

뇌파 데이터를 활용한 GRU 모델의 인간 감정 분류 성능 평가에 관한 연구

김윤상, 김고은, 문채은, 이나경, 김웅식*

건양대학교, *건양대학교

22806609@konyang.ac.kr, ssv591@naver.com, codms2288@naver.com,

skrud021012@naver.com, *wskim@konyang.ac.kr

A study on human Evaluation of human emotion classification performance of GRU model using EEG

Kim Youn Sang, Kim Go Eun, Moon Chae Eun, Lee Na Gyoung, Kim Woong Sik*

Konyang Univ. *Konyang Univ.

요약

선행 연구에서 Kaggle의 공공데이터셋을 활용해 4가지 인공지능 모델을 통해 뇌파 데이터의 감정 분류 연구를 진행하였다. 뇌파 감정 분류 연구 결과 GRU(Gated Recurrent Unit) 기반 인공지능 모델이 가장 뛰어난 성능을 보였다. 이후 분류 모델의 성능 평가 지표 중 Accuracy(정확도), Precision(정밀도), Recall(재현율), F1-Score, Normalized Confusion Matrix, ROC-Curve, AUC-Score로 뇌파 분류 모델의 성능을 검증하였다. 검증 결과 Accuracy는 0.9703, Precision은 0.9703, Recall은 0.9703, F1-Score는 0.9702이 나왔으며, AUC-Score는 0.9781을 보였다. 향후 연구로는 실제 측정된 뇌파를 활용하여 BCI 연구를 진행할 예정이다.

I. 서론

선행 연구에서는 뇌파 연구에 적합한 인공지능 모델의 선정을 위해 4가지 인공지능 모델을 사용하여 각 모델의 성능 비교를 진행하였다. 사용한 인공지능 모델은 LSTM(Long Short-Term Memory models), GRU(Gated Recurrent Unit), CNN(Convolution Neural Network), DNN(Deep Neural Network)이다. 연구 결과 LSTM의 Accuracy 97.188%, Loss 0.506, GRU의 Accuracy 97.812%, Loss 1.572, CNN의 Accuracy 97.812%, Loss 0.101, DNN의 Accuracy 96.875%, Loss 0.123로 GRU 모델이 가장 뛰어난 분류 성능을 보였다[1]. 이번 연구에서는 GRU 모델의 분류 성능에 대한 정확한 검증을 위해 분류 성능 평가 지표(Classification Evaluation Metrics)를 통해 GRU 모델의 성능 평가를 진행하였다. 사용한 성능 평가 지표는 Accuracy, Recall, Precision, F1-Score, Normalized Confusion Matrix, ROC-Curve(Receiver Operating Characteristic Curve), AUC Score(Area Under the Curve Score)이다.

II. 본론

본 연구에서는 Kaggle의 데이터 셋(EEG Brainwave Dataset: Feeling Emotions)을 사용하였다[2]. 연구의 데이터는 2명의 남녀에게 각 상황에 맞는 동영상을 보여주어 감정 자극을 유발하였다. 동시에 뇌파 측정기를 착용하여 TP9, AF7, AF8, TP10 EEG 배치의 뇌파가 수집되었다. 데이터는 $2,133 \times 2,549$ 으로 되어있다. 6분간의 휴식을 통한 중립 상태의 뇌파 데이터가 수집되었고 긍정적인 자극 감정 유발 동영상은 라라랜드(2016), Slow Life(2014), Funny Dogs(2015)이며, 부정적인 자극 감정 유발 동영상은 말리와 나(2008), 업(2009), 마이클(1991)이 사용되었다[3][4]. 뇌파 데이터는 시계열 배열의 CSV 파일로 구성되어 있으며, 데이터 활용을 위한 샘플링 전처리과정이 진행되었다. 모델의 학습 데이터는 7:3 비율로 나누어 학습 데이터(Training Data) 1,493개 테스트 데이터(Test Data) 640개로 구성하였으며, 학습 데이터는 7:3 비율로 1,045개의 학습 데이터와 448개의 학습 검증 데이터(Validation Data)로 나누어 구성하였다.

