Analysis and Information Retrieval from Electroencephalogram for Brain–Computer Interface Using WEKA

Jayesh Deep Dubey, Deepak Arora and Pooja Khanna

Abstract In the recent years, there has been huge development in the area of brain–computer interface. Brain–computer interface technology provides a new method of communication for people with neuromuscular damages and problems due to which they are unable to use usual communication methods. A brain–computer interface system makes it possible for an individual to direct the outside world through his thoughts without relying on muscle activity. Current brain–computer interface systems use either the EEG activity in which the brain signals are recorded through electrodes by placing them on the scalp of the individual or invasive procedures in which the electrodes are placed inside the brain. In this paper, we have used EEG data recorded from 8 selected channels out of 64 available channels. The eight channels were selected by using principal components analysis and association rule mining using apriori algorithm. K-means clustering was then implemented on the data obtained from eight selected channels and the number of cluster instances was set to three because the number of events observed in the data while observing it in the EEGLAB was three. The study analyzed electroencephalogram signal using WEKA and presented a set of information that can be retrieved. This information is important for the brain–computer interface systems.

Keywords EEG · BCI · PhysioBank · EEGLAB · EDFBrowser

J. D. Dubey · D. Arora · P. Khanna (✉)
Amity Institute of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India
e-mail: inkhanna@gmail.com

J. D. Dubey
e-mail: jayesh.d.dubey@gmail.com

D. Arora
e-mail: deepakarorainbox@gmail.com

M. L. Kolhe et al. (eds.), *Advances in Data and Information Sciences*, Lecture Notes in Networks and Systems 39, https://doi.org/10.1007/978-981-13-0277-0_11

1 Introduction

Brain–computer interface is defined as a technical system that provides direct interaction between the brain of a human being and external devices. BCI translates the user's objective which is reflected by brain signals into a desired output: controlling an external device or computer-based communication. The name "Brain–Computer Interface" was first used in the early 1970s by the works of Dr. J. Vidal, which was inspired by the optimism to create an alternate output method of communication for disabled individuals and enhancing human ability to control the external systems [1]. In brain–computer interface, there is no requirement for any kind of muscle activity for issuing the commands and completing the interaction. BCI was initially developed by research community for creating assistive devices for biomedical applications [2]. These devices have been able to restore the movement activity for people with physical disabilities and replacing lost motor functionality.

BCI systems bypass the muscular output pathways and use digital signal processing and machine learning to translate signals from the brain into actions [3]. A brain–computer interface system work by recognizing patterns in the data that has been extracted from the brain and associating those patterns with the commands [4]. The most important and interesting application of BCI is medical rehabilitation. Since a BCI system does not include any motor activities, people with severe motor damages [such as individuals suffering from paralysis] can also operate a BCI-based communication system. The objective of BCI this situation is to record the brain signals and convert them into a form such that computers and other devices can easily understand it.

2 Literature Review

Brain–computer interface is a system that is based on electroencephalography and is used to establish communication channel between brain of a human being and external environment. BCI systems use electrical activity in the brain to allow users to control external devices through their thoughts without relying on muscle activity [5]. These systems are very helpful in enhancing the quality of life of the disabled individuals who are suffering from motor disorders and neuromuscular damages due to which they are unable to use usual communication methods. BCI systems use EEG activity in which the brain signals are recorded through electrodes by either placing them on the scalp of the individuals according to 10-10 or 10-20 electrode placement system or by placing the electrodes inside the brain [6]. BCI systems work by recognizing patterns in the data that has been extracted from the brain and associating these patterns with the commands for the external devices.

In the recent years, there has been a great deal of research and development in the area of brain–computer interface. Brain–computer interface has been successfully applied in many fields such as military, entertainment, smart house, etc.

