

# 점진적 모델에 기반한 다채널 시계열 데이터 EEG의 특징 분석

김 선 희<sup>\*</sup> · 양 형 정<sup>\*\*</sup> · Kam Swee Ng<sup>\*\*\*</sup> · 정 종 문<sup>\*\*\*</sup>

## 요 약

BCI 기술은 생체신호인 뇌파를 수집하여 신호처리를 거친 후 실질적인 기기제어 및 통신 시스템 등을 제어하는 시스템 관련 기술이다. BCI 시스템 구현을 위해서는 뇌파의 특성을 실시간으로 분석하여 학습 시키고 학습된 뇌파의 특성을 적용하는 단계가 요구된다. 본 논문에서는 EEG 데이터를 효율적으로 분석하기 위해 점진적으로 갱신되는 주성분 분석을 이용하여 왼손/오른손 동작에 영향을 미치는 EEG 신호의 특징을 찾고, 이를 반영하여 데이터의 차원을 축소한다. 입력 자료의 특징을 충분히 포함하면서 낮은 차원을 가지는 데이터를 이용한다면 분류를 위한 계산량을 감소시킬 수 있을 뿐만 아니라 불필요한 특징을 제거함으로써 분류 성능을 향상시킬 수 있다. 본 논문에서는 점진적으로 갱신되는 주성분 분석을 이용하여 데이터의 차원을 축소하고 이에 대한 효율성을 검증하기 위해 K-NN분류기를 이용하여 분류 정확도 측정을 수행하였다. 그 결과 주성분 분석을 이용하여 특징을 추출하고 분류율을 측정할 경우보다 평균 5% 높은 분류 정확도를 보였다.

키워드 : 점진적 주성분 분석, 뇌파, 뇌-컴퓨터인터페이스, 특징 추출, 차원 축소

## Feature Analysis of Multi-Channel Time Series EEG Based on Incremental Model

Sun-Hee Kim<sup>\*</sup> · Hyung-Jeong Yang<sup>\*\*</sup> · Kam Swee Ng<sup>\*\*\*</sup> · Jong-Mun Jeong<sup>\*\*\*</sup>

## ABSTRACT

BCI technology is to control communication systems or machines by brain signal among biological signals followed by signal processing. For the implementation of BCI systems, it is required that the characteristics of brain signal are learned and analyzed in real-time and the learned characteristics are applied. In this paper, we detect feature vector of EEG signal on left and right hand movements based on incremental approach and dimension reduction using the detected feature vector. In addition, we show that the reduced dimension can improve the classification performance by removing unnecessary features. The processed data including sufficient features of input data can reduce the time of processing and boost performance of classification by removing unwanted features. Our experiments using K-NN classifier show the proposed approach 5% outperforms the PCA based dimension reduction.

Keywords : Incremental PCA, EEG, Brain-Computer Interface, Feature Selection, Dimension Reduction

## 1. 서 론

EEG(electroencephalography)란 뇌의 수많은 신경에서 발생하는 전기적인 뇌 표면의 신호를 전극을 이용하여 측정하는 것으로, 예측이 어려운 다변량 시계열 데이터이다. EEG 신호는 뇌의 활동, 측정시의 상태 및 뇌기능에 따라 시간영역에서의 특성과 뇌 기능과 관련된 공간적 특성을 함께 가지고

있다. 이러한 EEG는 뇌의 활동을 측정하는 비침습적 방법으로 적은 비용으로 손쉽게 측정할 수 있으며, 탁월한 시간 해상도를 지닌 데이터로서 BCI(Brain-Computer Interface) 시스템에 널리 이용되는 뇌 신호이다.

BCI는 인간과 컴퓨터 상호 작용을 연구하는 한 분야로, 뇌파 측정기를 통해 특정 상태의 뇌파 신호를 수집하여 특이점이나 특징을 추출하고 이를 분류한 후 일반적인 제어 신호로 변환하여 컴퓨터나 임의의 장치를 제어하는 시스템 관련 기술이다. 뇌파 신호를 이용한 BCI 시스템 구현을 위해서는 측정된 뇌 신호에 대한 신호처리 과정을 거친 후 실질적인 기기제어 및 통신 시스템에 대한 적용이 필요하다. 그러나 뇌파 신호는 측정 시 전극을 부착하는 위치와 방법, 실험 환경, 피험자의 움직임 등에 의하여 잡음이 발생하기

\* 이 논문은 2007년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-331-D11412)

† 준 회 원 : 전남대학교 전산학과 박사과정

‡ 종신회원 : 전남대학교 전자컴퓨터공학부 조교수

‡‡‡ 준 회 원 : 전남대학교 전자컴퓨터공학부 석사과정

논문접수 : 2008년 9월 8일

수 정 일 : 1차 2008년 11월 20일

심사완료 : 2008년 11월 21일

위우며, 이러한 잡음이 포함된 뇌파 신호의 경우에는 BCI시스템에 적용하기에 부적절하다. 따라서 효율적인 BCI 시스템을 구성하기 위해서는 뇌파 신호의 특징을 포함하면서 불필요한 특징을 제거한 신뢰성 있는 뇌파를 수집하여야 한다.

이를 위해 주성분 분석(PCA: Principal Component analysis)<sup>[18]</sup>, 웨이블릿 변환(wavelet transform)<sup>[12]</sup>, 적응적 AR 모델(AAR: adaptive Auto Regressive Model)<sup>[4]</sup>, 적응적 가우시안 표현(adaptive Gaussian representation)<sup>[3]</sup> 등 다양한 뇌파 신호의 특징을 분석하는 방법들이 제안 되었다.

