Predicting Subject Activity from EEG Using Machine Learning

Seran Chandrapalan

Ryerson University

July 27th, 2020

**Abstract**

Electroencephalography (EEG) is an increasingly common non-invasive method of tracking and recording brainwave patterns. In recent years, it has been making its way into consumer grade forms that produce progressively more consistent and reliable readings. This dataset is the result of employing one of these consumer grade devices, the Neurosky Mindwave Mobile, to measure the electrical activity of 30 participants in response to one of two 5 min presentations during which they are exposed to a variety of audio-visual stimuli manifesting in the form of activities. The result is a dataset consisting of nearly 30,000 rows and 20 columns with each row being the result of the condensation of roughly one second of raw EEG readings. This paper will attempt to classify and predict the activity of a participant based solely on the readings produced by this cost effective single-sensor (FP1 position) headset. It will be analyzed on both a within-subject and between-subject level due to the well documented interpersonal variability issues known to significantly reduce the generalizability of these devices. This paper will utilize various classification techniques including but not limited to logistic regression, random forest and K-Nearest Neighbours using Python in pursuit of accurately predicting participant activity. The commercial headset industry has progressed rapidly over the past decade and this paper seeks to quantify the extent of this improvement.

I.

INTRODUCTION

Electroencephalogram (EEG) have been a major part of neuroscience and psychology for decades. Traditionally, EEG measuring devices were large, bulky and expensive only used in academic or hospital settings. However nowadays with the seemingly exponential progress of technology we are seeing the commercialization of these devices for the average consumer. This new wave of consumer grade hardware is usually fitted with 3 or less electrodes (as opposed to the 32+ electrode medical-grade devices), are minimally invasive and are reaching sub 500-dollar price point. Additionally, these devices are becoming more and more accurate with their limited electrodes to the point where many of these devices are packaged with very simple virtual games that can be played with just the use of these devices. The idea being that these EEG headsets will be the foundation for brain computer interface technology in the future. This paper will examine the Mindwave Mobile, developed by Neurosky, a major company in this market. This headset boasts just one sensor on the forehead and is being sold at the 100 USD price point making it one of the cheapest devices on the market. The goal of this paper is to use classification analysis to quantify the accuracy of this consumer grade and determine its level of quality and reliability over the different tasks.

The objective of this study can be summarized as below:

*Research Question: Can we use machine learning models to accurately and reliably classify subject actions solely based on their EEG readings?*

The Neurosky Mindwave Mobile device that was used to collect the data in this dataset outputs both raw EEG readings and specific readings of wavelengths known to be important in correlating with various activities. These waves have been known to correspond with tasks such as sleeping, relaxing, intense concentration or even in more specific tasks such as recognizing a mistake.

This dataset is filled with data labelled by the second as subjects perform a variety of tasks. We are then going to use the eight main wavelengths outputted by the device to see if we can predict the activity a participant is performing at any point in time. This will involve using machine learning models such as K-Nearest Neighbours, Random Forest and Logistic Regression to determine.

The scope of this project will fall under classification and will ideally help realize the accuracy of brainwave analysis and its potential future uses.

There will also be a significant look at the differences in model prediction between al subjects and within subjects. EEG readings are known to be volatile and significantly change not only from person to person but over time in the same person as well. This is one of the main factors holding back the advances of EEG technology and a problem that must be solved before we can move onto further horizons.

Currently, there are a couple popular methods being used to account for this problem. In this study we will be looking at the method that aims to achieve a new baseline every time the device is adorned. In this method, the device will automatically achieve a baseline EEG level when it is worn however this baseline usually takes anywhere from 5-10 minutes. As a result, there are now different activities being developed that aim to help the devices achieve a baseline as quickly as possible. This is where this dataset comes into play.

Originally, this dataset was used as part of a larger study to investigate the efficacy of such activities. In this activity the subjects go through a 5-minute session in which they perform a variety of contrasting actions which should theoretically produce varying levels of EEG activity in the brain.

III.

