Reducing Electrodes based on Decision Tree Classification for EEG Motor Movement Data

Jayesh Deep Dubey[1], Deepak Arora[2], Pooja Khanna[3]

[1,2,3]Department of Computer Science and Engineering, Amity University, Lucknow
*Corresponding Author Email: [1]jayesh.d.dubey@gmail.com,
[2]deepakarorainbox@gmail.com, [3]pkhanna@lko.amity.edu

Abstract

Analysis of EEG data is one of the most important parts of Brain Computer Interface systems because EEG data consists of a substantial amount of crucial information that can be used for better study and improvements in BCI system. One of the problems with the analysis of EEG is the large amount of data that is produced, some of which might not be useful for the analysis. Therefore identifying the relevant data from the large amount of EEG data is important for better analysis. The objective of this study is to find out the performance of Random Forest classifier on the motor movement EEG data and reducing the number of electrodes that are considered in the EEG recording and analysis so that the amount of data that is produced through EEG recording is reduced and only relevant electrodes are considered in the analysis. The dataset used in the study is Physionet motor movement/imagery data which consists of EEG recordings obtained using 64 electrodes. These 64 electrodes were ranked based on their information gain with respect to the class using Info Gain attribute selection algorithm. The electrodes were then divided into 4 lists. List 1 consists of top 18 ranked electrodes and number of electrodes was increased by 15 [in ranked order] in each subsequent list. List 2, 3 and 4 consists of top 33, 48 and 64 electrodes respectively. The accuracy of random forest classifier for each of the list was compared with the accuracy of the classifier for the List 4 which consists of all the 64 electrodes. The additional electrodes in the List 4 were rejected because the accuracy of the classifier was almost same for List 4 and List3. Through this method we were able to reduce the electrodes from 64 to 48 with an average decrease of only 0.9% in the accuracy of the classifier. This reduction in the electrode can substantially reduce the time and effort required for analysis of EEG data.

Keywords: BCI, EEG, Annotations, EDF Browser, Physionet, Random Forest.

1. Introduction

Brain Computer Interface is a technical system that establishes a channel of communication between human brain and external devices or external environment. Brain Computer Interface system uses the activity of the human brain to control the external devices and environment without the participation of any kind of muscle activity thus allowing direct interaction between the human brain and external devices. The main purpose of the BCI system is to create an alternate communication channel for the disabled and paralyzed individuals and enabling them to control the external devices and external environment and replacing their lost motor functionality [1]. BCI systems do not require any kind of muscle activity for implementing the interaction or completing the commands. The initial objective of the research community behind developing BCI system was development of assistive devices for medical application [2]. However, the development of BCI devices for entertainment purpose is also a major topic of research nowadays and a good deal of research is being done in this field.

BCI system makes use of machine learning and digital signal processing to capture and transform brain signals into actions or commands for the external devices [3].

The working of BCI system is based on the extraction and recognition of patterns in the data obtained from the brain and association of these patterns with the commands for the external devices [4]. Medical rehabilitation is the most crucial application of BCI. BCI system makes use of EEG signals recorded from the brain. EEG or electroencephalography is a technique for recording the electrical activity within the brain with the help of electrodes that are placed on the head of the individual. There are two methods for recording the electroencephalographic signals. The first method is noninvasive EEG in which the electrodes are placed on the scalp of the individuals according to 10-10 or 10-20 electrode placement system .The second method is invasive EEG in which the electrodes are placed inside the brain.

Brain Computer Interface system consists of three main components: recorder module, preprocessing module and classifier module. The recorder module records the EEG signals from the brain, preprocessor module converts the input signal into a form that makes them easy to classify, and classification module is based on self learning software and makes use of feature to classify the signals into different categories.

