BCI에서 기계 학습을 위한 간질 뇌파 특징 선택을 통한 차원 감소 방법 분석
양 통* · Ibrahim Aliyu* · 임창균**

Analysis of Dimensionality Reduction Methods Through Epileptic EEG Feature Selection for Machine Learning in BCI
Yang Tong* · Ibrahim Aliyu* · Chang-Gyoon Lim**

요 약
지금까지 뇌파(Electroencephalography - EEG)는 뇌전증 진단 및 치료를 위한 가장 중요하고 편리한 방법이었다. 그러나 뇌전증 뇌파 신호의 파형 특성은 매우 약하고 비 정지 상태이며 배경 노이즈가 강하기 때문에 식별하기가 어렵다. 이 논문에서는 간질 뇌파의 특징 선택을 통한 차원 감소를 통한 분류 방법의 효과를 분석한다. 우리는 차원 감소의 성능 분석을 위해 Support Vector Machine (SVM), Logistic Regression (LR), K-Nearestneighbor (K-NN), Decision Tree (DR), Random Forest (RF) 분류 방법들을 사용하였다. 차원 감소 방법의 성능 분석을 위해 Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA) and Linear Discriminant Analysis (LDA) were investigated. The performance of each method was evaluated by using Support Vector Machine (SVM), Logistic Regression (LR), K-Nearestneighbor (K-NN), Decision Tree (DR) and Random Forest (RF). From the experimental result, PCA recorded 75% of highest accuracy in SVM, LR and K-NN. KPCA recorded 85% of best performance in SVM and K-NN while LDA achieved 100% accuracy in K-NN. Thus, LDA dimensionality reduction is found to provide the best classification result for epileptic EEG signal.

ABSTRACT
Until now, Electroencephalography (EEG) has been the most important and convenient method for the diagnosis and treatment of epilepsy. However, it is difficult to identify the wave characteristics of an epileptic EEG signals because it is very weak, non-stationary and has strong background noise. In this paper, we analyse the effect of dimensionality reduction methods on Epileptic EEG feature selection and classification. Three dimensionality reduction methods: Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA) and Linear Discriminant Analysis (LDA) were investigated. The performance of each method was evaluated by using Support Vector Machine (SVM), Logistic Regression (LR), K-Nearestneighbor (K-NN), Decision Tree (DR) and Random Forest (RF). From the experimental result, PCA recorded 75% of highest accuracy in SVM, LR and K-NN. KPCA recorded 85% of best performance in SVM and K-NN while LDA achieved 100% accuracy in K-NN. Thus, LDA dimensionality reduction is found to provide the best classification result for epileptic EEG signal.

키워드
BCI (Brain Computer Interface), EEG (Electroencephalography), Machine Learning, Epilepsy
뇌 컴퓨터 인터페이스, 뇌파, 기계 학습, 간질
I. INTRODUCTION

Epilepsy is a syndrome of chronic brain dysfunction caused by a variety of causes. Abnormal brain discharge leads to abnormal brain function. It often manifests as sudden loss of consciousness, hard or twitching of the whole body, or short-term trance, or the rapid contraction of muscles and other body[1]. About 50 million people have epilepsy and about 2.4 million are diagnosed with the disease each year around the world. In some developed countries, epilepsy has become the second most common disease of the nervous system after cerebrovascular disease [2]. The onset of epilepsy is sudden and difficult to predict to take appropriate measures. So far, there are no completely effective treatment methods, resulting in poor efficacy and prognosis.

EEG signals are made by a set of specific brain waves that contains brain function information. The amplitude of these waves changes with the brain waves while doing any mental tasks such as motor imagery or other cognitive tasks. This made EEG detection to become an indispensable item for clinical brain examination on brain functions and diagnosis of brain disease [3].

EEG can effectively help the medical staff to locate the epileptogenic lesions in the brain and to facilitate the embedding or surgical resection of lesions and other methods of treatment as well as the prediction of the disease from onset. With the development of digital signal processing technology and the widespread use of computers, the acquisition of EEG has become more convenient. In the near future, when medical physics develops to a certain stage, bioelectrical feedback may also be treated.

An efficient network model using 3-dimensional concept of drones can eliminate the restrictions. It would be useful to reduce some restrictions for collecting and send EEG signals in 2-dimensional space in real-time [4]. It is a quite good approach to observe the changes of EEG signals in the process for solving the problems in concentration analysis of Concentration-Related EEG Component Due to Smartphone [5].

Therefore, analysis and detection of EEG are very important. The most commonly used method for analyzing and processing electroencephalograms is to examine the characteristics of epileptic wave forms in electroencephalograms by experienced medical workers through the naked eye. However, due to some subjective factors, there have been several report of doctors’ inconsistency in judgments. Also, EEG signals are very weak, non-stationary chaotic and have strong background noise. But, the unprecedented breakthrough in computer technology over the years have aided the development of computer-assisted automatic analysis methods that provide a good auxiliary tool for analyzing EEG signals.

