

면접 코칭을 위한 전이 학습 기반 표정 인식 모델: 사전 학습 모델과 DenseNet의 결합을 통한 Grad-CAM 시각화

김서현, 박현희
명지대학교
{tjgus6602, hhpark*}@mju.ac.kr

Transfer learning-based facial expression recognition model for interview coaching: Grad-CAM visualisation with DenseNet combined with pre-trained models

Seohyeon Kim, Hyunhee Park *
Myongji Univ.

요약

본 논문은 면접 코칭 서비스에서 면접관에게 자신감, 감정 상태 등을 전달하는 중요한 수단인 비언어적 요소 중 얼굴 표정을 인식하기 위해 사전 학습 모델과 DenseNet 모델을 결합한 전이 학습 모델을 제안한다. 이 모델은 사용자의 얼굴 표정을 긍정, 중립, 부정의 세 가지 범주로 분류하여 이를 기반으로 피드백을 제공한다. 사전 학습 모델 이후 global average pooling에서 추출된 특징 값에 히트맵(Heatmap)을 적용한다. 이를 통해 사용자가 표정 인식 과정을 시각적, 직관적으로 이해할 수 있도록 한다.

I. 서론

면접에서 비언어적 요소, 특히 표정은 의사소통에서 중요한 역할을 한다. 면접관은 지원자의 답변뿐만 아니라 표정, 제스처, 시선과 같은 비언어적 표현을 통해 지원자의 의사소통 능력과 감정 상태를 파악한다. 이러한 비언어적 요소는 지원자의 자신감, 태도, 신뢰도를 반영하기 때문에 면접 성공에 큰 영향을 미칠 수 있다. 따라서 면접 준비 과정에서 비언어적 요소를 정확히 인식하고 개선하는 것이 중요하다.

본 논문에서는 면접 시 사용자의 비언어적 요소, 특히 얼굴 표정을 인식하여 보다 객관적인 피드백을 제공할 수 있는 모델을 개발하고자 한다. 사용자가 자신의 표정을 분석할 수 있도록 함으로써, 무의식적으로 나타나는 부정적인 표정이나 불안한 감정 표현을 인식하고 이를 개선할 수 있는 기회를 제공하는 것을 목표로 한다.

연구에서 널리 사용되는 사전 학습 모델은 얼굴 표정 인식에 효과적이지만, 더 깊은 특징 추출과 표현을 위해 DenseNet[1]을 결합한 전이 학습 모델을 제안한다. 사전 학습 모델인 VGGNet[2]은 이미지에서 주요 특징을 추출하는 데 우수한 성능을 보이지만, DenseNet의 강력한 연결 구조를 활용하여 보다 상세한 감성 분류가 가능하도록 한다. 제안하는 모델은 면접 영상 속 얼굴 표정을 긍정, 중립, 부정이라는 세 가지 클래스로 분류하고, 분류 결과에 대한 시각적 피드백을 제공하기 위해 히트맵(Heatmap)을 적용한다.

II. 본론

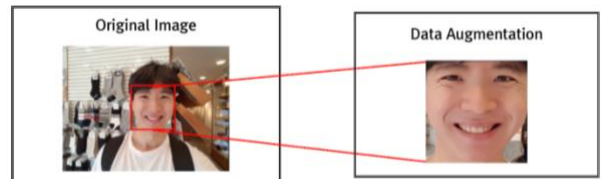
본 논문은 AI Hub의 한국인 감정 인식 데이터[3]를 활용하여 비대면 면접 코칭 서비스에서 사용할 수 있는 얼굴 표정 인식 모델을 학습한다. ¹ 해당 데이터셋은 기쁨, 당황, 슬픔, 분노, 불안, 상처, 중립이라는 7개의 감정 클래스로 구성되어 있다. 그러나 실제 면접 상황에서는 모든 감정이 중요성을 가지는 것은 아니므로, 면접 시 사용

자가 경험할 가능성이 높은 감정에 대해서만 학습을 진행한다. 이를 위해 감정 클래스를 3가지로 줄이고, 기쁨은 긍정으로, 중립은 그대로 중립으로, 그리고 당황과 불안은 부정으로 매핑하여 데이터를 재구성한다. 이렇게 분류된 3가지의 감정 클래스를 기반으로 면접 코칭 서비스에 사용한다.