[18]에서는 왼쪽과 오른쪽 손가락을 임의로 움직이는 동안 단일 채널로 기록된 EEG 데이터에 주성분 분석을 적용하여 특징을 추출하고 분류 학습을 수행하였다. Hazarika는 단일 채널에서 측정된 뇌파 신호를 4-레벨의 웨이블릿 변환을 한 다음 각 레벨에서 제일 큰 웨이블릿 계수와 두 번째로 큰 계수를 선택한 후, 이들을 입력으로 학습데이터를 구성하였다<sup>[12]</sup>. 또한 [2]에서도 웨이블릿 변환을 이용하여 특징을 추출한 후 recurrent 신경망(RNNs)을 이용하여 학습 시킴으로써 간질 발작에 관한 예측 문제를 해결하고자 하였다. [4]에서는 오른쪽, 왼쪽 손의 움직임을 상상할 때 측정된 데이터에 적응적 AR 모델을 적용하여 데이터의 특징을 찾고 이를 분류 학습에 이용하였다. 움직임, 의식 상태의 변화, 정신 활동 및 지각 자극 등의 상태를 반영하는 뇌의 영역은 특정 부위에 한정된 것이 아니라 뇌의 영역에 걸쳐서 이루어진다. 따라서 위와 같은 연구들의 경우 단일 채널에서 획득한 뇌파의 신호를 이용하였기 때문에 뇌의 기능을 판별하기에 부적합하다.

이를 해결하기 위해 [5]와 [13]에서는 다채널을 이용하여 측정된 데이터에 웨이블릿 변환을 적용하였고, [10]과 [22]에서는 주성분 분석을 이용하였다. [21]에서는 Zhang[20]에 의해 제안된 웨이블릿 신경망(Wavelet Network)을 EEG 신호의 특징을 추출하기 위해 적용하였다. 그러나 웨이블릿 신경망은 필요의 이상 많은 은닉층 노드로 구성되어 자원의 낭비뿐만 아니라 학습 속도에도 영향을 미친다.

위와 같은 연구들은 일반적인 뇌파 분석을 통해 이루어진 연구로, 예측이 어려운 시계열 데이터로서 시간에 따라 특성이 변하는 뇌파 데이터의 경우에 그 특성이 충분히 고려되지 않는 문제점을 가지고 있다. 이러한 이유로 특정 상태에 따른 뇌파 분류 정확도 문제에서도 어려움을 겪고 있을 뿐만 아니라 실시간 분석 및 예측이 불가능하다.

본 논문에서는 뇌파의 불필요한 특징을 제거하고 분류 학습의 계산량을 줄이기 위해 점진적 방법을 적용한다. 본 논문에서 적용된 방법은 시간적 특성을 고려하여 주성분 분석을 확장한 것으로 왼손/오른손 동작에 영향을 미치는 뇌파 신호의 특징을 찾고, 이를 반영하여 데이터의 차원을 축소한다. 이렇게 축소된 뇌파 데이터를 이용하여 분류 모델을 만들고 데이터 차원 축소의 효율성을 검증한다. 이를 위해 기존의 특징 추출 방법으로 잘 알려진 PCA<sup>[9]</sup>, DWT<sup>[19]</sup>, DFT<sup>[16]</sup>를 이용하여 전 처리한 후 K-NN 분류기를 이용하여 점진적 주성분 분석을 이용한 경우와 비교 하였다. 그 결과

다른 특징 추출 방법보다 점진적 주성분 분석을 이용하였을 경우 약 5% 높은 분류율을 보였다.

논문의 구성은 다음과 같다. 2장에서는 기존의 뇌파 분석에 관한 연구를 살펴보고, 3장에서는 본 논문에서 사용하는 점진적 주성분 분석에 대해 기술한다. 4장에서는 실험에 사용된 데이터에 대한 설명과 차원 축소에 대한 분류 실험 및 결과를 보인다. 5장에서는 결론 및 향후 연구에 대하여 기술한다.

## 2. 관련 연구

BCI 연구는 전처리 과정과 특징추출, 패턴 인식 등의 다양한 기반 기술로 구분된다. 측정된 뇌파 데이터에는 전극의 부착 위치나 피험자의 작은 움직임 등에 의해 발생하는 여러 가지 잡음이 섞여 측정되기 쉽다. 따라서 이러한 잡음을 제거하기 위한 전처리 과정이 수행되어야 한다. BCI 시스템을 구성하기 위한 전처리 과정으로는 자기상관함수(Auto Correlation Function)<sup>[11]</sup>, 독립성분분석(ICA: Independent Component Analysis)<sup>[14]</sup>, Bandpass-Filtering과 Notch-Filtering<sup>[6]</sup> 등이 있다. [11]에서는 신호 사이의 상관성을 나타내는 것으로 한 신호와 그 신호만큼의 시간 지연을 시킨 신호를 매치 시키는 자기상관함수를 뇌파 분석에 적용하였다. Scott Makeig은 신호에 섞여 있는 성분을 각각의 독립된 성분으로 보고 추출 되어진 뇌파 신호에서 안전도 및 근전도를 제거 함으로 원하는 신호만을 얻기 위해 독립성분분석을 사용하였다<sup>[14]</sup>. 또한 주변의 잡음을 걸러주기 위해 Bandpass 필터나 notch 필터를 사용하여 잡음을 제거하였다<sup>[6]</sup>.

특징 추출은 신호를 분류할 때 주어진 입력 자료 보다 적은 차원을 가지면서 동시에 자료를 분류하기 위한 특징을 충분히 포함하는 신호의 특징을 찾아내는 과정이다. 뇌파 신호 분석에서 파워스펙트럼은 오랜 시간 동안 중요하게 사용되어 왔다. 이 방법은 고속 푸리에 변환(FFT: Fast Fourier Transform)을 통한 주파수계열(Frequency Series) 분석법으로 시계열 신호의 해석이 어려울 때 시계열 뇌파 값을 주파수계열로 변환하여 밴드 별로 진폭의 세기를 비교 분석하는 방법이다<sup>[17]</sup>. 또한 웨이블릿 변환은 시간과 주파수의 두 가지 특성을 모두 반영하는 방법으로 뇌파의 특징 추출에 효과적인 것으로 알려져 있다<sup>[8]</sup>. 그러나 이러한 방법들은 고전적인 선형 분석법으로 수많은 신경세포에서 발생하는 복잡한 신호들을 정밀 분석하는데 한계가 있다.

