

주성분 분석과 수면 2기를 이용한 수면 장애 분류

Sleep Disturbance Classification Using PCA and Sleep Stage 2

신동근
삼육대학교 컴퓨터학부

Dong-Kun Shin(dkshin@syu.ac.kr)

요약

본 논문은 수면 2기의 EEG 신호와 주성분 분석(principle component analysis)을 이용하여 수면 장애를 분류하는 방안을 제안하고 있다. 초기 특징을 추출하기 위해서 첫 번째 단계에서는 수면 2기의 EEG 신호가 고속 푸리에 변환(fast Fourier transforms)에 의해서 잡음을 제거하는 과정이 수행되었다. 잡음이 제거된 EEG 신호를 두 번째 단계에서는 주성분 분석을 이용하여 5개의 차원으로 축소하였다. 마지막 단계에서는 축소된 5개의 차원을 가중 퍼지소속함수 기반 신경망(neural network with weighted fuzzy membership functions, NEWFM)의 입력으로 사용하여 분류성능을 측정하였다. 분류성능에 있어서 정확도(accuracy), 특이도(specificity), 민감도(sensitivity)가 모두 100%로 나타났다.

■ 중심어 : | 수면기 | 고속 푸리에 변환 | 주성분 분석 | 퍼지신경망 | 수면 장애 |

Abstract

This paper presents a methodology for classifying sleep disturbance using electroencephalogram (EEG) signal at sleep stage 2 and principal component analysis. For extracting initial features, fast Fourier transforms(FFT) were carried out to remove some noise from EEG signal at sleep stage 2. In the second phase, we used principal component analysis to reduction from EEG signal that was removed some noise by FFT to 5 features. In the final phase, 5 features were used as inputs of NEWFM to get performance results. The proposed methodology shows that accuracy rate, specificity rate, and sensitivity were all 100%.

■ keyword : | Sleep Stage | FFT | PCA | Fuzzy Neural Network | Sleep Disturbance |

I. 서론

수면은 여러 단계로 나뉘지는데, 크게 안구가 빠르게 마구 움직이는 REM (rapid eye movement) 수면과 그렇지 않은 N-REM (non-rapid eye movement) 수면으로 나뉜다[1][2]. 급속안구운동이 일어나는 REM 수면은 꿈을 꾸면서 자는 수면으로, 심박동과 호흡이 불규

칙하게 나타나는 특성이 있다. N-REM 수면은 다시 수면의 깊이에 따라 얇은 잠인 수면 1기, 수면 2기, 수면 3기, 수면 4기로 분류된다. 수면장애는 신체적, 정신적, 환경적 요인으로 수면의 양과 질에 영향을 받는 것을 말하는 것으로 수면이 규칙적이지 못하거나 수면과 관련한 비정상적인 생리적 현상을 경험하는 것을 의미한다[12]. 따라서 수면에 대해서 연구한다는 것은 환자의

신체적, 정신적, 환경적 요인을 역으로 알아볼 수 있는 중요한 요소라고 할 수 있다. 수면 중에 나타날 수 있는 질병 중에 수면무호흡증(sleep apnea syndrome)과 수면장애(sleep disturbance)를 대표적으로 들 수 있다. 수면과 관련된 연구는 주로 수면 무호흡증과 연관되어 있다[3-5]. 또한 수면기(sleep stage)에 대한 연구는 피실험자 간의 비교가 아닌 동일 피실험자를 수면기에 대한 연구가 주로 이루어지고 있다[6][7]. 그리고 뇌파를 통하여 수면 장애를 겪고 있는 사람을 분류하는 작업은 숙련된 임상 의사의 경험에 의존해야 할 뿐만 아니라 그 판정 또한 주관적이다. 일정한 규칙에 따라 시스템이 판단하는 것이 아니라 판단하는 사람의 주관적인 의견에 따라 판단 결과가 달라질 수가 있다는 단점이 있다.