LITERATURE REVIEW

The literature in this field is extensive and continually updating and improving every year. The original paper from which this dataset was sourced investigated the efficacy of mental gestures1. The variety of mental gestures performed in this experiment was to quickly and efficiently calibrate consumer grade BCI sensors. The authors show that within 5 minutes (the length of the stimulus) 75% of users were calibrated to the Neurosky Mindwave and within 30 minutes 100% of users were calibrated to the headset. The authors conclude that this specific method of calibration provides significantly faster calibration times than traditional BCI calibration.

A pioneering paper by Vidaurre & Blankertz2 from 2010 first outlined the so called ‘BCI illiteracy’ problem. This paper seeks to understand and solve the problem of BCI Illiteracy which is defined as the inability of BCI control to work in an estimated 15-30% of users. They used machine learning techniques to eventually lower BCI calibration down to 3-6 minutes. Unlike our dataset, this paper does not use a single sensor, consumer grade EEG. This paper was one of the first to mention the idea of machine learning or ‘co-adaptive calibration’ rather than the traditional offline calibration that takes about 30 minutes before the start of the experiment.

Building upon this paper was Grierson & Kiefer 3 who took an early look at consumer grade devices and event related potential (P300). They use an older version of the Neurosky Mindset to examine its sensitivity to ERP and whether they could accurately predict the presence of an ERP. The experiment flashed a target red square at participants wearing the headset, which should theoretically elicit an ERP and simply classified the accuracy of the readings. They found that they could predict with 100% accuracy the presence of the red square solely based on a subject’s EEG readings.

Of course, BCI is not only used to predict human activity it has a wide range of uses including but not limited to the control of machines and robots. Larson (2011)4 took a closer look at the Neurosky Mindset and aimed to create a simple snake game solely controlled by the BCI. During the process of developing this game, the paper goes into detail about the different wavelengths and their reactions to different activities similar to our dataset, namely meditation and blinking. The result of this report was the development of an artificial neural network that could play the snake game for almost 7 minutes at an accuracy of 98%. The paper also goes into detail about the differences in performance when the person is in different situations such as focused or relaxed. This is something that the current paper will also touch on.

The field of BCI is burgeoning and will only continue growing as new companies enter the market and the BCI technology becomes more intuitive and consumer friendly. However, before that happens there are clearly critical issues surrounding the clunky calibration of these machines,

II.

DATA

This dataset was collected by the Master’s in Data Science (MIDS) class at the UC Berkeley School of Information over the course of one day. In this experiment two groups of 15 subjects were given two similar stimuli in the form of a five-minute video. Each group saw their respective stimuli simultaneously with their 15-person group and their EEG activity was measured for this experiment. During these videos, they were exposed to and asked to perform several tasks for which their brainwaves were recorded at one second intervals by the MindWave headset. The tasks in the two videos ordered chronologically are as follows: blinking task, relaxation task, instruction task, mental mathematics, listening to music, watching a Doritos super bowl advertisement (genre being comedy), thinking of items task and a colour counting task. Both groups performed the same tasks with the only differences being that the specifics of the task varied slightly, such as different math questions or listening to different music. The entire experiment took place over the course of an hour and all data from this experiment was outputted solely by the MindWave device which includes both raw data and programmatically calculated data such as their proprietary attention and relaxation sensor data. This data has also been preprocessed by the MIDS class to a certain extent in that the activity labels have already been synchronized with browser latency and reading times to ensure accurate activity labelling for each row in the dataset.

This dataset consists of three tables – EEG data, stimulus times and subject metadata. The bulk of the information lies in the EEG data table which consists of about 30,000 rows of data. Each row of data corresponds to one second of EEG recordings and captures a variety of information outputted by the Neurosky Mindwave Mobile device. The last column in each row is our classification label which describes the activity being performed by the subject during this moment of the recording. Again, each packet of data is spaced roughly one second apart although in practice the spacing was determined by the 512Hz frequency of the MindWave. This results in minute differences in spacing between rows on the order of milliseconds

The second table shows us the synchronized times of each activity. This table was not used during this analysis because the preprocessing had already been taken care of by the MIDS class.

The third and final table presents us with subject metadata. After participating in the experiment, the subjects were given a questionnaire to fill out which included the following questions: Have you seen the video before? What colour did you choose during the colour counting exercise? Did you see the hidden icon displayed during the colour counting exercise? What is your gender? Do you wear contacts?

