Journal of the Korea Industrial Information Systems Research Vol.26 No.1, Feb. 2021 :1-10
http://dx.doi.org/10.9723/jksiis.2021.26.1.001 ISSN:1229-3741

어텐션 메커니즘 기반 Long-Short Term Memory Network를 이용한 EEG 신호 기반의 감정 분류 기법+

(Emotion Classification based on EEG signals with LSTM deep learning method)

김 유 민[1], 최 아 영[2]*
(Youmin Kim and Ahyoung Choi)

요 약 본 연구에서는 EEG 신호를 기반으로 감정 인식에 유용한 딥러닝 기법을 제안한다. 감정이 시간에 따라 변화하는 특성을 반영하기 위해 Long-Short Term Memory 네트워크를 사용하였다. 또한, 특정 시점의 감정적 상태가 전체 감정 상태에 영향을 미친다는 이론을 기반으로 특정 순간의 감정 상태에 가중치를 주기 위해 어텐션 메커니즘을 적용했다. EEG 신호는 DEAP 데이터베이스를 사용하였으며, 감정은 긍정과 부정의 정도를 나타내는 정서가(Valence)와 감정의 정도를 나타내는 각성(Arousal) 모델을 사용하였다. 실험 결과 정서가(Valence)와 각성(Arousal)을 2단계(낮음, 높음)로 나누었을 때 분석 정확도는 정서가(Valence)의 경우 90.1%, 각성(Arousal)의 경우 88.1%이다. 낮음, 중간, 높음의 3단계로 감정을 구분한 경우 정서가(Valence)는 83.5%, 각성(Arousal)은 82.5%의 정확도를 보였다.

핵심주제어: EEG, 감정 분류, Long-Short Term Memory Network, 어텐션 메커니즘

Abstract This study proposed a Long-Short Term Memory network to consider changes in emotion over time, and applied an attention mechanism to give weights to the emotion states that appear at specific moments. We used 32 channel EEG data from DEAP database. A 2-level classification (Low and High) experiment and a 3-level classification experiment (Low, Middle, and High) were performed on Valence and Arousal emotion model. As a result, accuracy of the 2-level classification experiment was 90.1% for Valence and 88.1% for Arousal. The accuracy of 3-level classification was 83.5% for Valence and 82.5% for Arousal.

Keywords: EEG, Emotion classification, Long-Short Term Memory Network, Attention mechanism

1. 서 론

최근 우울증 환자가 증가하고 있는데, WHO의 발표에 따르면 전세계 2억 6400만명의 사람들이 우울증에 시달리고 있다(WHO, 2020). 우울증 환자를 모니터링하고 치료하기 위해서 모

* Corresponding Author: aychoi@gachon.ac.kr
+ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2019R1F1A1060731).
Manuscript received September 02, 2020 / revised December 14, 2020 / accepted December 29, 2020
1) 엠로, 제1저자
2) 가천대학교 소프트웨어학과, 제2저자, 교신저자

니터링이 필요한 요소 중 하나가 감정 분석이다. 감정을 분석하기 위한 감정 모델로는 연속적 모델과 비연속적 모델이 있는데, 많은 연구에서 비연속적인 정서가(Valence)-각성(Arousal) 모델을 사용하고 있다. 정서가(Valence)-각성(Arousal) 모델은 감정의 긍정도를 나타내는 정서가(Valence)와 감정의 각성 정도를 나타내는 각성(Arousal)으로 이루어진다(J. Posner et al. 2005). Fig. 1은 정서가(Valence)-각성(Arousal) 모델을 나타낸 그림이다. x축은 정서가(Valence), y축은 각성(Arousal)를 나타내고 있다. x축의 수가 커질수록 더 긍정적인 감정을 나타내고, 작아질수록 더 부정적인 감정을 나타낸다. y축의 수는 커질수록 더 강렬한 감정을, 작아질수록 잔잔한 감정을 나타낸다.

