Global Journal of Information Technology: Emerging Technologies

Volume 06, Issue 3, (2016) 158-165

www.gjit.eu

An efficient algoritm for classification of EEG eye state data

Efendi Nasibov*, Department of Computer Science, Dokuz Eylul University, 35390, Izmir, Turkey
Alican Dogan, Department of Computer Science, Dokuz Eylul University, 35390, Izmir, Turkey

Suggested Citation

Nasibov, E. & Dogan, A. (2016). An Efficient Algoritm for Classification of EEG Eye State Data. *Global Journal of Information Technology: Emerging Technologies. 6*(3), 158-165.

Received May 17, 2016; revised August 24, 2016; accepted September 5, 2016.
Selection and peer review under responsibility of Prof. Dr. Dogan Ibrahim, Near East University, Cyprus.

Abstract

A recent technology which makes possible for us to interact with automated systems without using any body part is called Brain Computer Interface (BCI). In its concrete applications, electroencephalogram (EEG) is benefited by a BCI environment for being capable of obtaining brain waves. In our study, evaluation of success rates of the predictions made by C x k - Nearest Neighborhood (Cxk-NN) Algorithm for EEG Eye State Data whose states are called "Opened Eye" and "Closed Eye" is applied. This EEG Eye State dataset is obtained from UCI Machine Learning Repository on the web and it is a highly-used benchmark data on this field. As there are only two classes of the signals, we test binary classification performance of our classification algorithm (Cxk –NN). Comparison of those values with the ones obtained by the other successful classification algorithms in the literature applied on the same data set also take place in our study. Cxk-NN is an instance-based classification method advanced from simple k – Nearest Neighborhood Algorithm, and improved success results are observed when it is compared with k-NN.

Keywords: brain computer interface (bci), classificaiton, eeg, c x k - nearest neighborhood algorihtm.

*ADDRESS FOR CORRESPONDENCE: **Efendi Nasibov**, Department of Computer Science, Dokuz Eylul University, 35390, Izmir, Turkey.
E-mail address: efendi.nasibov@deu.edu.tr

I. INTRODUCTION

The brain which is the center of decision and balance has a complicated structure. Especially, human brain has a more difficult to understand structure than other animal brains. The systems that enable people to give instructions to digital computerized mechanisms without using anybody's limb, are aimed at forming a communication bridge between human brain and the device being commanded others (He, 2005; Schalk, McFarland, Hinterberger, Birbaumer & Wolpaw, 2016; Wolpaw, Birbaumer, Heetderks, McFarland, Peckham, Schalk & Vaughan, 2000).

The main input elements of this research field which is called as generally BCI (Brain Computer Interface) in short in the literature are human thoughts. The idea of being able to read those thoughts as inputs is based on the emergence of electrical signals in some parts of the human brain when a person imagines something in his mind. Consequently, those types of signals should be recognized in an electronic environment. These systems are quite beneficial especially for patients of ALS and people suffering from some sort of strokes others (He, 2005; Schalk, McFarland, Hinterberger, Birbaumer & Wolpaw, 2016; Wolpaw, Birbaumer, Heetderks, McFarland, Peckham, Schalk & Vaughan, 2000).

In today's world, a lot of methods such as magneto encephalography, electrocorticography, near infrared spectroscopy, local area potentials, functional magnetic resonance visualization, EEG and single cell records are used in BCI. Among those techniques, undoubtedly, EEG is the most widespread one because it provides more ease-of-use than the others (He, 2005).

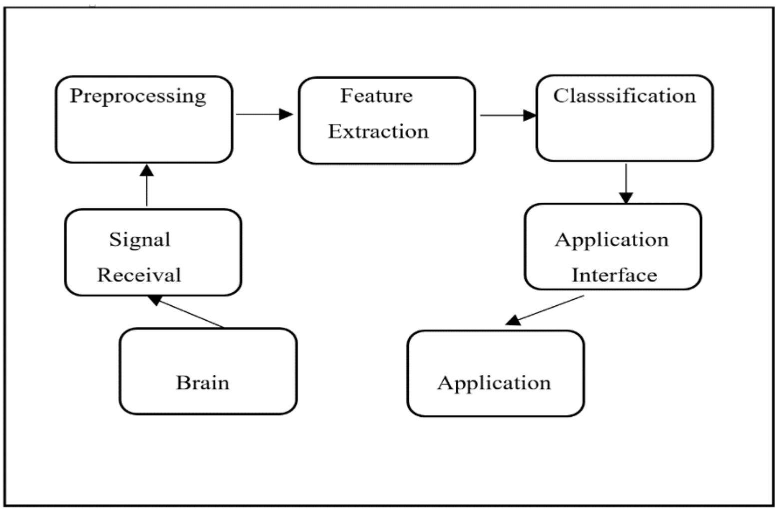
EEG has a capacity to make measurements 1000 times in a second (once in a millisecond). It detects electrical change in channels linked to the electrodes placed over some parts of the sculp. The Potential Difference between two extreme sides of electrodes those are measured as the unit of mV (millivolt) and this alteration is recorded depending on time. Brain signals are taken over hair skin since it does not cause any pain on the subject others (He, 2005).