These questions were chosen based on them possibly having some sort of effect on their EEG wave levels and possibly performance in the exercise.

IV.

METHODS

Given the nature of the data and its origins much of the difficult calibration was already finished. EEG readings change very quickly with many spikes in waves happening over a couple hundred milliseconds or less. Certain spikes also occur with a certain predictable delay after the onset of an event. In general, in an experiment such as this one, after collecting the data each row would have to be labelled correctly while accounting for various factors such as delay in packet reading from the device, delay in browser latency and general delay in the processing software. Fortunately, with the help of the MIDS class and the MindWave all of these calibration issues have been resolved.

However, there was still a significant amount of data cleaning and preprocessing that needed to be done before analysis. The general approach of the entire analysis has been outlined below.

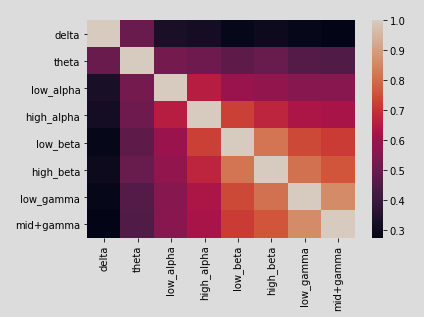
Step 1: Cleaning and Preprocessing

The first step in the analysis was to clean the data and remove unneeded information from the EEG data table. First order of business will be to remove all the redundant date time columns (excluding ‘indra\_time’) because all the time synchronization calculations have already been performed by the dataset creators. Next, we will be removing the rows in the table with a signal quality of 100 or higher because these packets of data provide next to no value (a lower signal quality is better). Finally, we removed all unlabelled data in the table; these rows of data were collected at a time when the participant is not taking part in the experiment. The experiment was set up in a manner where everyone was fitted with the headset at the same time, then group 1 watched stimulus 1 while group 2 was waiting outside the room and vice versa. This results in a fair amount of unlabelled data that we cannot and do not want to use. In addition, there is a baseline setting activity presented at the beginning of each stimulus so establishing a baseline will not be a problem going forward even with the removal of this data. Next, we separated the ‘EEG power’ column into 8 different columns corresponding to each of the wavelengths represented by each item in the array. This will be important later when looking for changes in individual wavelengths during specific tasks, as well as monitoring certain wavelengths. Lastly, all activity labels were then grouped together into a higher-level label, for example, ‘math1’ and ‘math2’ were grouped into a label called math. This is to decrease the number of classification labels to classify and hopefully increase the strength of our classifiers. Finally after all of this cleaning we merged the main EEG dataset with the subject metadata dataset to give us one Master Dataset with which the upcoming analyses can be run upon.

Step 2: Exploratory Data Analysis

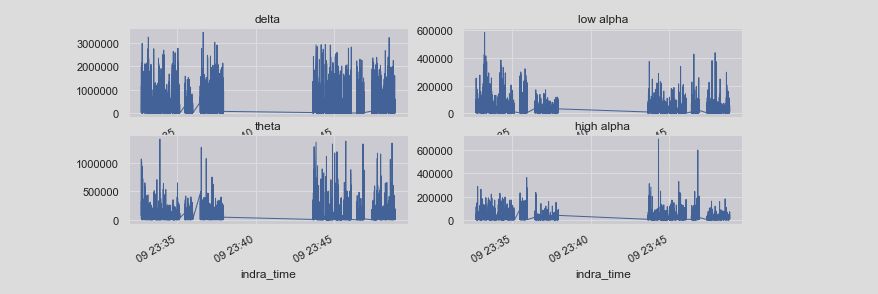
Next we performed some simple exploratory data analysis on our dataset to determine the shape and size of the data. We also looked at the correlation between the eight different wavelengths to make sure there was no heavy correlation occurring. We did found varying levels of correlation between wavelengths but this is to be expected with EEG wavelengths especially since they have been divided into subsections of wavelengths (aka high beta vs low beta). It is common practice to leave correlating waves alone when building machine learning models with EEG4. However, if computing problems do present in the future, we will not hesitate to remove these wavelengths.

