Grasp-and-Lift EEG Detection

Aaraadhya Narra
M.S. Computer Science & Engineering
Santa Clara University
Santa Clara, CA 95053
anarra@scu.edu

Bhargav Pabbisetti
M.S. Computer Science & Engineering
Santa Clara University
Santa Clara, CA 95053
bpabbisetti@scu.edu

Abstract

This project focuses on understanding the relationships between EEG signals and hand motions by using Pattern Recognition and Data Mining techniques.

1 Introduction

1.1 Motivation

Patients without arms wake up to harsh reality everyday. They can't do some of the most basic tasks like making coffee, drinking or locking the door. BCI(Brain Computer Interface) prosthetic device would greatly increase their independence and quality of life

"EEG is an electrophysiological monitoring method to record electrical activity of the brain." (Electroencephalography) So, these EEG Recordings also monitor the mental process before doing a particular hand movement. Because human beings can control the EEG signals through imaginary mental tasks, we could use the EEG data to predict the hand movement. *"EEG-based BMI has benefited not only healthy people but also patients suffering from severe motor impairments or disabilities, such as amyotrophic lateral sclerosis (ALS), brainstem stroke, cerebral palsy, and numerous other diseases of which the more traditional methods cannot be used because only cognitive function remains intact."* (Phothisonothai & Nakagawa). By using EEG based non-invasive BCI devices, we could provide an affordable, low-risk option for these patients to control the external prosthetics with their brain activity.

1.2 Problem Statement

Providing affordable, low-risk, non-invasive BCI devices is dependent on further advancements in interpreting EEG(Electroencephalography) signals. Better understanding the relationship between EEG signals and hand movement is critical to give individuals with neurological disabilities to navigate through this world. Our goal is to identify when a hand is grasping lifting and replacing an object using the EEG data that was taken from healthy subjects as they performed these activities.

1.3 Relevant work

https://hal.inria.fr/hal-01055103/file/lotte_EEGSignalProcessing.pdf

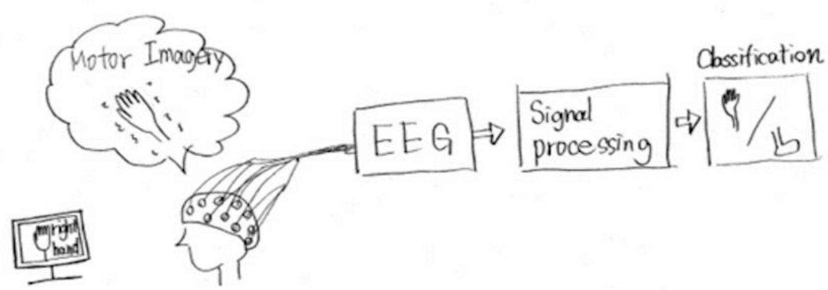
http://sccn.ucsd.edu/wiki/Introduction_To_Modern_Brain-Computer_Interface_Design

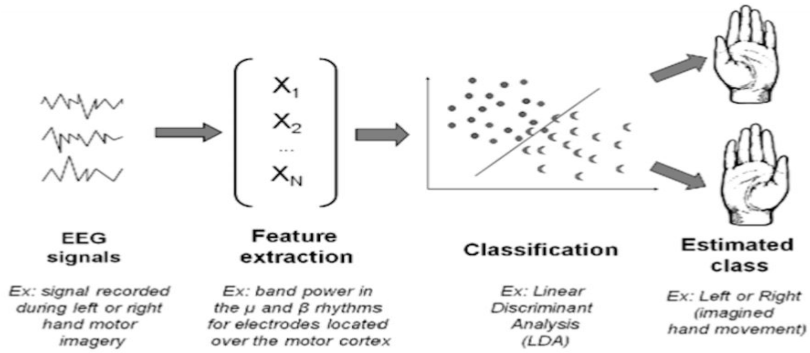
2 Data Set

Data contains EEG recordings of subjects performing grasp and lift(GAL) trails. There are 12 subjects in total, 10 series of trials for each subject. Test set contains the 9th and 10th series. Total test data is around 400mb, consisting of different subjects/series and training data is around 2GB with all subjects/series. Data from each subject/series is around 19mb. For each GAL, our task is to detect 6 different events(hand start, liftoff, replace, both released, first digit touch and both start load phase).

Each series data consist of 32 attributes and these attributes correspond to different EEG channels. First, we eliminated columns that weren't in the interest of computation and then we had to transform and merge the original data using python machine learning libraries. Transforming(columns to RawArray) was necessary to do computations. Merging was necessary because we had to combine all 12 subjects with 8 series each in the training set. We will need do this for test set also which contains 12 subjects with series 9 and 10 for each subject. We also have butterworth bandpass filter implemented to filter to obtain signal of interest.