However, the most important and interesting application of BCI is medical rehabilitation. Researchers at the University of Minnesota have achieved a breakthrough that enables individuals to control a robotic arm through their minds. This invention can help paralyzed patients and those suffering from neurodegenerative disorder [7]. In one of the researches, the research team placed a brain implant in a patient suffering from ALS disease which prohibited the patient to move and speak. The brain implant enabled the patient to operate a speech computer through her mind and she was able to spell her first word [8]. Researchers at the University of Utah and Oregon State University College of Engineering have carried out a research that has created hope for patients suffering from spinal cord injuries. The research suggests that a wearable, smartphone-sized control box might enable at least some level of movements in these patients by delivering impulses to implant electrodes in their peripheral nervous system [9]. Researchers have developed advanced electrodes known as “glassy carbon” electrodes that are capable of transmitting clearer, robust signals than the currently used electrodes. These electrodes could allow restoration of movements in patients with damaged spinal cords [10].

3 Dataset Description

The data used in this research article has been obtained from PhysioNet database. The data is electroencephalographic motor movement/imagery data, and the dataset contains about 1500 electroencephalographic recordings of 1 and 2 min duration. These recordings have been procured from 109 subjects. The experiment conducted to obtain the data has been described below [11, 12].

Different motor/imagery tasks were performed by the subjects, while 64-channel electroencephalography was recorded using the BCI2000 system. 14 tasks were carried out by each subject: two baseline tasks of 1 min each (one in which subject kept his eyes opened and the other in which subject’s eyes were closed), and three 2-min tasks of each of the following:

1. A cue is displayed on either the left or right side of the monitor. The subject opens and closes the corresponding hand until the cue vanishes. Then, the subject relaxes.
2. A cue is displayed on either the left or right side of the monitor. The subject imagines opening and closing the corresponding hand until the cue vanishes. Then, the subject relaxes.
3. A cue is displayed on either the bottom or top of the monitor. The subject opens and closes either both feet (if the target is on the bottom) or both fists (if the target is on the top) until the cue vanishes. Then, the subject relaxes.
4. A cue is displayed on either the bottom or top of the monitor. The subject imagines opening and closing either both feet (if the target is on the bottom) or both fists (if the target is on the top) until the cue vanishes. Then, the subject relaxes.

14 tasks were conducted for each subject but in our research work, we have considered 4 tasks for each subject which is described as follows.

1. Task 1 (opening and closing left or right fist),
2. Task 2 (imagine opening and closing left or right fist),
3. Task 3 (opening and closing both fists or both feet), and
4. Task 4 (imagine opening and closing both fists or both feet).

In our study, we have used data from first 14 subjects out of the 109 subjects available in the dataset.

4 Data Extraction and Preprocessing for Experimental Setup

We have used **PhysioBank ATM** online toolbox to export the data in EDF (European data format) format. PhysioBank ATM is an online toolbox for exploring PhysioBank through web browser. The toolbox contains software that can display annotated waveforms, RR interval time series, and histograms, and convert WFDB signal files to text, CSV, EDF or .mat files, and more.

After obtaining the data in EDF format, we imported it in **EEGLAB** toolbox. EEGLAB is a Matlab toolbox that is used for processing continuous and event-related electroencephalographic, magnetoencephalographic, and electrophysiological data. EEGLAB operates under Windows, Linux, UNIX, and Mac OS. It provides graphical user interface that enables users to process EEG and other brain data. After importing the data in EEGLAB, the data was bandpass filtered between 3 and 30 Hz to remove unwanted frequency components from the data. The data was filtered using EEGLAB, and then **EDFBrowser** was used to convert the data from EDF format to text format.

5 Analysis of EEG Data

5.1 Channel Selection

The EEG data obtained from PhysioBank ATM was recorded from 64 channels according to 10-10 electrode system. In order to obtain the channels that are capturing majority of the motor imaginary data, we have used principal component analysis for each of the runs of all the 14 subjects. This resulted in a list of channels for each of the experimental tasks.

