

뇌파 기반 실시간 집중 상태 분류 기술의 타당성 연구*

한구현⁰¹ 안단태¹ 송현주² 김보형³ 서진욱¹

¹서울대학교 컴퓨터공학부

²송실대학교 컴퓨터학부

³한국외국어대학교 바이오메디컬공학부

ghhan@hcil.snu.ac.kr, dtan@hcil.snu.ac.kr, hsong@ssu.ac.kr, bkim@huf.ac.kr, jseo@snu.ac.kr

Feasibility of EEG-based Real-time Attention State Classification

GuHyun Han⁰¹ Dantae An¹ Hyunjoo Song² Bohyoung Kim³ Jinwook Seo¹

¹Department of Computer Science & Engineering, Seoul National University

²School of Computer Science & Engineering, Soongsil University

³Division of Biomedical Engineering, Hankuk University of Foreign Studies

요약

본 연구는 뇌파를 활용하여 사용자의 실시간 집중 상태를 분류하는 기술의 타당성과 일상 환경에의 적용 가능성을 탐구하고자 했다. 먼저, 20~30대 성인 25명을 대상으로 책상에 앉아 휴식 중일 때와 영상 클립(자연경관 및 토론)을 시청할 때의 뇌전도를 측정하여 19채널 뇌파 데이터를 확보했다. 다음으로, 반복 측정 분산 분석을 수행하여 태스크 종류 및 뇌엽 위치에 따른 뇌파 주파수 밴드 특성값의 차이를 검증했다. 그다음으로, 유의한 분석 결과와 선행 연구를 참고하여 특성 집합을 선정한 뒤 현재 수행 중인 태스크를 초 단위로 예측하도록 개인화된 SVM 모델을 훈련했다. 테스트 결과, 개인화된 SVM들의 평균 f1-점수는 2-클래스 분류 시 0.95, 3-클래스 분류 시 0.80, 5-클래스 분류 시 0.57이었다.

1. 서론

본 연구는 BCI (Brain-Computer Interface) 방법론을 활용하여 사용자의 실시간 집중상태를 예측하는 기술의 타당성을 확인하고자 했다. 먼저, 요구하는 집중 상태가 서로 다른 동영상 시청 태스크들을 수행할 때의 뇌전도를 수집하는 실험을 수행하여 자체적인 뇌파 데이터를 확보했다. 다음으로, 사용자·태스크·뇌엽(brain lobe)을 요인으로 주파수 영역의 뇌파 특성을 관찰하고 통계적으로 비교 분석했다. 그다음, 분석 결과를 기준으로 특성 집합(feature set)을 선정하여 집중상태를 분류하는 기계학습 모델을 훈련했다. 마지막으로, 집중 상태 구분의 해상도 측면에서 분류 모델의 성능을 평가했다.

2. 관련 연구

뇌파는 뇌신경 내에 발생하는 전기적 흐름을 말한다. 뇌전도(EEG) 측정으로 수집한 데이터를 관찰하여 뇌파를 분석할 수 있다. EEG 데이터 내의 복잡한 파형을 이해하기 위해 주파수 영역으로 변환한 뒤에 스펙트럼 밀도(Spectral density)를 분석하는 것이 일반적이다. 분석에 자주 활용되는 주파수 영역에는 델타(1-4Hz), 세타(4-8Hz), 알파(8-12Hz), 베타(13-30Hz), 감마(30Hz 이

상)가 있다. 대뇌피질은 영역별로 주요한 담당 기능이 다르며, 이로 인해 EEG 데이터의 주파수 특성도 측정 부위 별로 다른 양상을 보이기도 한다.

인간-컴퓨터 상호작용 분야에서는 뇌파를 활용한 시스템 인터페이스 및 상호작용 방법론을 연구해왔으며, 뇌-컴퓨터 인터페이스(Brain-Computer Interface; BCI)라고 한다. 해당 분야에서는 뇌파로부터 인지, 운동, 감정 상태 등을 추론하여 사용성(Usability)을 개선하거나 새로운 사용 경험(User Experience)을 제공하는 방법을 연구한다. 측정자의 집중상태 정보를 활용하는 BCI 연구들도 활발하게 이루어지고 있다. 운전자의 주변 인식(Peripheral awareness) 능력을 판단하여 가속 여부를 보조하는 전기 자전거[1], 청취자의 태스크 관여(Task Engagement) 정도에 반응하는 스토리텔링 로봇[2] 등을 사례로 들 수 있다.

