

# EEG 데이터의 고차원 하이퍼에지에서의 주성분 분석

<sup>0</sup>김준식<sup>1</sup>, 이충연<sup>2</sup>, 장병탁<sup>1,2</sup>

<sup>1</sup>바이오지능 연구실, 서울대학교 컴퓨터공학부

<sup>2</sup>서울대학교 뇌과학협동과정

{jskim, cylee, btzhang}@bi.snu.ac.kr

## Principal Component Analysis of Higher-Order Hyperedges in EEG Data

<sup>0</sup>Joon Shik Kim<sup>1</sup>, Chung-Yeon Lee<sup>2</sup>, Byoung-Tak Zhang<sup>1,2</sup>

<sup>1</sup>Biointelligence Lab, School of Computer Science and Engineering,

<sup>2</sup>Interdisciplinary Program in Brain Science,

Seoul National University

### 요 약

고차 주성분 방법으로는 텐서 분석이 있었다. Electroencephalography(EEG) 데이터나 Social Network 데이터에 텐서 분석이 적용되어 주요한 성분들을 찾는 연구들이 있었다. 그러나 텐서 분석은 직관적으로 이해하기에 어려움이 있으며 중요한 노드를 찾는 데에는 다소 어려움이 있다. 본 논문에서는 고차 하이퍼에지 이차원 행렬을 만들고 주성분분석법을 이용하여 중요한 노드를 찾는 새로운 방법론을 제시한다. 데이터로는 Multimodal Memory Game(MMG) 수행시 촬영한 EEG 데이터를 사용하였다. MMG는 TV 드라마 기반의 기억인출게임이다. 베타파의 Power Spectrum Density(PSD)는 각 위치의 채널들의 활성도를 나타내는 지표이다. 우리는 Random Sampling을 바탕으로 PSD 상위 50%의 채널들간의 전이행렬을 구하였다. 그 후 고유치와 고유벡터를 구하였다. 가장 큰 고유치의 고유벡터는 주성분을 나타내며 고유벡터의 각 원소들은 중요도를 나타내는 centrality 이다. 세 명의 피험자에 대한 centrality 상위 30개의 중요한 채널들을 구하였고 세 명에 공통적으로 포함되는 채널을 확인하였다.

### 1. 서론

EEG는 우리말로 뇌파 혹은 뇌전도로 표현된다. 피험자가 특정 과업을 수행할 때 관련된 뇌 영역에서 전류가 발생되어 EEG 전극에서 측정된다. 해상도가 높은 그림과 낮은 그림을 볼 때 EEG 신호의 차이를 보는 연구가 있었다[1]. EEG 데이터를 space-time-frequency의 3차원 행렬로 만들고 tensor 분석을 통하여 성분들을 찾아내는 연구도 있었다[2]. Tensor 분석에서는 고차원 행렬을 outer product 연산을 통하여 성분간의 관계를 나타내는 core와 각 성분들로 분해한다[3,4]. 텐서를 이용하여 연상기억을 구현하는 연구는 인공지능 분야에서 시도되었었다[5].

최근에 하이퍼에지를 이용하여 고차원 상관관계를 나타내려는 시도가 있었다[6]. 그러나 아직까지 하이퍼에지들의 행렬에서 주성분을 찾는 연구는 없었다. 본 논문의 접근은 성분분석을 통하여 차원을 축소하는 기존의 연구[7-9]를 하이퍼에지 공간으로 차원을 확장한다는 데 의미가 있다.

### 2. 연구방법

MMG에서는 미국 드라마 “Friends”를 20분간 시청한 뒤 memory recall 게임을 수행한다. Text-2-Image(T2I) 게임에서는

cue presentation(단서제시화면)에서 비디오의 대사 하나가 음성과 함께 제시되고 answer 단계에서 제시되는 네 개의 이미지 중에서 앞서 제시된 대사와 관련된 화면을 고른다. 마지막으로 check 단계에서는 대사와 관련 이미지가 함께 주어져서 정답을 알려준다. 이 과정을 20번 반복하여 진행한다. 이미지를 제시하고 관련 대사를 고르는 Image-2-Text(I2T) 게임도 수행하였다.