본 논문에서는 Fpz-Cz 채널과 Pz-Oz 채널의 EEG 신호를 이용하여 숙면을 취하는 사람과 수면 장애가 있는 사람을 자동으로 분류하기 위해서 수면 2기의 EEG 신호와 주성분 분석(principle component analysis)를 이용하였다. NEWFM에서 입력으로 사용할 특징을 추출하기 위해서 첫 번째 단계에서는 Fpz-Cz 채널과 Pz-Oz 채널의 수면 2기 EEG 신호가 웨이블릿 변환(wavelet transforms) 또는 고속 푸리에 변환(fast Fourier transforms)에 의해서 잡음을 제거하는 과정이 수행되었다. 두 번째 단계에서는 잡음이 제거된 수면 2기 신호를 웨이블릿 기반의 특징 추출에서 사용되는 통계적 기법[11][13] 또는 주성분 분석을 이용하여 각각 80개의 특징과 5개의 차원으로 축소하였다. 마지막 단계에서는 축소된 80개의 특징과 5개의 차원을 가중 퍼지소속함수 기반 신경망(neural network with weighted fuzzy membership functions, NEWFM) [8-10]의 입력으로 사용하여 숙면을 취하는 사람과 그렇지 못하는 사람의 분류성능을 측정하였다. 웨이블릿 기반의 특징 추출 방법에 의한 분류성능에 있어서 정확도(accuracy), 특이도(specificity), 민감도(sensitivity)가 각각 91.70%, 91.73%, 91.67%로 나타났다. 고속 푸리에 변환과 주성분 분석에 의한 분류성능에 있어서 정확도(accuracy), 특이도(specificity), 민감도(sensitivity)가 모두 100%로 나타났다.

II. 수면 장애(Sleep Disturbance) 분류 모델의 설계

[그림 1]은 숙면을 취하는 사람과 수면 장애가 있는 사람의 분류 모델 다이어그램에서 알 수 있듯이 본 논문에서는 EEG 신호를 입력으로 웨이블릿 변환 또는 고속 푸리에 변환을 수행한 후에 웨이블릿 기반의 통계적 기법 또는 주성분 분석을 이용하여 특징을 추출하였다. 이렇게 추출된 특징은 NEWFM의 입력으로 사용되어 숙면을 취하는 사람과 수면 장애가 있는 사람을 분류하였다.

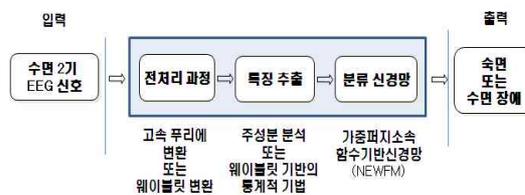
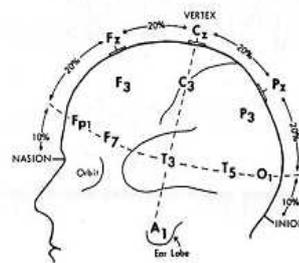


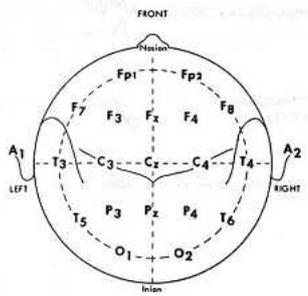
그림 1. 수면 장애 분류 모델 다이어그램

2.1 실험 데이터

본 논문에서는 PhysioBank(<http://www.physionet.org/physiobank/database/sleep-edf/>)에서 제공하는 EEG 신호를 사용하였다. EEG 신호는 21세부터 35세 사이의 백인 남녀로부터 얻었다. 이들의 EEG 신호는 그림 2에서의 Fpz-Cz 채널과 Pz-Oz 채널에서 얻었고 샘플링 주파수는 100Hz이다. 이러한 EEG 신호에는 수면기(sleep stage)로써 각성, 수면 1기, 수면 2기, 수면 3기, 수면 4기, REM이 0, 1, 2, 3, 4, 5로 표시되어 있다.