이러한 선형 분석법이 가지는 문제점을 해결하고자 비선형 분석들을 뇌파 분석에 적용하는 연구가 이루어 졌다. 그 중 [1]에서는 다차원 특징 벡터로 이루어진 데이터에 대하여 높은 차원에서의 정보를 유지하면서 낮은 차원으로 차원을 축소시키는 다변량 데이터 처리 방법 인 주성분 분석을 이용하여 각 전극에서 측정된 뇌파를 전극 수와 같은 개수의 성분으로 분리하였다<sup>[1]</sup>. [22]에서는 주성분 분석을 이용하여 특징을 추출한 후 LDS(linear dynamic system)와 HMM(hidden Markov model)을 이용하여 뇌파 데이터의 분류율

을 측정하였다. 위에 소개된 대부분의 방법들은 일반적인 뇌파 분석을 위한 특징 추출 방법으로 예측이 어려운 시계열 데이터인 뇌파 데이터의 경우 그 특성을 충분히 고려하지 못한다.

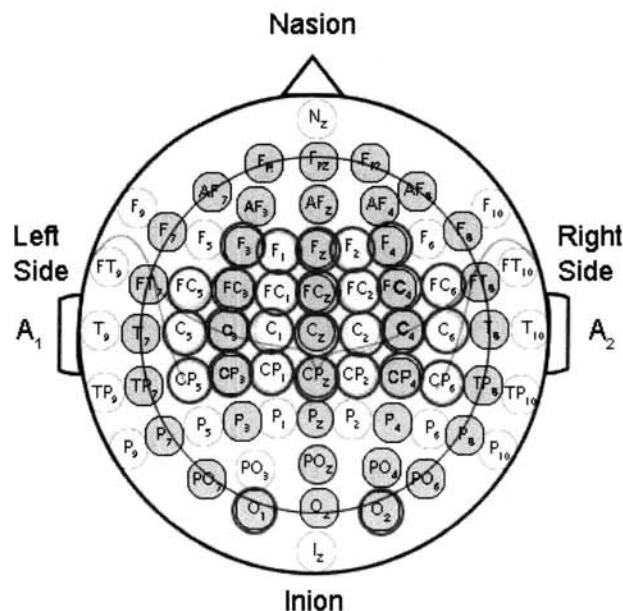
BCI 시스템에 이용되는 대다수의 뇌파 데이터는 컴퓨터의 저장 매체에 저장되며, 저장된 데이터에서 특징을 추출하고 분류 알고리즘을 학습시키는데 사용된다. 만약 측정된 데이터가 여러 요인으로 인해 발생하는 잡음이 많이 포함되어 있다면 좋은 분류기를 이용하더라도 높은 분류율을 기대하기는 매우 어려울 것이다. 따라서 분류를 위한 계산량을 줄이고 불필요한 특징을 제거하는 과정이 필요하다.

본 논문에서는 시계열 데이터로서 시간에 따라 변할 수 있는 뇌파의 특성을 실시간으로 분석하기 위해 데이터 특징 추출을 통한 데이터 차원 축소 방법으로 점진적 방법<sup>[15]</sup>을 적용한다.

### 3. 점진적 주성분 분석을 이용한 데이터 차원 축소

뇌파 연구에서 일반적으로 사용하고 있는 두피상 전극의 위치는 국제 표준 규격인 10-20 시스템 [7]에 기초한 것으로, 뇌에 대한 해학적 위치와 대응되는 부분이다. 더 정확한 위치 판별이 가능하도록 10-20시스템의 전극들 사이에 하나씩 전극들을 더 부착하여 이를 Modified Combinatorial Nomenclature 명명법이라고 하였다. (그림 1)은 Modified Combinatorial Nomenclature 명명법을 적용한 전극의 명칭이며, 진하게 표시된 부분이 국제 전극 배치 10-20시스템과 다르게 명명된 전극이다. 또한 붉은 선으로 표기된 부분은 본 논문에서 이용한 뇌파 데이터를 얻기 위해 부착된 전극의 위치이다.

뇌파 측정은 두피로부터 대뇌피질의 신경세포군에서 발생



(그림 1) Modified Combinatorial Nomenclature 전극 배치 법

되는 미세한 전기적 파동을 체외로 도출하고 이를 증폭하여 전위를 종축으로 시간을 횡축으로 하여 기록한다. 즉 뇌파기로 측정된 뇌파 데이터  $X$ 는  $n \times t$  행렬로 구성된다. 뇌파 데이터는 시계열 데이터로 시간적인 의존성이 존재한다. 주성분 분석의 경우 의존성이 존재하는 데이터로부터 뇌파의 서로 다른 성분을 분리하는데 사용되어 왔다<sup>[1]</sup>.

주성분 분석은 주어진 데이터의 정보를 잘 표현하면서 정보량을 줄일 수 있는 효과적인 방법 중 하나이다. 주성분 분석은  $n$ 차원을 갖는 데이터  $X$ 의 고유값(eigenvalue)과 고유벡터(eigenvector) 행렬을 얻어 이를 기반으로 주성분을 찾고 원래 데이터를 더 잘 나타내는 새로운 축을 구성한다.

본 논문에서는 점진적으로 갱신되는 주성분 분석을 이용하여 시계열 데이터 은닉 변수를 찾고 이를 기반으로 EEG 데이터의 차원을 축소한다. 이를 위해 측정된 EEG 데이터를  $n$ -차원의 특징 벡터  $X_t = (x_{1,t}, x_{2,t}, \dots, x_{n,t})$ 로 표현한다.

모든 특징 벡터는  $n$ -차원 특징 벡터 공간 내에 점으로 표현되며 벡터 공간에서 모든 특징 벡터들은  $n$ -차원 벡터들의 집합에 의해 채워지고 벡터 공간에서 각각의 축 방향을 가리키는 단위 벡터들을 기저 벡터  $V_t = \{\vec{v}_{1,t}, \vec{v}_{2,t}, \dots, \vec{v}_{n,t}\}$ 로 나타낸다. 즉, 특징 벡터 공간에 모든 특징 벡터들이 위치하는 것을 식 (1)과 같이 기저벡터의 선형 조합 형태로 표현할 수 있다.

$$X_t = x_{1,t} \vec{v}_{1,t} + x_{2,t} \vec{v}_{2,t} + \dots + x_{n,t} \vec{v}_{n,t} \quad (1)$$

식 (1)에서  $x_{i,t}$ 는  $X_t$ 의 좌표이다. 즉, 좌표벡터에 구성요소라 할 수 있다. 일반적으로 기저벡터는 표준 기저벡터를 사용한다. 그러나 시계열 EEG 데이터의 경우 일정한 측정 시점에 따라 측정된 데이터이므로 데이터를 표현하는데 있어 표준 기저벡터는 적절하지 못하다. 따라서 시계열 데이터의 측정 시점에 따른 특성을 잘 표현 하기 위해 새로운 기저값이 요구 된다.