2. Literature Survey

In the recent years, Brain Computer Interface has grown as a highly researched field. BCI has wide variety of applications such as medical rehabilitation, neuroergonomics and smart environment, games and entertainment. There has been a great deal of research in all these fields. Encephalophone is a thought

controlled music device created by neurologist Dr. Deuel in collaboration with Dr. Felix Darvas. This device will enable patients with motor impairments to play music without the need for any kind of movement. Encephalophone is based on electroencephalography and works by transforming the electrical signals obtained from the brain into musical notes [5].In 2007, F.Lotte et al presented a paper on review of classification algorithms for EEG based BCI systems. This paper proposed that band power features are efficient for classification of motor imagery tasks [6]. Ali Bashashati et al carried out a study in which they surveyed signal processing algorithms in EEG based Brain Computer Interface systems. This survey gave information about signal processing method used for certain BCI system [7]. C. Guger et al carried out a study to demonstrate the technique of common spatial patterns for analysis of EEG in real time and giving feedback to the subject. The technique was used to enable accurate and fast feedback while carrying out a imagery left and right hand movement task [8]. YI Fang et al presented a study on improved classification methods for BCI systems. In this study they used event related synchronization and event related desynchronization as features for classification of EEG data of left and right hand motor imagery task [9]. GertPfurtscheller et al published a study on trends in BCI research which involved classification of electroencephalographic patterns while performing five different mental tasks.ERD and ERS patterns were detected during the imagery motor task [10].

3. Dataset Description

The data used in the study was obtained from the physionet database which is an online repository consisting of various types of data for study and research. The data is defined as EEG motor movement/imagery data and it consists of two minutes EEG recordings of 109 subjects. These subjects were asked to perform different motor/imagery tasks and EEG signals were recorded using 64 electrodes that were placed on the scalp of the individual [11-12]. The subjects performed various tasks during the experiment but in our study we have used the data for the following tasks:

1. A cue is displayed on either the left side of the monitor or the right side of the monitor. The subject opens and closes the left hand if cue is on the left side and right hand if the cue is on the right side. When the cue vanishes the subject gets into a relaxed position.
2. A cue is displayed on either the left side of the monitor or the right side of the monitor. The subject imagines opening and closing the left hand if cue is on the left side and right hand if cue is on the right side. When the cue vanishes the subject gets into relaxed position.

In this study we have used data for the first forty subjects out of the 109 subjects. These subjects are marked as S1, S2, S3 and so on in the paper.

4. Event Extraction

Motor movement/imagery EEG data was downloaded in CSV format from the Physiobank ATM. It is an online toolbox to access the data stored in the Physionet database using web browser.

The dataset consists of annotations which correspond to the types of events performed by the subjects during the experiment. There are three types of annotations [T0,T1,T2] which are defined as follows:
- **T0** corresponds to the rest event [i.e. subject is at complete rest and not performing or imagining any type of motor movement task].

- **T1** represents the start of actual or imagined motion of the left hand [i.e. the subject performs or imagines performing movement of left hand].
- **T2** represents the start of actual or imagined motion of the right hand [i.e. the subject performs or imagines performing movement of right hand].

EDF Browser was used as a tool to extract all the annotations/events from the two minute EEG recording of all the forty subjects. EDF Browser provided following information about the events/annotations:
- The time period of the start of each of the annotation
- The time duration for each of the annotation.

5. Classification

The annotations obtained using the EDF Browser were used to create the output label for running the classifier for each of the forty subjects. The three types of classes for the classifier are **T0, T1, and T2**. The data from all the 64 electrodes was used for the classification task and the column containing the annotations was used as the output channel for the classifier. **Random Forest classifier** was used for the classification task. Random Forest classifier is an ensemble algorithm that works by creating a set of decision trees from the subset of training dataset [13]. This subset is selected randomly. The votes for each of the decision tree are calculated and then the votes from different decision trees are aggregated to decide the final class of the object under test. Random Forest classifier was used in the study because of the following reasons:

It is one of the most accurate algorithms and produces high accuracy for different types of datasets.
- Its efficiency on large datasets is high
- It is capable of handling many input variables
- It also gives estimation of important variables for the classification task.