(a) Left side of head



(b) Top of head

그림 2. 국제 10-20 표준전극

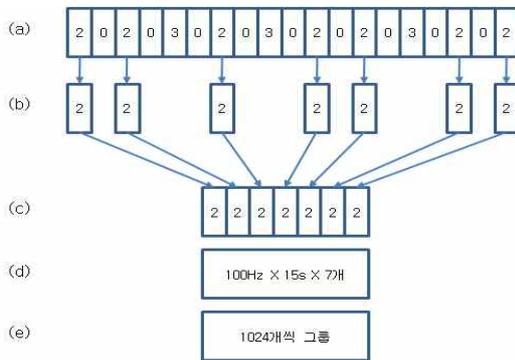


그림 3. 저장되어 있는 EEG 신호로부터 실험용 데이터를 추출하는 과정

수집된 EEG 신호를 보면 [그림 3]의 (a)와 같이 수면기가 표기되어 있다. [그림 3]의 (b)를 보면 알 수 있듯이 수면 2기의 EEG 신호만을 추출하였다. 이렇게 추출된 수면 2기의 EEG 신호를 [그림 3]의 (c)와 같이 연속적으로 합쳤다. 본 실험에서는 하나의 수면기가 15초 동안의 EEG 신호를 기반으로 표기되어 있다. 따라서 주파수가 100Hz인 신호를 15초 동안을 하나의 수면기로 표기하였기 때문에 하나의 수면기는 1500개의 EEG 신호로 그룹 되어 있다. [그림 3]의 (d)는 (c)에 대해서 계산한 수면 2기의 전체 EEG 신호를 계산한 결과이다.

본 논문에서는 수면 2기의 EEG 신호를 사용하였는데 그 이유는 수면 1기의 EEG 신호를 사용하여도 되지만 수면 1기에서는 느린 안구 운동들이 나타날 수 있고 수면 2기에서는 이러한 안구 운동이 없기 때문에 본 실험에서는 수면 2기의 EEG 신호를 선택하였다.

2.2 웨이블릿 변환(wavelet transforms) 기반의 통계적 기법

본 논문에서는 [그림 3]의 (d)에서 처리된 수면 2기의 전체 EEG 신호를 [11][13]에서 사용한 스케일 레벨 5인 이분 비연속 Haar 웨이블릿 변환을 수행하였다. 이렇게 수행된 레벨 2부터 레벨 5까지의 웨이블릿 계수인 detail coefficient와 approximation coefficient를 [11][13]에서 사용한 웨이블릿 기반의 통계적 기법을 이용하여 Fpz-Cz 채널과 Pz-Oz 채널에서 각각의 40개의 입력을 다음과 같이 추출하였다.

- (1) 각 레벨 안에 있는 모든 계수들에 대한 절대값의 평균값
- (2) 각 레벨 안에 있는 모든 계수들을 제공하여 구한 평균값
- (3) 각 레벨 안에 있는 모든 계수들의 중앙값
- (4) 각 레벨 안에 있는 모든 계수들의 표준편차
- (5) 인접한 레벨간의 레벨 안에 있는 모든 계수들에 대한 평균값의 절대값 비율

위에서 언급한 통계적 기법 (1), (2), (3)은 신호에 대한 주파수 분포를 의미한다. 또한 통계적 기법 (4), (5)는 신호에 대한 주파수 변동량을 의미한다. 본 논문에서는 통계적 기법에 의해서 추출된 총 80개의 입력을 구성원으로 하는 실험군을 구성하였다. 숙면을 취하는 사람의 2982개 실험군과 수면 장애 환자의 2321개 실험군을 이용하여 [표 1]과 같이 5대5의 비율로 훈련 집합(training sets)과 테스트 집합(test sets)으로 나누었다.