After obtaining the list of channels, we have applied association rule mining to obtain correlated channels. Apriori algorithm was used for this task. Weka was used as a tool to employ principal components analysis and apriori algorithm. The channels that we have obtained on the basis of this analysis are **FC5, FC3, FC1, FCz, FC2,**

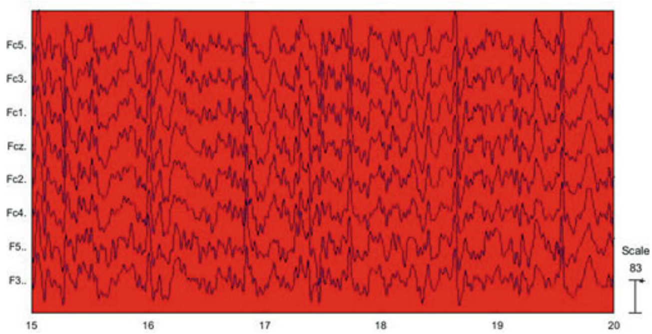


Fig. 1 EEGLAB plot for baseline run with only one type of event marked in red color

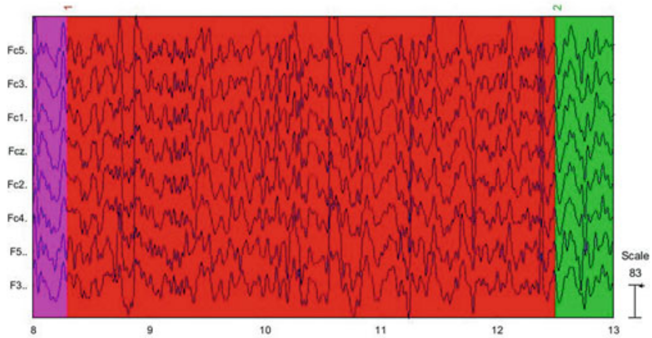


Fig. 2 EEGLAB plot for motor movement task with three types of events marked in red, pink, and green colors

FC4, F3, and F5. We will use the data recorded from the above mentioned channels for further analysis.

5.2 Data Analysis in the EEGLAB

When data for the baseline run [in which the subject is in relaxed state and not performing any movement] was analyzed in the EEGLAB, then only one type of event was observed which corresponds to the relaxed state (Fig. 1). Then, we performed analysis of data for Task 1, Task 2, Task 3, and Task 4 as mentioned above in the dataset description. For this data, three types of events were observed, one of which is the same event as observed in the baseline run which corresponds to relaxed state and two more events were observed which corresponds to the movement or imagination of movements of fist or feet (Fig. 2).

Table 1 Clusters obtained for each of the 4 tasks for all the 14 subjects

Subject	Task	Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)
Subject 1	Task 1	22	48	30
Subject 1	Task 2	30	22	47
Subject 1	Task 3	20	30	50
Subject 1	Task 4	21	49	30
Subject 2	Task 1	26	48	26
Subject 2	Task 2	25	47	28
Subject 2	Task 3	52	23	25
Subject 2	Task 4	26	26	48
Subject 3	Task 1	35	52	13
Subject 3	Task 2	21	29	50
Subject 3	Task 3	8	57	35
Subject 3	Task 4	37	57	6
Subject 4	Task 1	37	2	61
Subject 4	Task 2	36	3	61
Subject 4	Task 3	3	75	22
Subject 4	Task 4	56	42	2
Subject 5	Task 1	22	50	28
Subject 5	Task 2	54	29	17
Subject 5	Task 3	48	27	26
Subject 5	Task 4	26	24	50
Subject 6	Task 1	32	14	54
Subject 6	Task 2	28	53	19
Subject 6	Task 3	32	55	13
Subject 6	Task 4	51	28	21
Subject 7	Task 1	54	9	37
Subject 7	Task 2	27	52	21
Subject 7	Task 3	10	32	58
Subject 7	Task 4	18	53	29
Subject 8	Task 1	9	37	54
Subject 8	Task 2	30	54	16
Subject 8	Task 3	5	39	56
Subject 8	Task 4	19	28	53
Subject 9	Task 1	11	35	54
Subject 9	Task 2	31	53	17
Subject 9	Task 3	37	9	54
Subject 9	Task 4	17	30	53