이 실험에서는 세 명의 피험자에 대하여 128 채널 EEG 장비를 이용하여 MMG 수행 시 뇌전도 데이터를 측정하였다. 베타 영역(13-30 Hz)의 PSD를 구하고 그 로그 값으로 각 채널들의 신호 강도를 구하였다. 베타파를 선택한 것은 베타파가 작업기억의 인지와 지각에 관련이 있기 때문이다. Ground, Electrooculography (EOG) 채널들 그리고 노이즈가 있는 채널들을 제거하여 3명의 피험자에 공통된 115 채널들을 선택하였다. 본 연구의 목적은 MMG T2I 작업 수행에 중요한 채널을 식별하는 것이므로 구해진  $\log(\text{PSD})$  값의 각 작업별 상위 50%는 1로 하고 하위 50%는 0으로 두었다. 여기서 작업은 각 세션의 cue, answer, check 단계들을 의미한다. Epoch 1의 cue, answer, check, epoch 2의 cue, answer, check, ..., epoch 20의 cue, answer, check 순으로 작업이 진행되었다. Random sampling을 앞에 기술된 작업의 진행 순서로 실시하여 전이행렬  $G$ 를 구하였다. 선행 작업에서 115 채널 중 1-order 혹은 2-order의 하이퍼에지를 선택하고 바로 다음에 오는 후행

작업에서 1-, 혹은 2-order의 하이퍼에지를 선택하였다. 전이행렬 G의 차원은 6670×6670 이다. 행과 열의 처음 115개가 1-order 하이퍼에지들을 나타낸다. 116부터 6670은 2-order 하이퍼에지들을 나타낸다. 선택된 선행 하이퍼에지 i와 후행 하이퍼에지 j의 값이 모두 1일 때 G의 (j,i) 값이 1 만큼 증가된다. 물론 처음에 G의 모든 행렬 원소값은 0 이었다. Sampling은 총 222444500번 시행하였다. Sampling 과정은 그림 1에서 나타내었다. G를 구한 후 G의 모든 원소를 그 중 최대값으로 나누어서 normalization을 하였다. 그 후 고유치와 고유벡터를 구하였다. 가장 큰 고유치에 해당하는 고유벡터가 가장 많이 기여하는 성분이며 그 성분의 원소값이 각 원소의 중요도를 나타내는 centrality를 나타낸다[10].

### Sampling for Constructing a Transition Matrix

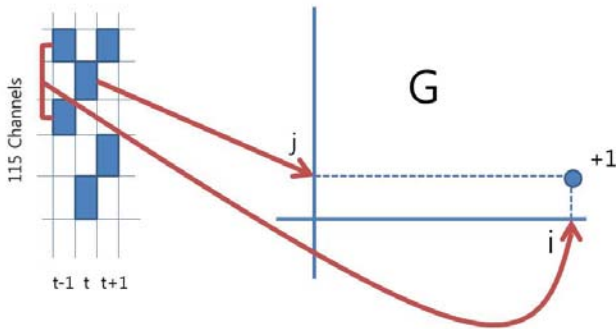


그림 1. 전이행렬 G를 sampling을 통하여 만드는 과정

### 3. 결과

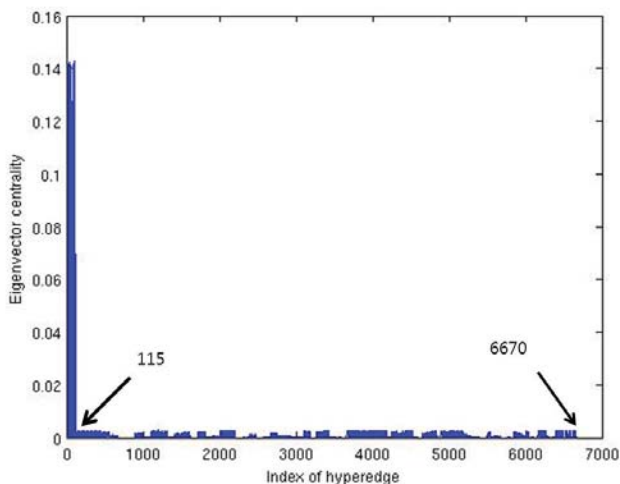


그림 2. 피험자 1에서의 주성분의 eigenvector centrality(EC). x축에서 1에서 115의 1-order 하이퍼에지들의 EC가 나머지 6670까지의 2-order 하이퍼에지들보다 월등하게 큼을 볼 수 있다.