2.3 고속 푸리에 변환(fast Fourier transforms)

고속 푸리에 변환(fast Fourier transforms)은 이산 푸리에 변환(discrete Fourier transforms)을 고속으로 산출하기 위한 알고리즘이다. 고속 푸리에 변환은 디지털 신호 처리에서 편미분 방정식의 근을 구하는 알고리즘에 이르기까지 많은 분야에서 사용한다. x_0, \dots, x_{n-1} 이 복소수라고 가정할 때, 이산 푸리에 변환은 다음과 같이 정의한다.

$$f_j = \sum_{k=0}^{n-1} x_k e^{-\frac{2\pi i}{n}jk} \quad (j = 0, \dots, n-1) \quad (1)$$

이 식을 정의에 따라 계산하면 $O(n^2)$ 의 연산이 필요하지만, 고속 푸리에 변환을 이용하면 $O(n \log n)$ 의 연산만으로 가능하다. 따라서 n 의 값이 크면 클수록 고속 푸리에 변환은 더욱더 효과적이다.

[그림 3]의 (d)에서 처리된 수면 2기의 전체 EEG 신호를 고속 푸리에 변환을 하기 위해서 1024개씩 그룹화하였다. 본 실험에서는 Fpz-Cz 채널과 Pz-Oz 채널에서 각각 수집한 1024개의 EEG 신호를 고속 푸리에 변환에 의해서 각각 513개씩의 특징을 추출하였다.

2.4 주성분 분석(principle component analysis)

주성분 분석은 n개의 주어진 변수들의 일차결합으로 주성분을 구성하고 변동의 설명에 대한 기여도가 큰 순서대로 제1주성분, 제2주성분, ..., 제m주성분으로 구하여 m개의 주성분으로 전체의 변동을 설명하고자 하는 것이 목적이다. EEG 신호의 개수가 N개라고 가정하자.

$$X = (x_1, x_2, \dots, x_n) \tag{2}$$

$$S = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)(z_i - \mu)^T \tag{3}$$

여기서 S는 공분산행렬, μ 는 EEG 신호에 대한 평균을 의미한다. $W = (w_1, w_2, \dots, w_n)$ 를 m개의 가장 큰 고유값에 대응하는 m개의 고유벡터라 하자. 그러면 원래의 샘플 X의 집합에 대해서 그에 대응하는 특징벡터 $Y = (y_1, y_2, \dots, y_n)$ 는 다음과 같이 구할수 있다.

$$y_i = W^T(x_i - \mu) \tag{4}$$

본 실험에서는 Fpz-Cz 채널과 Pz-Oz 채널에서 각각 수집한 1024개의 EEG 신호를 고속 푸리에 변환에 의해서 각각 513개씩의 특징을 연결하여 1026개의 특징으로 만들었다. 이렇게 연결한 1026개의 특징을 주성분 분석을 이용하여 가장 높은 정확도를 보이고 있는 최소의 차원인 5개의 차원으로 축소하였다.

표 1. 수면 장애 분류에 사용한 실험군 (5대5 비율)

클래스	훈련 집합	테스트 집합	전체 개수
수면 장애	1161개	1160개	2321개
숙면	1491개	1491개	2982개
전체 개수	2652개	2651개	5303개

본 논문에서는 주성분 분석에 의해서 추출된 5개의 차원을 입력으로 하는 실험군을 구성하였다. 숙면을 취하는 사람의 2982개 실험군과 수면 장애 환자의 2321개 실험군을 이용하여 [표 1]과 같이 5대5의 비율로 훈련 집합(training sets)과 테스트 집합(test sets)으로 나누었다.

III. 가중 퍼지소속함수 기반 신경망 (Neural Network with Weighted Fuzzy Membership Function, NEWFM)

수면 2기의 EEG 신호로부터 숙면을 취하는 사람과 그렇지 못하는 사람을 분류하기 위해서 본 논문에서 사용한 NEWFM은 특징으로부터 학습된 가중 퍼지소속함수의 경계합(bounded sum of weighted fuzzy membership function, BSWFM)을 이용하여 클래스 분류를 하는 지도학습(supervised) 퍼지 신경망이다 [8-10]. 웨이블릿 기반의 통계적 기법에 의해서 추출된 80개의 특징 또는 주성분 분석에 의해서 축소된 5개의 차원을 그림 4에서 n개의 특징을 갖는 l번째 입력 $A_l = (a_1, a_2, \dots, a_n)$ 에 대한 NEWFM의 입력으로 사용되어진다.