기존의 주성분 분석에 시계열의 특성을 반영하기 위해 점진적 주성분 분석은 새로운 기저 값을 원래의 고차원 데이터를 투영하는 기저 벡터  $V$ 를 점진적으로 갱신함으로써 찾는다. 즉, 주성분 구성요소 분석을 기반으로 주성분 구성요소를 찾고, 평균 재현 에러율(average reconstruction error)을 최소화하는 주성분 구성요소의 기저벡터  $V_{ij}$  ( $i = 1 \dots p, j = 1 \dots t$ )를 점진적으로 갱신하면서 새로운 기저 벡터를 찾는다.  $p$ 는 주성분 구성요소를 의미한다. 따라서 점진적 주성분 분석은  $p$ 개의 기저 벡터를 필요로 하며, 주성분 구성요소  $\vec{v}_i = [v_{i,1}, \dots, v_{i,t}]^T$  ( $i=1 \dots p$ )를 가지고 원소  $V_{ij}$ 의 값을 갱신한다.

본 논문에서 은닉 변수의 수는  $h$ 로 표기하며 초기화 값은 3으로 설정하였다. 그러면 특징벡터  $X_t$ 는 새로운 공분산 행렬  $3 \times t_t$ 의 크기를 갖는다. 첫 번째로 요약된 기저벡터를 이용하여 평균 재현 에러율과 시간에 따른 주성분 구성요소의

수를 점진적으로 결정하기 위해 구성요소의 중요도 값을 평가한다. 중요도 임계치는 최소 95%, 최대98%로 [15]에서 정의된 표준 값을 이용하였다. 즉, 시간  $t$ 에 대한  $X_t$ 의 구성요소의 중요도가 중요도 최소 임계치 값보다 작으면 구성요소의 수를 증가 시키고, 반면에 최대 임계치 값보다 커지면 구성요소의 수를 감소시킨다.

기저 벡터 갱신 시 이용되는 평균 재현 에러율은 아래의 식 (2)와 같다.

$$ARE = \sum_i ||\tilde{X}_i - X_i|| \quad (2)$$

식 (2)에서  $\tilde{X}_i$ 는  $p$ 차원 공간에  $X_i$ 를 투영한 후에 재구성된 점을 의미하며,  $X_i$ 의 재구성은 식 (3)과 같이 나타낸다.

$$\tilde{X}_{k,t} = \sum_{j=1}^n v_{i,j} \times A_{k,j} \quad (k=1 \dots n, j=1 \dots t) \quad (3)$$

위 식에서와 같이  $A_{k,j}$ 는  $x_{k,j}$ 와 기저벡터의 선형 조합 형태를 의미한다. 즉  $A_{k,j}$ 는  $n$ 개의 특징 벡터들에 각각의 기저벡터를 곱한 값을 나타내며,  $A_k$ 는 식 (4)와 같이 표현하였다.

$$A_{k,j} = \sum_{i=1}^n v_{i,j} \times x_{k,i} \quad (i=1 \dots p, k=1 \dots n, j=1 \dots t) \quad (4)$$

위의 식 (4)을 이용하여 구해진 값에 대한 중요도는 시간에 따른 주성분 구성요소의 수를 점진적으로 결정하기 위해 식 (5)와 같이 나타낸다.

$$M_{k,t} = \frac{1}{t} \sum_{r=1}^t A_{k,r}^2 \quad (5)$$

시간 변화에 따른 새로운 샘플  $X_{t+1}$ 이 주어졌을 때 아래와 같은 단계로 기저 벡터를 새로이 갱신하여 점진적으로 주성분의 구성요소를 변경한다.

1. 기저벡터 초기화( $v_1=[100 \dots 0]^T$ ,  $v_2=[010 \dots 0]^T$ ,  $v_3=[001 \dots 0]^T$ )  
 → 은닉변수 수  $H_i$  정의( $i = 1 \dots h$ , 본 논문에서  $h = 3$ 으로 초기화)
2. 기저 벡터 갱신  
 → 기저벡터( $v_i$ )을 이용하여  $X_{t+1}$ 를 투영( $A_{i,t+1} := v_i^T \times x_{i,t+1}$ )  
 → 중요도를 평가하여 은닉 변수 변경( $H_i \leftarrow \lambda \cdot H_i + A_{i,t+1}^2$ )  
 → ARE(평균 재현 에러율) 평가( $e_i := x_{i,t+1} - A_{i,t+1} \cdot v_i$ )  
 → 주성분 구성요소를 평가하여 기저벡터 갱신( $v_i \leftarrow v_i + (1/H_i) \cdot A_{i,t+1} \cdot e_i$ )  
 → 갱신된  $X_{t+2}$ 를 가지고 반복( $X_{t+2} := x_{i,t+1} - A_{i,t+1} \cdot v_i$ )

위의 단계에서 시간 변화에 따른 뇌파 신호를 적용하기 위한 가중치( $\lambda$ )값은 0.96으로 설정하였다.( $0 < \lambda \leq 1$ )

본 논문에서는 위와 같은 방법을 EEG 데이터에 적용하

여 시간 변화에 따른 데이터의 특성을 포함하고 있는 특징을 추출하였다. 또한 분류 학습을 위하여 추출된 특징을 기반으로 데이터의 차원을 축소하여 분류 학습데이터를 구성하였다. 점진적 주성분 분석은 EEG 데이터의 특징을 충분히 포함하면서 입력 데이터 보다 적은 차원으로 데이터를 축소 함으로써 많은 저장 공간을 필요로 하지 않고 분류 계산량을 감소 시킬 수 있다.