II. THE PHASES OF BCI

A. Preprocessing

Preprocessing is one the basic processes of BCI (Fig. 1). The signals taken from the brain are not at the sufficient level to be ready for feature extraction phase as preprocessing operations are inevitable. EEG signals do not only transmit neurological information produced in thinking process, but they also convey data from the sources in which we are not interested and these data are known as noises. It is quite difficult to make an operation on the data combining of many sources of noise. Noises involved in EEG signals should be eliminated by being exposed to a preprocessing operation (Aydemir & Kayikcioglu, 2011).

Figure 1. The Basic Functional Elements of BCI others (Aydemir & Kayikcioglu, 2011).



Because of the fact that EEG signals have very low amplitudes, noises easily come to appear on those signals. They get exposed to be affected by external elements like the existence of devices which emit electromagnetic waves in the environment or light spreading from a lamp and physical movements such as moving an arm from left to right, heartbeats and eye-blinking.

In the literature, there are plenty of efficient noise elimination techniques. The most common ones among them are Independent Component Analysis, Linear or Non-linear Filtering, Wavelet Transformation, Principal Component Analysis, Fourier Transformation and Source Dipol Analysis. The most critical source of noise in BCI is based on eye-blinking. This type of noise is especially detected on the electrodes located in brain rear lobes which are near to the eyes others (He, 2005).

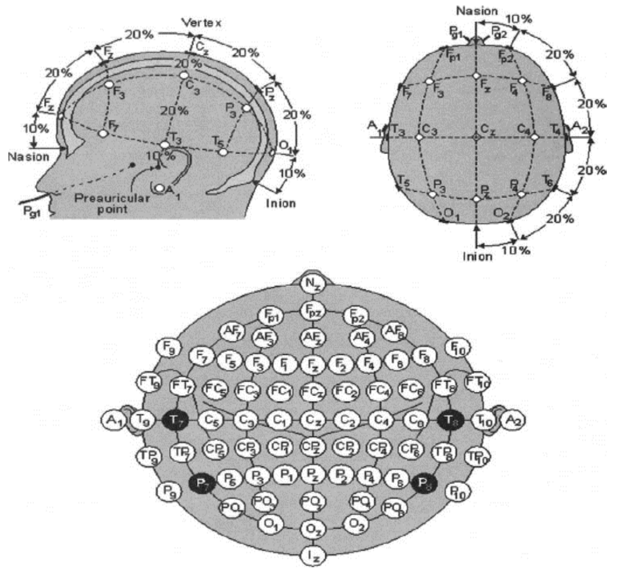
The other preprocessing operations apart from noise elimination methods are taking average, determining a threshold value, and normalization.

B. Feature Extraction

Feature extraction is the operation of obtaining features after operation of determining important attributes of recorded EEG signals when thoughts differ. For example, when a disabled individual who cannot use his hands in a normal way thinks of changing position of a cursor on computer, how EEG signals get in shape becomes a feature extraction operation. Each generated feature vector can complete a signal, but in the classification phase; none of them may not give a correct result. Therefore, distinct characteristics should be used together to acquire a high success ratio. Each extracted feature may not define a signal and none of them may give 100% correct response in the classification phase. For this reason, different features should be used together in order to obtain a high success ratio others (He, 2005).

Another fundamental point is to decide to choose the most suitable electrode in order to extract features when making an analysis on EEG data recorded via multiple electrodes. Different parts of the brain are more sensible for different sorts of task. For instance, hearing senses are processed in the occipital lobe of the brain. As a result, it would be a better idea to utilize data taken from the parts of the brain where cerebral activity is much more intensive instead of benefiting from the whole data coming from multiple electrodes covering all parts of the brain because this situation reduces time spent in the classification phase when decisions should be taken and provide opportunity to find the best features for classification.