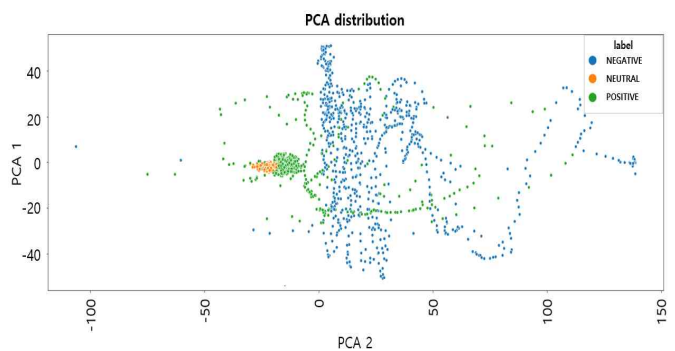


그림 1. PCA 기법으로 확인한 데이터 분포

그림1은 데이터셋의 분포를 알아보기 위해 PCA(Principal Component Analysis) 기법에 적용하여 데이터를 분석했다. PCA기법은 고차원의 데이터의 계산과 시각화가 어려워 분석하기가 쉽지 않기 때문에 원데이터의 분포를 최대한 보존하며 고차원 공간의 데이터를 저차원 공간의 데이터로 변환하는 것이다. 데이터의 기존 변수를 조합하여 변수를 만드는데 이를 주성분(Principal Component)이라 한다. 주성분끼리는 서로 연관성이 없다. 주성분을 사용하여 저차원으로 데이터를 축소를 하게 되면 기존 변수들의 데이터 성질을 최대한 보존하고 있어 원본 데이터의 성질을 최대한 반영할 수 있다[6]. 신경과학에서 PCA 기법은 뉴런의 활동 전위 모양으로 특정 뉴런을 알아차리는데 사용된다. 뉴런은 ‘스파이크’라고 하는 활동 전위를 생성하는데 다른 뉴런에 스파이크를 할당하는 스파이크 정렬(Spike sorting)을 진행하기 위해 PCA기법이 기반이 된다[7]. 그림에서 노란색 점으로 표현된 데이터는 중립 감정 데이터로 y축 0, x축 -30 범위 내에서 군집화되어 분포되어있다. 초록색 점은 긍정적인 감정 데이터이며 중립 데이터의 근처에 y축 0, x축 -20 ~ 0 사이에 군집화 되어있고 일부 성분들이 다양한 범위에 퍼져있다. 파란색 점은 부정적인 감정 데이터이며 x축의 0 ~ 50까지 범위에 주로 분포되어있다.

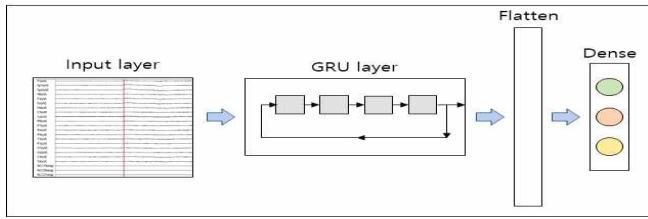


그림 2. 뇌파 분류 GRU 모델 구조

그림2는 뇌파 분류 GRU 모델의 구조이다[5]. 데이터를 전처리 진행 과정을 걸쳐 Input Layer에 입력된다. 데이터는 GRU 모델을 걸쳐 평탄화(Flatten) 단계를 걸친다. 이후 Dense Layer를 통과하며 3가지 분류로 나누어 뇌파 감정 분류가 진행되었다.

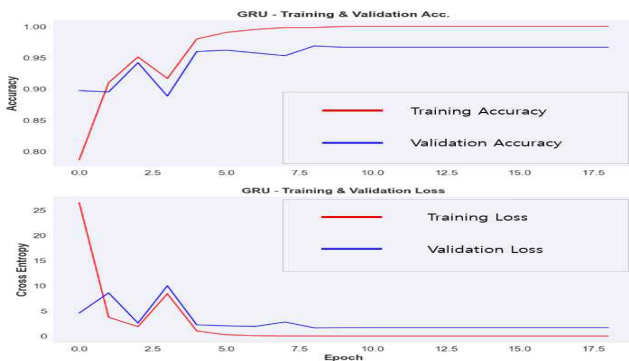


그림 3 GRU 모델 Training & Validation Acc, Loss

그림3은 연구에서 사용한 모델의 Training & Validation Accuracy, Loss의 그래프이다. GRU 모델의 Training Accuracy 평균은 97.031%이며 Loss는 1.339가 나왔다.