#### 4. 실험

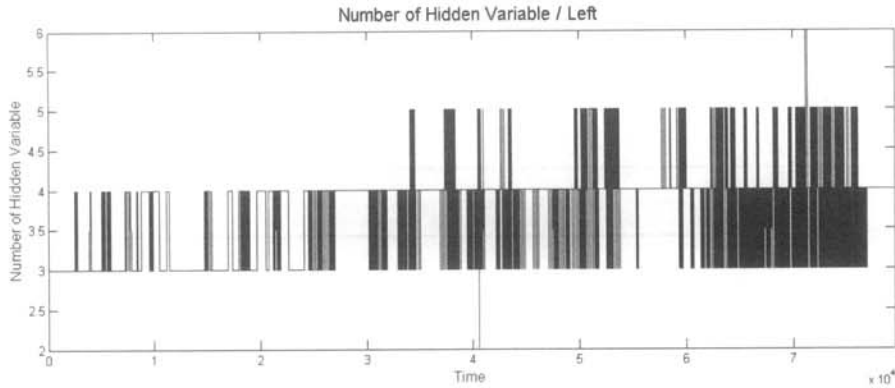
본 논문에서는 [23]에 의해 제공되는 EEG 데이터에 점진적으로 갱신되는 주성분 분석을 적용하였다. 이 데이터는 피드백이 없는 일반적인 피험자로부터 기록된 것으로 피험자가 아무런 외부 자극 없이 임의로 오른손 또는 왼손 약지 손가락으로 키보드 키를 누르며 손동작을 행하는 동안에 기록된 EEG신호이다. 이 실험은 6분 동안에 걸쳐 수행 되었으며, 타이핑의 평균 속도는 약 1초 간격으로 이루어 졌다.

뇌파 측정은 은-염화은(Ag/AgCl) 전극 캡과 뉴로스캔 증폭기를 이용하여 측정 되었으며, 국제적 표준 전극 10-20배치법의 위치에 따라 28개의 채널에서 전극을 통하여 뇌파 신호를 기록 하였다. 샘플링 율은 1000Hz이며, 밴드패스 필터링은 0.02~200Hz이다. 본 논문에서는 위와 같은 방법으로 제공된 EEG데이터를 실험에 이용하였다. 뇌파 패턴 인식에 있어서 중요한 영향을 미치는 특징을 점진적 주성분 분석을 이용하여 찾고, 이를 기반으로 추출된 특징들을 학습 데이터로 구성하여 학습 시킨 후 테스트를 거쳐 분류 학습을 수행하였다.

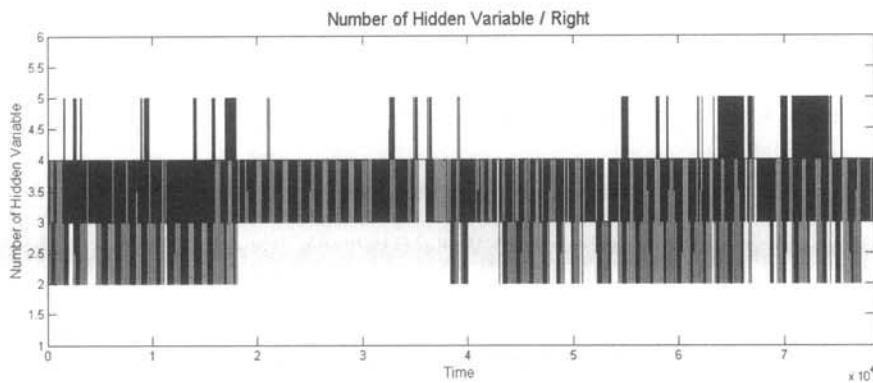
(그림 2)와 (그림 3)은 본 연구에서 이용한 점진적 주성분 분석을 이용하여 왼손과 오른손의 움직임을 기록한 뇌파에서 찾아낸 시간에 따른 은닉 변수의 개수 변화를 나타낸 것이다. 본 논문에서는  $h$ 개의 은닉 변수를 이용하여 EEG 데이터를 요약하였으며( $h < n$ (전극수)), 왼손과 오른손의 경우 마지막 시점에서 각각 4개의 은닉변수를 발견 하였다. 이 실험에서 왼손을 움직이는 경우 은닉변수의 수가 시간에 따라 최소 2개에서 최대 6개까지 변화하는 것을 볼 수 있으며, 오른손의 경우는 최소 2개에서 최대 5개까지 변화하는 것을 볼 수 있다. (그림 4)는 점진적 주성분 분석으로 찾아낸 왼손과 오른손 약지의 움직임에 대한 첫 번째 은닉 변수이다. (그림 4)에서와 같이 첫 번째 은닉 변수를 통해 서로 다른 움직임을 명확하게 구분할 수 있었다.

위와 같이 데이터의 특성을 잘 나타내는 은닉 변수를 이용하여 데이터를 축소한다면 계산량을 감소시킬 수 있을 뿐만 아니라 불필요한 특징을 제거함으로써 분류성능을 향상시킬 수 있다. 따라서 왼손/오른손 동작에 영향을 미치는 EEG 데이터의 특징을 찾고, 이를 반영하여 데이터의 차원을 축소하였다.

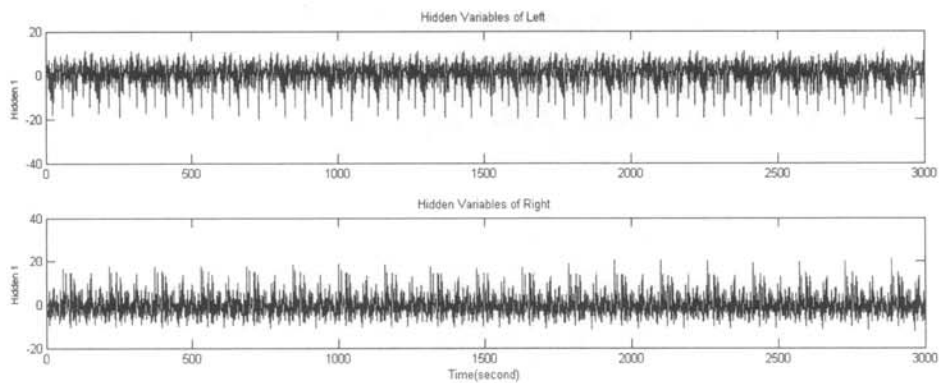
<표 1>은 데이터의 차원 축소 시 주성분 구성요소의 수와 그에 따른 재현 에러율을 식 (2)을 이용하여 주성분 분석 방법과 점진적 주성분 분석 방법(iPCA, incremental