표 1. 모델 학습에 사용한 데이터셋

| 클래스 | 개수 |
|-----|--------|
| 기쁨 | 3,000 |
| 당황 | 3,000 |
| 불안 | 3,000 |
| 중립 | 3,000 |
| 합계 | 12,000 |

데이터 전처리 과정에서는 데이터의 다양성을 높이고, 모델이 다양한 환경에서도 높은 일반화 성능을 보일 수 있도록 데이터 증강 기법을 적용한다. 구체적으로 Crop(자르기)과 Rotation(회전)과 같은 데이터 증강 기법을 활용하여 다양한 각도와 크기의 이미지를 생성하였다. 최종적으로 구축된 데이터셋은 각 클래스의 균형을 맞추고, 다양한 상황에서의 얼굴 표정을 학습할 수 있도록 설계한다.



본 논문에서는 모델 학습에서 이미지 분류에 널리 사용되는 사전 학습 모델과 DenseNet을 결합한 전이 학습

그림 1. 데이터 증강 기법을 적용한 이미지 예시

¹ 이 연구는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 구축된 "한국인 감정인식을 위한 복합 영상"을 활용하여 수행된 연구입니다. 본 연구에 활용된 데이터는 AI 허브(aihub.or.kr)에

서 다운로드 받으실 수 있습니다.

방식을 제안한다. 사전 학습 모델(VGGNet)은 합성곱 신경망(CNN)의 구조 중 하나로, 이미지에서 저수준의 세부 정보부터 고수준의 의미 있는 특징까지 추출할 수 있는 능력이 뛰어나지만, 네트워크의 깊이가 깊어질수록 학습 속도가 느려지고, 매개변수의 수가 기하급수적으로 증가하는 문제가 발생할 수 있다. 이러한 문제를 해결하기 위해 DenseNet 을 결합한 새로운 모델을 설계한다. DenseNet 은 각 층이 이전 모든 층의 출력을 재사용하여 효율적인 학습을 가능하게 하며, 네트워크의 학습 과정에서 중복되는 정보의 계산을 줄여준다. 이로 인해 DenseNet 은 매개변수의 수를 감소시키면서도 정보 흐름을 극대화하여 모델이 더 적은 데이터와 자원으로도 높은 성능을 유지할 수 있게 한다. 이러한 전이 학습 모델을 통해 학습 시간은 단축되면서도 높은 성능을 유지할

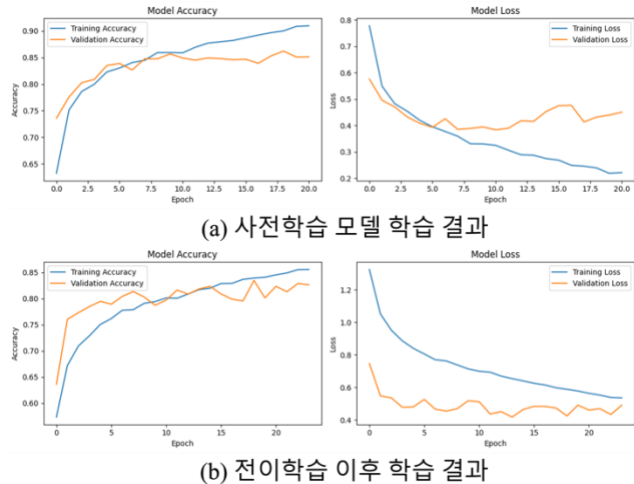


그림 3. 사전학습 모델과 전이학습 모델의 학습 결과 비교

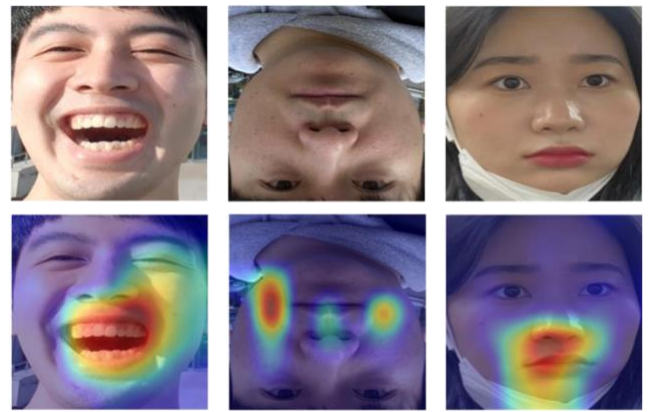
수 있다.