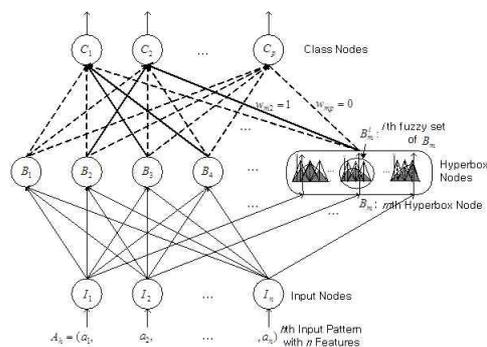


그림 4. 가중 퍼지소속함수 신경망(NEWFM)의 구조

IV. 실험 결과 (Experimental Results)

[표 2][표 3][표 4]에서는 [표 1]에서 제시한 훈련 집합과 테스트 집합의 비율이 5대 5인 실험군을 사용한 분류 성능을 보여주고 있다. 여기서 TP(True Positive)는 수면 장애가 있는 사람을 수면 장애가 있는 사람으로 분류되는 것을 말하고 TN(True Negative)은 숙면을 취하는 사람을 숙면을 취하는 사람으로 분류되는 것을 의미한다. 또한 FP(False Positive)는 수면 장애가 있는 사람을 숙면을 취하는 사람으로 분류되는 것을, FN(False Negative)은 숙면을 취하는 사람을 수면 장애가 있는 사람으로 분류되는 것을 의미한다. [표 4]에서 구한 민감도(sensitivity), 특이도(specificity), 정확도(accuracy)는 식 (5)와 같이 정의된다. 또한 [표 4]에서는 웨이블릿 변환과 고속 푸리에 변환의 분류성능을 비교하였다.

$$\begin{aligned}
 \text{민감도} &= \frac{TP}{TP+FN} \times 100 \\
 \text{특이도} &= \frac{TN}{TN+FP} \times 100 \\
 \text{정확도} &= \frac{TP+TN}{TP+FN+TN+FP} \times 100
 \end{aligned}
 \tag{5}$$

표 2. 웨이블릿 변환을 이용한 Confusion Matrix

클래스	결과(개)	
	TP	FP
수면 장애	1034	126
숙면	FN	TN
	94	1397

표 3. 고속 푸리에 변환을 이용한 Confusion Matrix

클래스	결과(개)	
	TP	FP
수면 장애	1160	0
숙면	FN	TN
	0	1491

표 4. 웨이블릿 변환과 고속 푸리에 변환의 분류성능 비교

알고리즘	정확도	특이도	민감도
웨이블릿 변환	91.70%	91.73%	91.67%
고속 푸리에 변환	100%	100%	100%

V. 결 론

본 논문에서는 Fpz-Cz 채널과 Pz-Oz 채널의 수면 2기의 EEG 신호를 이용하여 숙면을 취하는 사람과 수면 장애가 있는 사람을 분류하였다. 수면 2기의 EEG 신호를 사용한 이유는 수면 1기의 EEG 신호를 사용하여도 되지만 수면 1기에서는 느린 안구 운동들이 나타날 수 있고 수면 2기에서는 이러한 안구 운동이 없기 때문에 본 논문에서는 수면 2기의 EEG 신호를 선택하였다. 또한 본 논문에서는 수면 장애 분류를 위해서 웨이블릿 변환(wavelet transforms)과 고속 푸리에 변환(fast Fourier transforms)을 비교 실험하였다. 수면 2기의 EEG 신호는 시간의 흐름에 대한 주파수의 분석이 아니라 특정 시점에 대한 주파수의 분석이므로 고속 푸리에 변환을 사용한 분류 성능이 웨이블릿 변환을 사용한 분류 성능보다 높게 나타났다. 이러한 실험결과를 통하여 수면 분석에는 고속 푸리에 변환이 웨이블릿 변환 보다 효과적임을 알 수가 있었다. 본 논문에서는 수면 2기에 대해서만 실험하였지만 향후에는 각각의 수면기(sleep stages)간의 비교 실험을 통하여 수면의 질을 연구하는 실험을 할 예정이다.