(그림 2) 왼손 약지를 움직였을 때 시간에 따른 은닉 변수의 개수의 변화



(그림 3) 오른손 약지를 움직였을 때 시간에 따른 은닉 변수의 개수



(그림 4) 왼손과 오른손 약지 움직임에 대한 첫 번째 은닉변수

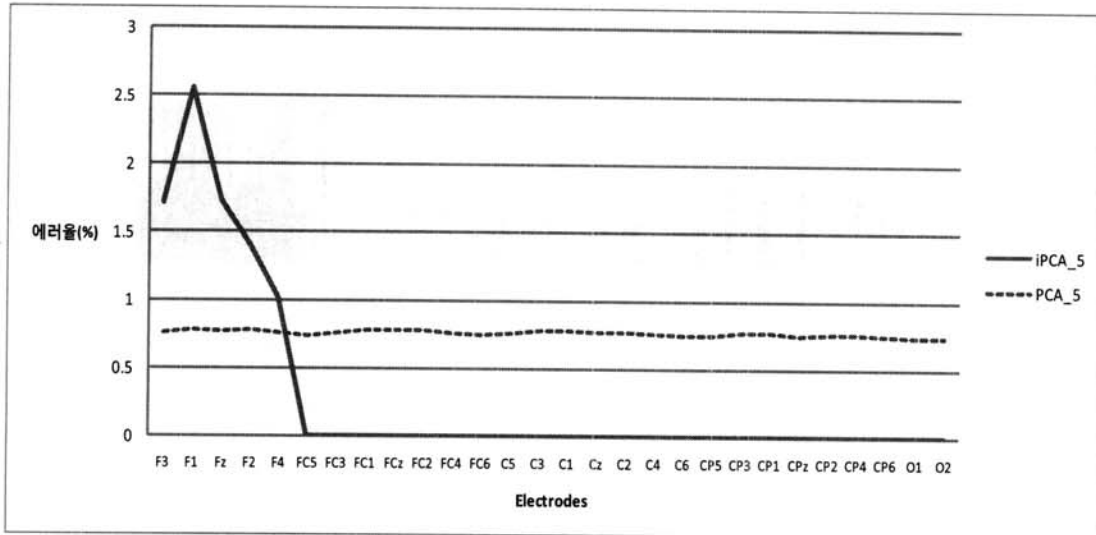
<표 1> 주성분 구성요소 개수에 따른 재현 에러율(단위:%)

방법	PC 개수								
	5	7	9	11	...	19	21	23	25
PCA	0.7597	0.7649	0.7672	0.7684	...	0.7710	0.7714	0.7716	0.7718
iPCA	0.3006	0.3103	0.3476	0.3504	...	0.3773	0.3931	0.4293	0.4446

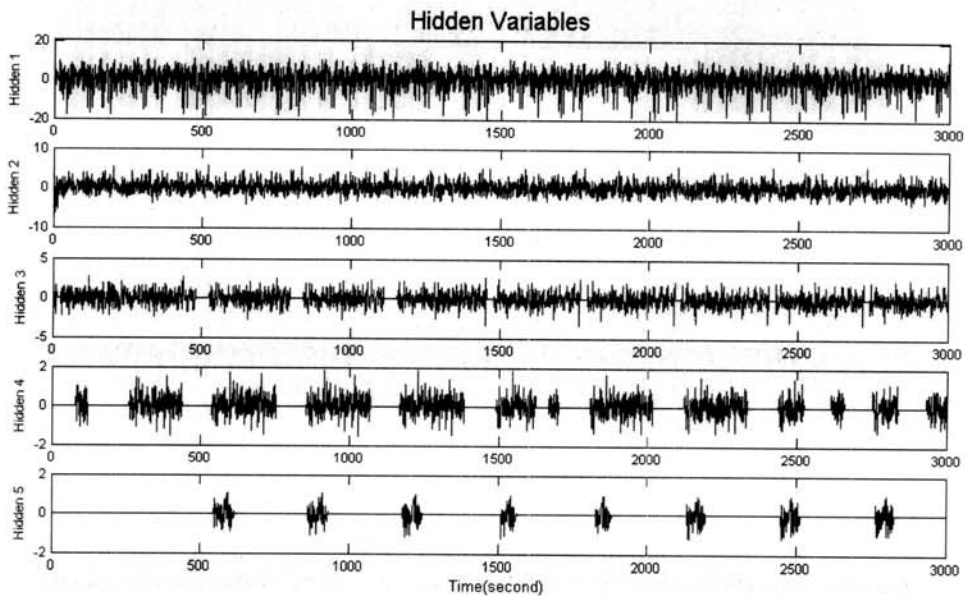
PCA)을 비교하였다. 평균 재현 에러율은 원래의 데이터 값을  $p$ 차원 공간에 투영한 후 재구성된 데이터 값에서 원래의 데이터를 뺀 값들의 합이다. 그 결과 주성분 분석과 점진적 주성분 분석 모두의 경우 주성분 구성요소의 개수가 증가할 수록 재현 에러율의 값도 증가하는 것을 볼 수 있다. 이 결과를 통해 실험에 이용되는 뇌파 데이터에 불필요한 데이터가 포함되어 있음을 가정하고 가장 낮은 에러율을 가지는

차원으로 데이터를 축소한다. 또한 점진적 주성분 분석을 이용하여 데이터의 차원을 축소하였을 시 주성분 분석을 이용한 경우 보다 낮은 평균 재현 에러율을 나타내는 것을 볼 수 있었다.