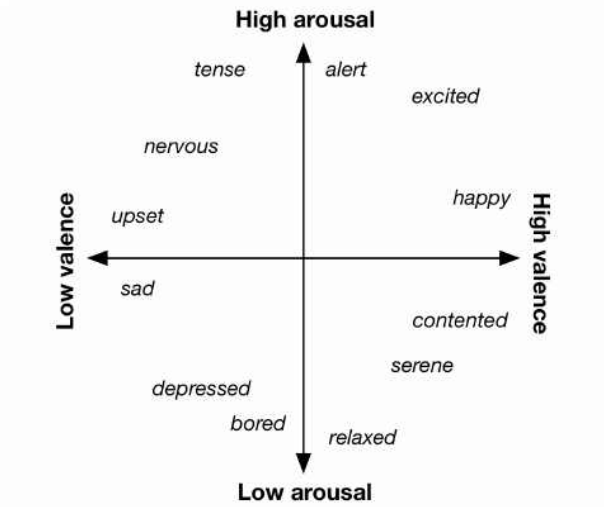

Fig. 1 Valence-Arousal emotion model

감정 분석을 위한 방식으로 표정 및 행동 인식을 통한 감정 분석 연구와 사용자의 생체 신호인 ECG(Electrocardiogram, 심전도) 신호와 EMG(Electromyogram, 근전도) 신호를 활용한 분석 방법이 연구되어왔다(Huang et al. 2016; Hemanth et al. 2018). 특히, 뇌파 신호인 EEG (Electroencephalogram) 신호를 기반으로 한 감정 인식 연구는 사용자가 느끼는 자연스러운 감정을 인식할 수 있는 장점이 있어 많은 연구에서 활용되고 있다.

기존의 EEG 신호를 이용한 감정 분석에서는 특징 추출을 통한 분석이 주를 이뤘다(Al-Fahoum and Al-Fraihat, 2014). 특징 추출 방법으로는 Fast Fourier Transform, Wavelet Transform, Eigenvector, Autoregressive Method 등이 있다. 추출한 특징들은 Support Vector Machine, Decision Tree 등의 머신 러닝 모델과 Artificial Neural Network, Recurrent Neural Network 등의 딥러닝 모델 등에 따라 분류되었다.

최근에는 많은 연구에서 딥러닝을 기반으로 하여 연구자가 연관성 있는 특징을 선택하는 것이 아닌 필터링 된 시계열 입력 신호를 분석 모델에 적용하여 특징 추출부터 인식까지 end-to-end 모델의 형태로 분석하는 방법이 제안되었다(Al-Nafjan et al. 2017; Alhagry et al. 2017). 그러나 기존 감정 인식 연구에서는 분석하고자 하는 클래스의 수가 많아질수록 감정 인식 정확도가 떨어지는 문제점을 가지고 있다. 또한, 연구자마다 감정에 영향을 미치는 특징이 다양하고, 일반화되어 있지 않아 실험 환경에 따라 분석 정확도가 다른 문제점을 가지고 있다.

본 논문에서는 EEG 신호를 기반으로 한 감정 인식 모델을 제안하기 위해 특징 추출을 하지 않고 필터링 된 EEG 신호를 입력으로 하는 딥러닝 기반의 감정 인식 분석 방법을 제안한다. 또한, 분석하고자 하는 감정의 수가 늘어날수록 정확도를 보장하기 위해서, 심리학 이론을 기반으로 한 딥러닝 모델을 제안한다. 본 연구에 적용된 심리학 이론인 peak-and-end rule 이론은 Fredrickson 등이 제안한 이론으로, 가장 강렬하게 감정을 느낀 장면과 가장 마지막에 봤던 장면에서 느꼈던 감정이 어떤 사건에 대해 느끼는 전체의 감정을 결정한다는 이론이다 (Fredrickson et al., 1993). Fredrickson 등은 영화 클립을 보고 감정을 기록하는 실험에서 영화 클립의 전체 시간보다 특정 기억에 의해 감정이 정해진다는 것을 실험을 통해 증명하였다. 따라서 본 논문에서는 시계열 데이터인 EEG 신호를 입력으로 하여 감정 인식의 정확도 향상을 위하여 기존 상태가 이후의 상태에 영향을 미칠 수 있도록 LSTM 모델을 설계하였다. 또한 은닉상태에서 감정 상태에 영향을 많이 끼치는 부분의 가중치를 더하기 위해 어텐션 메커니즘을 적용

하였다. 제안한 방법은 공용 데이터베이스인 DEAP 데이터베이스를 사용하여 32명의 사용자 데이터를 기반으로 분석하였다. 분석하고자 하는 감정의 수에 따라 정확도가 떨어지지 않는지 확인하기 위하여 2단계 감정 분석 (Low, High)과 3단계 감정 분석 (Low, Middle, High)을 4 fold validation을 적용하여 검증하였다.