참 고 문 헌

[1] D. J. Dijk, T. L. Shanahan, J. F. Duffy, J. M. Ronda, and C. A. Czeisler, "Variation of electroencephalographic activity during non-rapid eye movement and rapid eye movement sleep with phase of circadian melatonin rhythm in humans," *J. Physiol.* 505, pp.851-858, 1997.

[2] E. O. Bixler, A. Kales, J. A. Jacoby, C. R. Soldatos, and A. Vela-Bueno, "Nocturnal sleep and wakefulness: effects of age and sex in normal sleeper," *Int. J. Neurosci.* 23, pp.33-42, 1984.

[3] J. M. Lee, D. J. Kim, I. Y. Kim, K. S. Park, and S. I. Kim, "Detrended fluctuation analysis of

- EEG in sleep apnea using MIT-BIH polysomnography data," *Computers in Biology and Medicine* 32, pp.37-47, 2002.
- [4] Y. S. Chung, "Pathophysiology and Diagnosis of Sleep Apnea," *The KJAsEM*, Vol.20, No.1, 2010.
- [5] Elif Derya Übeyli, Dean Cvetkovic, Gerard Holland, and Irena Cosic, "Adaptive neuro-fuzzy inference system employing wavelet coefficients for detection of alterations in sleep EEG activity during hypopnoea episodes," *Digital Signal Processing* 20, pp.678-691, 2010.
- [6] Rajendra Acharya, Oliver Faust, N. Kannathal, TjiLeng Chua, and Swamy Laxminarayanb, "Non-linear analysis of EEG signals at various sleep stages," *Computer Methods and Programs in Biomedicine* 80, pp.37-45, 2005.
- [7] Salih Güneş, Kemal Polat, and Şebnem Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting," *Expert Systems with Applications* 37, pp.7922-7928, 2010.
- [8] S. H. Lee and J. S. Lim, "Forecasting KOSPI based on a neural network with weighted fuzzy membership functions," *Expert Systems with Applications* 38, pp.4259-4263, 2011.
- [9] J. S. Lim, "Finding Features for Real-Time Premature Ventricular Contraction Detection Using a Fuzzy Neural Network System," *IEEE TRANSACTIONS ON NEURAL NETWORKS* 20, pp.522-527, 2009.
- [10] J. S. Lim, D. Wang, Y.-S. Kim, and S. Gupta, "A neuro-fuzzy approach for diagnosis of antibody deficiency syndrome," *Neurocomputing* 69, pp.969-974, 2006.
- [11] 이상홍, 신동근, 임준식, "운동 형상 분류를 위한 웨이블릿 기반 최소의 특징 선택", *한국콘텐츠학*

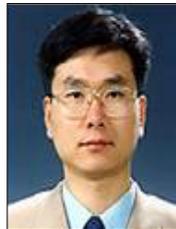
회논문지, Vol.10, No.6, pp.27-34, 2010.

- [12] T. L. Baker, "Medical clinics of north america," Philadelphia; Saunders, 1985.
- [13] 이상홍, 임준식, 신동근, "걸음걸이 분석 기반의 파킨슨병 분류를 위한 특징 추출," *한국인터넷정보학회논문지*, Vol.11, No.6, pp.13-20, 2010.

저 자 소 개

신 동 근(Dong-Kun Shin)

정회원



- 1986년 : 인하대학교 전자계산학과 졸업(학사)
 - 1996년 : 동국대학교 대학원 전자계산학과 졸업(공학석사)
 - 2010년 : 인하대학교 대학원 컴퓨터정보공학과 졸업(공학박사)
 - 1997년 9월 ~ 2006년 2월 : 삼육의명대학 컴퓨터정보과 교수
 - 2006년 3월 ~ 현재 : 삼육대학교 컴퓨터학부 교수
- <관심분야> : neuro-fuzzy System, HCI, Software Engineering