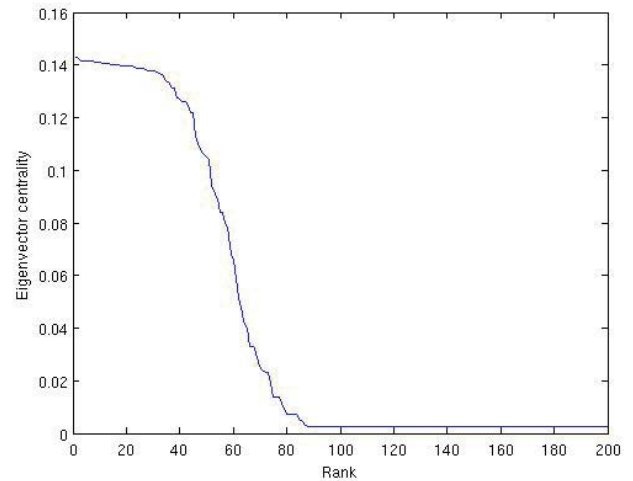


그림 3. 피험자 1의 EC의 내림차순 정렬 분포. 순위 1부터 40까지의 채널들이 주요한 기여를 하는 것을 알 수 있다.

그림 2는 피험자 1의 eigenvector centrality(EC)를 보여준다. 가장 큰 eigenvalue의 eigenvector 원소들을 그림으로 그렸다.

그림 3에서는 subject 1의 EC의 내림차순 정렬 결과를 순위 1부터 200까지 보여준다. 순위 1부터 40까지의 채널들이 주요한 기여를 하는 것을 알 수 있다.

표 1은 세 사람의 피험자의 EC 상위 30개의 채널번호를 나타낸다.

표 1. EC 상위 30개의 채널번호

Subject Number	Channel Index
1	105, 38, 102, 98, 40, 92, 19, 93, 43, 49, 96, 85, 110, 34, 86, 112, 60, 20, 109, 94, 54, 10, 44, 48, 59, 108, 84, 58, 3, 99
2	71, 16, 87, 89, 14, 7, 103, 24, 69, 2, 76, 41, 44, 86, 21, 59, 85, 81, 80, 36, 61, 63, 19, 23, 31, 11, 75, 56, 51, 65
3	25, 88, 2, 26, 33, 19, 80, 43, 17, 41, 5, 32, 65, 63, 28, 11, 58, 1, 77, 4, 76, 112, 64, 30, 6, 75, 12, 115, 114, 109

그림 4에서는 채널들의 분포를 보여준다. 앞쪽은 전두엽(frontal), 뒤쪽은 후두엽(occipital), 옆쪽은 측두엽(temporal), 머리 윗쪽은 두정엽(parietal) 이다. 우반구는 짝수로 좌반구는 홀수로 나타낸다.

표 2는 세 피험자간의 상위 EC 30 채널들의 공통적으로 포함되는 채널의 갯수를 나타낸다. 표 1의 세 명에 대한 채널번호에서 세 명 모두에 포함되는 채널은 19이며 이는 FCZ이다. 그림 4에서 중앙 위에서 세 번째 채널이 FCZ임을 확인 할 수 있다.

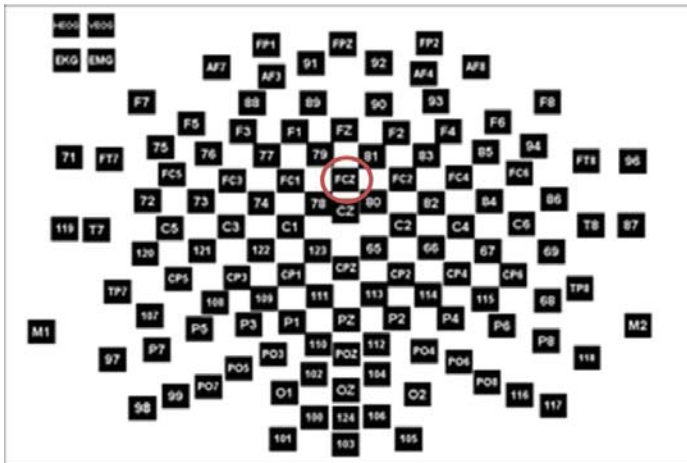


그림 4. EEG 채널들의 위치 분포. FCZ가 붉은색 원으로 표시되어 있다.