2. 관련 연구

Table 1에서 기존 연구 방식을 요약하여 나타내었다. 최근 EEG 기반의 감정 인식을 위한 모델로 머신 러닝 기법을 적용하여 정서가(Valence)와 각성(Arousal)을 High와 Low 또는 High, Middle, Low로 분류를 하는 연구가 진행되었다(Chung and Yoon, 2012; Koelstra et al. 2012; Zhang et al. 2013; Atkinson and Campos, 2016; Jirayucharoensak et al. 2014). Chung 등의 연구에서는 DEAP 데이터베이스의 32개의 EEG 채널과 bipolar montage 기법으로 생성한 가상의 61개의 EEG 채널을 사용하였다. 이 신호로부터 power spectral density와 power asymmetry를 추출하고 Bayes 분류기를 사용하여 정서가(Valence)와 각성(Arousal) 공간을 두 가지 클래스로 분류하였다. 결과적으로 정서가(Valence)와 각성(Arousal)에 대해 53.4%의 정확도를 보였다. S. Koelstra 등의 연구에서는 EEG 신호로부터 power spectral density와 power asymmetry 추출하여 Naïve Bayes 분류기를 이용하여 정서가(Valence)와 각성(Arousal) 공간을 두 가지 클래스로 분류하였다. 결과는 정서가(Valence)에 대해 57.0%, 각성(Arousal)에 대해 62.0%의 정확도를 보였다. Zhang 등의 연구에서는 DEAP 데이터베이스로부터 PSD(Power Spectral Density)를 추출하여 ontological model을 사용했다. 실험 결과 정서가(Valence)에서는 75.19%, 각성(Arousal)에서는 81.74%의 정확도를 기록했다. Atkinson과 Campos가 제안한 방법에서는 DEAP 데이터베이스에서 EEG 채널 중 14개의 감정과 가장 밀접한 채널을 선택하여 통계적 특징값, band power, Hjorth parameter, fractal dimension의 특징을 추출하였다. 이 특징들을 입력으로 사용한 Support Vector Machine을 학습시켰다. 이는 두 가지의 클래스로 분류한 결과 정서가(Valence)에서는 73.14%, 각성(Arousal)에서는 73.06%의 결과를 보였다. 하지만 기존 연구에서는 시간에 따른 변화가 아닌 단일 시간의 신호 패턴으로 감정을 구분하는 연구를 주로 하였고, 앞뒤 신호 간의 연관 관계를 고려한 분석 방법이 적용되지 않았다.

Table 1 Related works of emotion analysis

	Research	Method	Accuracy (%)	
			Valence	Arousal
two-level classification	Chung et al., 2012	Bayes	53.4	53.4
	Koelstra et al. 2012	Naive Bayes	57.0	62.0
	Zhang et al. 2013	Ontological model	75.19	81.74
	Atkinson, 2016	Support Vector Machine	73.14	73.06
	Al-Nafjan et al. 2017	Deep Nueral Network	82.0	82.0
	Alhagry et al. 2017	LSTM	85.4	85.6
three-level classification	Chung et al., 2012	Bayes	51.0	51.0
	Jirayucharoensak et al. 2014	Deep Learning Network + Stacked Auto Encoder	49.52	46.03
	Atkinson et al., 2016	SVM	60.7	62.33

최근에는 딥러닝을 활용한 감정 인식 연구도 활발히 진행되고 있다(Alhagry et al. 2017;

