

감성 분석의 다양한 모델들에 대한 실용성 평가

김민구(기계공학과) , 송유지(소프트웨어학부), 신재현(컴퓨터공학), 유승태(컴퓨터공학), 이민정(전기전자공학부)

2021 CUA이 중앙대학교 인공지능 학회 동계 컨퍼런스
2021 Chung-Ang University Artificial Intelligence Society's Winter Conference

Abstract

상황에 따라 행동이나 말을 결정하고 그 속에서 묻어나는 감정을 파악하는 것은 인간 고유의 영역이었지만, 인공지능 기술이 발전됨에 따라 기계가 사람의 감정을 파악하는 기술 또한 더 이상 불가능한 일이 아니게 되었다. 이러한 기술은 활용성이 높아 필요성이 대두되고 있다.

우리 팀은 위와 같은 감정 분석에 관련한 다양한 접근 방식을 직접 현실 세계를 보다 잘 반영한 데이터를 통해 이들 기술이 얼마나 발전되었는지 파악하고, 그리고 어떤 부분에 있어 아직 발전 가능성이 존재하는지 분석한다.

Introduction

인간이 감정을 파악하는 것에 있어 가장 많이 사용되는 데이터는 표정, 어조, 발화 내용 등이다. 따라서 우리는 이미지, 음성, 텍스트를 활용한 감정 인식 방법에 중점을 두고 조사 및 테스트를 진행하였다.

이미지를 활용한 분석에서는 기본적인 Deep Convolutional Neural Network를 활용한 연구부터 시작하여, 이들만으로 해결할 수 없는 표정의 불명확함을 해결할 수 있는 새로운 방법론을 조사하였다.

텍스트 감성 분석의 경우 BERT를 사용하여 진행한 연구에 대해 주로 조사하였다.

음성을 통한 감성 분석은 머신러닝과 딥러닝의 기술 발전과 더불어 빠르게 발전하고 있으며, 현재 많은 연구는 MFCC와 Mel Spectrogram을 활용한 RNN/CNN 모델들을 사용하고 있다.

다양한 모델들의 실용성을 확인하기 위해 기존의 벤치마크를 위한 데이터를 활용하는 것은 실제 유효성을 확인하기 어렵다고 판단되어 AI-HUB의 ‘멀티모달 영상 데이터셋’을 활용하였다. 해당 데이터셋은 110시간 분량의 총 6천개 영상 클립으로 구성되어 있고 클립 당 1-3분 내외의 대화를 포함하는 영상 파일과 영상을 설명하는 메타파일로 구성되어 있다. 발화 정보에는 학습 과정에서 정답 라벨로 활용할 수 있는 감정과 발화별 대화 의도 및 대화전략 정보가 부착되어 있다.

Materials and Methods

1.텍스트

전처리. Script 감정 라벨에 대하여 분노, 경멸, 혐오, 공포, 행복, 슬픔, 놀람, 순서대로 라벨 인코딩을 진행하였다.

모델. 감정 분류 모델로는 ktrain 패키지의 ALBERT 모델을 사용하였다. albert-large 모델의 pretrained weight을 활용하였다.

2.이미지

전처리. 텍스트와 동일한 라벨을 사용하였고 학습은 FER-2013 데이터셋으로 하고 테스트로 AI-HUB 데이터셋을 사용하였다. 테스트 이미지는 DLib에서 제공하는 Face Detector로 얼굴 영역만 48x48 해상도로 추출하였다.

모델. 케글에 공개된 표정 인식 관련 소스를 활용하여 6개의 Convolutional Layer를 활용한 DCNN 모델을 활용.

3.음성

전처리. 영상 내 각 대화별로 음성 데이터를 추출하고 zero padding을 실시하였다. 이후 데이터 증강을 하고 MFCC 또는 Mel Spectrogram으로 변환하였다.

모델. Mel Spectrogram을 활용하여 bidirectional CRNN 모델을 기반으로 학습하였다.