Figure 2. The Placement of Electrodes to Receive Non-invasive EEG Signals others (He, 2005).



In Figure 2, the application of International 10/20 System in the operation of placing electrodes is given. The letters written on each electrode represent a special sub-lobe. Those letters and their meaning are listed below.

- FP: Prefrontal Lobe
- F: Frontal Lobe
- T: Temporal Lobe
- C: Central Lobe
- P: Parietal Lobe
- O: Occipital Lobe.

The number that was shown at sub index, points out the location of electrode in brain hemisphere. Z symbolizes the exact middle line of brain, odd numbers symbolize somewhere in left hemisphere and even numbers symbolize a place over right hemisphere. The nearer the distance between the electrode and middle line, the less number gets closer to 0 others (He, 2005).

C. Classification

Dividing EEG signals recorded in diverse visual and mental situations into classes according to their feature vectors is a significant issue in BCI. The decision of which signal should belong to which class must be rapid and correct. There are many different classification methods in the literature. Among the most widely used ones of them, Support Vector Machines, k-Nearest Neighborhood, Linear Discriminant Analysis, Neural Networks and Bayesian Classification take place (Aydemir & Kayikcioglu, 2011).

We have made a classification over EEG signals labeled to Eye-Blinking Situation signals during an experiment of a subject and recorded using an EEG signal acquisition device called Emotiv having 14 channels that are found in UCI Machine Learning Repository by applying C x K – Nearest Neighborhood Algorithm proposed by Ulutagay and Nasibov which is similar to k-Nearest Neighborhood Algorithm in this study. We have detected the ratio of belonging to correct class of the data by sending single signals that we had already know to which class they belong to C x K – Nearest Neighborhood Algorithm as test data. We have got successful results in these data including binary classification for EEG signals. The aim of this research we made is to demonstrate that C x K – Nearest Algorithm can make a more successful classification than k-Nearest Neighborhood Algorithm, which is a basic algorithm, when classifying binary EEG signals. At the same time, we aimed at testing reliability of the algorithm which we will require for multiclassification over EEG signals in our other studies (Dogan & Nasibov, 2016; Nasibov, Ozgoren, Ulutagay, Oniz & Kocaaslan, 2010; Nasibov & Ulutagay, 2010; Ulutagay & Nasibov, 2016; Ulutagay & Nasibov, 2009).

III. C x K - NEAREST NEIGHBORHOOD ALGORITM

The C x K – Nearest Neighborhood approach is an improved version of the well-known k- Nearest Neighborhood (KNN) algorithm. In this approach, the value whose class is not obvious is put into a class by taking the distances of its k-nearest neighbors from each class into consideration. First of all, all distances among all nodes are calculated mathematically, so how close or how far a node is to one another becomes clear. *K* nearest samples from each class to node having no determined class, totally *C x K* number of nodes are analyzed. The letter C here represents the number of classes. The name of the algorithm comes from this expression. The average of distances between a node to be classified and k-nearest nodes to this node is calculated. Thus, average distance between the node that does not have a clear class yet and each class is obtained. Among those distances, the class having the nearest distance is labeled as new class of the node to be classified (Dogan & Nasibov, 2016; Nasibov &Ulutagay, 2010).

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of n labeled samples, $C_j = \{x_1^j, x_2^j, \ldots, x_{n_j}^j\}$ $j = 1, 2, \ldots, C$ be the priori known classes, n_j be the number of elements in class C_j where $n_1 + n_2 + \cdots + n_c = n$. The pseudo-code of C x K - Nearest Neighborhood Algorithm is given below (Ulutagay & Nasibov, 2016; 2009).

CxKNN Algorithm:
BEGIN
Input *x* of unclassified data,
Set *K*, $1 \le K \le n$
FOR EACH CLASS C_j
DO UNTIL (*K* – nearest neighbor found in class C_j)
Set $i = 1$.
Calculate the distance between x and
IF ($i \le K$) **THEN**
Assign $x_i^{c_j}$ to the set of *K* – nearest neighbors in C_j
ELSE IF ($x_i^{c_j}$ is closer to *y* than any other previous neighbor)
THEN
Delete the farthest sample in the set of *K* nearest nighbors.
Assign $x_i^{c_j}$ to the set of *K* – nearest neighbors in C_j.
END IF
$i = i + 1$
END DO
Calculate the average distance d_j of *x* from *K* – nearest neighbors of class C_j.
END FOR
Mark the class with minimum distance $d_x = min_j d_j$;
Classify *x* in the class *r* of the last minimum found.
END (Ulutagay & Nasibov, 2016; 2009).