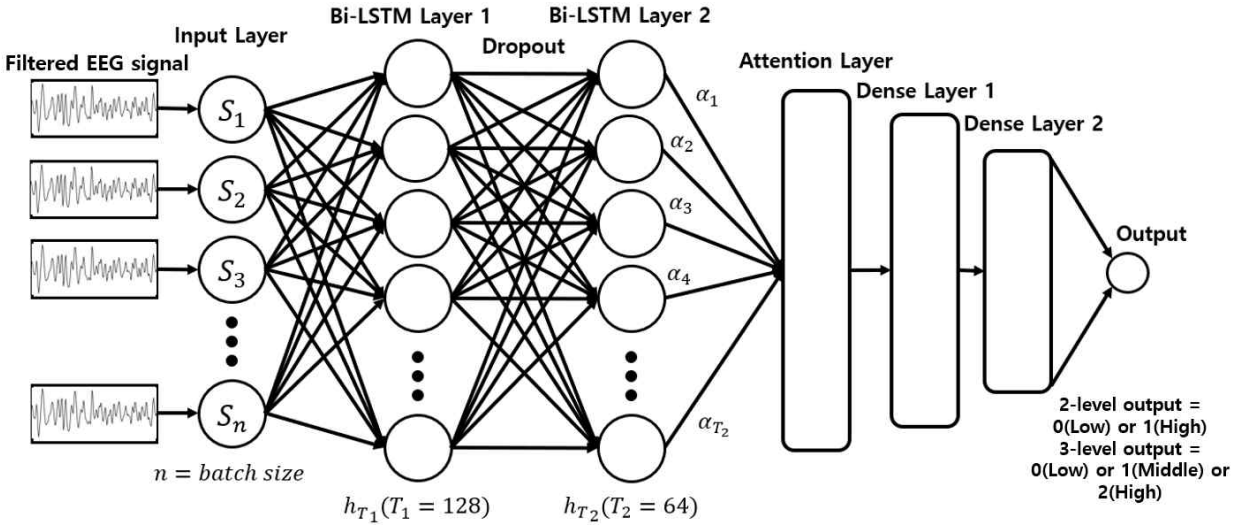

Fig. 2 LSTM network structure with attention mechanism

Al-Nafjan et al. 2017; Jirayucharoensak et al. 2014). Al-Nafjan 등은 EEG 채널로부터 PSD(Power Spectral Density)와 Frontal asymmetry를 추출하여 Deep Neural Network를 통해 정서가(Valence)와 각성(Arousal)에 대해 4가지의 클래스로 분류하는 방법을 제안하였다. 최종 결과는 정서가(Valence)와 각성(Arousal) 모두 82.0%의 정확도를 보였다. Alhagry 등은 EEG와 같은 시계열 데이터에 적합한 LSTM(Long-Short Term Memory) network를 사용하여 raw EEG signal을 12개의 block으로 분할해 입력으로 사용하였다. 출력으로는 2단계 분류를 사용하여 분석하였다. 결과는 정서가(Valence)에 대해 85.4%, 각성(Arousal)에 대해 85.6%의 정확도를 보였다. Jirayucharoensak 등은 DEAP 데이터베이스에 존재하는 EEG 신호에서 32채널에 대해 PSD를 추출하여 가장 중요한 특징을 정하는 PCA(Principal Component Analysis)를 적용하여 특징을 추출하였다. 뽑은 특징들을 입력으로 하고 특징을 계층적으로 할 수 있는 Stacked Auto Encoder를 적용한 Deep Neural Network를 설계하여 3단계 분류를 수행했다. 결과는 정서가(Valence)와 각성(Arousal)에 대해 각각 49.52%와 46.03%의 정확도를 보였다.

그러나 딥러닝을 사용한 감정 분석의 경우에도 감정 인식의 정확도가 상황에 따라 유지되지 않은 문제점을 가지고 있다. Alhagry 등이 딥러닝 기반의 LSTM 네트워크를 사용하여 EEG 신호 기반으로 감정 분석을 수행하였으나, 정확도가 85.4%와 85.6%로 나타났다. 특히, 분석하고자 하는 클래스의 수가 많아질수록 감정 인식 정확도가 떨어지는 문제점을 가지고 있다. 또한, 연구자마다 감정에 영향을 미치는 특징이 다양해서 일반화된 분석 방법 적용에 문제가 있다.

3. 제안하는 딥러닝 기반 감정 분석 모델

3.1 시스템 구성

본 논문에서는 이전의 학습 정보가 현재에 영향을 미치도록 하기 위하여 Recurrent Neural Network 모델을 사용했으며, 그 중 장기 의존성 문제를 해결하기 위하여 Schmidhuber 등이 제안한 Long-Short Term Memory 모델을 적용하여 감정 분석을 수행하였다(Hochreiter and Schmidhuber, 1997). 본 논문에서는 양방향

LSTM layer를 사용하였는데, 일반적으로 어텐션 메커니즘을 사용할 때 단방향 LSTM 층보다 전후 은닉 상태를 모두 고려할 수 있는 양방향 LSTM 층을 사용했을 때 성능이 높다고 알려져 있다. 또한, 특정 시간에서 발생한 감정이 전체 감정에 영향을 미치는 것을 모델링하기 위해서 LSTM layer 이후에 어텐션 메커니즘을 적용하였다. 어텐션 메커니즘은 들어온 입력에 대해 출력에 가장 영향을 많이 미치는 부분에 가중치를 주어 학습시키는 방법이다.

제안하는 모델의 전체 구조는 Fig. 2와 같다. 첫 번째 은닉 층으로는 128개의 뉴런을 가졌고 활성화 함수로 ReLU를 사용하는 양방향 LSTM layer를 사용한다. 그 후 과적합 문제를 방지하기 위하여 0.2의 확률을 가진 Dropout을 실행하고, 64개의 뉴런을 가졌고 활성화 함수가 ReLU인 두 번째 양방향 LSTM layer에 통과시킨다. 두 번째 layer의 출력인 은닉 상태로 어텐션 가중치를 계산한 후 16의 출력 크기와 활성화 함수 ReLU를 사용하는 첫 번째 Dense layer를 통과시켜 차원을 축소한다. 마지막으로 출력 크기가 1이고 활성화 함수로 2단계 분류에서는 Sigmoid, 3단계 분류에서는 Softmax를 사용하는 두 번째 Dense layer를 통과시켜 최종 출력을 구한다. 2단계 분류에서는 정서가(Valence) 또는 각성(Arousal)에 대해 Low, High에 대한 결과이고, 3단계 분류에서는 Low, Middle, High에 대한 결과가 나온다.