Feature extraction is a method of transforming a measured value of a pattern to highlight a representative feature of the pattern. Its also refers to the method and process of using a computer to extract characteristic information in an image [6]. In this work, we analyze three feature selection methods for epileptic EEG recognition based on five classifiers and compare the results to identify the best method for Epilepsy classification.

II. FEATURE EXTRACTION

2.1 Epilepsy EEG

Brain waves are spontaneous and rhythmic neural electrical activities. They generally have a sine wave rhythm. Their frequency ranges from 0.5Hz to 100Hz. In clinical practice, brain electricity is usually divided into 4 bands based on frequency, that is, δ Wave (0.5Hz–4 Hz), θ Wave (4Hz – 8Hz), α Wave (8Hz to 13Hz) and β Waves (13Hz
BCI에서 기계 학습을 위한 간질 뇌파 특징 선택을 통한 차원 감소 방법 분석

The difference between the normal and abnormal brain waves can be substantiated through the frequencies [7]. Fig. 1 shows two normal brain waves.

During seizures, changes in the pathology or function of the brain tissue occur, the curve changes accordingly [8]. Fig. 2 shows a typical EEG during seizures.

From the graphs, it can be observed that the amplitude of brain waves during this period of seizures is more rhythmic and much greater than the amplitude of normal brain waves. In addition, there are many characteristics of brain waves during seizures that are different from those of normal brain waves.

2.2 EEG Dataset

The data used in this study is a publicly available data from the Epilepsy Research Center at Bonn University in Germany [9]. The EEG data set consists of five subsets: denoted A~E, and each subset contains 100 segments of EEG signals. Each segment consists of 4097 sampling points. The sampling time and rates are 23.6s and 173.61Hz. A and B contain EEG data of five healthy individuals in two states: blink and eyes closed state. C and D contain epileptic EEG collect from lesion and intralesion respectively. E is the EEG data collected by five epileptic patients during the episode. A-D is considered normal EEG and E is classified as epileptic data. Each category has 100 samples. 80 samples were randomly selected from each class as training and 20 samples were used as test sets. Table 1 shows the composition of the dataset.

<table>
<thead>
<tr>
<th></th>
<th>Training set (# of samples)</th>
<th>Test set (# of samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>4*80</td>
<td>80</td>
</tr>
<tr>
<td>Patient</td>
<td>4*20</td>
<td>20</td>
</tr>
</tbody>
</table>

2.3 Feature Extraction from EEG using Wavelet

Feature extraction is a method of transforming a measured value of a pattern to highlight a representative feature of the pattern [6]. The selection of suitable wavelet function is basically determine by experiment as different wavelet function provide different performance. In this work, Daubechies–2 wavelet function $\psi(t)$ is adopted. This is because EEG signals are similar to that of spike wave [10]. Also, Daubechies sequence has a better scalability as well as providing a more flexible way of weighing boundary problems.

It has the following features:
1) The effective support length of wavelet function $\psi$ and scale function $\Phi$ is $2^N - 1$, where $N$ is the order of the wavelet function.
and the vanishing moment order of wavelet function $\Psi$ is $N$.

2) Most of dbN have no symmetry. For some wavelet functions, asymmetry is very obvious.

3) Regularity increases with the number $N$.

4) Function has orthogonality.

2.4 System Structure

Daubechies-2 wavelet transformation function was first applied to remove noise and extract 20 eigenvalues features. The transformed EEG was then subjected separately to three different dimensionality reduction methods to reduce the dimension of the EEG data. The reduction methods include Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA) and Linear Discriminant Analysis (LDA). Then each reduced data were then used for classification using Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbor(KNN), Decision Tree (TR) and Random Forest(RF). Fig. 3. shows the structure of the system.

III. FEATURE SELECTION BY REDUCING DIMENSION OF FEATURE

3.1 Principal Component Analysis (PCA)

Principal component analysis, abbreviated as PCA, is an unsupervised learning algorithm that is often used to perform data dimensionality reduction, lossy data compression, feature extraction, and data visualization. It is also called the Karhunen-Loève transform. The idea of PCA is to map $n$-dimensional features to $k$-dimensional space ($k < n$). This $k$-dimensional feature is a completely new orthogonal feature. Thus, PCA can be defined as an orthogonal projection of data in a low-dimensional linear space [11].

Input: $n$ dimension sample set $D = (x^{(1)}, x^{(2)}, \ldots, x^{(m)})$, dimension reduced to dimension $n'$.