For example, according to the algorithm, if we have two classes with the names of C1 and C2 and k = 3, then we look at 3 data which are nearest to the node from class C1 and 3 nearest data from class C2, 6 nodes in total. We take the average of distances from the node whose class is unknown to the 3 nearest nodes from each class. If the one belonging to C1 from those averages we found is less the new class of the node becomes C1. Otherwise, it becomes C2. The functioning principle of C x K – Nearest Neighborhood Algorithm is in this way. If k = 1, there will be no need to take average. The class having a nearest node becomes the new class of the node. That is to say, this class becomes the class that contains this node. In this case, the same results with k – Nearest Neighborhood Algorithm are acquired.

IV. DATASET

Our dataset is EEG Eye State Dataset (https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State#) taken from UCI Machine Learning Repository and comprises EEG signals recorded during a specific time period (117 seconds) when an individual subject opens and closes his eyes. “Eye State” attribute is added in the last column of these EEG data and the value of this attribute with respect to each single data is assigned to 0 if subject's eyes are open, to 1 if they are closed (Nasibov, Ozgoren, Ulutagay, Oniz & Kocaaslan, 2010).

EEG signals are obtained as saved records to a text file including many rows. Each row refers to a single node. Totally 14980 data are recorded consecutively to a text file depending on time. In coding process, we have converted these data on a text file into a proper format in order for the values of each signal in different electrodes to be read in a correct way. As explained in UCI Machine Learning Data Repository, these signals are acquired by utilizing Emotiv EEG Signal Record Device (Nasibov, Ozgoren, Ulutagay, Oniz & Kocaaslan, 2010).

V. EXPERIMENTAL STUDY

At first, learning data, independent from test data, are divided into 10 parts having equal sizes randomly and in such a way that they constitute a homogenous distribution from all data belonging to both two classes for 10-folds cross validation process. Afterwards, data from each fold are considered like test data and success ratios using C x K – Nearest Neighborhood and K – Nearest Neighborhood algorithms with best possible k values. In Table 1, the results which belong to C x K are demonstrated and in Table 2, the ones for K-Nearest for the test dataset are shown. A comparative result for folds averages is given in Figure 4 (Dogan & Nasibov, 2016).

Table 1. The success ratios for each fold during 10-folds cross validation phase for all k values between 1 and 10 using C x K – NN Algorithm.

Folds/K	1	2	3	4	5	6	7	8	9	10	Avg.
1	0,958	0,968	0,971	0,972	0,956	0,971	0,972	0,962	0,971	0,966	0,967
2	0,966	0,971	0,970	0,972	0,958	0,970	0,978	0,967	0,975	0,967	0,969
3	0,967	0,969	0,971	0,969	0,961	0,970	0,977	0,968	0,971	0,968	0,969
4	0,970	0,971	0,969	0,964	0,960	0,969	0,978	0,970	0,972	0,965	0,969
5	0,966	0,972	0,967	0,962	0,959	0,968	0,977	0,969	0,971	0,963	0,967
6	0,968	0,970	0,965	0,958	0,958	0,967	0,974	0,970	0,971	0,963	0,966
7	0,968	0,969	0,963	0,962	0,957	0,966	0,973	0,967	0,971	0,962	0,966
8	0,961	0,967	0,961	0,959	0,957	0,968	0,971	0,965	0,967	0,962	0,964
9	0,958	0,965	0,959	0,958	0,957	0,964	0,967	0,960	0,964	0,960	0,961
10	0,957	0,965	0,955	0,956	0,943	0,964	0,964	0,959	0,958	0,960	0,958

Table 2. The success ratios for each fold during 10-folds cross validation phase for all k values between 1 and 10 using K–NN Algorithm.

Folds/K	1	2	3	4	5	6	7	8	9	10	Avg.
1	0,958	0,968	0,971	0,972	0,956	0,971	0,972	0,962	0,971	0,966	0,967
2	0,957	0,952	0,952	0,961	0,944	0,955	0,965	0,953	0,964	0,950	0,955
3	0,959	0,967	0,964	0,964	0,956	0,964	0,975	0,962	0,956	0,961	0,963
4	0,953	0,958	0,954	0,953	0,950	0,960	0,962	0,950	0,946	0,954	0,954
5	0,951	0,963	0,961	0,951	0,955	0,960	0,967	0,962	0,952	0,956	0,958
6	0,946	0,955	0,948	0,943	0,946	0,946	0,957	0,952	0,949	0,955	0,950
7	0,947	0,960	0,948	0,944	0,943	0,956	0,960	0,956	0,947	0,952	0,951
8	0,945	0,952	0,940	0,937	0,939	0,953	0,951	0,952	0,948	0,947	0,946
9	0,941	0,961	0,937	0,938	0,946	0,954	0,957	0,950	0,949	0,954	0,949
10	0,935	0,951	0,932	0,930	0,943	0,948	0,952	0,948	0,942	0,947	0,943