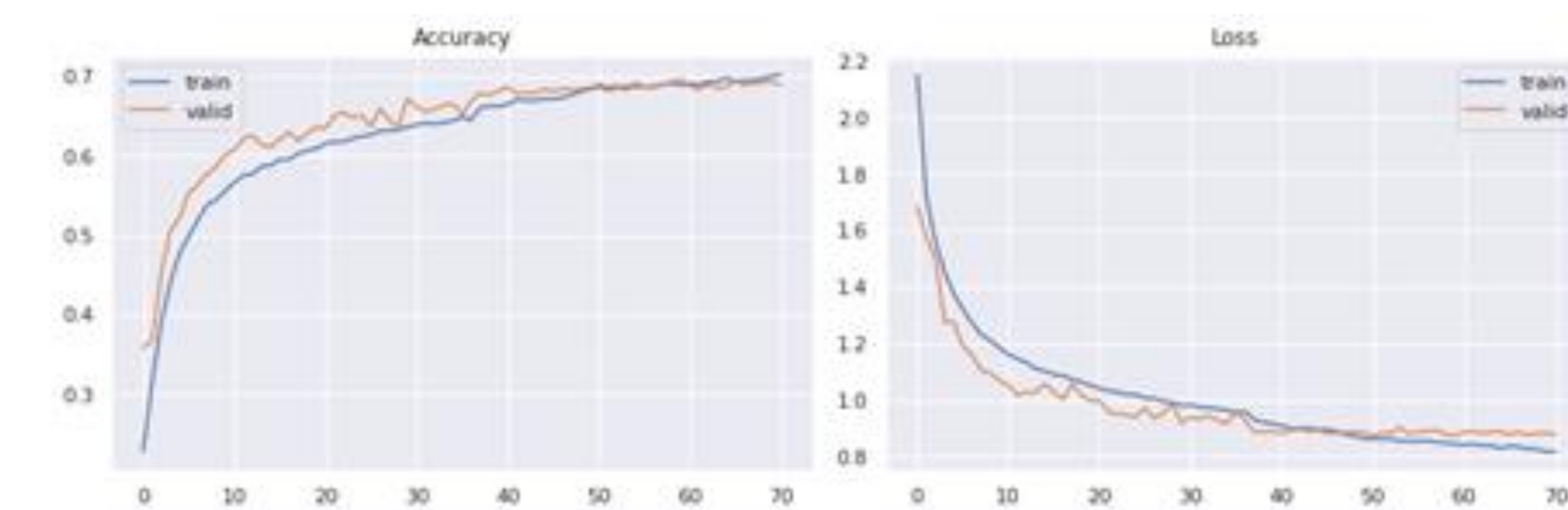
Conclusion

다양한 도메인에서 공통적으로 보여지는 문제점은 복합적인 감정을 표현할 수 없는 라벨링으로 보여진다. 멀티모달 데이터셋에서도 표정, 음성, 텍스트가 모두 다른 감정으로 라벨링되어있는 경우도 심심치 않게 보여진다. 따라서 복잡한 감정을 학습할 수 있도록하는 라벨링의 필요성이 보여진다.

