

Attention 기반의 장단기적 시간 정보를 활용한 EEG 감정인식 모델

김하늘^{*1}, 김호재^{*2}, 우다연^{*3}, 윤성호^{*4}, *진창균⁵

고려대학교 건축사회환경공학부¹, 성균관대학교 소프트웨어학과², 성균관대학교
데이터사이언스융합전공³, 서강대학교 수학과⁴, 서울과학기술대학교 인공지능응용학과⁵ 소속
e-mail : adsky0309@korea.ac.kr¹, ghghghost@g.skku.edu², dyeonwu@gmail.com³,
jason1009@sogang.ac.kr⁴, jcg.brainai@gmail.com⁵

EEG emotion recognition model using attention based long- short term
temporal information

Kim Haneul^{*1}, Kim Hojae^{*2}, Woo Dayeon^{*3}, Yoon Sungho^{*4}, *Jin Chang-gyun⁵
School of Civil, Environmental and Architectural Engineering, Korea University¹, Department of
Computer Science and Engineering, Sungkyunkwan University², Department of Data Science,
Sungkyunkwan University³, Department of Mathematics, Sogang University⁴, Applied of
Artificial Intelligence, Seoul National University of Science & Technology⁵

Abstract

EEG-based emotion recognition stands out from facial expression or speech data in that emotions cannot be intentionally concealed. However, existing studies either learn only a local range of temporal information or do not consider label noise in emotion recognition data. To solve these problems, we propose a temporal convolutional attention with top-k(TCAT) model that learns through a top-k voting method based on multiple instance learning(MIL) using features of short and long term temporal changes. The proposed model achieves state of the art(SOTA) performance with

an accuracy of 98.67% in experiments using the leave-one-subject-out(LOSO) method to evaluate its generalization performance on the SEED dataset, a representative open dataset in the field of EEG emotion recognition.

I. 서론

Brain computer interface(BCI) 분야의 발전과 함께, EEG 데이터는 비침습적이며 휴대가 가능하다는 장점으로 인해 감정인식^[1], motor imagery^[2], 질병진단^[3] 등 다양한 응용분야에서 널리 활용되고 있다. 특히 감정인식의 경우, 기존에 사용되던 표정이나 음성 데이터와는 달리 피험자가 자신의 감정을 의도적으로 숨길 수 없다는 점에서 EEG의 활용도가 높아 활발한 연구가 진행되고 있다^{[4] [5]}.