어텐션 가중치는 Fig. 3과 같이 계산된다. 입력의 어느 부분이 출력에 영향을 가장 많이 미치는지 어텐션 가중치를 계산하여, 해당 입력 부분의 가중치가 높을수록 네트워크가 훈련할 때 더 값을 크게 하는 방식이다. 계산 순서는 다음과 같다. 2번째 양방향 LSTM layer를 통해 계산된 은닉 상태 벡터는 무작위로 초기화된 어텐션 가중치와 곱해주는데, 어텐션 가중치의 길이는 은닉 상태 벡터의 길이와 같다. 두 번째 양방향 LSTM layer의 출력 크기는 64이며, 양방향이기 때문에 이의 2배인 128이다. 따라서 어텐션 가중치의 길이도 이와 같은 128이다. 이 계산을 통해 나온 결과값은 softmax를 통해 확률값으로 변환되고, 변환된 어텐션 벡터는 처음 계산된 은닉 상태 벡터와 함께 합쳐져 최종 어텐션 출력으로 계산된다. Attention layer와 연결되어있는 Dense layer 1에서는 해당 어텐션 출력을 입력으로 받아, 어느 부분의 가중치가 가장 중요한지를 앞으로의 학습에 반영하여 더 정확한 결과를 낼 수 있도록 한다.

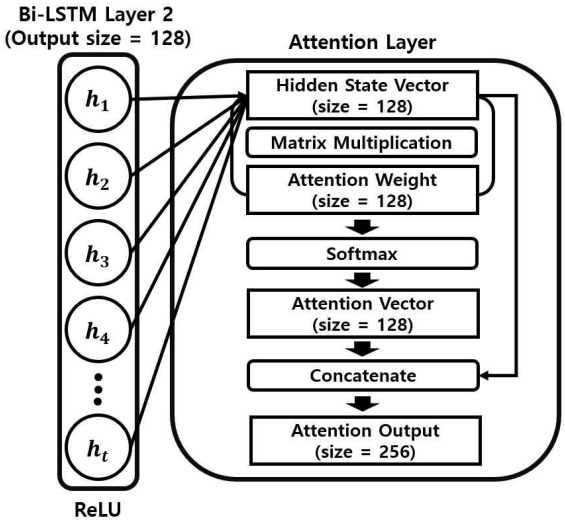

Fig. 3 Attention network architecture

최적화 함수로는 Adam optimizer를 사용하였고, learning rate는 0.001로 설정하였다. 손실 함수로는 이진 분류 작업에 적합한 cross entropy 손실 함수를 사용하였다. 정확도를 측정하기 위해서 Stratified K-fold cross validation 방법을 사용했으며 fold의 개수는 총 4개를 사용하였다. 이 방법을 통해 label이 각 fold에 균형 있게 분배되어 학습할 수 있도록 하였고, 한 번의 fold를 학습시킬 때 마다 데이터의 75%는 training set, 25%는 test set으로 사용되었다. Epoch는 30으로 설정하여 각 fold가 30번을 돌아가며 학습하도록 하였고, 입력에 대한 batch size는 32로 설정하였다. 각 하이퍼파라미터들은 실험을 통하여 최적화 과정을 거쳤다.

4. 실험 환경 및 결과

4.1 실험 환경 및 사용 데이터

트레이닝 및 테스트 환경은 Intel(R) Core(TM) i7-6850K CPU @ 3.60GHz, 6 코어, 12 논리 프로세서와 64GB의 RAM, NVIDIA GeForce GTX 1080 Ti 2대를 탑재한 PC에서 진행되었다. 딥러닝 모델 구현은 파이썬 딥러닝 라이브러리인 Keras로 구현을 해 진행하였다.

본 논문에서는 입력 데이터로 DEAP 데이터베이스를 이용했다(S. Koelstra et al. 2012). DEAP 데이터베이스는 피험자들의 감정 점수와 EEG 신호를 기록한 데이터베이스이다. 평균 나이가 26.9세인 32명의 참가자는 슬픔, 즐거움, 만족, 흥미, 분노 등의 다양한 장르를 가진 1분의 뮤직비디오를 시청하고 감정을 평가하는 것에 사용되는 설문지인 Self-Assessment Manikin (SAM)을 이용하여 정서가(Valence), 각성(Arousal), Dominance, Liking, Familiarity를 측정하였다. 각 평가 항목은 1 ~ 9의 점수로 이루어져 있으며, 각각의 참가자들은 40번의 실험을 거쳤다. 32명 중 22명에 대해서는 정면 얼굴 비디오를 기록했다. 실험에서는 48개의 채널(32개의 EEG 채널, 12개의 peripheral 채널, 3개의 unused 채널, 1개의 status 채널)이 측정되었으며, sampling rate는 512Hz이다. 데이터는 사용되기 전 전처리 과정을 거쳤는데 Fig. 4는 그 과정을 그림으로 나타낸 것이다.

입력 신호는 128Hz로 다운샘플링 되었고, 각 데이터는 3초의 베이스 라인과 60초의 실험 동안의 EEG 측정값을 가진다. Fp1 채널에서 EOG (Electrooculography) 신호가 제거되었고, 4.0-45.0Hz의 대역폭을 가진 band pass filter가 적용되었다. 각 EEG 채널들은 8,064 값으로 이루어졌다.