(continued)

Table 1 (continued)

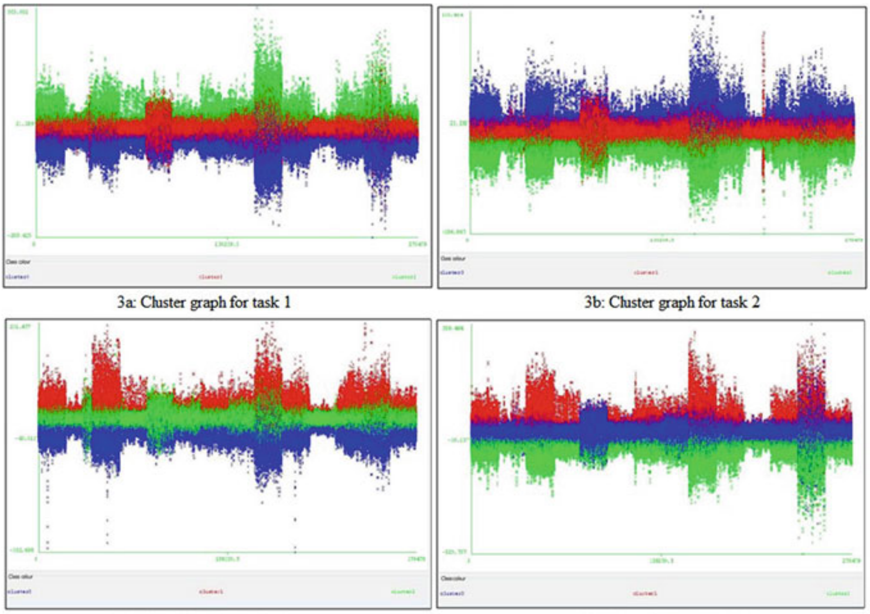
Subject	Task	Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)
Subject 10	Task 1	48	29	23
Subject 10	Task 2	52	29	19
Subject 10	Task 3	27	51	22
Subject 10	Task 4	51	22	27
Subject 11	Task 1	28	51	22
Subject 11	Task 2	49	23	28
Subject 11	Task 3	23	29	48
Subject 11	Task 4	28	24	48
Subject 12	Task 1	40	55	4
Subject 12	Task 2	38	4	58
Subject 12	Task 3	6	61	34
Subject 12	Task 4	31	61	8
Subject 13	Task 1	20	28	52
Subject 13	Task 2	28	22	50
Subject 13	Task 3	28	53	19
Subject 13	Task 4	50	29	21
Subject 14	Task 1	47	23	30
Subject 14	Task 2	30	48	22
Subject 14	Task 3	26	50	24
Subject 14	Task 4	28	23	49

5.3 Clustering of Data

WEKA was used as a tool to run **K-means clustering** algorithm on the data recorded from the selected channels as mentioned above. The k-means algorithm was run for each of the four tasks for all the subjects. Since we have considered 14 subjects and 4 experimental tasks for each subject, the algorithm was run for 56 times. The data was divided into three clusters because during analysis of the data in EEGLAB, three types of events were observed. The clusters that we have obtained can be seen in the table below. It can be observed in Table 1 that for most of the subjects and tasks two clusters are uniformly distributed, while the third cluster is almost twice the size of other two clusters. For example, for task 1 of subject 2, the clusters obtained are 26, 48, and 26%; for task 1 of subject 10, the clusters are 48, 29, and 23%; and for task 4 of subject 14, the clusters are 28, 23, and 49%. Similar pattern can be found for many tasks for the subjects in the table.

Figure 3 shows the graphical illustration of the three types of clusters for task 1, task 2, task 3, and task 4 for all the 14 subjects as described below.