6. Results and Discussions

The result of the Random Forest Classifier for each of the 40 subjects is described in the table below. The values of all the parameters in the table are weighted average of all the three classes **T0**, **T1** and **T2**.

Table 1: Performance of Random Forest Classifier for all the classes

Subject	Accuracy	TP Rate	Precision	ROC Area
S1	75.95%	0.760	0.803	0.929
S2	56.07%	0.561	0.598	0.727
S3	84.26%	0.843	0.962	0.968
S4	98.13%	0.982	0.982	0.999
S5	53.25%	0.533	0.574	0.682
S6	55.25%	0.552	0.557	0.72
S7	53.75%	0.538	0.51	0.679
S8	93.80%	0.938	0.939	0.992
S9	76.23%	0.762	0.790	0.920
S10	98.65%	0.987	0.987	1
S11	97.82%	0.978	0.979	0.999
S12	79.23%	0.792	0.808	0.932
S13	58.29%	0.583	0.593	0.75
S14	55.49%	0.555	0.591	0.739
S15	60.44%	0.604	0.627	0.792
S16	94.01%	0.940	0.943	0.994
S17	90.19%	0.902	0.909	0.983
S18	88.19%	0.882	0.887	0.973
S19	58.12%	0.581	0.588	0.766
S20	97.01%	0.97	0.971	0.998
S21	59.39%	0.594	0.642	0.753
S22	89.42%	0.894	0.902	0.983
S23	57.55%	0.576	0.629	0.742
S24	51.73%	0.517	0.485	0.64

S25	94.92%	0.949	0.951	0.994
S26	88.69%	0.887	0.896	0.978
S27	96.65%	0.967	0.967	0.998
S28	94.12%	0.941	0.943	0.992
S29	83.29%	0.833	0.841	0.954
S30	84.96%	0.85	0.861	0.962
S31	52.12%	0.521	0.492	0.653
S32	67.12%	0.671	0.708	0.858
S33	60.87%	0.609	0.654	0.801
S34	98.49%	0.985	0.985	0.999
S35	52.58%	0.526	0.488	0.627
S36	63.07%	0.631	0.642	0.816
S37	84%	0.840	0.859	0.962
S38	53.69%	0.537	0.611	0.710
S39	97.24%	0.972	0.972	0.998
S40	96.65%	0.967	0.968	0.999

Precision: Precision is defined as the fraction of relevant instances among the total number of selected instances. For example, in the table 1 precision for subject 3 is 0.962. It means that 96.2% of selected instances are relevant.

TP Rate: True positive rate corresponds to the fraction of the positive instances that are correctly considered as positive with respect to all the positive instances. For example, the TP rate for subject16 is 0.94. It means that 94% of instances were correctly classified.

ROC Area: It is the area under the ROC curve and it is the measure of the accuracy of the classifier. High ROC area represents high and it also gives information about the worthiness of the test. If ROC area is 1 it a perfect test and if ROC area is 0.5 it is a worthless test. In the table above roc area was less than 0.7 in 5 subjects out of 40. It means that for these 5 subjects the classification test performance is poor. For other subjects the test performance is good. The test did not fail for any of the subject as the ROC area was greater than 0.6 for all the subjects. Figure 1 shows the graph of the accuracy of random forest classifier for all the 40 subjects.
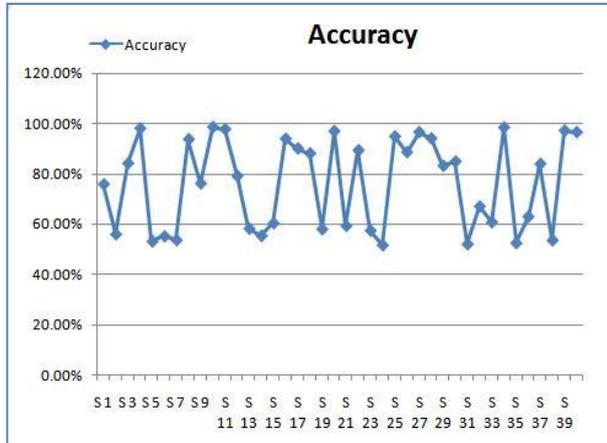

Fig. 1: Accuracy of Random forest classifier for all the 40 subjects

Info Gain attribute selection algorithm was used to reduce the number of electrodes. After running the algorithm all the 64 electrodes were ranked from 1 to 64.In the List 1 top 18 ranked electrodes were considered. Similarly 3 other lists were prepared. List 2 consists of top 33 electrodes, List 3 consists of top 48 electrodes and List 4 consists of all the 64 electrodes[i.e. in each subsequent list 15 electrodes were added based on their rank in the result of attribute selection algorithm].The details of all the 4 lists are given in the Table 2.

Table 2: Information about the electrodes for all the four lists

List	Electrodes in the list	Number of electrodes
List 1	Fp1,Fpz,Fp2,Af7,Af3,Afz,Af4,Af8,F7,F5,F3,F1,Fz,F2,F4,F6,F8,T9	18
List 2	Fc5,Fc3,Fc1,Fc2,C5,C3,Cp1,Fp1,Fpz,Fp2,Af7,Af3,Afz,Af4,Af8,F7,F5,F3,F	33
	1,Fz,F2,F4,F6,F8,Ft7,T7,T9,T10,Tp7,Tp8,P8,Po7,O2	
List 3	Fc5,Fc3,Fc1,Fc2,C5,C3,C1,Cp3,Cp1,Cpz,Cp2,Cp4,Fp1,Fpz,Fp2,Af7,Af3,Afz,Af4,Af8,F7,F5,F3,F1,Fz,F2,F4,F6,F8,Ft7,Ft8,T7,T8,T9,T10,Tp7,Tp8,P5,P3,P2.P6,P8,Po7,Poz,O1,Oz,O2,Iz	48
List 4	All the 64 electrodes	64

Random forest classifier was implemented for all the 4 lists and the result of the classifier for all of the subjects is given in the Table 3. The accuracy of the classifier for List 3 and List 4 is almost same for most of the subjects whereas there is some amount of difference between the accuracy of the classifier for List 2 and List4. Figure 2 and 3 shows the comparison between accuracy of the classifier between List 4 and List3 and List 4 and List2 respectively. The average difference of the accuracy of the classifier is **0.9%** between List 4 and List 3 and **2.4%** between List 4 and List 2. The difference between the accuracies is very minimal for List 3and List 4. This indicates that the random forest classifier is able to give almost the same accuracy even if the data recorded from the electrodes in the List 3are considered and remaining electrodes are rejected. Therefore the numbers of electrodes were reduced from 64 to 48. This reduction in the number of electrodes can reduce the amount of data and time required to pre-process and classify that data.
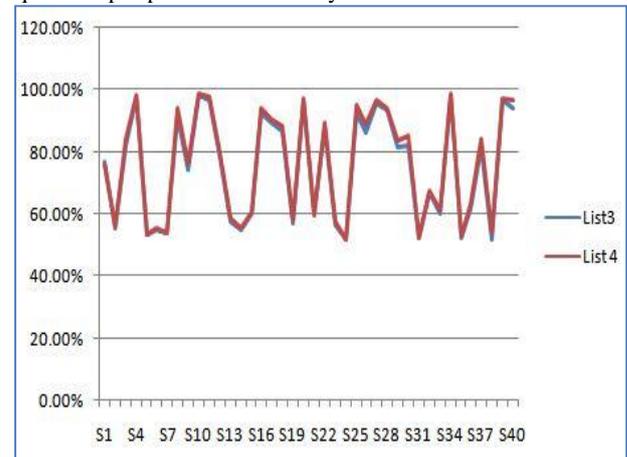