3. EEG 데이터 수집 및 분석

3.1. 실험 설계

책상에 앉아 영상(시청각 자극)을 시청하는 태스크 수행 중의 EEG 데이터를 측정하는 실험을 진행했다. 실험 계획 및 참가자 모집은 연구윤리위원회의 승인 아래 진행되었다. 참가자는 별도로 마련된 실험 공간 내에서 건식 전극(Dry electrode) 방식의 19채널 EEG 측정 장비를 착용하고 책상 앞에 앉아 뇌전도를 측정했다. 참가자는

* 이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2019R1A2C1088900).

서로 다른 집중 정도를 요구하는 세 종류의 상황에서 총 다섯 번 뇌전도를 측정했다. 먼저, 눈을 감고 3분간 휴식하는 동안 신호를 측정했다. 그다음, 눈을 뜨고 4분 길이의 자연경관 영상을 감상하는 동안 신호를 측정했다. 참가자는 편안한 자세를 유지한 상태로 해당 영상을 감상했다. 마지막으로, 토론 영상을 시청하는 동안 신호를 측정했다. 영상은 세 종류(여성가족부 정책, 가상화폐, 기본소득)였고, 각 영상의 길이는 6분 내외였으며, 제공 순서는 카운터밸런싱(counterbalancing)했다. 토론 영상 시청 전후에 주제 관심도, 주제에 대한 의견 등을 묻는 객관식 문항으로 설문을 진행했다. 시청 중에는 편안한 자세를 유지했다. 실험 설명, 장비 착용, 측정 간 휴식을 포함하여 위의 프로토콜 수행에 1시간 내외 소요되었다.

3.2. 참가자 모집

대학교 학내 온라인 커뮤니티를 통해 뇌 질환, 혹은 인지 기능 장애를 겪지 않은 20~30대의 성인 남녀를 모집했다. 총 27명이 지원했고, 기술적 문제로 인해 뇌파가 정상적으로 측정되지 않은 두 명을 제외한 25명(남=13명, 여=12명)이 측정을 완료했다. 측정을 완료한 참가자에게는 사례비로 1만 3천 원을 지급했다.

3.3. 수집 데이터 전처리

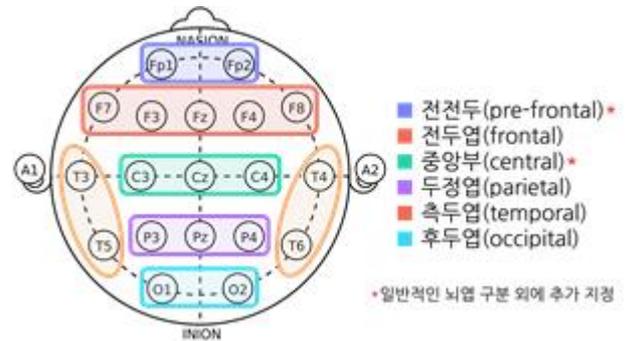
EEG 데이터에 특화된 전처리에는 MNE-python을 사용했다. 먼저, 하이패스, 로우패스, 노치 필터를 적용한 뒤 ICA (Independent-Component Analysis)를 적용하여 아티팩트를 제거했다. 다음으로, 노이즈 경감 및 계산 효율성 증대를 위해 다운 샘플링(500Hz→250Hz)한 뒤 1초 윈도우 단위로 쪼갰다. 마지막으로, 각 윈도우에 Welch의 파워 스펙트럼 밀도 계산 알고리즘을 적용하여 델타·세타·알파·베타·감마 밴드의 파워 스펙트럼 밀도를 계산했다.

4. 뇌파 데이터 통계 분석

4.1. 분석 방법

전처리가 완료된 EEG 데이터를 참가자·태스크·뇌엽 기준으로 집계하여 통계 분석용 데이터 테이블을 생성했다. 뇌엽 변수는 총 6개의 범주로 분류했다. 그림 1처럼 10-20 시스템에 근거하여 전두엽, 두정엽, 측두엽, 후두엽에 위치한 전극들을 네 영역으로 묶어 구분했다. 전전두엽(Fp1, Fp2) 영역 및 두상 중앙(Cz, C3, C4)영역 전극들을 묶어 영역을 추가로 구분했다. 참가자별로 태스크 및 뇌엽 단위 데이터 평균을 집계하여 최종적으로 참가자(25) * 태스크(5) * 뇌엽(6) * 주파수 밴드(5) 형태로 정리했다.

