www.dergipark.gov.tr
ISSN:2148-3736
El-Cezerî Fen ve Mühendislik Dergisi
Cilt: 5, No: 2, 2018 (591-595)

El-Cezerî Journal of Science and Engineering
Vol: 5, No: 2, 2018 (591-595)

ECJSE

Research Paper / Makale

Examining the Effect of Dimension Reduction on EEG Signals by K-Nearest Neighbors Algorithm

Duygu Kaya[1*], Mustafa Türk[2], Turgay Kaya[3]

[1,2,3] Firat University, Faculty of Engineering, Department of Electrical and Electronics Engineering, Elazig, Turkey
dgur@firat.edu.tr, mturk@firat.edu.tr, tkaya@firat.edu.tr,

Received/Geliş: 29.01.2018 **Revised/Düzeltme:** - **Accepted/Kabul:** 01.02.2018

Abstract: Machine learning which a paradigm of methods that makes inferences from existing data using mathematical and statistical methods and is inferred to be unknown. The proposed method in this paper, supervised learning algorithm is applied to EEG (electroencephalography) data and classification algorithm performance is analyzed and results are examined in MATLAB. K-Nearest Neighbors Algorithm (k-NN) is used in this paper as algorithm. This classification was evaluated in two stages, with and without Principal Component Analysis (PCA). Dimension reduction is the process of reducing the size of dimension of the data. By reducing the size of the data set with PCA, it is expected to protect important data features. KNN has given results that can be regarded as prudent in terms of classification accuracy. The results of the present work showed that appropriate features combined with classifier can be done significant classification for different bioelectrical signal.

Keywords: Supervised learning algorithms; k-nearest neighbors algorithm (kNN); Principal Component Analysis (PCA); Dimension reduction; EEG

En Yakın Komşu Algoritması Kullanılarak EEG Sinyallerine Boyut Azaltmanın Etkilerinin İncelenmesi

Öz: Makine öğrenmesi, var olan verilerin çıkarımlarını matematiksel ve istatistiksel yöntemlerle yapan ve bilinmeyen bir yöntem paradigmasıdır. Bu çalışmada, denetimli öğrenme algoritması, EEG (elektroensefalografi) verilerine uygulanmış, sınıflandırma algoritması performans analiz sonuçları MATLAB ile incelenmiştir. Bu çalışmada, algoritma olarak en yakın komşu algoritması (k-NN) kullanılmıştır. Bu sınıflandırma, Temel Bileşen Analizinin (TBA) kullanıldığı ve kullanılmadığı durumlar için iki aşamada değerlendirilmiştir. Boyut azaltma, verilerin boyut boyutunu küçültme işlemidir. TBA ile veri kümesinin boyutunun azaltılarak, önemli veri özelliklerini korunması beklenir. KNN, sınıflandırma doğruluğu açısından önemli sayılabilecek sonuçlar vermiştir. Mevcut çalışma, farklı biyoelektriksel sinyaller için uygun özelliklerin uygun bir sınıflandırıcı ile kombine edildiğinde anlamlı bir sınıflandırma yapılabileceğini göstermiştir.

Anahtar kelimeler: Denetimli öğrenme algoritması; En yakın komşu algoritması (kNN); Temel Bileşen Analizi; Boyut Azaltma; EEG

1. Introduction

The brain is the center of the nervous system, which insures body control and regulation of the activities, interprets sensory orders and is transmitted to the relevant muscles and organs. During the

How to cite this article
Kaya D. Turk M., Kaya T., "*Examining the Effect of Dimension Reduction on EEG Signals by K-Nearest Neighbors Algorithm*" El-Cezerî Journal of Science and Engineering, 2018, 5(2); 591-595.