3 Preprocessing

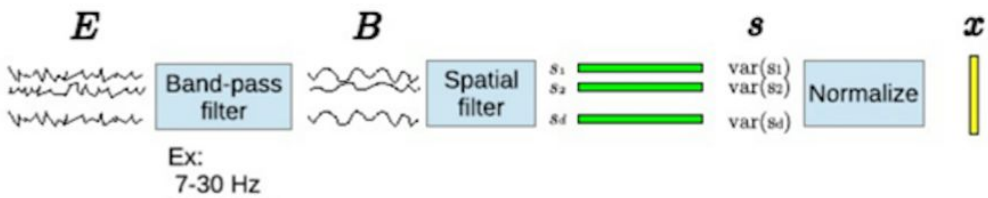




Two pictures above show overall view of how we will be able to predict appropriate classes(hand motions) from EEG signals. Several major steps need to be followed to successfully predict the outcome.

<u>EEG Signals to Feature extraction</u>

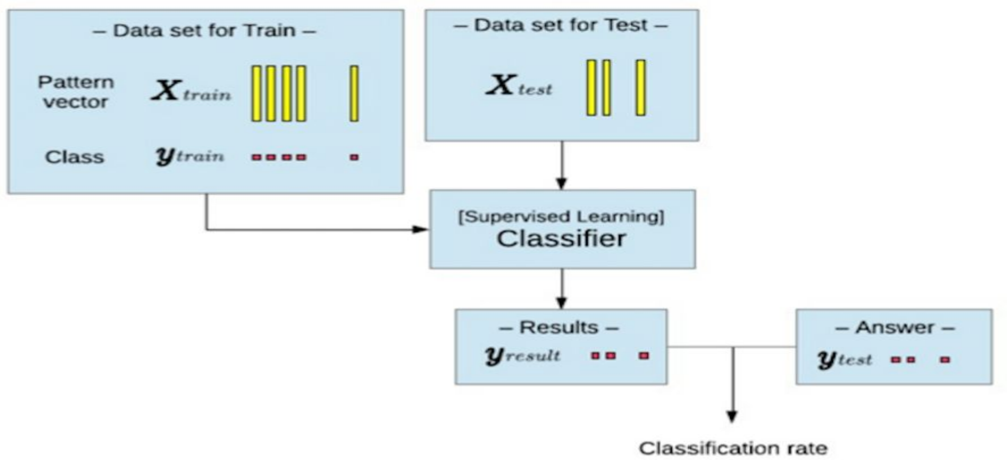
Preprocessing



- After transforming and merging original data, filter it through butterworth bandpass filter, frequencies between 7Hz and 30Hz. Main reason why we do this is because certain band of frequencies is best for classification. EEG signals are also very noisy so noise reduction needs to be done to do proper data analysis.
- We used PCA as our dimensionality reduction algorithm. The reason why we need to do this is to filter out channels that aren’t important for classification.

Feature extraction to Classification

Classification



4 Models Used

We used logistic regression without any data preprocessing for our base model. We included Band-pass filtering and PCA to preprocess our model and then applied logistic regression to get better results.

4.1 PCA instead of CSP

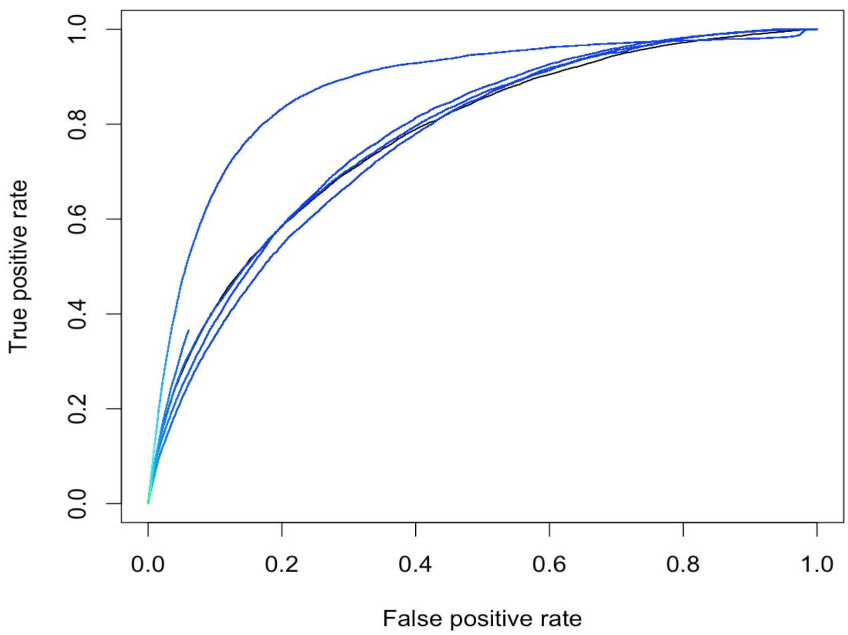
- We have used PCA(Principal Component Analysis) instead of CSP algorithm to do dimensionality reduction. Initially we wanted to use CSP because it is commonly used in EEG data preprocessing, however due to steep learning curve and complexity, we didn't use CSP. We were more familiar with PCA feature extraction and we have successfully implemented PCA to extract important channels to do predictive analysis.