위 데이터 테이블을 활용하여 뇌파 밴드별 상대 파워(relative power)를 통계적으로 분석했다. 먼저, 뇌파 크기의 개인차를 보정하기 위해 참가자 별로 각 뇌파 밴드의 상대 파워를 계산한 뒤 태스크 및 뇌엽에 의한 차이를 확인했다. 실험 설계상 태스크 및 뇌엽 모두 동일 참가자로부터 반복 측정한 변수이므로 분석 방법은 2-way Repeated Measure ANOVA를 선택했다. 구형성(sphericity) 보정을 위해 Greenhouse-Geisser 보정 p-value로 검정했다. 사후 검정은 가능한 모든 조합에 대해 paired t-test를 수행한 후, 본페로니 교정하여 신뢰도 95% 수준에서 검정했다.



(그림 1) 10-20 시스템 기준 뇌엽 변수 설정

4.2. 뇌파 밴드별 분석 결과

표 1과 같이 모든 뇌파 밴드에서 상대 파워 비율의 평균 측정값이 태스크 및 뇌엽에 따라 유의한 차이가 있었다. 또한, 상대적 저주파 밴드(델타, 세타, 알파)에서 태스크와 뇌엽 간의 유의한 상호작용 효과를 확인했다. 태스크 변수를 사후 검정한 결과, 모든 밴드 영역에서 휴식상태와 영상 시청 태스크(자연경관, 토론 영상 3종) 간 유의한 차이가 있었다. 알파 밴드와 타 뇌파 밴드가 대조적인 결과를 보였다. 델타, 세타, 베타, 감마 밴드는 휴식상태 측정 평균값이 다른 시청 태스크보다 유의하게 작았으나 알파 밴드에서는 반대로 유의하게 컸기 때문이다. 추가로, 세타 밴드에서만 자연경관 영상 시청 태스크 측정 평균값이 토론 영상 시청 태스크 측정 평균값보다 유의하게 작았다.

뇌파 밴드별 분석 결과 아래와 같은 통찰을 얻을 수 있었다. 먼저, 뇌파 밴드 간 상대 파워 비율이 측정자의 상태를 이진 분류하는 데에 활용 가능한 특성임을 확인했다. 특히, 알파 밴드는 두 상태 간 측정값 평균 차이가 크고 부호 또한 다른 밴드와 정반대이므로 이진 분류 시 주요한 영향을 미칠 것으로 예상했다. 두 번째로, 세타 밴드의 뇌파 특성값은 다른 밴드보다 한 단계 더 세밀한 상태 구분(휴식-자연경관 영상-토론 영상)에 활용 가능하리라 예상할 수 있었다. 좀 더 높은 집중도를 요

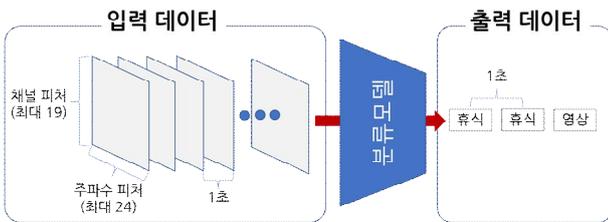
구하는 태스크를 수행할수록 측정값의 평균이 유의하게 증가했기 때문이다. 마지막으로, 측정 뇌엽을 제한했을 때 상태 분류 능력에 변동이 생길 수 있음을 확인했다. 알파 밴드는 뇌의 후면부에서 측정 시 더 높은 파워 비율을 점유했으나 다른 밴드는 전면부에서 더 높은 비율을 차지했기 때문이다.

(표 1) 뇌파 밴드별 ANOVA 검정결과

뇌파	요인	검정 결과	유의
델타	태스크	$F_{(4,96)} = 39.42, p < 0.01$	✓
	뇌엽	$F_{(5,120)} = 6.40, p < 0.01$	✓
	상호작용	$F_{(20,48)} = 9.13, p < 0.01$	✓
세타	태스크	$F_{(4,96)} = 24.83, p < 0.01$	✓
	뇌엽	$F_{(5,120)} = 3.32, p < 0.05$	✓
	상호작용	$F_{(20,48)} = 17.68, p < 0.01$	✓
알파	태스크	$F_{(4,96)} = 126.28, p < 0.01$	✓
	뇌엽	$F_{(5,120)} = 11.51, p < 0.01$	✓
	상호작용	$F_{(20,48)} = 22.11, p < 0.01$	✓
베타	태스크	$F_{(4,96)} = 14.22, p < 0.01$	✓
	뇌엽	$F_{(5,120)} = 8.91, p < 0.01$	✓
	상호작용	$F_{(20,48)} = 2.72, p > 0.05$	
감마	태스크	$F_{(4,96)} = 15.36, p < 0.01$	✓
	뇌엽	$F_{(5,120)} = 7.73, p < 0.01$	✓
	상호작용	$F_{(20,48)} = 2.15, p > 0.05$	