Bu makaleye atıf yapmak için
Kaya D. Turk M., Kaya T,, "En Yakın Komşu Algoritması Kullanılarak EEG Sinyallerine Boyut Azaltmanın Etkilerinin İncelenmesi" El-Cezerî Fen ve Mühendislik Dergisi 2018, 5(2); 591-595.

brain activity, both continuous rhythmic electrical potentials and receptor activity caused by electrical potentials occurs. The recording process of this electrical potential is called Electroencephalography (EEG). EEG signals can provide information on the neurological diseases [1-4], so that are used diagnostic indicator to investigate physiological conditions [5]. People who suffer from epilepsy to decide whether or not based on observation with long-term experience [6,7]. This process will take a lot of time in order to facilitate the interpretation of the signals. Also, the frequency spectrum of EEG signals contains important information for the interpretation of the signal. For classifying EEG signals, there are Neural Networks based works and machine learning algorithms in the literature [8-16]. The aim of study is to examine the effect of dimension reduction algorithms on EEG data to be classified by KNN.

2. Material and Method

Machine learning constitutes another branch of artificial neural networks and aims to identify unknown samples by learning from known sample. It is examined in two main topics as supervised learning algorithms and unsupervised learning algorithms. Each algorithm has its advantages and disadvantages in the way it is used [11-18]. Principle shame of supervised learning algorithm is shown in Figure 1. Data is firstly separated into to train and test sets, algorithm is chosen, data is tested and finally results are evaluated.

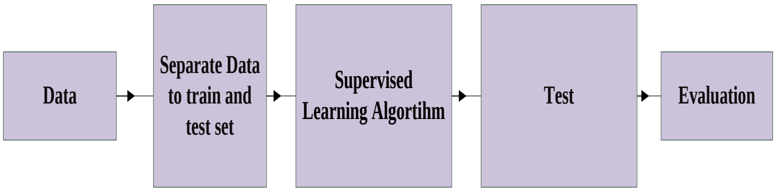

Figure 1. Principle shame of supervised learning algorithm

Dimension reduction is the process of reducing the size of dimension of the data. By reducing the size of the data set with dimension reduction algorithm, it is expected to protect important data features. Dimension reduction is an application for identifying and disposing of attributes that are unnecessary, while preserving the properties of the given data [19-21]. PCA is the most used method for dimension reduction as linear [24]. Results of PCA, actual size is determined. With PCA dimension is reduced, but there is no data loss. Principle shame of dimension reduction algorithm is shown in Figure 2.

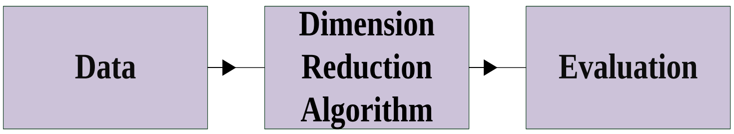

Figure 2. Principle shame of dimension reduction algorithm

There have been kinds of supervised learning algorithms, but in this study we use k-Nearest Neighbors Algorithm (k-NN) because of that's popularity in literature. K-NN is from the extracted characteristics, the closeness of the new individual desired to be categorized to k of the previous individuals is examined. The aim is to classify the existing learning data when a new instance arrives. When a new sample arrives, it looks at its nearest neighbor and decides for instance the class.

If we want to look at k number of samples for a new classifier, the closest k is taken from the previously classified elements. This distance, Euclidean, Manhattan, Chebyshev, Minkowski et al. distance formulas are calculated and the euclidean distance is usually used for real valued input variables. d (p, q) denoting the distance between two points [p,q],

$$d_{öklid}(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (1)$$

$$d_{Manhattan}(p,q) = \sum_{i=1}^{n}|p_i - q_i| \quad (2)$$

$$d_{Chebyshev}(p,q) = \max_i(|p_i - q_i|) \quad (3)$$

$$d_{Minkowski}(p,q) = \sqrt[m]{\sum_{i=1}^{n}|p_i - q_i|^m} \quad (4)$$

is expressed.

The distance of the test data to all learning samples is calculated, sorted by minimum distance, the class values to which the new sample belongs, and the dominant class is selected [22,23].