표 2. 세 피험자간에 공통적으로 포함되는 채널의 수

관련된 피험자	공통적으로 포함된 채널의 수	공통된 채널
1, 2	5	19, 85, 86, 44, 59
2, 3	10	2, 76, 41, 80, 63, 19, 23, 11, 75, 65
1, 3	5	19, 43, 58, 112, 109
1, 2, 3	1	19

#### 4. 결론

그림 2에서 볼 때, 상위 115개의 1-order 하이퍼에지의 EC가 그 이후의 2-order 하이퍼에지의 centrality보다 월등하게 큰 값을 가진다. 이는 random sampling 시에 second order 하이퍼에지의 선택이 그 만큼 적었음을 의미한다. Sampling 과정에서 1-order와 2-order의 선택을 동등하게 하였다. 1-order는 115개이고 2-order는 6555개 이므로 각각의 2-order 하이퍼에지가 1-order 하이퍼에지에 비하여 적게 선택되었다. 이런 이유로 2-order 하이퍼에지의 EC가 작게 나왔다고 추론할 수 있다.

표 2에서 두 명의 피험자에 공통적으로 포함되는 EC 상위 30개의 채널들은 각각 5, 10, 5로서 어느 정도 중요한 채널들이 피험자들에 대하여 공통적이라고 하겠다. 이 결과는 EC 기반으로 중요한 채널을 찾는 본 논문의 방법이 유효함을 보여준다.

세 명의 피험자에 공통적으로 포함된 FCZ는 앞쪽 두정엽 중앙에 위치한 채널로서 기억 인출시에 머리의 앞쪽 부분이 중요한 역할을 수행함을 추측할 수 있다[11].

#### 5. 감사의 글

이 논문은 교육과학기술부의 재원으로 국가연구재단의 지원을 받아 수행한 연구(2011-0018299, BrainNet 및 2011-0016483, Videome)이며, 지식경제부 산업원천기술개발사업(10035348, mLIFE) 및 교육과학기술부의 BK21-IT사업에 의해 일부 지원되었음.

#### 참고문헌

- [1] S. Scholler, S. Bosse, M.S. Treder, B. Blankertz, G. Curio, K.-R. Müller, T. Wiegand, "Towards a direct measure of video quality perception using EEG," *IEEE Trans Image Process*, 2012.
- [2] F. Miwakeichi, E. Martínez-Montes, P.A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, Y. Yamaguchi, "Decomposing EEG data into space-time-frequency components using parallel factor analysis," *Neuroimage*, Vol. 22, pp. 1035-1045, 2004.
- [3] T.G. Kolda, B.W. Bader, "Tensor decompositions and applications," *SIAM Review*, Vol. 51, pp. 455-500, 2009.
- [4] D.G. Leibovici, "Spatio-temporal multiway decompositions using principal tensor analysis on k-modes: The R package PTak," *Journal of Statistical Software*, Vol. 34, pp. 1-34, 2010.
- [5] P. Smolensky, "Tensor product variable binding and the representation of symbolic structures in connectionist systems," *Artificial Intelligence*, Vol. 46, pp. 159-216, 1990.
- [6] B.-T. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory," *IEEE Computational Intelligence Magazine*, Vol. 3, pp. 49-63, 2008.
- [7] M. Belkin, P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, Vol. 15, pp. 1373-1396, 2003.
- [8] J. Dien, D.J. Beal, P. Berg, "Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations," *Clinical Neurophysiology*, Vol. 116, pp. 1808-1825, 2005.
- [9] C.E. Tenke, J. Kayser, "Reference-free quantification of EEG spectra: Combining current source density (CSD) and frequency principal components analysis (fPCA)," *Clinical Neurophysiology*, Vol. 116, pp. 2826-2846, 2005.
- [10] P. Bonacich, A.C. Holdren, M. Johnston, "Hyper-edges and multidimensional centrality," *Social Networks*, Vol. 26, pp. 189-203, 2004.
- [11] J. Onton, A. Delorme, S. Makeig, "Frontal midline EEG dynamics during working memory," *Neuroimage*, Vol. 27, pp 341-356, 2005.