Miasya Bulger
MAIS 202
February 20th, 2019

Project Deliverable 2:
EEG Classifier Preliminary Results

Problem Statement

Originally, to combine my interest in biology and computer science, I selected an EEG dataset comprised of motor control data corresponding to grasp, lift, drop, and intermediate actions. However, after the early stages of preprocessing, I discovered that the dataset provided was extremely large and beyond my desired level of complexity for this project given the timeframe and my current rate of skill development. As a result, I migrated my attention towards a simpler dataset also related to EEG signal interpretation; EEG emotion classification, which is described in more detail below.

Data Preprocessing

i. Original Processing

The dataset selected, "EEG Brainwave Dataset: Feeling Emotions" was originally uploaded on Kaggle in December 2018 by Jordan J. Bird, PhD candidate of Aston University.[1] It contains raw EEG data measured from the TP-9, AF-7, AF-8, and TP-10 electrodes of two subjects, male and female, while watching various emotion-provoking clips (e.g. "Marley and Me" death scene). This data was split into time windows to reduce the impact of the noisy nature of biological data. In addition, features related to statistical methods (such as mean, standard deviation, min, and max) as well as processed features (such as fast Fourier transforms and eigenvalues) in order to create a total of 2131 points each containing 2547 features each. Classes for each of the respective datasets, titled NEGATIVE, NEUTRAL, or POSITIVE, were also assigned as appropriate.

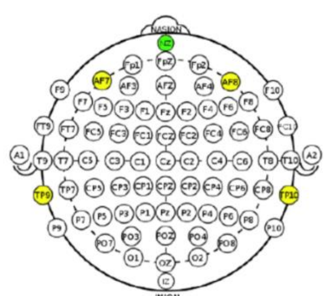

Figure 1. TP-9, AF-7, AF-8, TP-10 electrodes on the international standard EEG placement system. R. Rojas, *"Adaboost and the super bowl of classifiers a tutorial introduction to adaptive boosting*," Freie University, Berlin, Tech. Rep, 2009

[1] J. J. Bird, A. Ekart, C. D. Buckingham, and D. R. Faria, *"Mental emotional sentiment classification with an eeg-based brain-machine interface,"* in The International Conference on Digital Image and Signal Processing (DISP'19), Springer, 2019.

Miasya Bulger
MAIS 202
February 20th, 2019

ii. Additional Processing

Additional processing upon acquiring the dataset was limited and was reserved to separating the features and performing a train, test, validation split. A 60-20-20 split was selected to ensure enough datapoints existed in the test and validation sets.

Link to dataset: https://www.kaggle.com/birdy654/eeg-brainwave-dataset-feeling-emotions

Machine Learning Model

i. Current Models

Random Forest Classifier

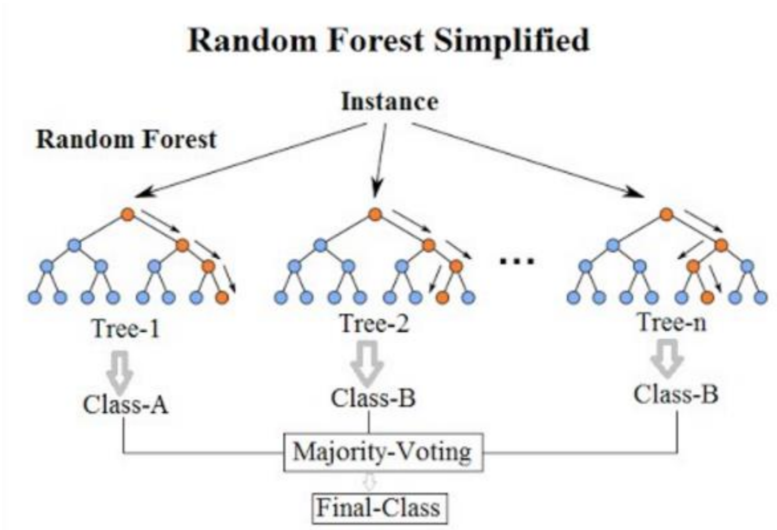

Figure 2. Random Forest Diagram. Koehrsen, Will. *"Random Forest Simple Explanation."* From: Medium. https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d. (accessed February 20, 2019).

Before diving into any complex, difficult-to-implement models, I implemented a Random Forest Classifier from Scikit-learn. After fitting my model to my training dataset, I remarked that my accuracy values against my validation set immediately ranged from 95-98% using 100 estimators and a 3 min sample split. In all honesty, these outputs, while consistent, raised immediate suspicions as to the integrity of my model. I spent the next several hours scouring through the dataset selectively picking and choosing features to make sure there were no class "giveaways" hidden in the set. I also verified that there was no class imbalance (715 NEUTRAL, 708 NEGATIVE, 708 POSITIVE). However, upon the discovery of paper based on this dataset, I realized I had obtained very similar results to those published by James J. Bird. More interestingly, this type of model had actually been the highest performing model among those tested. This new information eased my concern and I later evaluated the performance of the Random Forest Classifier with various estimators and max features.