(그림 5)는 주성분 구성요소의 개수를 5로 하였을 시 각 전극 위치에 대한 재구성 에러율을 나타내는 것으로 점진적 주성분 분석을 이용한 경우 전극 F3부터 F4까지를 제외한



(그림 5) 5개의 주성분 구성요소 대한 주성분 분석과 점진적 주성분 분석의 에러율 비교



(그림 6) 점진적 주성분을 이용한 은닉변수

나머지 전극의 경우 0에 가까운 에러율을 보였다. 주성분 분석의 경우 차원 축소 시 전체 데이터를 대상으로 주성분 구성요소를 찾음으로써 비교적 일정한 에러율을 보인다. 그에 비해 점진적 주성분 분석의 경우는 매 측정 시점마다 기저 벡터를 갱신하여 새로운 기저 벡터 값을 다음 측정 시점 데이터에 적용함으로써 점차적으로 에러율이 0에 가깝게 줄어드는 것을 볼 수 있다. 이는 시계열 데이터의 경우 데이터 전체에 주성분 분석을 적용하여 차원을 축소하는 방법보다 점진적 주성분 분석을 이용하여 데이터의 차원을 축소하였을 시 시계열 데이터의 특성을 더 잘 반영할 수 있음을 보여준다.

(그림 6)은 점진적 주성분 분석을 이용하여 재현 에러율을 최소로 하여 찾아낸 5개의 은닉 변수이다. 이를 기반으로 뇌파 데이터의 차원을 축소하였다.

본 연구에서는 데이터 차원 축소에 대한 효율성을 검증하

기 위해 K-NN분류기를 이용하여 분류 정확도 측정을 수행하였다. 그 결과 <표 2>에서와 같이 많은 특징 추출 기법 중 EEG 신호 분석에 많이 사용되는 이산 웨이블릿 변환(DWT: Discrete Wavelet Transform)<sup>[19]</sup>과 이산 푸리에 변환(DFT: Discrete Fourier Transform)<sup>[16]</sup>, 주성분 분석(Principal component analysis)<sup>[9]</sup>을 이용하여 특징을 추출한 후 이를 이용하여 분류 결과를 비교 측정하였다. 웨이블릿 변환은 복잡하고 비정상적인 신호들의 검출방법으로 선호되고, 시계열 분석과 신호 저장 방법으로 사용되며 생체신호 분야에서 신호처리에 유용한 방법이다. 또한 이산 푸리에 변환은 디지털 신호나 시계열(time series) 데이터를 주기 함수로 변환하여 신호 속에 내포되어 있는 주파수를 찾는 것으로 임의의 신호의 주파수 특성을 이산 신호로부터 얻어내기 위한 변환방법이다. 주성분 분석은 고차원 데이터로부터 데이터의 구조를 밝히거나, 데이터의 차원을 낮추는 데 많이

이용되는 다변량 통계 분석 방법이다. 즉 주어진 데이터를 분산이 최대가 되는 축으로 변환하는 것으로, 새로운 차원에서의 데이터의 벡터들을 주성분 이라 한다. 이때 분산이 작은 성분을 제거함으로써, 데이터의 차원을 줄이는 동시에 데이터에 포함되어 있던 잡음(noise)을 제거 할 수 있다. 각각의 방법을 이용하여 특징을 추출하고 이를 분류 학습에 적용하였다.

그 결과 점진적 주성분 분석을 이용하여 은닉 변수를 찾고 이를 이용하여 학습 데이터로 구성된 후 분류 정확도를 측정 비교한 결과 3가지 특징 추출 방법 중 가장 높은 분류율을 가지는 주성분 분석을 특징 추출에 이용한 경우 보다 약 5% 더 높은 분류 정확도를 보였다.

또한 본 논문에서는 비교를 위해 [22]에서 사용된 뇌파 데이터에 본 논문에서 제안하는 방법과 기존연구에서 제안된 특징 추출 방법을 적용한 후 분류율을 측정하였다.

그 결과 <표 3>에서와 같이 주성분 분석을 적용한 경우보다 점진적 주성분 분석을 적용하여 차원을 축소하고 분류기 K-NN을 이용하여 분류 정확도를 측정하였을 시 점진적 주성분 분석을 이용한 방법이 평균 3% 더 높은 분류율을 보였다.

따라서 본 과제에서 제안한 방법은 대용량의 시계열 데이터를 실시간으로 효과적이고 효율적으로 요약 할 수 있다. 또한 점진적 주성분 분석을 이용하여 데이터의 특징을 지닌 은닉 변수를 찾아 시계열 데이터를 축소함으로써 한정된 메모리의 문제와 처리시간 등의 문제를 해결 할 수 있다.

<표 2> K-NN을 이용한 분류 정확도

	iPCA	PCA	DWT	DFT
3-NN	70.2329 %	64.6519 %	60.8127 %	50.3696 %
5-NN	70.1006 %	65.193 %	61.743 %	50.4076 %
7-NN	69.8665 %	65.5918 %	62.388 %	50.3627 %
평균	70.0667 %	65.1456 %	61.6479 %	50.3800 %

<표 3> [23] 데이터를 이용한 K-NN 분류 정확도

	iPCA	PCA	DWT	DFT
3-NN	76.4063 %	71.875 %	69.3125 %	66.5625 %
5-NN	70.3125 %	66.25 %	68.1563 %	62.1875 %
7-NN	65.7813 %	65.4688 %	66.0938 %	60.625 %
평균	70.8337 %	67.8646 %	67.8542 %	63.125 %

## 5. 결 론

본 논문에서는 시간의 연속성을 가지는 뇌파 데이터에 시간적 특성을 고려하는 점진적인 방법으로 데이터의 차원을 축소하였다. 방향성을 띄는 실제 동작을 행하는 동안 동작과 관련된 EEG데이터에 점진적 주성분 분석을 적용하여 동작과 관련된 특징을 포함하는 은닉변수를 찾고, 이를 기반으로 데이터의 차원을 축소하였다. 제안된 방법은 저장공간의 필요량을 줄일 수 있었으며, 데이터 차원 축소에 대한

효율성을 증명하기 위해 분류 정확도 측정을 수행 한 결과 기존 연구에 적용되었던 특징 추출 방법보다 분류학습 성능을 약5%정도 향상 시킬 수 있었다. 따라서 대량의 시계열 데이터는 주요 특징을 나타내는 몇 개의 은닉 변수로 축약될 수 있으며, 이를 이용함으로써 데이터 처리의 복잡도를 줄일 수 있을 뿐만 아니라 분류 성능 향상 및 예측 분석도 가능하다. 향후 연구로는 PCA가 데이터의 직교를 가정해야 하는 제약을 해결하는 연구가 필요하며, BCI에 바로 적용 가능하도록 더 정확한 분류 학습 연구가 필요하다.

