms from each creative task, we can gain some understanding about how the model is making predictions. It appears the idea generation class may be activated by the high-power regions of spectrograms, while the idea evolution and evaluation classes seemed to be activated by different low-power, background regions of the spectrograms.

4 Future Research Directions

Given the initial success of this method for using EEG data to distinguish between very similar tasks, future efforts could be put toward a similar approach for more generalized EEG spectrogram classification. This could involve training a much more complicated CNN model to classify between a far more diverse set of tasks. Similar to other large deep learning models, this would likely allow for transfer learning for faster, more accurate performance on a specific task with less training time and data.

Future studies could also be done on the latent representations produced by these models processing EEG spectrograms to see if the relationship between latent representations could be attributed to high-level information about the different activities. For example, could clusters in the latent representation of different activities represent, for example, creative, analytical, and play-related types of tasks?

5 Discussion

In this study, two approaches were explored for using EEG signal to infer the creative task being completed. We saw that with the manual extraction of features that were demonstrated to be useful by subject matter experts, we were not able to reliably distinguish between the three creative tasks. However, by simply changing the format of the signal and framing the task as an image classification problem with EEG spectrograms, we were able to identify what creative task was being completed with over 82% accuracy using only a 10-second segment of EEG signal.

While the manually extracted features are certainly helpful in understanding the neural activity involved for different creativity-related tasks, the findings of this study present a familiar theme. That is, while we have a tendency to look for specific features that neatly explain different phenomena, even if these features are useful, we lose a lot of information by reducing our original data to a small subset of features. Many machine learning applications seem to perform well using the greatest amount of information. By transforming the raw EEG signal to a spectrogram representation, we were able to retain all of the information about the original signal and re-frame the problem as one that we know deep learning methods are well-suited for: image classification. Findings like this may also provide intuition about the information that the brain uses to perform similar tasks. If models are more effective with these information-rich representations of data, it is likely a better approximation of related neural processes as well.

References

1. Jia, W., Zeng, Y. EEG signals respond differently to idea generation, idea evolution and evaluation in a loosely controlled creativity experiment. Sci Rep 11, 2119 (2021). https://doi.org/10.1038/s41598-021-81655-0
2. Rashid, M., Sulaiman, N., P P Abdul Majeed, A., Musa, R. M., Ab Nasir, A. F., Bari, B. S., Khatun, S. (2020). Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. Frontiers in neurorobotics, 14, 25. https://doi.org/-10.3389/fnbot.2020.00025
3. Michel, C.M., Koenig, T. (2017). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. Neuroimage, 180, 577–593.
4. Wmvanvliet. (2018). Wmvanvliet/mne-microstates: Microstate Analysis for use with MNE-python. GitHub., https://github.com/wmvanvliet/mne-microstates
5. Scipy. Fundamental algorithms for scientific computing in Python. (2022), https://scipy.org
6. Signal processing (scipy.signal) - SciPy v1.8.0 Manual. (2022)., https://docs.scipy.org/doc/scipy/reference/signal.html
7. Craik, A., He, Y., Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. Journal of neural engineering, 16(3), 031001. https://doi.org/10.1088/1741-2552/ab0ab5
8. Simonyan, K., Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556.
9. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.