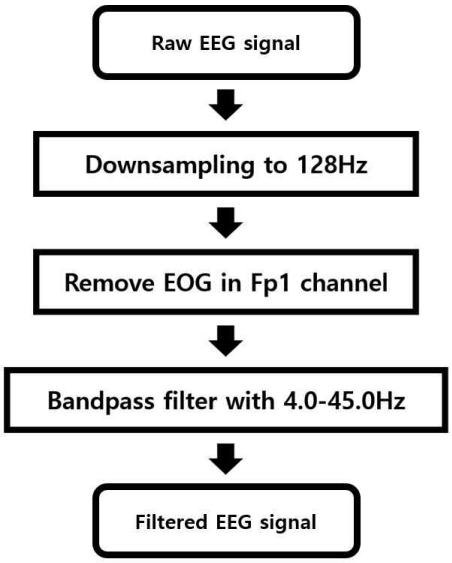

Fig. 4 Preprocessing procedure of raw EEG signal

결과 분석은 정서가(Valence), 각성(Arousal)에 대해서만 이루어졌으며, 정서가(Valence)와 각성(Arousal) 모델을 사용했다. 2단계 분류에서는 Low(1~5)와 High(5~9) 클래스로, 3단계 분류에서는 Low(1~3), Middle(3~6), High(6~9) 클래스로 분류하였다. 각 클래스는 전처리 과정에서 점수 범위에 따라 2단계 분류에서는 Low가 0, High가 1로, 3단계 분류에서는 Low가 0, Middle이 1, High가 2로 각각 분류되었다. EEG 신호의 각 32개의 채널을 약 15초 간격으로 세그멘테이션하여 총 163,840개의 데이터로 정리하였고, 각 데이터는 2,016개의 값을 가졌다. 결과적으로 2단계 분류에서 수행되는 클래스 각각의 데이터의 수는 Low가 59,840개이고 High가 104,000개이다. 3단계 분류에서 수행되는 클래스들은 Low가 20,160개, Middle이 60,160개, High가 83,520개이다.

4.2 결과 분석

Table 2는 제안한 방법을 기존 연구들과 비교한 결과이다. 각 연구의 정확도는 기존 논문의 수치를 인용하였다. 기존 연구와의 비교는

DEAP database를 이용하여 정서가(Valence)와 각성(Arousal)에 대해 2단계 분류를 한 연구들과 했다. 최종 정확도는 4개의 fold로부터 나온 정확도의 산술 평균을 사용하였다. 가장 정확도가 높은 기존 연구와 비교했을 때, 정서가(Valence) 분류 결과는 이전보다 4.8%가 증가한 평균 90.2%의 성능을 보였고, 각성(Arousal) 분류 결과는 이전보다 2.5%가 증가한 평균 88.1%의 성능을 보였다.

Table 2 2-level Classification Result

Research	Method	Valence	Arousal
Chung et al, 2012	Bayes	53.4	53.4
Koelstra et al. 2012	Naive Bayes	57.0	62.0
Zhang et al. 2013	Ontological model	75.19	81.74
Atkinson et al, 2016	Support Vector Machine	73.14	73.06
Al-Nafjan et al. 2017	Deep Neural Network	82.0	82.0
Alhagry et al. 2017	LSTM	85.4	85.6
Proposed Method	**LSTM+ Attention**	**90.2**	**88.1**

2단계 분석의 결과를 그룹별로 확인하기 위하여 컨퓨전 매트릭스를 Fig. 5와 같이 나타내었다. 실험에서 사용된 4-fold cross validation의 결과 중 가장 정확도가 높게 나온 fold에 대해 분석하였으며 Fig 5는 네 번째 fold로 평균 정확도는 91.98%이다. x축은 예측된 결과를 나타내고, y축은 실제 타겟 클래스를 나타낸다. 컨퓨전 매트릭스 분석 결과 Low 클래스에 대해서는 14,960개 중 11,810개를 올바르게 분류하여 78.94%의 정확도를 보였고, High 클래스에 대해서는 26,000개 중 25,865개를 올바르게 분류하여 99.48%의 높은 정확도를 보였다. Low 클래스보다는 high 클래스의 정확도가 높게 나타났는데, 이는 low와 high 클래스의 데이터 개수가 약 2배 차이가 나면서 모델이 특정 클래스에 바이어스 되어 학습이 되었을 가능성이 있다.