3. An Application

Data is evaluated by supervised machine learning algorithms. KNN is used to analyze the data. This work composes of two stages. Classification was done without PCA, later with PCA. The effects of PCA on the KNN algorithm have been investigated and the effectiveness of the k-NN algorithm in classification for the data used has been investigated.

Firstly, Confusion matrix and ROC curve are used to analyze classification accuracy. Confusion matrix shows training, validation and test results. Percentage error refers to the not classification of the sample rate. For ROC curve, the best diagnostic test, closer the area to the underground area (AUC-Area Under Curve) 1 is, the better the classification is [24].

For confusion matrix, as shown in Table 1, PCA is quite effective for classification. %87.5 and %81.3 success are achieved with PCA and without PCA, respectively. For ROC curve, as shown in Table 2 PCA is quite effective for this. 0.88 and 0.81 success are achieved with PCA and without PCA, respectively.

Likewise, the performance of this model can be assessed by taking into account the number of samples assigned to correct and incorrect classifications. In this study, accuracy, precision, sensitivity and error rate for model performance are discussed. The TP (true positive), FP (false positive), FN (false negative), TN (true negative) values are specified for the calculation. Accuracy, precision, sensitivity and error rate were evaluated to evaluate the performance of the supervised machine algorithm (Table 3).

Table 1: Confusion matrix results

	With PCA (%)	**Without PCA (%)**
KNN	87.5	81.3

Table 2: ROC Curve results

	With PCA	**Without PCA**
KNN	0.88	0.81

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (7)$$

$$\text{Error rate} = \frac{FP+FN}{TP+TN+FN+FP} \quad (8)$$

Table 3: KNN Evaluation

	Accuracy	
	With PCA(%)	**Without PCA(%)**
KNN	% 87.5	% 81.25
	Precision	
	With PCA(%)	**Without PCA(%)**
KNN	% 87.5	% 100
	Sensitivity	
	With PCA(%)	**Without PCA(%)**
KNN	% 87.8	% 72
	Error Rate	
	With PCA(%)	**Without PCA(%)**
KNN	**% 12.5**	**% 18.75**

4. Discussion

The proposed method in this paper, K-Nearest Neighbors Algorithm (k-NN), which is one of the supervised learning algorithms, is applied to EEG data. The data used for the study were taken from the database of Bonn University [25]. This classification was evaluated in two stages, with and without Principal Component Analysis (PCA). KNN has given results that can be regarded as prudent in terms of classification accuracy. The results of the present work showed that appropriate features combined with classifier can be done significant classification. The PCA has increased the classification percentage by reducing the size of the data while preserving features that can be considered important. Thus, it is thought that it will contribute considerably to the classification of complex data with a very large size.

With reference to this study, we can examine the advantages and disadvantages of each other by using supervised (SVM, kNN) and unsupervised (k-means) classification algorithms in LabVIEW in future studies.

References

[1] Shaker M. M., “EEG Waves Classifier using Wavelet Transformand Fourier Transform”, Int. Journal of Biological and Life Sciences, 2005, 1(3):85-90.

[2] Bhattacharya J. and Petsche H. “Universality in the brain while listening to music”, Proc. Royal Society Lond. B., 2001, 268(1484):2423-2433.

[3] Stam C. J., Pijn J.P.M., Suffczynski P. and Lopes da Silva F.H.. “Dynamics of the human alpha rhythm: evidence for non-linearity”, Clinical Neurophysiology, 1999, 110(10):18011813.

[4] Buzsaki G.. Rhythms of the Brain Oxford University Press, Oxford, (2006).

