

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

A Report On

**Classification Algorithms for Brain Computer Interface**

Prepared By

**Ritik Rohit Bavdekar**                         **2017B4A70349P**

Supervisor

**Dr Vandana Agarwal**

Computer Science Department

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

(Aug 2020-Dec 2020)

# Acknowledgements

I would like to thank Dr Vandana Agarwal for giving me the opportunity to work in this new and exciting field. She constantly taught me, guided me and used her vast experience to help me with the project. I would also like to thank the CS department for giving opportunities like Study Oriented Projects. I want to express gratitude towards my parents who have been providing for internet, electricity backup and mental support in these tough lockdown times.

# Introduction

Brain Computer Interface is hardware and software that helps people communicate with and control machines using cerebral activities. The main aim of BCI devices is to provide a means of communication for people with total paralysis or neuromuscular disorders.

A BCI is a system that can recognise certain brain signals using patterns. These signal patterns are used by the system to perform different BCI actions. The stages of BCI are : signal acquisition, signal preprocessing ,feature extraction, classification and control interface. The two major steps which are worked on in this study are feature extraction and classification.

Before feature extraction is applied several artefact removal algorithms are applied to clean raw data. Feature extraction involves using several dimension reduction algorithms like Discrete Cosine Transformation, Principal Component Analysis and Independent Component Analysis. It also involves using feature transformation techniques like Fast Fourier Transformation and band pass filters to sample signals of a required range of frequencies. There are also algorithms like Cognitive Spatial Patterns(CSP) which are specially made for feature extraction in BCI.

Classification involves using the feature extracted processed data to identify particular thought classes. These require machine learning algorithms like Support Vector Machines and Linear Discriminant Analysis. Nowadays Neural networks are also picking up. All these algorithms’ potentials are limited by the amount of data present. As it is a comparatively new field and requires strenuous mental work by the subject, data is limited.

For many years BCI technology was considered strange and unattractive. People didn’t like the idea of being able to decipher their own thoughts and intentions using computers. In the last 20 years multidisciplinary studies with researchers from neuroscience , psychology , computer science and other disciplines have been performed.

# Brain Computer Interface Signals

There are many different brain signals which can be measured for BCI classification. There are 2 types of different signals i)electrophysiological and ii)hemodynamic

Electrophysiological signals are generated by the firing of neurons. The current movement is measured. The magnetic fields generated by the current can also be measured. The activity is measured by EEG, ECoG, MEG etc.

Hemodynamic measurement is based on the fact that blood releases glucose at higher rate to the active neurons compared to inactive neurons. The veins of active area have surplus of oxyhemoglobin . The changes are measured by neuroimaging methods like near infrared spectroscopy and functional magnetic resonance. These measures are called indirect since this is a response to the change in electric signals in the neurons. Direct methods involve measurement of the electric signals.[1]

EEG is the most common brain signal used for classification. EEG has low cost, wide temporal resolution, good portability and very little risks to users. These also have more artefacts due to smearing effect of the skull. MEG measures magnetic activity , is non-invasive and non portable.

ECoG is portable but invasive so is used less. Since it’s closer to the neurons there are lesser errors as hair and skull doesn’t come in between to dissipate the current. Indirect methods are used rarely as they provide scope for even more error.

# Different EEG wave frequencies

EEG is used to measure the activity in the brain due to the flow of electric currents. This happens during synaptic excitations of the dendrites. EEG signals are sensitive to secondary currents. The EEG signal is as potential difference between the signal and a reference electrode. The electrodes are usually made of silver chloride(AgCl). There is usually a multichannel configuration which can comprise of 128-256 electrodes. Also a gel is used on the subjects scalp to get good quality signals. The newer electrodes do not require this gel and are called dry electrodes.

EEG signal comprises of different frequency bands which have different biological significances. The 5 common frequency bands are: delta, theta, alpha, beta and gamma.

The delta band lies between below 4Hz and the amplitude of delta waves is detected mainly in babies and decreases with age. These rhythms are observed in adults only in deep sleep state.[1]

Theta waves lie within 4Hz to 7Hz range. In a normal awake adult very little theta waves are observed. Theta waves are generally observed in drowsy or meditative adults and babies.

Alpha waves lie within 8Hz to 12Hz range. Their amplitude increases when eyes are closed and body relaxes. They are attenuated when mental effort is applied. It may also be related to the memory function if the brain. These signals primarily reflect visual processing.

Mu rhythms are in the same region as alpha rhythms however they are physiologically different. They have high correlation with the beta rhythms and are related to motor activity.