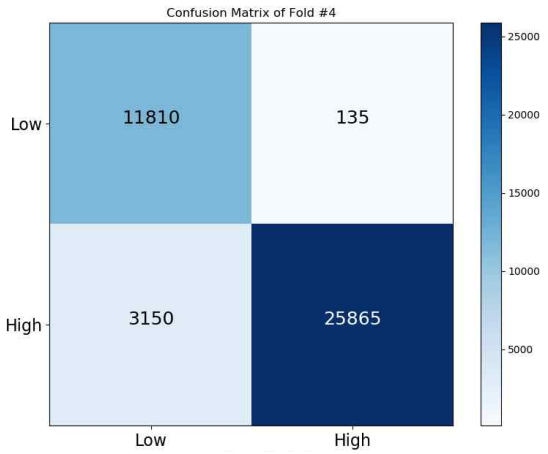

Fig. 5 Confusion Matrix of 2-level Classification Result using Proposed Method about Valence

또한 정서가(Valence) 예측이 각성(Arousal) 예측보다 정확도가 높게 나타났다. Howe 등의 연구에 따르면 사람이 단어를 기억할 때 얼마나 강하게 긍정적이고 약하게 긍정적인지보다 긍정적인 것과 부정적인 것으로 나누어 기억을 할 때 더 오래 기억한다는 것이 증명되었다(Howe, 2011). 따라서 정서가(Valence)에 대한 점수를 레이블링 하는 것이 더욱 직관적이므로 딥러닝을 통한 학습에서 더욱 뚜렷한 패턴을 찾을 수 있었기 때문으로 추측한다.

Table 3는 3단계 분류 작업을 수행한 기존 연구들과 비교한 결과이다. 가장 정확도가 높은 기존 연구와 비교했을 때 정서가(Valence)를 분류한 결과는 22.8%가 높은 83.5%의 성능을 보였고, 각성(Arousal)을 분류한 결과는 20.17%가 높은 82.5%의 결과를 보였다. Fig. 6는 제안된 모델로 정서가(Valence)를 3단계 분류한 결과를 컨퓨전 매트릭스로 나타낸 것이다. x축은 예측 결과, y축은 실제 클래스를 나타낸다.

이는 결과의 네 번째 fold이며, 해당 fold의 정확도는 87.84%이다. Low 클래스에 대해 5,053개 중 3,699개를 옳게 분류하여 73.2%의 정확도를, Middle 클래스에 대해 15,009개의 중 11,625개를 옳게 분류하여 77.45%의 정확도를 보였다. 가장 정확도가 높은 클래스는 High 클래스로, 20,898개 중 20,655개를 옳게 분류하여 98.83%의 정확도를 보였다. 각 클래스의 정확도가 차이나는 이유는 2단계 분석과 유사하게 데이터의 대부분이 High 클래스이기 때문에 상대적으로 수가 적은 Low와 Middle 클래스보다 학습이 잘 되기 때문에 발생할 수 있다.

Table 3 3-level Classification Result

Research	Method	Valence	Arousal
Chung et al, 2012	Bayes	51.0	51.0
Jirayucharoensak et al. 2014	Deep Learning Network + Stacked AutoEncoder	49.52	46.03
Atkinson and Campos, 2016	SVM	60.7	62.33
Proposed Method	**LSTM+ Attention**	**83.5**	**82.5**

실험을 통해 분석한 결과 기존 연구들은 2단계 분류에서는 약 85% 성능을 보였으나 3단계 분류에서는 46%~62%의 분석 정확도를 보였다. 그러나 제안하는 네트워크를 통한 학습 결과, 2단계 분류에서는 기존 연구 중 가장 정확도가 높았던 Alhagry 등의 연구와 비교 했을 때 3~5% 가량의 성능 향상을 보였고, 3단계 분류에서는 Atkinson 등의 연구에서 보였던 성능 대비 20~22%의 성능 향상을 보였다. 3단계 분류에서 Low 클래스에 대한 정확도가 Middle, High 클래스에 대한 정확도보다 약 4%, 25% 가량이 낮은데 이는 데이터 불균형에서 야기된 것으로 추측한다. 따라서 이를 해결할 수 있는 데이터 샘플링, 가중치 조절 방법 등을 적용하여, 보다 정확한 성능을 낼 수 있도록 할 수 있을 것이다.

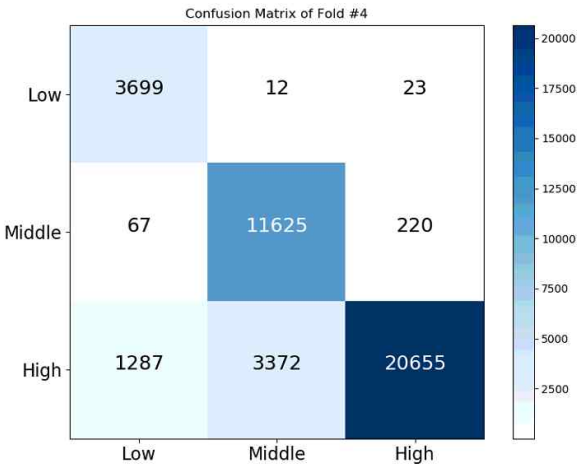

Fig. 6 Confusion Matrix of 3-level Classification Result using Proposed Method about Valence