Output: Sample Set $D'$ after dimension reduction

1) Centralize all samples:

   $$x^{(i)} = x^{(i)} - \frac{1}{m} \sum_{j=1}^{m} x^{(j)}$$

2) Calculate the sample’s covariance matrix $XX^T$

3) Eigenvalue decomposition of matrix $XX^T$

4) The eigenvector $(w_1, w_2, \ldots, w_{n'})$ corresponding to the largest $n'$ eigenvalues is taken out. After all the feature vectors are normalized, a feature vector matrix $W$ is formed.

5) For each sample $x^{(i)}$ in the sample set, a new sample $z^{(i)} = W^T x^{(i)}$ is converted.

6) Get output sample set $D' = (z^{(1)}, z^{(2)}, \ldots, z^{(m)})$

In this work an epilepsy detection algorithm based on the PCA dimension reduction method and five different classifiers. First we determine the
input \( N \)-dimensional sample \( D \) and the dimension \( N' \) to be reduced. Then the sample \( D \) is centered, the sample’s covariance matrix is calculated, and the matrix is subjected to eigenvalue decomposition. Then the eigenvectors corresponding to the largest \( N' \) eigenvalues were taken out, and we normalize the eigenvectors to form the eigenvector matrix. The centered sample is transformed to obtain the output sample set. Finally, the output sample set is input into five different classifiers for detection. In this group of EEG data tests, the method achieved better detection performance.

3.2 Kernel Principal Component Analysis (KPCA)

Another Epilepsy detection algorithm employed is KPCA dimension reduction method with five different classifiers. KPCA, is a nonlinear extension of the PCA algorithm [12]. Unlike KPCA, PCA is linear and it tends to be powerless for nonlinear data. For example, face images between different people must have a non-linear relationship. The recognition rate that PCA can achieve is only 88\%, while KPCA algorithm, which is also unsupervised learning, can easily achieve a recognition rate of about 93\%. A large part of this is because KPCA can tap into the non-linear information contained in the data set.

First, a batch of data obtained for \( n \) indexes (\( m \) samples for each index) is written as a \((m \times n)\) dimensional data matrix. The parameters in the Gaussian radial kernel function are selected and the kernel matrix \( K \) is calculated. And then correct the kernel matrix to get \( KL \). The Jacobi iterative method is used to calculate the eigenvalues of \( KL \lambda_1, \ldots, \lambda_n \), that is, the corresponding feature vectors \( v_1, \ldots, v_n \). We sorted the eigenvalues in descending order (by selecting sorting) yields \( \lambda_1' > \ldots > \lambda_n' \) and adjusted the eigenvector accordingly to \( v_1', \ldots, v_n' \). \( \alpha_1, \ldots, \alpha_n \) is obtained by normalizing the eigenvectors with the Schmidt normalization method. Then we calculated the cumulative contribution rate \( B_1, \ldots, B_n \) of the feature value. According to the given extraction efficiency \( p \), if \( B_i \geq p, \) \( t \) principal components \( \alpha_1, \ldots, \alpha_t \) are extracted. Then, the projection \( Y = KL \cdot \alpha \) of the corrected kernel matrix \( X \) on the extracted feature vector is calculated. The resulting projection \( Y \) is the data obtained after the data is reduced by KPCA. The algorithm is implemented as follows:

\[
\sum_{i=1}^{m} x^{(i)} x^{(i)^T} W = \lambda W
\]

Map as:

\[
\sum_{i=1}^{m} \phi(x^{(i)}) \phi(x^{(i)})^T W = \lambda W
\]

The eigenvalue decomposition of the covariance matrix is performed in the middle of the high dimensionality. Then, we use the same method as PCA for dimensionality reduction. In general, the mapping \( \phi \) does not need to show calculations, but it is done by a kernel function when calculations are needed.

3.3 Linear Discriminant Analysis (LDA)

Lastly, an epilepsy detection algorithm based on the LDA dimension reduction method and five different classifiers were experimented. LDA is a dimensionality reduction technique for supervised learning, which means that each sample of its
First, we determined the dimension $d$ of the dimension reduction for the datasets. Then the intra-class and inter-class divergence matrix was calculated. Thereafter, we calculated the eigenvectors corresponding to the eigenvalues to obtain the projection matrix. Finally, each sample feature is transformed into a new sample and an output sample set is obtained. LDA algorithm is as follows:

Input: Dataset
\[ D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}, \]
where any sample $x_i$ is an $n$-dimensional vector, $y_i \in \{C_1, C_2, \ldots, C_k\}$, reduced dimension to dimension $d$.

Output: Sample Set $D'$ after dimension reduction.