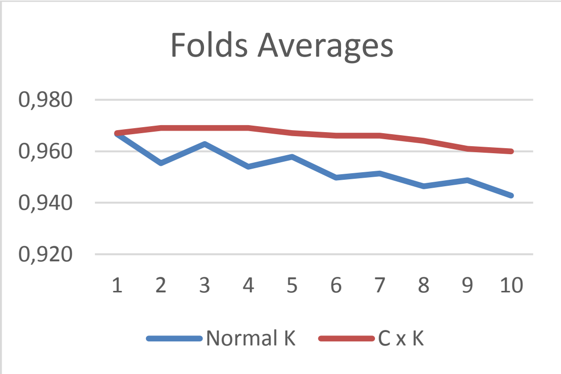
As can be seen clearly in Table 1, Table 2 and Fig. 3, although “Open Eye” data outnumbers the other class, data from each class are taken in similar numbers to each other and homogenous distribution is not ignored in this way.

When we observe the result for other k values, we notice that C x K - NN algorithm provides an improvement having an increase up to 2% in the comparison of success results. Also, when we take a look at the best k values, in the process of cross validation, the values of 2, 3 and 4 with the success ratio of 0,969 for C x K-NN algorithm and the value of 1 with the success ratio of 0,967 for K-Nearest Neighborhood Algorithm are seen clearly as the best values (Dogan & Nasibov, 2016).

VI. CONCLUSION

We have noticed in our study that C x K – Nearest Neighborhood Algorithm has given a quite successful results for Eye State Dataset which is a sample of a binary EEG signals. Those promising outputs we obtained in this experiment has motivated us to get significant classification success with C x K – NN algorithm for other BCI studies.

Figure 3. The graph making a comparison for average success ratios in all folds between K – NN and C x K – NN algorithms during 10-folds cross validation phase for all k values between 1 and 10.



REFERENCES

Aydemir, O., & Kayikcioglu, T. (2011). Classification of EEG signals recorded during right/left hand movement imagery using Fourier Transform based features. Signal Processing and Communications Applications (SIU), *IEEE 19th Conference on. IEEE*, 415-418.

Dogan, A., & Nasibov, E. (2016). *Classification of EEG Eye State Signals Using C x K – Nearest Neighborhood Algorithm,* In Proceedings of International Conference on Computer Science and Engineering, 20-23 October Tekirdag, Turkey, 696-701.

He, B. (2005). *Neural Engineering.* Kluwer Academic / Plenum Publishers, Minnesota, 85-95.

Machine Learning Repository.(2013)Retrieved from; https://archive.ics.uci.edu/ml/datasets/EEG+Eye+State#, “Last Access: 20.06.2016”,

Nasibov E., Ozgoren M., Ulutagay G., Oniz A., & Kocaaslan S. (2010). On the analysis of BIS stage epochs via fuzzy clustering. *Biomedizinische Technik/Biomedical Engineering, 55*(3), 147–153.

Nasibov,E., & Ulutagay,G. (2010). Comparative clustering analysis of bispectral index series of brain activity, Expert Systems with Applications, 37, 2495-2504.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer,N., & Wolpaw, J. R. (2016). BCI2000: a general-purpose brain.

Ulutagay,G., & Nasibov, E. (2016). C x K Nearest Neighbor Classification with Ordered Weighted Average Distance. Novel Applications of Intelligent Systems, 586, 105-122.

Ulutagay G., & Nasibov, E. (2009). Detection of BIS stage levels via fuzzy clustering approach, Biomedical Engineering Meeting (BIYOMUT), IEEE Xplore Digital Library

Ulutagay, G., & Nasibov, E. (2012). *A New CxK-Nearest Neighbor Linkage Approach to The Classification Problem.* In Proceedings of the 10th International FLINS Conference, Istanbul, Turkey, 1, 471-476.

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., & Vaughan, T. M. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE transactions on rehabilitation engineering, 8*(2), 164-173.