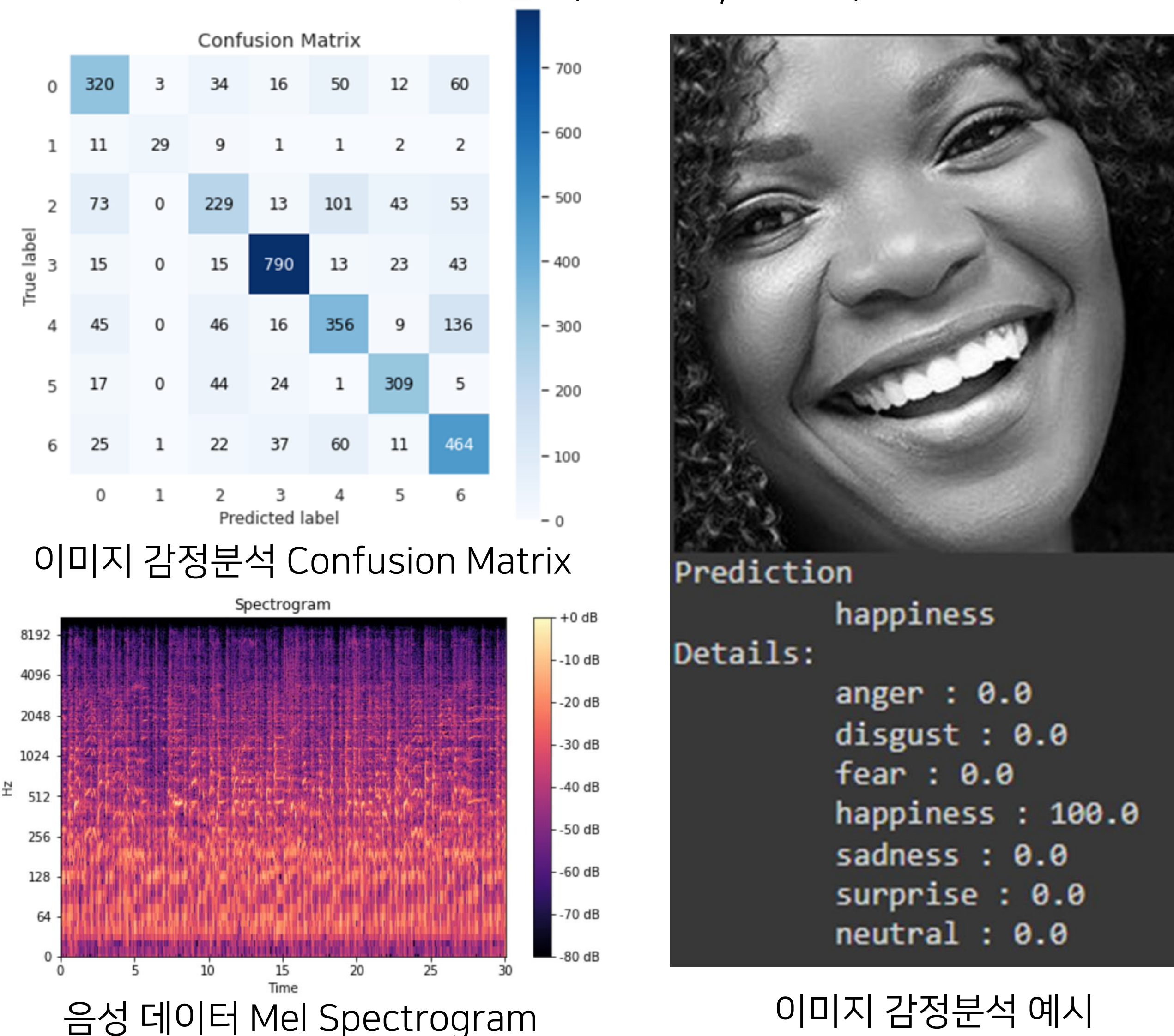
Results

	precision	recall	f1-score	support
fear	0.37	0.13	0.19	175
angry	0.00	0.00	0.00	74
contempt	0.31	0.03	0.05	393
sad	0.42	0.56	0.48	685
surprise	0.29	0.65	0.40	474
dislike	0.37	0.29	0.32	435
happy	0.58	0.23	0.33	310
accuracy			0.36	2546
macro avg	0.33	0.27	0.25	2546
weighted avg	0.37	0.36	0.32	2546

ktrain ALBERT 모델 결과	ktrain ALBERT 모델 confusion matrix
---------------------	-----------------------------------



DCNN 학습결과(Accuracy & Loss)



음성 데이터 Mel Spectrogram

이미지 감정분석 예시

Reference

- [1] “음성,이미지, 텍스트를 동시 인식하는 AI 플랫폼 설명”. Available: <http://www.aitimes.kr/news/articleView.html?idxno=24067>
- [2] “Github Link for Facial Expression Recognition”. <https://github.com/WuJie1010/Facial-Expression-Recognition.Pytorch>
- [3] Zhou Yue, Feng Yanyan, and Zeng Shangyou “Facial Expression Recognition Based on Convolutional Neural Network”, IEEE International Conference on Software Engineering and Service Sciences, Oct. 2019
- [4] AI-HUB 데이터 출처 “멀티모달 영상” <https://aihub.or.kr/aidata/137>
- [5] Yuhang Zhang, Chengrui Wang, and Weihong Deng “Relative Uncertainty Learning for Facial Expression Recognition” 35th Conference on Neural Information Processing Systems (NeurIPS 2021).
- [6] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
- [7] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.
- [8] Lan, Zhenzhong, et al. "Albert: A lite bert for self-supervised learning of language representations." arXiv preprint arXiv:1909.11942 (2019).