1) Calculate the divergence matrix $S_w$ in the class
2) Calculate the divergence matrix $S_b$ between the class
3) Calculate matrix $S_w^{-1}S_b$
4) Calculate the largest $d$ eigenvalues of the $S_w^{-1}S_b$ and Corresponding $d$ eigenvectors $(w_1, w_2, \ldots, w_d)$, get the projection matrix
5) Each sample feature $x_i$ in the sample set is converted to a new sample $z_i = W^T x_i$
6) Get output sample set
\[ D' = \{(x_1', y_1'), (x_2', y_2'), \ldots, (x_m', y_m')\} \]

IV. EXPERIMENTAL RESULT AND DISCUSSION

4.1 Experiment Results with Dimensionality reduction methods

In our work, we used the discrete wavelet transform to preserve the effective information of EEG signals and filter out the noise. Each instance gets 20 eigenvalues. We use PCA, KPCA and LDA for dimensionality reduction methods to keep the most effective eigenvalues. Finally, we used the training set to train the classifiers (i.e SVM, LR, KNN, DT, RF) and input the test set for classification, as shown in Fig. 4.

Table 2 shows the accuracy of the dimensionality reduction methods with the classifiers. In PCA, the accuracy for normal people by all the classifier is high while that of the patients is low. KPCA shows a better result for patient than PCA. On the other hand, LDA gives a better result in both the normal detection rate and the disease detection rate.

![General flow chart. The features from the dataset are extracted using discrete wavelet transform then feature selection were conducted using three dimensionality reduction methods before passing the data to classifiers.](image-url)
### 4.2 Experiment Results without Dimensionality reduction methods

Fig 5 shows the result of classification with the dimensionality reduction methods against that without dimensionality reduction (DWT). From the result, LDA gives the best result against all the other reduction method and DWT.

<table>
<thead>
<tr>
<th>F. S.</th>
<th>CLASSIFIERS</th>
<th>SVM (NORMAL)</th>
<th>LR (PATIENT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>SVM</td>
<td>98.75</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>98.75</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>98.75</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>98.75</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>RT</td>
<td>98.75</td>
<td>60</td>
</tr>
<tr>
<td>KPCA</td>
<td>SVM</td>
<td>97.50</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>98.75</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>97.50</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>96.25</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>96.25</td>
<td>80</td>
</tr>
<tr>
<td>LDA</td>
<td>SVM</td>
<td>98.75</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>100</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>98.75</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>97.50</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>97.50</td>
<td>95</td>
</tr>
</tbody>
</table>

- F. S.: Feature Selection

Fig. 5 Comparison of the performance of dimensionality reduction methods and descret wave transformation in classification

Fig. 6 Comparing epileptic EEG classification results with three dimension reduction methods
Furthermore, we compared the overall results of epilepsy recognition after three dimensionality reduction methods, as shown in Fig. 6. From the result of this comparison, LDA also gives the best overall result. Thus, the classification results of LDA dimensionality reduction are excellent for this data set.

V. CONCLUSION AND FUTURE WORKS

In this research, we have successful demonstrated that employing dimensionality reduction methods to EEG data can improve the performance of epilepsy classification. PCA, KPCA and LDA reduction methods were applied to the data after discrete wave transformation of the data. Then the reduced EEG data were each classified using Support Vector Machine (SVM), Logistic Regression (LR), K-Nearestneighbor (K-NN), Decision Tree (DR) and Random Forest (RF). From the experimental result, PCA recorded 75% highest accuracy in SVM, LR and K-NN. KPCA recorded 0.85% best performance in SVM and K-KNN while LDA achieved 100% accuracy in K-NN. Thus, LDA dimensionality reduction is found to provide the best classification result for epileptic EEG signal.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2017R1D1A1B03035968).

References


저자 소개

Yang Tong
2016 : BEIJING INSTITUTE OF PETROCHEMICAL TECHNOLOGY(B.S.)
2018년 : Chonnam National Univ.Dept. of Computer Engineering(M.E.)
※ 관심분야 : Machine Learning, wireless Communication

Ibrahim Aliyu
2014: Federal University of Technology, Minna (B.Eng)
2018: Federal University of Technology, Minna (M.Eng)
※ 관심분야 : Machine Learning, Internet security, Software Defined, networking, Cloud Computing, Internet of Things

임창균 (Chang-Gyoon Lim)
1990 : Chosun Univversity (M.S.)
1997년 Wayne State University, Computer Engineering(Ph. D.)
2011년~2014 광주테크노파크 가전 로봇지원센터 센터장
1997년~현재 전남대학교 공학대학 전기전자통신컴퓨터공학부 교수
※ 관심분야 : 인공지능, 지능형로봇, 기계학습, 패턴인식, 클라우드 컴퓨팅, 블록체인