5. 결 론

본 연구에서는 peak-and-end rule 이론에 기반하여 감정 분석에 적합한 최신 기술인 어텐션 메커니즘을 적용하여 정서가(Valence)와 각성(Arousal) 기반의 감정 모델을 효과적으로 분류하는 어텐션 기반 LSTM 딥 러닝 모델을 제안하였다. 제안된 모델은 별도의 특징 추출 과정이 없이 EEG 신호를 사용하여 2단계 분석의 경우 90.2%, 88.1%의 정확도를 보임을 확인하였다. 3단계 분석의 경우 83.5%, 82.5%의 분석 정확도를 보였다. 따라서 클래스 수의 증가에 따라 정확도가 크게 변하지 않음을 확인하였다. 이를 통해 실생활에서도 실시간으로 사람의 감정 상태를 분류할 수 있을 것으로 예상된다.

향후 연구에서는 데이터 불균형을 해결할 방법을 적용하여 감정 분석의 정확도를 low 클래스에서도 향상시키고자 한다. 또한, 연속적 감정 모델의 경우에는 감정의 정도를 나타내는 것이 중요하므로 딥러닝 기반 감정 인식 모델의 마지

막 단에 regression 모델을 적용하여 감정의 정도를 다양하게 분석되도록 하고자 한다.

References

Al-Fahoum, S., and Al-Fraihat, A. (2014). Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. ISRN Neuroscience, 2014, 1 - 7. https://doi.org/10.1155/2014/730218

Al-Nafjan, A., Hosny, M., Al-Wabil, A., and Al-Ohali, Y. (2017). Classification of Human Emotions from Electroencephalogram (EEG) Signal using Deep Neural Network. International Journal of Advanced Computer Science and Applications, 8(9), 419 - 425. https://doi.org/10.14569/ijacsa.2017.080955

Alhagry, S., Aly, A., and A., R. (2017). Emotion Recognition based on EEG using LSTM Recurrent Neural Network. International Journal of Advanced Computer Science and Applications, 8(10). https://doi.org/10.14569/ijacsa.2017.081046

Atkinson, J., and Campos, D. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. Expert Systems with Applications, 47, 35 - 41. https://doi.org/10.1016/j.eswa.2015.10.049

Chung, S. Y., and Yoon, H. J. (2012). Affective classification using Bayesian classifier and supervised learning. International Conference on Control, Automation and Systems, 1768 - 1771.

Fredrickson, B. L., and Kahneman, D. (1993). Duration Neglect in Retrospective Evaluations of Affective Episodes. Journal of Personality and Social Psychology, 65(1), 45 - 55. https://doi.org/10.1037/0022-3514.65.1.45

Hemanth, D. J., Anitha, J., and Son, L. H. (2018). Brain signal based human emotion analysis by circular back propagation and Deep Kohonen Neural Networks. Computers and Electrical Engineering, 68, 170 - 180. https://doi.org/10.1016/j.compeleceng.2018.04.006

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735 - 1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, X., Kortelainen, J., Zhao, G., Li, X., Moilanen, A., Seppänen, T., and Pietikäinen, M. (2016). Multi-modal emotion analysis from facial expressions and electroencephalogram. Computer Vision and Image Understanding, 147, 114 - 124. https://doi.org/10.1016/j.cviu.2015.09.015

Jirayucharoensak, S., Pan-Ngum, S., and Israsena, P. (2014). EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation. Scientific World Journal, 2014. https://doi.org/10.1155/2014/627892

Koelstra, S., Mühl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2012). DEAP: A database for emotion analysis; Using physiological signals. IEEE Transactions on Affective Computing, 3(1), 18 - 31. https://doi.org/10.1109/T-AFFC.2011.15

Mark L. Howe (2011). The Nature of Ealry Memory: An Adaptive Theory of the Genesis and Development of Memory, Oxford University Press.

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology, 17(3), 715 - 734. https://doi.org/10.1017/S0954579405050340

World Health Organization (2020). Depression, https://www.who.int/news-room/fact-sheets/

detail/depression(Accessed on Jul. 16th, 2020)

Zhang, X., Hu, B., Chen, J., and Moore, P. (2013). Ontology-based context modeling for emotion recognition in an intelligent web. World Wide Web, 16(4), 497 - 513. https://doi.org/10.1007/s11280-012-0181-5

김 유 민 (Youmin Kim)

- 가천대학교 소프트웨어학과 공학사
- 가천대학교 IT융합학과 공학석사
- (현재) 엠로, 연구원
- 관심분야: 뇌파 분석, 머신 러닝, 딥러닝

최 아 영 (Ahyoung Choi)

- 광주과학기술원 정보통신학과 공학박사
- 삼성전자 무선사업부 책임연구원
- (현재) 가천대학교 IT대학 소프트웨어학과 조교수
- 관심분야: 생체신호처리, 모바일 헬스케어, 머신러닝, 딥러닝