## 참 고 문 헌

- [1] Akrami A, Solhjoo S, Motie-Nasrabadi A and Hashemi-Golpayegani M.R, "EEG-Based Mental Task Classification: Linear and Nonlinear Classification of Movement Imagery," Annual International Conference of the IEEE, pp.4626-4629, 2005.
- [2] Arthur P., Danil V.P., Richard H, Richard D. and Donald C.W., "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG," Neurocomputing pp.201-218, 2000.
- [3] Costa EJ and Cabral EF Jr, "EEG-based discrimination between imagination of left and right hand movements using Adaptive Gaussian Representation," Med Eng Phys, pp.345-348, 2000.
- [4] Dharwarkar G.S and Basir O, "Enhancing Temporal Classification of AAR Parameters in EEG single-trial analysis for Brain-Computer Interfacing," IEEE-Engineering in Medicine and Biology Society, pp.5358-5361, 2005.
- [5] Folkers A, Mosch F, Malina T and Hofmann U.G, "Realtime bioelectrical data acquisition and processing from 128 channels utilizing the wavelet-transformation," Neurocomputing, pp.247-254, 2003.
- [6] Gasser B, Toussant M, Luthringer R, Macher J.P and Cerf R, "Wave Separation versus Bandpass Filtering: a Comparative Non-linear Analysis of Brain  $\alpha$ -EEG Signals with and without Psychotropic Drug Treatment," Journal of Biological Physics, pp.209-225, 1996.
- [7] Harner P.K and Sannit T, "A review of the international ten-twenty system of electrode placement," Quincy, MA: Grass Instrument, 1974.
- [8] Li Y and Zhang S, "Apply wavelet transform to analyse EEG signal," Annual International Conference of the IEEE, pp. 1007-1008, 1996.
- [9] Lindsay I Smith, "A tutorial on Principal Components Analysis," 2002.
- [10] Lee H and Choi S, "PCA-based linear dynamical systems for multichannel EEG classification," Neural Information Processing, pp.745-749, 2002.

- [11] Michael D and Houchin J. "Automatic EEG analysis: a segmentation procedure based on the autocorrelation function," *Electroencephalogr Clin Neurophysiol*, pp.232-5, 1979.
- [12] Neep H, Jean Z.C, Ah C.T and Alex S, "Classification of EEG signals using the wavelet transform," *Signal Processing*, pp.61-72, 1997.
- [13] Pari J, Vassilis K and Kenneth R, "EEG Signal Classification Using Wavelet Feature Extraction and Neural Networks," *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing*, pp.120-124, 2006.
- [14] Scott M, Anthony J.B, Tzyy-Ping J and Terrence JS, "Independent Component Analysis of Electroencephalographic Data," *Advances in Neural Information Processing Systems*, pp.145-151, 1996.
- [15] Spiros P, Jimeng S and Christos F, "Streaming pattern discovery in multiple time-series," In *Proceeding of the 31st VLDB Conference*, pp.697-708, 2005.
- [16] Shaker M.M, "EEG waves classifier using wavelet transform and Fourier transform," *IJBS* pp.85-90, 2006.
- [17] Walter D.O and Leuchter A.F, "A Tourial on Classical computer analysis of EEGs: Spectra and Cohereneces in analysis of the Electrical Activity of the Brain," By Angeleri R, Butler S, Giaquinto S, Majkowski J. *Wiley & Sons*, pp.105-124.1997.
- [18] Yong L, Xiaorong G, Hesheng L and Shangkai G, "Classification of single-trial electroencephalogram during finger movement," *IEEE- Biomedical Engineering*, pp.1019-1025, 2004.
- [19] Zarjam P and Mesbah M, "Discrete wavelet transform based seizure detection in newborns EEG signals," *Signal Processing and Its Applications*, pp.459-462, 2003.
- [20] Zhang Q and Benveniste A, "Wavelet Networks," *IEEE Trans, Neural Networks*, pp.889-898, 1992.
- [21] 임성길, 박찬호, 이현수, "웨이블릿 신경망을 이용한 패턴 분류 시스템 설계 및 EEG 신호 분류에 대한 연구", *전자공학회논문지-CI* 제39권 3호 pp.32-43, 2002.
- [22] 이해경, 최승진, "다채널 뇌파 분류를 위한 주성분 분석 기반 선형동적시스템", *한국뇌학회지*, Vol.2, No.1, pp.53-59, 2002.
- [23] <http://ida.first.fraunhofer.de/>



### 김 선 희

e-mail : wkdal749@hanmail.net

2003년 한국교육개발원 졸업(학사)

2006년 동국대학교 컴퓨터공학과 졸업(석사)

2006년~현 재 전남대학교 전산학과 박사 과정

관심분야 : 데이터 마이닝, 센서 마이닝, 스트림 마이닝



### 양 형 정

e-mail : hjyang@chonnam.ac.kr

1991년 전북대학교 전산통계학과 졸업(학사)

1993년 전북대학교 전산통계학과 졸업(석사)

1998년 전북대학교 전산통계학과 졸업(박사)

2003~2005년 카네기멜런 대학교 연구원

2005~현 재 전남대학교 전자컴퓨터공학부 조교수

관심분야 : e-Design, 데이터 마이닝, e-Learning



### Kam Swee Ng

e-mail : kamswee@gmail.com

2005년 University Technology of Malaysia (Bachelor Degree of Computer Science)

2008년~현 재 Chonnam National University (Master in Computer Science)

관심분야 : Bioinformatics, Data Mining



### 정 종 문

e-mail : amiris@nate.com

2008년 전남대학교 전자컴퓨터공학과 졸업(학사)

2008년~현 재 전남대학교 전자컴퓨터공학부 석사과정

관심분야 : 멀티미디어, 데이터 마이닝, e-Learning