Miasya Bulger
MAIS 202
February 20th, 2019

ii. Future Models

Support Vector Classifier (SVC)

My experience with the Random Forest Classifier prompted the question, "which features are most important in determining emotional classification?" This question strikes me as important because answer not only satisfies a curiosity, but potentially withholds information regarding the most effective strategy for potential future healthcare technologies requiring the monitoring of EEG waves for diagnostic interpretation. Naturally, the next step in answering this question therefore relates to dimensionality reduction. Among various possible options, an SVC was selected to find the ideal separating hyperplane through the data. More specifically, I'm looking into potentially using the Radial Basis Function Kernel (RBF) because our data is very complex and requires an equally complex kernel. However, this is still to be determined and I'm seeking the advice of more experienced students before going forward in this direction.

Neural Net

In the cited paper, classification was also performed via fully-connected feed forward deep neural net with an accuracy 94%.[2] I would personally like to compare and see if a different type of neural net could perform at the same standard as other models or potentially outperform them. Also, neural nets of all types have been found to be useful for a variety of applications, and I personally would love an excuse to further develop this skill.

Preliminary Results

i. Validation Accuracy vs Max Features (RFC)

A Random Forest Classifier validation accuracy plot by max number of features is shown below. In this trial, the number of estimators and min sample split were kept constant at 100 and 3. The training accuracy plot is omitted because it scored consistently approximately 100% and did not vary. The validation accuracy plot provides two notable insights. Firstly, the minimum number of max features (i.e. one feature) allowed the model to automatically score approximately 92.4%, which allows us to hypothesize that a singular feature among our 2547 is incredibly significant. Secondly, we note that after approximately 25 features, the validation accuracy fails to change, so we can select 25 as our max features for all future runs.

[2] Ibid.

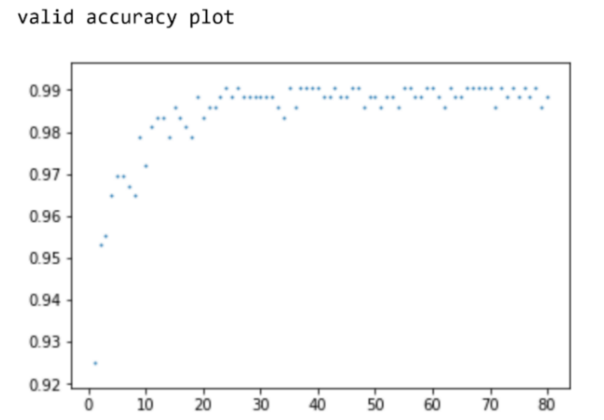

Figure 3. Validation accuracy as a function of max features plot generated using mathplotlib.

i. Confusion Matrix using 25 Max Features (RCF)

On the confusion matrix, the most errors arose at (2, 0), which signifies that positive experiences were falsely classified as negative (0, 1, 2 corresponding to negative, neutral, and positive respectively). The converse was also observable. I found these results to be extremely fascinating, because what I had previously considered as opposite emotions were getting confused. Earlier, I had personally wondered why we were approaching this task as a classification problem as opposed to a regression problem, because one belief is that emotions can be considered as a spectrum, as opposed to discrete categories. In this study, the classification approach was a reasonable method both due to the means of production of the data (using sad vs happy film clips as stimulus), and the limitations of the degree of which one can truly know how an individual subject is feeling.

```
Valid accuracy:  0.9859154929577465
Confusion Matrix:

<matplotlib.axes._subplots.AxesSubplot at 0x159b031bb70>
```

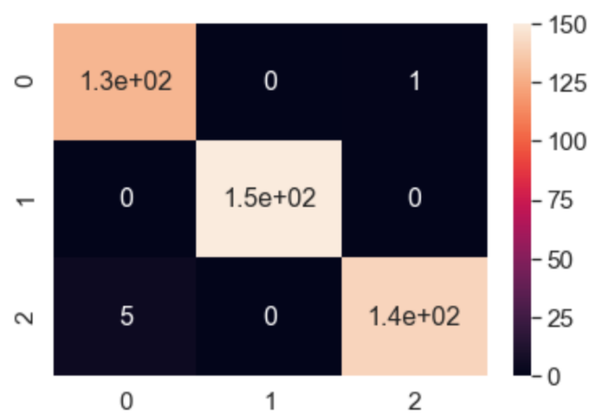

Figure 4. RFC confusion matrix created using scikit-learn and seaborn.

Miasya Bulger
MAIS 202
February 20th, 2019

Going forward, possible explanations of the confusion matrix results could be attributed to the staying power of emotional experiences, a potential similar degree of intensity of emotion, or due to the natural wandering of the mind in human cognition. Limited means to discern these possible explanations force us to declare the source of error as inconclusive, but we can leave open as a topic for further contemplation.

Next Steps

i. New Models

As discussed in the “Machine Learning Models” section of this project progress report, I am curious about potentially experimenting with other machine learning models, specifically one wherein I could isolate the most important feature(s) using my own methods or create a previously untested model.

ii. Future Idea Proposals

<u>Acquisition of Original Data</u>

One underlying sentiment I experienced during the early phases of this project was that my work exclusively served to solidify the ideas expressed during the cited paper, verifying them only in slightly different methods. An approach I can use to remediate this would be to produce original data in a similar fashion with similar external stimuli but a fresh hypothesis.

<u>Back to the Drawing Board</u>

Alternatively, I can seek an alternative dataset related to signals produced by the human physiology. In this case, I would need to make sure to change gears immediately, for the sake of avoiding lost time due to an extended sense of investment into a sub-optimal project. Proposed ideas include: the grasp-lift EEG dataset (investigated earlier), the image classification of cells using a convolutional neural net (CNN), or the analysis of other publicly available datasets.