5. 집중상태 분류 기계학습 모델 도출

5.1. 모델 선정



(그림 2) SVM 모델을 활용한 실시간 분류

학습데이터가 상대적으로 적은 경우에도 효과적인 학습이 가능하고, 잡음에 강건하다고 알려진 서포트 벡터 머신(Support Vector Machine; SVM)을 선택했다. LibSVM 기반의 C-Support Vector Classification 모델을 사용했으며 C regularization parameter는 1.0, 커널 함수는 radial basis function을 사용했다. 통계 분석 결과를 기반으로 19개의 채널에서 24종의 채널 별 주파수 영역 피쳐들을 선정했다. 주요 피쳐들은 상대 파워, 최대 파워, 최대 파워 주파수, 파워 비율이었다. 총 456개(19 * 24)의 피쳐를 입력받아 초 단위로 현재 수행 중인 태스크를 분류할 수 있도록 설계했다(그림 2). 태스크 구분의 해상도를 더욱 세밀하게 파악하기 위해 2-클래스(휴식, 영상 시청), 3-클래스(휴식, 자연경관 영상, 토론 영상), 5-클래스(휴식, 자연경관 영상, 여가부 토론, 비트코인 토론, 기본소득 토론) 분류 모델을 각각 구성했다.

식, 영상 시청), 3-클래스(휴식, 자연경관 영상, 토론 영상), 5-클래스(휴식, 자연경관 영상, 여가부 토론, 비트코인 토론, 기본소득 토론) 분류 모델을 각각 구성했다.

5.2. 훈련 및 성능 평가

각 참가자의 태스크 세션별 데이터 내에서 비복원 무작위 추출하여 80%를 훈련용, 나머지 20%를 테스트용 데이터로 분리했다. 이를 활용하여 참가자 별로 개인화된 모델을 훈련했고 테스트했다. 2-클래스, 3-클래스로 분류할 경우 클래스 간 불균형(2-클래스의 영상 시청, 3-클래스의 토론 영상 3종)을 상쇄하기 위해 훈련 데이터의 클래스 빈도에 반비례하는 클래스 가중치를 주어 훈련했다. 성능 평가를 위해 각 훈련모델의 분류 클래스별로 precision, recall, f1-점수를 계산했으며, 모델 전체 성능은 클래스별 성능지표에 테스트 셋의 클래스 빈도를 반영한 가중평균값으로 평가했다.

5.3. 결과

모든 분류 모델이 무작위 추측(1/2, 1/3, 1/5)에 비해 매우 높은 성능을 보였고, 특히 2-클래스 분류 모델이 신뢰할만한 성능을 보였다. 동영상 태스크(자연경관 및 토론 영상)를 세부 분류할수록 성능이 떨어졌다. 3-클래스 모델에서는 자연경관 영상 클래스의 평균 분류 성능(f1=0.51)이 휴식(f1=0.83) 및 토론 영상 3종(f1=0.85)보다 낮았다. 5-클래스 모델에서는 클래스별로 휴식(f1=0.83), 여가부 토론(f1=0.58), 기본소득 토론(f1=0.54), 자연경관 영상(f1=0.51), 비트코인 토론(f1=0.48) 순의 평균 분류 성능을 보였다.

(표 2) 개인화 분류 모델 해상도별 평균 성능

분류모델	Precision	Recall	F1
2-클래스	0.9583	0.9532	0.9548
3-클래스	0.8281	0.7884	0.8010
5-클래스	0.5781	0.5696	0.5692

참고문헌

- [1] Andres, Josh and schraefel, m.c. and Semertzidis, Nathan and Dwivedi, Brahma and Kulwe, Yutika C. and von Kaenel, Juerg and Mueller, Florian Floyd, "Introducing Peripheral Awareness as a Neurological State for Human-Computer Integration", Proceedings of the CHI Conference on Human Factors in Computing Systems, pp.1-13, 2020.
- [2] Szafir, Daniel and Mutlu, Bilge, "Pay Attention! Designing Adaptive Agents That Monitor and Improve User Engagement", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.11-20, 2012.