1. 3a is the cluster graph for task 1 for all the 14 subjects.



3a: Cluster graph for task 1

3b: Cluster graph for task 2

3c: Cluster graph for task 3

3d: Cluster graph for task 4

Fig. 3 Cluster graphs for each of the 4 tasks for all the 14 subjects, **a** is the cluster graph for task 1 for all the 14 subjects, **b** is the cluster graph for task 2 for all the 14 subjects, **c** is the cluster graph for task 3 for all the 14 subjects, **d** is the cluster graph for task 4 for all the 14 subjects

2. 3b is the cluster graph for task 2 for all the 14 subjects.
3. 3c is the cluster graph for task 3 for all the 14 subjects.
4. 3d is the cluster graph for task 4 for all the 14 subjects.

6 Conclusions

Analysis of EEG data is the most important part of BCI and many different methods have been used by researchers around the world for analysis of EEG data. In our experiment, we have used Weka as a tool for analysis of executed and imagined EEG data. The data from eight selected channels [out of 64 channels] has been considered in the experiment. The channel selection was performed using principal component analysis and apriori algorithm. K-means clustering algorithm has been used for clustering the data into three clusters. During the analysis of data in EEGLAB, three types of events were observed, and therefore we have selected the number of clusters as three while performing k-means clustering. EEG signal can be clustered on the basis of type of task brain is performing. Though EEG signals are captured using/from many electrodes point, few plays significant role in task identification.

However, further analysis of data is required for exact classification of data into left- and right-hand motor movements which will be performed in future experiments.

References

1. Vidal JJ (1977) Real-time detection of brain events in EEG. Proc IEEE 65(5):633–641
2. Rao RPN, Scherer R (July 2010) Brain-computer interfacing [In the Spotlight]. IEEE Signal Process Mag 27(4):152–150. https://doi.org/10.1109/msp.2010.936774, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5484181&isnumber=5484155
3. Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckhalm PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM Brain computer interface technology: a review of the first international meeting
4. Rohani DA, Henning WS, Thomsen CE, Kjaer TW, Puthusserypady S, Helge BD (2013) Sorensen BCI using imaginary movements: the simulator. Comput Methods Programs Biomed 111:300–307
5. Abdulkader SN, Atia A, Mostafa M-SM Brain computer interfacing: applications and challenges. HCL-LAB, Department of computer Science, Faculty of computers and information, Helwan University, Cairo, Egypt
6. Niedermeyer E (1999) The normal EEG of the waking adult, Niedermeyer E, Lopes da Silva FH (eds) Baltimore, MD, Williams and Wilkins
7. Meng J, Zhang S, Bekyo A, Olsoe J, Baxter B, He B (2016) Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks. Sci Reports 6:38565. https://doi.org/10.1038/srep38565
8. Vansteensel MJ, Pels EGM, Bleichner MG, Branco MP, Denison T, Freudenburg ZV, Gosselaar P, Leinders S, Ottens TH, Van Den Boom MA, Van Rijen PC, Aarnoutse EJ, Ramsey NF (2016) Fully implanted brain–computer interface in a locked-in patient with ALS. New Engl J Med. https://doi.org/10.1056/nejmoa1608085
9. Frankel MA, Mathews VJ, Clark GA, Normann RA, Meek SG (2016) Control of dynamic limb motion using fatigue-resistant asynchronous intrafascicular multi-electrode stimulation. Front Neurosci. https://doi.org/10.3389/fnins.2016.00414
10. Vomero M, Castagnola, FC, Maggiolini E, Goshi N, Zucchini E, Carli S, Fadiga L, Kassegne S, Ricci D (2017) Highly stable glassy carbon interfaces for long-term neural stimulation and low-noise recording of brain activity. Sci Reports 7:40332. https://doi.org/10.1038/srep40332
11. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals
12. Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR (2004) BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans Biomed Eng 51(6):1034–1043