[†] : Equally Contribution

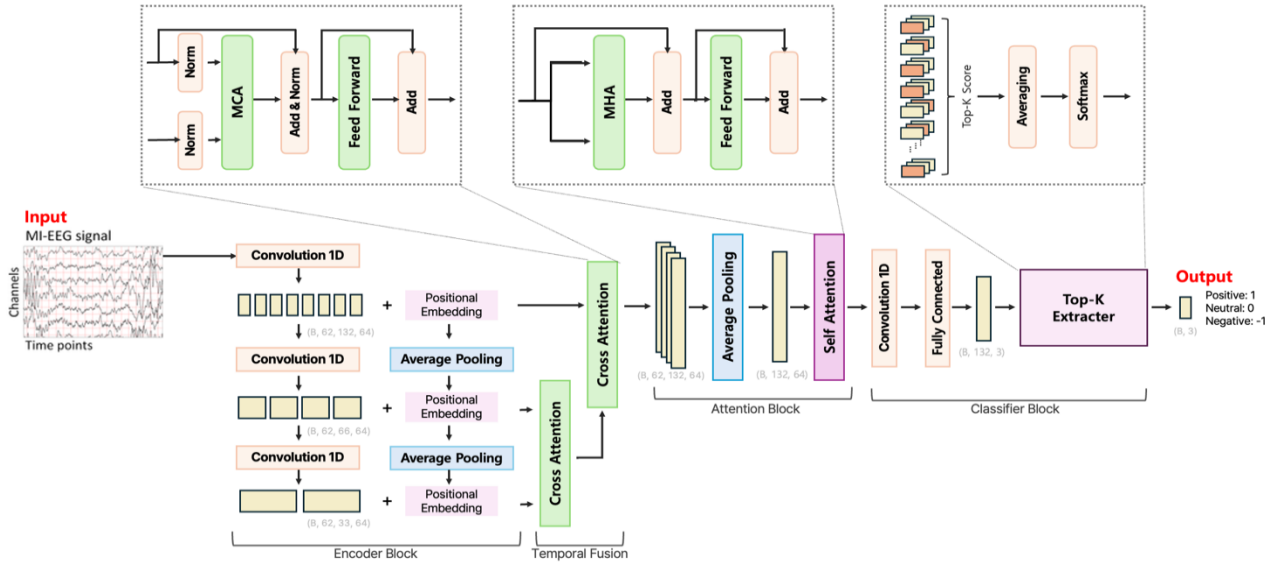


그림 1. TCAT 아키텍처의 전체 구조도

기존 연구에서는 EEG의 공간적 정보를 활용하는 graph neural network(GNN)기반 모델^[4]과 시간적 정보를 중점적으로 사용하는 convolution neural network 기반 모델^[1]이 제안되어 왔다. 그러나 GNN의 경우 최적의 인접 행렬을 구하는 방법이 명확하게 확립되지 않았고, CNN 기반 모델은 국소적인 시간정보에만 집중하여 성능의 한계가 있었다. 또한 EEG 감정인식 데이터가 가지는 문제점 중 하나인 라벨 노이즈^[6]에 대한 고려가 이루어지지 않았다. 긴 시간의 영상에 대해 하나의 클래스로 라벨링하는 감정인식 데이터의 특성상 라벨 노이즈는 안정적인 학습을 위해 반드시 고려되어야 하는 부분이다.

본 논문에서는 이러한 문제를 해결하기 위해 TCAT 모델을 제안한다. 제안 모델은 encoder, temporal fusion, attention, classifier block으로 구성되며, 다양한 시간 변화에 따른 특징을 추출하고 cross attention을 통해 이를 융합하여 EEG 신호의 장단기적 특성을 모두 활용할 수 있도록 설계하였다. 또한 라벨 노이즈 문제를 해결하기 위해 multiple instance learning 방법 중 하나인 top-k voting 방식을 적용하여 프레임 단위로 분류를 진행함으로써 안정적인 학습을 가능하게 하였다.

실험은 모델의 일반화 성능을 확인하기 위해 학습에 참여하지 않은 새로운 피험자 데이터로 검증하는 LOSO 방식을 사용하였다. 제안 모델은 EEG기반 감정 인식의 대표적인 오픈 데이터셋인 SEED^[7]에서 98.67%의 정확도로 state of the art(SOTA) 성능을 달성하였다.

본 논문의 순서는 다음과 같다. 2장에서는 제안 모델인 TCAT에 대해 자세히 설명하고, 3장에서는 실험 결과를 기반으로 우수성을 검증한다. 마지막 4장에서는 결론과 향후 연구계획으로 마무리한다.

II. 제안 방법

제안 모델의 전체적인 overview는 그림 1에서 확인할 수 있다.

2.1 Encoder Block

Encoder block은 EEG 데이터의 장단기적 시간 정보를 추출하기 위한 모듈로, convolution layer와 average pooling layer로 구성된다.

먼저 3개의 convolutional 1d(Conv1D) layer는 각 convolution 연산에서 kernel 및 stride의 크기를 모두 2로 설정하여, layer를 거칠 때마다 프레임 수를 절반으로 압축한다. 이러한 계층적 필터링 과정을 통해 시간 변화에 따른 EEG의 특징을 효과적으로 추출할 수 있다.

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d}) \quad (1)$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d}) \quad (2)$$

Positional embedding(PE)의 경우 입력 데이터와 동일한 크기의 벡터를 생성하기 위해 식 (1)과 (2)에서 정의된 바와 같이 \sin , \cos 함수를 이용하여

계산된다. 여기서 pos 는 위치, i 는 현재 차원, d 는 전체 차원의 개수이다. 이렇게 계산된 PE가 모델에 입력된 이후에는 kernel과 stride의 크기를 2로 설정한 세 번의 average pooling을 통해 각 Conv1D layer에서 추출된 벡터와 동일한 크기로 압축한다. 최종적으로, 이렇게 얻어진 벡터들을 더하여 시간적 특징과 시점별 순서 정보를 결합한다.

2.2. Temporal Fusion

Temporal fusion block에서는 encoder block에서 추출된 각기 다른 시간의 변화에 따른 정보를 융합하기 위해 두 번의 cross attention 연산을 수행한다. 이때 얇은 계층에서 얻어진 단기적 시간 정보는 세부적인 시간변화에 민감한 특성을 반영하고, 깊은 계층에서 얻어진 장기적 시간 정보는 전체적인 시계열 데이터의 패턴을 포착하는데 중점을 둔다. Cross attention을 통해 이 두 가지 정보를 상호 참조하여 융합함으로써, 다양한 스케일에서 나타나는 EEG의 특성을 효과적으로 통합할 수 있다.

2.3. Attention Block

Attention block에서는 피험자마다 전극 부착 위치가 다소 차이가 있을 수 있음을 고려하여 여러 전극 채널에 나타나는 공간적 특징을 average pooling을 통해 하나의 벡터로 압축함으로써 시간 정보에 집중하였다. 이후, self-attention을 거쳐 각 타임 스템프가 최종 출력에 미치는 중요도를 평가하여, 높은 attention score를 가질수록 감정 분류에 큰 영향을 미치도록 학습한다.