Beta rhythms are in 12 to 30 Hz range. They are related to motor activities. They are mainly attenuated in the central and frontal regions of the brain. Beta waves are symmetric in nature when there is no motor activity. During active movement , beta waves attenuate and the symmetric distribution is disturbed. [1]

Gamma rhythms are in 30 to 100 Hz range. Gamma rhythms are related to a few motor functions and perceptions in healthy adults. There is also a relation between gamma waves and maximal muscle contraction.

# Control Signals in BCI

There are many different control signals generated by the brain with specific characteristics. These characteristics can be used to design Brain Computer interface systems. There are 4 major control signals used in BCI systems: i)VEP ii)SCP iii)P300 and iv)Sensorimotor rhythms.

Visual Evoked Potentials(VEPs) are modulations generated after visual stimulus [1,2]. The amplitude of these modulations increases as the stimulus gets closer to central visual field. [1,3]. Specialised versions of VEP allow users to select targets using eye gaze.

Slow Cortical Potentials(SCPs) are brain signals with slow voltage shifts in EEG. SCPs is part of the EEG signals below 1 Hz [1,4]. Change in cortical activity results in SCP change. An positive SCP corresponds to increase in neuronal activity and vice versa[1,4].SCP signals can be used by all types of users to control external BCI systems. SCP shift can be used to control a simple Input devices like mouse[1,5] .People can be trained to modulate their SCP shifts. However it requires months of training and hence is not practical for patients with neurological problems.

P300 Evoked Potentials are positive peaks in EEG signals caused by unexpected stimulus. Stimulus could be visual, auditory or somatosensory .The more unexpected the stimulus the higher the amplitude. It does not require training. Common uses of P300 signal based systems involves flashing random lights on the subject and searching for P300 signals in the EEG data. P300 speller is a common BCI system using these signals. However when the user gets used to random stimulus the performance decreases due to decrease in amplitude of P300 potentials.

Mu and beta rhythms are the most important for sensorimotor BCI systems. The amplitudes of the rhythms vary with motor tasks. Actual movement is not required and motor imagery is adequate for generation of the signals. Sensorimotor rhythms have two different kinds of amplitude modulation. Event related synchronisation(ERS) and event related desynchronisation(ERD).

# Feature Extraction

Feature extraction involves converting the raw EEG data into compressed relevant features which can be used as dataset for machine learning. [6] Feature extraction is necessary in BCI systems due to the curse of dimensionality and limited size of datasets. It is recommended to use 5-10 times training examples per class as input vector feature dimensions. This is a huge task for the subject.

There are 3 different types of information to be extracted in EEG signals:

1. Spatial Information: These features define the origin of signals from specific spatial regions of the brain. This involves selecting channels which have a high correlation with the specific spatial region.
2. Spectral Information: These describe variation of power in specific frequency bands like mu bands or beta bands.
3. Temporal Information: These describe the variation of signals with time.

BCI for motor imagery tasks (or oscillatory activity) depend mainly on spatial and spectral information. BCI systems based on event related potentials need spatial and temporal information.

BCI based on oscillatory activity depend on change in power in specific frequency bands. A simple design could involve using band-pass filters to select frequency bands of interest and selecting the spatial information manually using neurological theory.eg- channel C3 and C4 correspond to right hand movement and left hand movement respectively.

For modern Oscillatory activity brain computer interfaces it is suggested to have a minimum of 8 channels to record enough spatial information and optimal performance is achieved with 48 channels. To select correct features many approaches are applied. They involve selecting best feature selections, channel selections and spatial filtering algortihms.

Spatial filtering is a technique which creates a small number of new channels as a linear combination of the original channels. This helps in decreasing the smearing effect of the skull. There are spatial filters with neurophysiological significance like the Laplacian and Bipolar filter.

Apart from neurophysiological filters there are also data driven filters. These include Principal Component Analysis(PCA) and Independent Component Analysis(ICA). There is also a spatial filter algorithm called Cognitive Spatial Patterns algorithm which is used for all BCI based EEG oscillatory activity. This feature selection algorithm has helped in boosting all present day BCI systems.

# Classification Algorithms

With dimension-reduced, artefact free and feature selected data classification is performed. This involves selecting different classes for different signals at different times. There are mainly 4 different approaches to classify algorithms i) Bayesian analysis ii) Linear Discriminant Analysis iii) Support Vector Machines iv)k-NNC v)Artificial Neural Networks

Linear Discriminant Analysis is a simple yet effective model. This provides good accuracy without very high computation costs. LDA is good for designing online BCI platforms.