[5] Kumar S. P., Sriraam N., Benakop P.G. and Jinaga B.C. “Entropies based detection of epileptic seizures with artificial neural network classifiers”, Expert Syst. Appl., 2010, 37(4):3284–3291.
[6] Iasemidis L.D., “Epileptic seizure prediction and control”, IEEE Trans Biomed Eng. 2003, 50(5):549-558.
[7] Haydari Z., Zhang Y., and Zadeh H.S., “SemiAutomatic Epilepsy Spike Detection from EEG Signal Using Genetic Algorithm and Wavelet Transform”, IEEE International Conference on Bioinformatics and Biomedicine Workshops, Atlanta, GA, USA, (2011).
[8] Park H. S., Lee Y. H., Kim N. G., Lee D.S. and. Kim S. I. “Detection of epileptic form activities in the EEG using neural network and expert system”, Studies in health technology and informatics, (1998), 9(2):1255–1259.
[9] Kaya D., Türk M., “Biyoelektriksel Kökenli İşaretlerde Rahatsızlık Teşhisinin Yorumlanması”, Fırat Üniversitesi Mühendislik Bilimleri Dergisi, Cilt 29, Sayı 1, (2017).
[10] Kaya T. and İnce M.C., “The Obtaining of Window Function Having Useful Spectral Parameters by Helping of Genetic Algorithm”, Procedia-Social and Behavioral Journal, Elsevier, 2012, 83:563-568,.
[11] Makinac M., “Support Vector Machine Approach for Classification of Cancerous Prostate Regions”, Proceedings of World Academy of Science, Engineering and Technology, 2005, vol(7):1307-6884.
[12] Christopher J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition”, Data Mining and Knowledge Discovery, 1998, 2(2): 121-167,
[13] Chandaka S., Chatterjee A., Munshi S., “Cross-correlation aided support vector machine classifier for classification of EEG signals”, Expert Systems with Applications: An International Journal, 2009, 36(2):1329-1336.
[14] Silver A.E., Lungren M.P., Johnson M.E., O'Driscoll, S.W., An, K.N. and Hughes, R.E., “Using support vector machines to optimally classify rotator cuff strength data and quantify post-operative strength in rotator cuff tear patients”, J. Biomech. Vol. 39, 973-979.
[15] Fan J., Shao C., Ouyang Y., Wang J., Li S., Wang Z., “Automatic seizure detection based on support vector machines with genetic algorithms”, Proceedings of the 6th international conference on Simulated Evolution And Learning, Hefei, China. October 15-18, (2006).
[16] Siuly Y. Li and Wen P., “EEG signal classification based on simple random sampling technique with least square support vector machines”, Int. J. Biomed. Engineering and Technology (2010),7(4):390-409.
[17] Chen G. and Hou R., “A New Machine Double-Layer Learning Method and Its Application in Non-Linear Time Series Forecasting,” in International Conference on Mechatronics and Automation, ICMA, 795 –799, (2007).
[18] Labview machine learning toolkit user manual.
[19] Bishop C.M., “Pattern Recognition and Machine Learning”, Springer, New York (2006).
[20] Hotelling H., “Analysis of a complex of statistical variables into principal components”. Journal of Educational Psychology, (1933), 24:417–441, and 498–520.
[21] Hotelling H., “Relations between two sets of variates”, Biometrika, 1936, vol.28: 321–377.
[22] Türk Ö., Özerdem M. S, “EEG İşaretlerinin k-NN ile Sınıflandırılmasında Dalgacıklara İlişkin Performansların Karşılaştırılması, TIPTEKNO-14, KAPADOKYA.
[23] Lee S., Kang P., Cho S., “Probabilistic local reconstruction for k-NN regression and its application to virtual metrology in semiductor manufacturing”, Neurocomputing, 2014, 131: 427– 439,
[24] Tomak L. and Bek Y. “İşlem Karakteristik Eğrisi Analizi ve Eğri Altında Kalan Alanların Karşılaştırılması”, Journal of Experimental and Clinical Medicine, 2009, 27(2): 58-65.
[25] EEG Data http://www.meb.uni- bonn.de/ epileptologie/science/ physik/eegdata.html.