2.4. Classifier Block

단일 이미지가 아닌 시간에 따른 감정을 표현한 데이터의 특성상, 하나의 데이터 내에서 감정 상태가 일관되지 않은 라벨 노이즈가 발생할 수 있다. 명확한 프레임 수준의 레이블이 부족하기 때문에 본 논문에서는 top-k voting 방식을 사용하였다. 이를 위해 classifier block에서는 프레임 수준의 예측을 진행한다. 이 과정에서 linear layer를 통해 추출되는 벡터의 크기는 입력 데이터 프레임 수의 절반과 클래스 개수와 동일하다. 이후, 각 클래스별로 프레임 점수가 가장 높은 k개의 프레임을 선택하고 평균을 계산하여 최종 예측 클래스로서 softmax 확률을 출력한다.

III. 실험

3.1 실험 환경

한 명의 피험자 데이터를 평가에 사용하고, 나머지 피험자 데이터를 학습에 사용하는 LOSO 방식은 EEG 분야에서 모델의 일반화 성능을 검증하기 위한 대표적인 방법이다. 본 논문에서는 SEED 데이터셋에서 제공하는 15명의 피험자 데이터로 실험을 진행하였으며, 각 피험자 데이터에 대한 정확도의 평균(ACC)과 표준편차(STD)를 기준으로 성능을 평가하였다.

실험은 NVIDIA GeForce RTX 3090을 사용하여 수행하였다. 하이퍼 파라미터에서 batch 크기는 32, attention을 위한 head 수는 8, 에포크는 300, k는 32로 설정하였다. 손실 함수는 cross entropy loss를 적용하였으며, AdamW 옵티마이저를 사용해 모델을 최적화했다.

3.2 실험 결과

표 1. Comparison

Model	ACC(%) / STD
RGNN ^[4]	85.30 / 6.72
DPGAT ^[5]	87.07 / 6.31
STFCGAT ^[8]	94.83 / 3.41
ATCNet + 2phase ^[1]	95.71 / 3.58
TCAT (Ours)	98.67 / 1.53

표 1에서는 제안 모델과 기존 모델의 성능을 비교하였다. 기존 SOTA 성능을 기록한 ATCNet과 2phase 결합 모델^[1]은 95.71%의 정확도를 보인 반면, 제안 모델은 98.67%로 약 2.96% 향상된 성능을 나타냈다. 또한, 피험자 간 정확도의 표준편차가 1.53으로 가장 낮은 값을 기록하며 피험자별 모델의 성능차이를 최소화한 것을 확인할 수 있다.

IV. 결론 및 향후 연구 방향

본 논문에서는 시계열 데이터인 EEG의 장단기적 시간 정보를 모두 반영하고 라벨 노이즈 문제를 해결하기 위한 TCAT 모델을 제안하였다. 제안 모델은 LOSO 방식에서 98.67%로 우수한 성능을 달성하였으며 피험자별 성능차이를 줄이는데 성공하였다. 그러나 단일 데이터셋에서만 실험을 진행하여 모델의 신뢰성을 충분히 확인하지 못하였다. 향후에는 다양한 EEG 감정인식 데이터셋에서의 실험을 진행하여 모델의 우수성 및 일반화 가능성을

참고문헌

- [1] D. Kim, H. Kim, C. Jin, and S.-E. Kim*, "Attention-Temporal Convolutional Networks for EEG-based Emotion Recognition," Proceedings of ICCE-Asia 2023, Oct 23-25, Busan, Korea.
- [2] Jenke, R., Peer, A., & Buss, M., "Feature extraction and selection for emotion recognition from EEG", IEEE Transactions on Affective computing, Vol.5, no. 3, pp. 327-339, 2014.
- [3] Tabar, Y. R., & Halici, U., "A novel deep learning approach for classification of EEG motor imagery signals", Journal of neural engineering", Vol. 14, no. 1, 016003, 2016.
- [4] Zhong, P., Wang, D., & Miao, C., "EEG-based emotion recognition using regularized graph neural networks", IEEE Transactions on Affective Computing, Vol. 13, no. 3, pp. 1290-1301, 2020.
- [5] Li, X., Li, J., Zhang, Y., & Tiwari, P., "Emotion recognition from multi-channel EEG data through a dual-pipeline graph attention network", In IEEE International Conference on Bioinformatics and Biomedicine, pp. 3642-3647, Houston, TX, USA, Dec. 2021.
- [6] Dauwels, J., Vialatte, F., & Cichocki, A., "Diagnosis of Alzheimer's disease from EEG signals: where are we standing?", Current Alzheimer Research, Vol. 7, no. 6, pp. 487-505, 2010.
- [7] BCMI in SJTU, "SEED dataset", Shanghai, China, May 2022, [Online]. Available: <https://bcmi.sjtu.edu.cn/home/seed/index.html>
- [8] Li, Z., Zhang, G., Wang, L., Wei, J., & Dang, J., "Emotion recognition using spatial-temporal EEG features through convolutional graph attention network", Journal of Neural Engineering, Vol. 20, no. 1, 016046, 2023.