Support Vector Machine is a machine learning supervised classification model. It is used for separating linear and non-linear data(provided it follows Mercer’s Theorem). It takes the help of kernels to project non-linear data into linear data in higher dimensions. In SVM a hyperplane is selected which best separates the two classes of the dataset. Support vectors are the closest vectors to the hyperplane. The margin is the sum of perpendicular distances between the support vector of both the classes and the hyperplane. This hyperplane is found by converting it into a primal optimisation problem. We notice that the primal cannot be solved and hence it is converted to its dual form. This dual form is solved and the results are returned.

SVMs are widely used in BCI systems as they provide quick results with low computation costs and high accuracies. It can be used for online BCI platforms as well.

Artificial Neural Networks involves finding a nonlinear decision plane by minimizing error in classifying training data. The user needs to subjectively select many parameters. ANNs are non linear classifiers. They are sensitive to overtraining and have high computation cost.

SVM with Gradient Descent optimizer , Genetic Algorithm optimizer and Particle Swarm Optimizer are used to find hyperplanes that separate the dataset well and give good results.

# Datasets

All the datasets have been taken from BCI competition 2008.

* **Dataset 2a** has EEG data of 9 subjects. It involves four different motor imagery tasks- Right hand movement , left hand movement , both feet movement and tongue movement.2 sessions were recorded on 2 different days for each subject. Each session comprises of 6 runs separated by short breaks. 1 run consists of 48 trials(12 of each class). They also contain EOG recording for 5 minutes in start to help remove EEG artefacts. It has 22 EEG components sampled at 250Hz with bandpass filters between 0.5Hz and 100Hz.Each run is separated by 100 missing values.
* **Dataset 2b** has the same bandpass filters and notch filters. It is sampled at 250Hz. EOG is included to help clean artefacts. 2 classes are there. Motor imagery movement of right and left hand. 2 sessions are taken without feedback,6 runs and 10 trials for each class. 3 sessions are taken with feedback, 4 runs and 20 trials, for each class
* **Dataset 3** has 2 subjects. MEG signals are sampled at 625Hz. It has 4 different classes depending on subjects wrist movements. The subject moves his joystick from center to a target direction using his right hand and wrist. The dataset has 10 MEG channels. There are bandpass filters at 0.5 and 100Hz.The signal is resampled to 400Hz.
* **Dataset 4** has 3 different subjects. ECoG reading has been taken for each subject. Bandpass filter of 0.15Hz and 200Hz is there and it is sampled at 1000Hz. The channels for subject 1 ,2 and 3 are 62,48 and 64 respectively. Subjects are cued to move a particular finger. For each cue subject moves his finger 3-5 times.30 stimulus are made for each finger for each subject.

# Implemented Code(MATLAB)

Support Vector Machines have been implemented from scratch. These can be used to create classify both linear and non-linear data. A few common kernel functions like rbf, polynomial and linear have also been implemented. The hyperparameter C is also included to make it a soft margin classifier. Quadprog is used for optimisation purposes.

![Chart, scatter chart

Description automatically generated]()

Linearly Separable Data

![Chart

Description automatically generated]()

Completely linearly separable data

![Chart, scatter chart

Description automatically generated]()

Linearly non-separable data

A few data plots for dataset 2a were also done:![Graphical user interface, text, application

Description automatically generated]()

Dataset EEG channels Plot

![A picture containing chart

Description automatically generated]()

Scatter Plot for 5 different classes with 2 feature vectors

![Graphical user interface, chart, line chart

Description automatically generated]()

Spectral Density Plot

![Graphical user interface, application

Description automatically generated]()

Stacked Channel EEG signals plot

The **Discrete Cosine Transformation** of data has been calculated. Functions to **remove artefacts** from the data have been implemented. The **Fast Fourier Transformation** was calculated and plotted for dataset 2a and 2b.

Discrete Fourier Transformation is a widely used signal processing tool. It decomposes a signal with noise into complex sinusoidal waves which can be superimposed to reconstruct the signal. The Y axis plots the amplitude of the wave and the X axis plots the frequency of the sin wave. Fast Fourier Transformation algorithm was used to calculate the DCT of the signal.

![Graphical user interface

Description automatically generated]()

FFT Component 1

![Graphical user interface, application

Description automatically generated]()

FFT of component 25

![Graphical user interface

Description automatically generated]()

Mean FFT

Discrete Cosine Transformation is a compression algorithm that decomposes a signal into its corresponding cosine waves with coefficients. The cosine waves with small coefficients are dropped thereby compressing the wave. These cosine waves can be superimposed to form the signal again.![Graphical user interface, chart

Description automatically generated]()

DCT of component 10

Manual Feature Extraction of dataset A04E.gdf.