4.2 Logistic Regression and Results

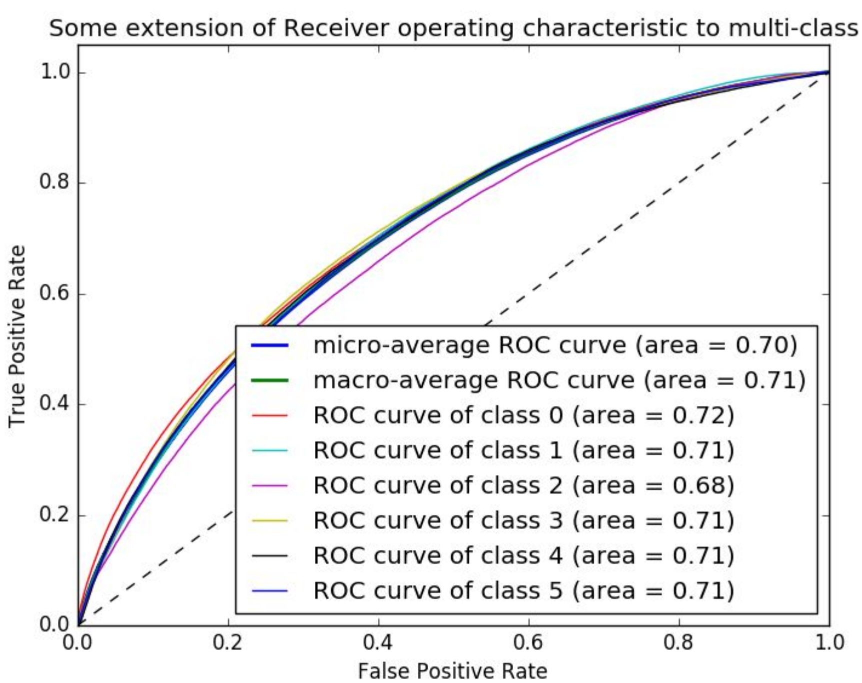
We split up our training set into a training and cross validation set. We did a 75% split on the column HandStart. We did this so that we can compare our model to the baseline.

This is the graph of our ROC Curve. We selected a random test subject and plotted it against our cross validation set.

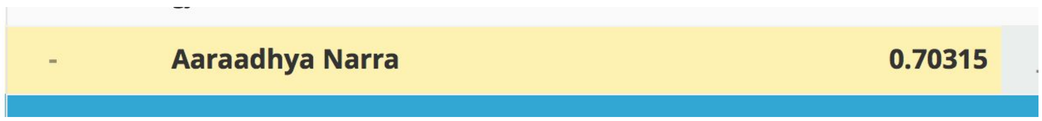
R Graph



Python Graph



Once we finished training and testing our data, we uploaded our results to kaggle to compare it to our baseline. The Baseline AUC was 0.65. The AUC we got was 0.70315. Below is the screenshot from our result.



5 Lessons learned

- Domain knowledge and preprocessing was the hardest part of this project. We needed to do full research on EEG signal analysis. We referred several research papers that gave us proper guidance to start our project. Reading up on domain knowledge helped us understand how EEG signals work, and which areas of the signals are significant for our particular project.
- During preprocessing part, we had to do signal processing. We have done research on how low pass, high pass and band pass filters work. We also didn't have any prior technical knowledge on any of these filters. This helped us get an insight of how digital filters work. We have learned and reviewed range of topics from brief introduction to sampling theorem to significance of FIR and IIR filters. Fourier Transform is key to filtering out raw data. This knowledge helped us look at right frequencies of the signal.
- Preprocessing is the hardest step and it is vital to do it properly because the entire outcome will be useless if preprocessing goes wrong. We spent a lot of time just reading up on how to preprocess EEG data. There are many other complex techniques that can enhance your result.
- One of the major lessons learned from this project is that programming in R is much slower than programming in Python. We have written rough code in python. Due to familiarity with R, we have chose R to be the final code. In the future, we will try to use Python because it is much faster.

6 Conclusion

Analyzing EEG brain signals gave us an opportunity to learn a lot about data analysis. In this project, having some domain knowledge could have been more helpful especially while preprocessing. However, we were still able to achieve AUC of 0.7. Implementing an ensemble model can definitely improve our accuracy to at least 0.9. We are planning to use combination of Support vector machine, neural network and logistic regression for future progression of this project. We are also planning on using CSP for data preprocessing.

7 References

[1] PHOTHISONOTHAI, MONTRI, and MASAHIRO NAKAGAWA. "A Classification Method Of Different Motor Imagery Tasks Based On Fractal Features For Brain-Machine Interface." *Journal Of Integrative Neuroscience* 8.1 (2009): 95-122. *Academic Search Complete.* Web. 16 Nov. 2015.

[2] "Electroencephalography." *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. 15 November 2015. Web. 16 Nov. 2015. < https://en.wikipedia.org/wiki/Electroencephalography>