On the other hand, the delta and theta wavelengths we particularly free of correlation with others. This is also to be expected seeing as the operate on a completely different frequency from the others.

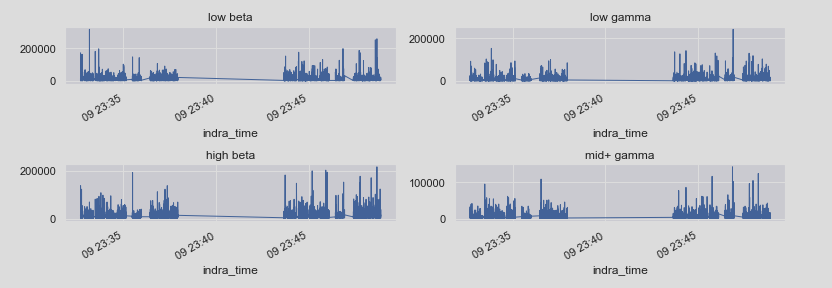


*Figure 1. Correlation Analysis between Wavelengths.*

Next, we took a visual look at the different wavelengths and the frequencies they operate at using time series graphs.



*Figure 2. Wavelength Time Series*



*Figure 3. Wavelength Time Series(Part 2)*

We can see here that the correlation chart is reflective of the time series graphs. Again, we see that delta and theta wavelengths are the most different.

Step 3: Initial Results

We performed a KNN classification on the dataset without splitting the subjects into groups of their own. This resulted in a surprisingly adequate accuracy in the high 60s however upon further inspection it was found that most of the accuracy was in correctly predicting when a datapoint was ‘unlabelled’ or not in the experiment. This was an obvious oversight and so the unlabelled data was also removed.

Furthermore, the classification of other labels was very poor so it was decided that we would consolidate all labels of the same activity into one group. For example, there were 12 math variables, so we consolidated that into on large math label. Once this was finished, we were left with 6 different labels as follows: blink, relax, math, video, colour and think of items. These are the six labels we will be working with for the rest of the analysis.

We then performed the KNN analysis once more and this time we were met with dismal results, an accuracy of about 30%. This obviously will not suffice, so we now decided to investigate the generalization problem with EEGs.

Step 4: Between vs Within Subject Analysis

Our next steps were to start dividing our data by subjects to see if we could achieve better results by looking within-subject. Our first results were not promising with accuracy’s sitting in that same range. It is important to note that our accuracy was marginally better than expected at random but far from what has been expressed in other studies.

Step 5: Math vs Relax

Finally, we decided to look just into the Math vs Relaxation aspect of the activities. It is a well-known phenomenon that EEG waves are best at distinguishing between wakefulness and sleep so in theory this should be the easiest classification to perfect. In fact, the Neurosky MindWave has a native reading that measures this aspect, respectively called ‘attention esense’ and ‘meditation esense.’

Step 6: Results and Analysis

In this part, we decided to bring in the other classifiers Random Forest and Logistic Regression to see if we could get measurably better results than those with the KNN. This will be expanded upon in this upcoming section.

V.

RESULTS

The final results of the experiment were not up to what was expected. It is difficult to say what may have been the underlying problem, but it made classification very difficult and cumbersome.

We will first start with K Nearest neighbours, when we first started analyzing between-subjects data, it gave us about 30% accuracy, we than added cross validation to look for any differences but no significant changes were made to the accuracy of the model. Keep in mind that this over 6 classification variables so it was still faring better than random.

We then moved onto within-subject classification and fared a little better but still hovered in the 30s. This shows some very weak proof that the within-subject classification fares better than the between-subject classification which would stay consistent with other literature in the field.

Once within-subject did not yield the results needed we explored the possibility of reducing the complexity of the classification and focusing solely on Math vs Relaxation categories in the hopes of salvaging some sort of result.

Fortunately, this did help quite a bit and we found that when classifying between these two labels our classification jumped up to mid 60s with similar increases in recall and precision. However, we must note that the data were working with was shrinking in size and these results may not be as robust compared to a between-subjects analysis.