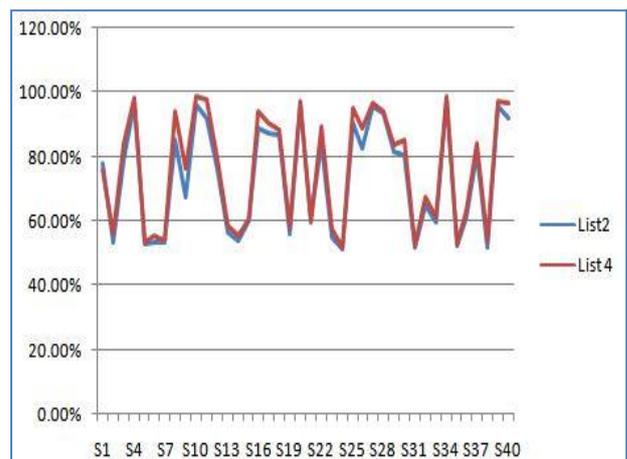**Fig. 2:** Comparison of the accuracy of Random forest classifier for List 3 and List 4

Fig. 3: Comparison of the accuracy of Random forest classifier for List 2 and List 4

It can be observed from Figure 2 that the accuracy of the classifier is same for List 3 and List 4 for almost the subjects whereas in Figure 3 the accuracy of the classifier varies for List 2 and List 4 for some of the subjects. On the basis of these comparisons and analysis 48 electrodes in the List 3 were considered and remaining electrodes of the List 4 were rejected.

Table 3: Accuracy of the Random Forest Classifier for all the four lists for all the 40 subjects

Subject	Accuracy of Random Forest Classifier [%]			
	List 1	List 2	List 3	List 4
S1	78.11	78.19	76.89	75.95
S2	53.13	53.7	55.39	56.07
S3	75.52	79.39	81.95	84.26
S4	94.27	96.89	97.91	98.13
S5	50.25	52.81	53.01	53.25
S6	50.97	53.35	54.74	55.25
S7	52.71	53.79	53.68	53.75
S8	72.29	85.26	91.81	93.8
S9	65.39	67.74	74.31	76.23
S10	91.51	96.37	98.46	98.65
S11	83.99	92.01	96.9	97.82
S12	70.46	76.36	78.61	79.23
S13	54.41	56.83	57.24	58.29
S14	51.57	54.02	54.58	55.49
S15	62.93	60.6	59.75	60.44
S16	84.5	89.22	92.72	94.01
S17	85.21	87.35	89.45	90.19
S18	84.46	86.71	87.12	88.19
S19	53.44	56.22	56.68	58.12
S20	95.08	96.2	96.71	97.01
S21	58.46	60.61	60.02	59.39
S22	72.45	84.36	87.83	89.42
S23	52.82	55.35	56.37	57.55
S24	50.81	51.47	51.54	51.73
S25	83.6	90.79	92.87	94.92
S26	76.86	82.88	86.42	88.65
S27	95.33	95.9	96.04	96.65
S28	93.99	93.52	93.96	94.12
S29	75.54	81.6	81.49	83.29
S30	76	80.82	82.19	84.96
S31	52.2	51.95	51.85	52.12
S32	64.51	65	66.61	67.12
S33	58.17	59.62	60	60.87
S34	94.99	97.92	98.31	98.49
S35	50.6	52.27	52.32	52.58
S36	60.37	61.58	61.74	63.07
S37	79.07	80.98	81.81	84
S38	50.41	51.78	51.9	53.69
S39	93.9	95.96	96.65	97.24
S40	84.29	92.31	94.49	96.65