Channels 1,3,6 and 11 were selected

Fft\_channels= 1 and 6

Dct\_channels= 1 and 6

These were selected as a feature vector for classification

![Chart

Description automatically generated]()

Channel-1 Signal Plot

![Chart

Description automatically generated]()

Channel-3 Signal Plot

![Graphical user interface, chart

Description automatically generated]()

Channel-6 Signal Plot

![Graphical user interface

Description automatically generated]()

Channel-11 Signal Plot

![A picture containing graphical user interface

Description automatically generated]()

Channel-1 FFT Plot

![Graphical user interface

Description automatically generated]()

Channel-6 FFT Plot

![Graphical user interface

Description automatically generated]()

Channel-1 DCT Plot

![Graphical user interface, chart

Description automatically generated]()

Channel-6 DCT Plot

Final Feature Extraction:

1. Power-Spectra with log scale
2. Single-Channel Cognitive Spatial Patterns
3. Gray Level Co Occurrence Matrix

Power Spectra with log scale: Fast Fourier Transformation of the signals were taken. The power was calculated using the following formulas![A picture containing text, watch, gauge

Description automatically generated]()![A picture containing calendar

Description automatically generated]()

Formulae from [7]

Single-Channel Cognitive Spatial Patterns(SC-CSP): CSP is a spatial filter search algorithm that finds filters that increases variance for one class and decreases for the other in the filtered signal [6]. This procedure helps improve the classification performance by a lot. For motor imagery classification EEG signals are first bandpass filtered between 8-30 Hz and then CSP is applied to find the best spatial filters. The values are then converted to log10 scale.

![Text, letter

Description automatically generated]()

Formulae from [6]

These final features were used on SVM with radial basis kernel function.

Training Accuracy =67.8571

Testing Accuracy=60.7143

SVM with nature inspired algorithms:![Diagram

Description automatically generated]()

SVM with genetic algorithm:![A picture containing diagram

Description automatically generated]()

![Chart, scatter chart

Description automatically generated]()

SVM with PSO and linearly separable data![Chart, scatter chart

Description automatically generated]()![Chart, scatter chart

Description automatically generated]()

SVM with PSO and non-linear data![Chart, scatter chart

Description automatically generated]()![Chart, scatter chart

Description automatically generated]()

PSO-SVM on BCI data

Hyperparameters:

* Agents=100
* Generations=10000

Training accuracy=0.6056

Testing accuracy=0.5417

# CONCLUSION

BCI motor imagery data is raw signals data with noise and of huge dimensions. Feature extraction and signal pre-processing are essential for classification. Feature extraction can be done using spatial filtering: neurophysiological or data-driven(Cognitive Spatial Patterns). After feature extraction algorithms like Support Vector Machines, Linear Discriminant Analysis and ANNs can be used for classification. Different types of SVMs were extensively experimented for motor imagery classification. Several different optimizers were used to get different results. SVMs can be used in the future for real-world on-line motor imagery task classification tasks.

# FUTURE SCOPE

The algorithms and code implemented can be used for testing on a lot of different datasets. The code can be extended for other algorithms like LDA and ANNs. The feature extraction is complete and can be used for any motor imagery based thought detection classification problem. Support Vector Machines code is also complete. The full pipeline can be used for testing different motor imagery classification datasets. These algorithms can be used to develop on-line classification. Also a study can be done to find accuracies of SVMs over different test subjects. This study was restricted to finding SVM parameters for a specific user. More studies can find out if universal SVM parameters can be tuned that work on different users.

# References

*[1] Nicolas-Alonso, Luis Fernando, and Jaime Gomez-Gil. "Brain computer interfaces, a review." Sensors 12.2, 1211-1279 (2012).*

*[2] Regan, D. Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine; Elsevier: New York, NY, USA, 1989.*

*[3] Yijun, W.; Ruiping, W.; Xiaorong, G.; Bo, H.; Shangkai, G. A practical VEP-based brain-computer interface. IEEE Trans. Neural Syst. Rehabil. Eng. 2006, 14, 234–240.*

*[4] Birbaumer, N.; Elbert, T.; Canavan, A.G.; Rockstroh, B. Slow potentials of the cerebral cortex and behavior. Physiol. Rev. 1990, 70, 1–41.*

*[5] Hinterberger, T.; Schmidt, S.; Neumann, N.; Mellinger, J.; Blankertz, B.; Curio, G.; Birbaumer, N. Brain-computer communication and slow cortical potentials. IEEE Trans. Biomed. Eng. 2004, 51, 1011–1018.*

*[6] Bi, Luzheng, Xin-An Fan, and Yili Liu. "EEG-based brain-controlled mobile robots: a survey." IEEE transactions on human-machine systems 43.2, 161-176 (2013).*

*[7] http://www.cmp.caltech.edu/~mcc/Chaos\_Course/Lesson6/Power.pdf*