표 1. GRU 모델 분류 성능 지표 결과

	Precision	Recall	F1-Score	Support
Negative	0.9641	0.9895	0.9766	190
Neutral	0.9784	0.9784	0.9784	231
Positive	0.9673	0.9452	0.9561	219
Accuracy			0.9703	640
Macro avg	0.9699	0.9710	0.9704	640
Weighted avg	0.9703	0.9703	0.9702	640

표1은 GRU 모델의 분류 성능 평가 지표 결과이다. 640개의 테스트 데이터를 사용하여 데이터 분류를 진행하였다. 성능 평가 지표의 최적화 목적 함수인 Accuracy는 0.9703이 나왔다. 본 연구는 Label 별로 분류하는 연구이기 때문에 연구의 목적성에 맞는 Weighted avg의 결과로 분류 성능 지표를 살펴보았을 때 Precision은 0.9703, Recall은 0.9703, F1-Score는 0.9702이 나왔다.

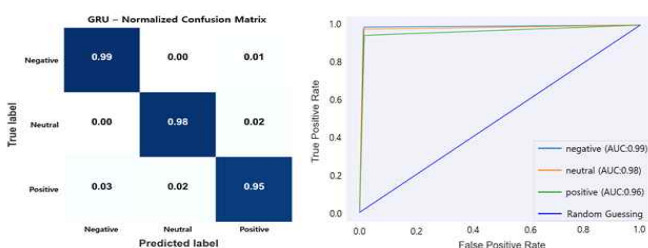


그림 4. Normalized Confusion Matrix 와 ROC-Curve

그림4는 Normalized Confusion Matrix와 ROC-Curve이다. Normalized Confusion Matrix에서 Negative, Neutral, Positive를 분류한 결과 Negative는 99%, Neutral은 98%, Positive는 95%로 나타났다. ROC-Curve를 보면 AUC가 Negative 0.99, Neutral 0.98, Positive 0.96이 기록되었으며 AUC-Score는 0.9781로 나타났다.

III. 결론

본 논문에서는 뇌파 데이터를 활용한 GRU 모델의 인간 감정 분류 성능 평가에 관한 연구를 진행하였다. 선행 연구에서 가장 뛰어난 성능을 보인 GRU 모델을 통해 뇌파 분류를 진행하였다. 이후 분류 성능 평가 지표를 통해 모델의 분류 성능 평가를 확인하였다. 연구에서 사용한 분류 성능 평가 지표는 Accuracy, Precision, Recall, PCA 기법, ROC-curve, Normalized Confusion Matrix, AUC-Score이다. 분류 성능 평가 결과 각 지표들 모두 뛰어난 분류 성능 결과가 나타났다. 향후 연구로 실제 인간을 대상으로 수집한 뇌파를 수집할 계획이며 수집된 데이터를 분류하여 BCI에 활용하는 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 성과는 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 3단계 산학연협력 선도대학 육성사업(LINC 3.0)의 연구결과입니다(NTIS 과제 번호. 1345356198).

참 고 문 헌

- [1] Younsang K., "A Study on the Artificial Intelligence Model of Emotional Recognition Classification Using EEG Signals", CEIC, 2022.
- [2] Kaggle, "EEG Brainwave Dataset: Feeling Emotions : Positive and Negative emotional experiences captured from the brain", (<https://www.kaggle.com/datasets/birdy654/eeg-brainwave-dataset-feeling-emotions>.)
- [3] Bird J. J., Manso L. J., Ribiero E. P., Ekart A., and Faria D. R., "A study on mental state classification using eeg-based brain-machine interface," in 9th International Conference on Intelligent Systems, IEEE, 2018.
- [4] Bird J. J., Ekart A., Buckingham C. D., and Faria D. R., "Mental emotional sentiment classification with an eeg-based brain-machine interface," in The International Conference on Digital Image and Signal Processing (DISP'19), Springer, 2019.
- [5] Kyunghyun C., Bart van M., Caglar G., Dzmitry B., Fethi B., Holger S., Yoshua B., "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv, 2014.
- [6] Abdi, Hervé, Lynne J. W., "Principal component analysis." Wiley interdisciplinary reviews: computational statistics 2.4 pp. 433-459, 2010.
- [7] Rey, Gonzalo H., Pedreira C., Quiroga R. Q., "Past, present and future of spike sorting techniques." Brain research bulletin 119 pp. 106-117, 2015.