7. Conclusion

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. The objective of this study was to find out the performance of the random forest classifier on the EEG data and reduce the number of electrodes so that the computational effort of the classifier can be reduced. In our experiment we have used Weka as a tool for analysis of executed and imagined EEG data. The performance/accuracy of the random forest classifier is found to be high in around **60%** of the total subjects under study and was better than other classifiers such as J48 and Random Tree. Further InfoGain attribute selection is used to reduce the electrodes. The algorithm ranked the attributes based on their information gain with respect to the class. The data set is then divided into four lists based on the order of the attributes in the result of attribute selection algorithm. List 1 consists of top 18 electrodes and next fifteen ranked electrodes were added in the subsequent lists. List 4 consists of all the 64 electrodes. Table 2 presents the details of all the lists. Classification accuracy using Random forest classifier each of the list was compared. Based on the results of classification accuracy study concludes that only the electrodes in List 3 can be considered for the classification task while rest of the 16 electrodes [i.e. electrodes present in the List 4 but not present in the List 3] can be ignored as the average difference of the accuracy of the classifier is **0.9** between List 4

and List 3 which is very nominal and no substantial change in the accuracy of the classifier was observed after the reduction of additional electrodes present in the List 4. Through this study authors reduced the number of electrodes from 64 to 48 without substantial reduction in the accuracy of the classifier.

References

[1] Vidal, J. J. "Real-time detection of brain events in EEG." Proceedings of the IEEE, Vol. 65, Iss. 5, pp. 633-641. (1977).

[2] R. P. N. Rao and R. Scherer, "Brain-Computer Interfacing [In the Spotlight]," in IEEE Signal Processing Magazine, vol. 27, no. 4, pp. 152-150, July 2010. doi: 10.1109/MSP.20 10.936774 URL: http://i eeexplore.ieee .org/stamp/s tamp. jsp?tp=& arnumber=54 84181&isnu mber= 5 484155

[3] Brain Computer Interface Technology: A review of the first international meeting by Jonathon R. Wolpaw, Niels Birbaumer, William J Heetderks, Dennis J.McFarland, P. Hunter Peckhalm, Gerwin Schalk, Emanuel Donchin, Louis A.Quatrano, Charles J. Robinson, Theresa M. Vaughan

[4] Darius A. Rohani, William S. Henning, Carsten E. Thomsen, Troels W.Kjaer, S. Puthusserypady, Helge B.D. "Sorensen BCI using imaginary movements: The simulator," Computer Methods and Programs in Biomedicine, vol. 111, pp. 300-307, August 2013

[5] Thomas A. Deuel, Juan Pampin, Jacob Sundstrom, Felix Darvas. The Encephalophone: A Novel Musical Biofeedback Device using Conscious Control of Electroencephalogram (EEG). *Frontiers in Human Neuroscience*, 2017; 11OI: 10.3389/fnhum.2017.00213

[6] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEGbased brain-computer interfaces. Journalof Neural Engineering, 4:R1–R13, 2007

[7] Ali Bashashati, Mehrdad Fatourechi, Rabab K Ward and Gary E Birch. A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals: J. Neural Eng. 4 (2007) R32– R57.

[8] C. Guger, H. Ramoser, and G. Pfurtscheller. Real-Time EEG Analysis with Subject-Specific Spatial Patterns for a Brain–Computer Interface (BCI): IEEE TRANSACTIONS ON REHABILITATION ENGINEERING, VOL. 8, NO. 4, DECEMBER 2000

[9] YI Fang, LI Hao and JIN Xiaojie. Improved Classification Methods for Brain Computer Interface System: I. J. Computer Network and Information Security, 2012, 2, 15-21.Published Online March 2012 in MECS (http://www.mecs-press.org/).

[10] GertPfurtscheller et al. "Current trends in Graz braincomputer interface BCI research". IEEE Transactions on Rehabilitation Engineering, 8(2):216–219, 2000

[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, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Transactions on Biomedical Engineering 51(6):1034-1043, 2004.

[13] Brieman L, Random Forests, (2001): Machine Learning, 45, 5-32.