Finally, once we got fairly satisfactory results, we decided to try other machine learning models, Logistic Regression and Random Forest.

The results here were very similar with Random Forest performing best as expected with a peak accuracy of about 70%. This is quite an improvement from our preliminary results!

VIII.

CONCLUSION

In conclusion, the analysis performed on the dataset did not give us significant results. When classifying all labels we achieved an accuracy just hovering in the mid 30s mark with certain labels being easier to classify than others, such as the colour counting task. Although this is better than random it is far from the potential results that we could have achieved. Once we moved our focus into looking solely at the Math vs Relax we fared a bit better. At this point we started getting accuracy at about the 70% mark which is significantly better than random. Although this is an improvement there is still much to go as others have achieved results in high 90s when working with similar data.

One possible hinderance in this experiment was that we had to split the dataset up into smaller and smaller chunks as the analysis progressed. By the end we were getting better classification results but our individual data for each subject was also miniscule. Therefore, it may have benefitted to have used a much stronger classifier such as SVM or some Linear Discriminant Analysis. Unfortunately, due to the situation at hand we were not able to get access to a powerful enough computer to handle such workloads.

One other technique to consider for the future would be to investigate using the raw data directly. It is a technique that has been discussed and used but involves much deeper understanding of EEG wavelengths and the neuronal properties behind them, such as their exponential or logarithmic firing nature. Nonetheless, it is a future experiment that may yield better results.

APPENDIX

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Column Name | Data Type | Data Format | Description | Example |  |
| ID | Integer |  | Unique ID for every row | 3730 |  |
| Subject Number | Integer |  | Subject ID | 12 |  |
| Indra time | Date Time | yyyy-mm-dd hh:mm:ss.ss | Time of sent data packet | 2015-05-09 23:13:42.281 |  |
| Browser Latency | Integer |  | Delay in Browser | 1461 |  |
| Reading Time | Date Time | yyyy-mm-dd hh:mm:ss.ss | Time of Reading | 2015-05-09 16:13:40.954 |  |
| Attention Esense | Integer |  | Aggregated Attention Value (Rating from 0 to 100) | 0 |  |
| Meditation Esense | Integer |  | Aggregated Meditation Value (0-100) | 0 |  |
| EEG Power | Array | [delta, theta, low-alpha, high-alpha, low-beta, high-beta, low-gamma, mid-gamma] | Array of Common EEG wavelength values | [944412.0, 111373.0, 52404.0, 28390.0, 3237.0, 32728.0, 4845.0, 2036.0] |  |
| Raw Values | Array | [value1, value2, …,  value512] | Array of Raw packet values (512Hz) | [-203.0, -202.0, -196.0, -185.0,  …,  145.0,] |  |
| Signal Quality | Integer |  | Signal quality ranging from 0-200 (lower value is better quality) | 200 |  |
| Created At | Date Time | yyyy-mm-dd hh:mm:ss.ss | Creation Time | 2015-05-09 23:13:39.550 |  |
| Updated At | Date Time | yyyy-mm-dd hh:mm:ss.ss | Update Time | 2015-05-09 23:13:39.549 |  |
| Label | String  (factor) | From event name in Stimulus Times table | Activity Label (from event name column) | unlabeled |  |

REFERENCES

Merrill, N., Maillart, T., Johnson, B., & Chuang, J. C. I. (2015, February). Improving Physiological Signal Classification Using Logarithmic Quantization and a Progressive Calibration Technique. In *PhyCS* (pp. 44-51).

Vidaurre, C., & Blankertz, B. (2010). Towards a cure for BCI illiteracy. *Brain topography*, *23*(2), 194–198. https://doi.org/10.1007/s10548-009-0121-6

Grierson, M., & Kiefer, C. (2011). Better brain interfacing for the masses: progress in event-related potential detection using commercial brain computer interfaces. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 1681-1686).

Larsen, E. A. (2011). *Classification of EEG signals in a brain-computer interface system* (Master's thesis, Institutt for datateknikk og informasjonsvitenskap).

Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces, a review. *Sensors (Basel)*. 2012;12(2):1211–1279. doi:10.3390/s120201211