DenseNet 의 출력이 나온 후에는 글로벌 평균 풀링(Global Average Pooling, GAP) 레이어를 적용한다. 이 레이어는 각 피쳐 맵(feature map)에서 평균 값을 계산하여, 네트워크의 출력 차원을 줄이고 과적합을 방지한다. 이를 통해 네트워크는 보다 간결한 형태로 학습할 수 있으며, 궁극적으로 높은 분류 성능을 발휘할 수 있다.

이후, Grad-CAM(Gradient-weighted Class Activation Mapping)을 적용하여 얼굴 표정을 인식할 때 모델이 어느 부분에 초점을 맞추었는지 시각적으로 보여준다. Grad-CAM 은 모델이 예측한 클래스에 대해 어느 부분이 중요한 역할을 했는지 시각적으로 나타내는 도구로, 히트맵 형태로 출력된다. 예를 들어, 사용자가 긍정적인 표정을 짓고 있을 때, 네트워크가 사용자의 입이나 눈과 같은 영역에 주목했는지 여부를 Grad-CAM 을 통해 확인할 수 있다. 이 시각적 피드백은 면접 준비를 하는 사용자에게 중요한 정보를 제공할 수 있다. 사용자는 자신의 표정 중 어느 부분이 부정적이거나 불안한 인상을 줄 수 있는지를 직관적으로 파악할 수 있으며, 이를 개선하기 위한 노력을 할 수 있다.

III. 결론

본 논문에서는 면접 코칭 서비스에서 비언어적 요소, 특히 얼굴 표정을 분석하고, 이를 기반으로 피드백을 제공할 수 있는 모델을 제안하였다. 기존 사전 학습 모델만을 사용한 경우 테스트 정확도는 84.67%를 기록한 반면, DenseNet 을 결합한 전이 학습 구조를 통해 성능을 개선한 결과, 테스트 정확도는 85.58%로 증가하였다. 이는 사전 학습 모델이 기본적인 얼굴 특징을 잘 포착하는 동



(a) 기쁨 - 긍정 (b) 무표정 - 중립 (c) 당황 - 부정
그림 2. Grad-CAM 을 적용하여 히트맵으로 시각화한 이미지 예시

시에, DenseNet 이 보다 세밀하고 복잡한 특징을 추가로 분석하여 상호 보완적인 시너지 효과를 발휘한 결과로, 모델의 일반화 성능이 향상되었음을 보여준다. 또한, Grad-CAM 기법을 사용하여, 각 표정의 분류 과정에서 모델이 주의를 기울인 부분을 히트맵으로 시각화한다. 이를 통해 사용자가 표정 인식의 과정을 직관적으로 이해할 수 있도록 한다.

향후 연구 방향에서는 긍정, 중립, 부정 세 가지 표정 클래스 외에도 더 다양한 감정 클래스를 추가하여, 보다 세밀한 표정 분석이 가능하게 할 예정이다. 또한, 실시간 성능을 도입하여 즉각적인 피드백을 제공함으로써, 면접 코칭 시스템의 사용자 경험을 더욱 향상시키는 방향으로 모델을 확장할 예정이다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1A2C2005705, 부산 머신러닝 기반 지능형 플라잉 기지국을 위한 AI-MAC 프로토타입)

참고 문헌

- [1] Gao Huang, Zhuang Liu, Laurens van der Maaten and Kilian Q. Weinberger. "Densely Connected Convolutional Networks." arXiv preprint arXiv:1608.06993 (2016).
- [2] Karen Simonyan, Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
- [3] AI Hub, 한국인 감정인식을 위한 복합 영상, <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=82>
- [4] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization