

## AI면접 대상자에 대한 다면적 평가방법론 -얼굴인식, 음성분석, 자연어처리 영역의 융합

지현욱<sup>o</sup>, 이상진<sup>o</sup>, 문성민\*, 이재열\*, 이동은\*, 임규상\*

<sup>o</sup>청년취업사관학교 새싹 금천3기 AI융합서비스 개발자양성과정,

\*청년취업사관학교 새싹 금천3기 AI융합서비스 개발자양성과정

e-mail: jk\_117@naver.com<sup>o</sup>, sjsj9620@gmail.com<sup>o</sup>

## Multifaceted Evaluation Methodology for AI Interview Candidates - Integration of Facial Recognition, Voice Analysis, and Natural Language Processing

Hyunwook Ji<sup>o</sup>, Sangjin Lee<sup>o</sup>, Seongmin Mun\*, Jaeyeol Lee\*, Dongeun Lee\*, kyusang Lim\*

<sup>o</sup>Seoul Software Academy, Keumcheon 3th AI Convergence Service Development Program,

\*Seoul Software Academy, Keumcheon 3th AI Convergence Service Development Program

### ● 요약 ●

최근 각 기업의 AI 면접시스템 도입이 증가하고 있으며, AI 면접에 대한 실효성 논란 또한 많은 상황이다. 본 논문에서는 AI 면접 과정에서 지원자를 평가하는 방식을 시각, 음성, 자연어처리 3영역에서 구현함으로써, 면접 지원자를 다방면으로 분석 방법론의 적절성에 대해 평가하고자 한다. 첫째, 시각적 측면에서, 면접 지원자의 감정을 인식하기 위해, 합성곱 신경망(CNN) 기법을 활용해, 지원자 얼굴에서 6가지 감정을 인식했으며, 지원자가 카메라를 응시하고 있는지를 시계열로 도출하였다. 이를 통해 지원자가 면접에 임하는 태도와 특히 얼굴에서 드러나는 감정을 분석하는 데 주력했다. 둘째, 시각적 효과만으로 면접자의 태도를 파악하는 데 한계가 있기 때문에, 지원자 음성을 주파수로 환산해 특성을 추출하고, Bidirectional LSTM을 활용해 훈련해 지원자 음성에 따른 6가지 감정을 추출했다. 셋째, 지원자의 발언 내용과 관련해 맥락적 의미를 파악해 지원자의 상태를 파악하기 위해, 음성을 STT(Speech-to-Text) 기법을 이용하여 텍스트로 변환하고, 사용 단어의 빈도를 분석하여 지원자의 언어 습관을 파악했다. 이와 함께, 지원자의 발언 내용에 대한 감정 분석을 위해 KoBERT 모델을 적용했으며, 지원자의 성격, 태도, 직무에 대한 이해도를 파악하기 위해 객관적인 평가지표를 제작하여 적용했다.

논문의 분석 결과 AI 면접의 다면적 평가시스템의 적절성과 관련해, 시각화 부분에서는 상당 부분 정확도가 객관적으로 입증되었다고 판단된다. 음성에서 감정분석 분야는 면접자가 제한된 시간에 모든 유형의 감정을 드러내지 않고, 또 유사한 톤의 말이 진행되다 보니 특정 감정을 나타내는 주파수가 다소 집중되는 현상이 나타났다. 마지막으로 자연어처리 영역은 면접자의 발언에서 나오는 말투, 특정 단어의 빈도수를 넘어, 전체적인 맥락과 느낌을 이해할 수 있는 자연어처리 분석모델의 필요성이 더욱 커졌음을 판단했다.

**키워드:** 인공지능(AI), 머신러닝(Machine learning), 합성곱 신경망(CNN), 장단기 기억 신경망(LSTM)

## I. Introduction

최근 인공지능(AI) 기술의 발전과 함께 기업들은 인재 선발 과정에 AI 면접시스템을 도입하고 있다. 이러한 변화는 면접 과정의 효율성과 정확성을 높이는 한편, AI 면접의 실효성과 공정성에 대한 논란도 함께 증가시켰다. 본 논문에서는 이러한 AI 면접시스템의 다양한

측면을 분석하고, 그 효과와 한계를 탐구하고자 한다.

AI 면접시스템은 지원자의 시각적, 음성적, 그리고 자연어 처리 측면에서 평가를 진행하는데, 이는 지원자의 감정, 태도, 그리고 의사소통 능력을 다방면으로 분석하려는 시도이다. 특히, 시각적 부분에서는

딥러닝 기술을 이용해 지원자의 표정과 시선 처리를 분석하고, 음성 부분에서는 음성의 주파수 변화를 통해 감정 상태를 추정한다. 본 논문에서는 AI 면접시스템의 각 부분에 대한 심층적인 분석을 통해 그 유효성을 검증하고자 한다.

이 논문은 AI 면접시스템의 여러 측면을 종합적으로 평가함으로써, AI 기술이 인재선발 과정에서 어떻게 활용될 수 있는지에 대한 중요한 시사점을 제공하고자 한다. 이는 향후 AI 기술의 발전과 함께 기업들의 인재선발 방식에 중대한 영향을 미칠 것이기 때문이다.

## II. Preliminaries

### 1. Related works

최근 인공지능(AI) 기술의 발전은 채용 과정에 혁신적인 변화를 가져왔습니다. 특히, AI 면접 도구의 사용은 전 세계적으로 빠르게 확산되고 있으며, 이는 채용 효율성을 높이고 편견 없는 평가를 가능하게 하는 주요 도구로 자리 잡고 있다. 최근의 연구들은 AI 면접 도구가 인사 담당자의 부담을 줄이고, 보다 객관적인 평가를 제공한다고 보고하고 있다.

하지만, 이러한 기술의 사용은 윤리적, 법적 측면에서도 중요한 논의를 필요로 한다. 개인 정보 보호, 데이터 보안, 그리고 AI의 결정에 대한 투명성은 중요한 이슈로 대두되고 있다. 특히, 알고리즘에 의한 편견 발생 가능성은 심각한 우려 사항으로 지적되고 있으며, 이를 방지하기 위한 연구와 정책 개발 또한 고려되어야 한다. 이미 실제 다수의 기업이 AI 면접을 채용 과정에서 사용하고 있는데, 문제는 이러한 AI 면접 방식에 대한 제대로 된 검증체계가 없다는 것이며, 또 이와 관련된 논문도 많지 않다는 것이다.

## III. The Proposed Scheme

이하에서는 3가지 분야에서 연구방법론, 이행방법과 연구의 결과 중심으로 AI 면접시스템에 대한 분석을 진행하고자 한다.

### 1. 면접자 얼굴인식 및 시선 처리 분야

#### 1.1 연구방법론

얼굴의 감정인식 분야는 컴퓨터 비전 분야에서 기계학습 알고리즘을 활용해 감정 상태를 식별할 수 있는데, 특히 합성곱 신경망(Convolutional Neural Networks, CNN)은 딥러닝의 한 분야로, 주로 이미지 인식과 처리에 사용된다. 이외에도 얼굴의 랜드마크(facial landmarks)를 표시해, 랜드마크의 움직임을 분석해 회로애라의 감정 상태를 식별할 수 있다. 논문에서는 이미지 분석 분야에서 보편적으로 활용되는 CNN을 활용해 얼굴의 감정(기쁨, 슬픔, 화남, 놀람, 자신감, 중립)을 분석했다.

- 데이터셋 출처 및 구성방법:

-참여자의 6명 얼굴 사진, AI허브 제공 감정 이미지, 온라인상 구글링 이미지, 증식 9회 실시, 총 108,000장 사용. 시선 처리의

20,000장 사용

-오픈 캠을 활용해 다양한 각도에서 다양한 감정 사진 캡처, 노이즈 방지를 위해 얼굴 부분만 잘라내기

- 데이터 전처리:

-이미지 데이터를 64\*64 크기, 배열로 전환 후 증식

-이미지 형성 과정에서 조명 밝기나 이미지 채도 등이 결괏값에 영향을 미치서, ±30% 범위 무작위 방식으로 밝기, 대조, 각도, 줌, 채도 등 다양한 상황 가정해 증식

- 모델 구성:

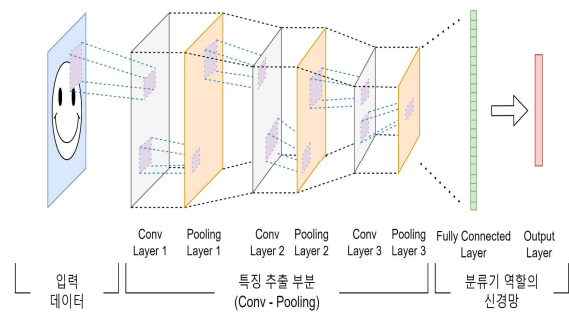


Fig. 1. 합성곱 신경망의 구조 및 작동 방식

-합성곱 신경망(CNN) 모델을 3개의 층으로 쌓고, 이미지가 입력되면 특징추출 부분에서 합성곱 연산 및 풀링 과정을 거쳐 이미지의 특징을 추출,

-이미지 감정인식은 다중 분류(6가지), 시선 처리는 이진 분류(2가지)로 구성해 결괏값 도출했으며, 동일한 CNN 모델을 적용했음.

### 1.2 연구결과

#### (1) 6가지 감정인식모델 결과

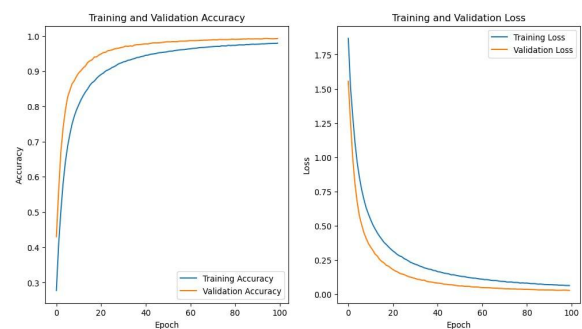


Fig. 2. 훈련 및 검증데이터 정확도 및 손실률  
accuracy: 0.9795 / val\_accuracy: 0.9927 / val\_loss: 0.0258

(2) 시선응시 모델결과

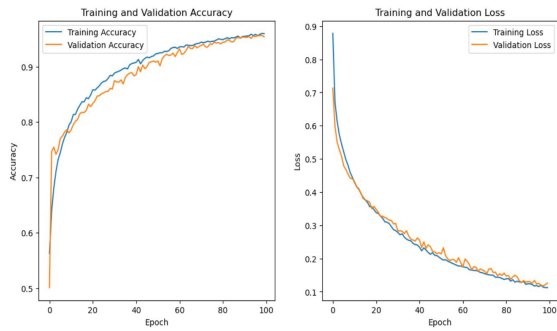


Fig. 3. 훈련 및 검증데이터 정확도 및 손실률  
accuracy: 0.9597 / val\_accuracy: 0.9542 / val\_loss: 0.1256

1.3 모델평가

CNN을 활용한 감정인식 판단 부분의 정확도는 97.95%, 시선 처리 여부의 정확도는 95.97%로 상당히 높게 나왔으며, CNN모델을 활용해 얼굴 감정 상태의 변화와 시선 처리 여부를 상당히 잘 판단하는 것으로 평가되었다. 다만, 얼굴 감정 상태만으로 면접자의 모든 성향을 판단하기 어렵기 때문에, 음성분석과 자연어처리 분석과 병행해 판단해야 할 것이다.

2. 음성 분석을 통한 감정인식

2.1 연구방법론

면접 과정에서 음성은 지원자의 감정 상태를 반영하는 중요한 요소이다. 이 연구는 AI-hub에서 제공하는 "감성 및 발화 스타일별 음성합성 데이터"를 활용하여, 성별에 따라 선별된 경우들이 녹음한 95,000개의 감정 발성 데이터를 분석하였다. 음성 데이터에서 MFCC(Mel-Frequency Cepstral Coefficients)를 추출하기 위해 librosa.feature.mfcc() 함수를 사용하였으며, 이를 특징 벡터로 변환하여 모델 학습에 사용하였다.

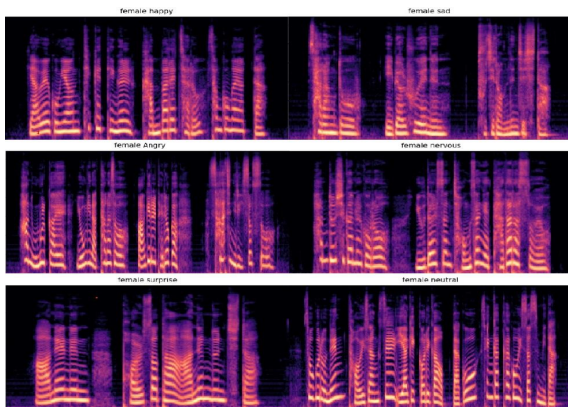


Fig. 4. 감정별 스펙트로그램

2.2 연구결과

감정 분류 모델은 음성의 스펙트로그램을 기반으로 학습되었고, Bidirectional LSTM 모델을 중심으로 4층 구조로 설계되었다. 학습된 모델은 혼동행렬을 사용하여 감정인식 성능을 평가하였고, 교차 검증을 통해 모델의 일반화 성능을 확인하였다.

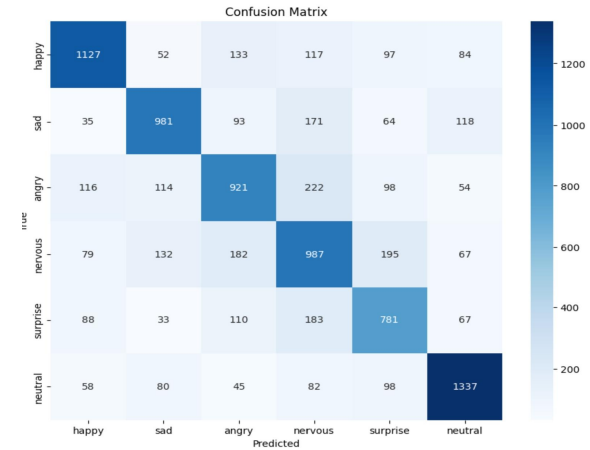


Fig. 5. Confusion Martrix를 이용한 감정 교차 검증

2.3 모델평가

평가 결과 Bidirectional LSTM 모델이 95%의 높은 정확도를 달성하였으며, 다른 모델들과 비교했을 때 상대적으로 우수한 성능을 보였다. LSTM 단독 모델은 65%, LSTM과 Transformer의 결합은 60%, 그리고 WaveNet과 CNN은 각각 55%의 정확도를 보였다. 이러한 결과는 Bidirectional LSTM이 음성 기반 감정인식에 있어서 효과적인 모델임을 시사한다.

Table 1. 음성 분석 모델별 정확도

모델	정확도
Bidirectional LSTM	95%
LSTM	65%
LSTM + Transformer	60%
WaveNet	55%
CNN	55%
RNN	55%
CNN + Transformer	50%
VGG 19	40%

3. 자연어 처리

3.1 연구방법론

자연어 처리를 통해 지원자의 발언을 분석하고, 지원자의 감정 상태와 적합성을 평가하는 방법을 제시한다. Whisper 기술을 활용한 STT(Speech To Text) 변환을 통해 얻은 텍스트 데이터를 분석하여, 지원자의 발언 중 빈도수가 높은 단어와 기업이 원하는 인재상에 맞는 문맥을 파악하였다. 이 연구에서는 특히 KoBERT를 사용하여 자연어의 감정분석을 수행했다.

- 데이터셋 출처:
  - 최근 2년간의 면접 영상 400개를 활용
  - AI 허브 감정 대화 말뭉치 및 영화 리뷰 데이터
- 데이터 전처리:
  - 합격자 불합격자 영상 라벨링, 면접자 인터뷰 영상에서 텍스트 문장 단위 추출
  - 형태소 분석(Okta)를 사용해 특수문자, 불용어 등을 제거하는 자연어 전처리 진행
  - 워드 임베딩 작업 및 라벨링
- 모델 구성:
  - Encoder-Decoder 구조의 seq2seq 모델을 구축하고, Attention 메커니즘을 통해 중요한 단어에 가중치를 부여했다.

### 3.2 연구결과

KoBERT를 활용한 자연어 처리 모델은 기계학습 기반의 기존 모델들에 비해 높은 정확도를 보였으며, 문맥 분석과 감정인식에서 좋은 성능을 나타냈다. 그러나 전체적인 평가에서 기대했던 75%의 정확도에는 미치지 못했다. 이는 문장 분석을 통한 지원자의 평가가 언어적 요소만으로는 한계가 있음을 시사한다. 따라서, 얼굴인식 및 시선 처리와 같은 다른 모달리티와의 결합을 통해 더욱 포괄적인 면접자 평가가 필요함을 제안한다.

### 3.3 모델평가

Table 2. 문장 분류 모델별 정확도

모델	정확도
KoBert	67.5%
RandomForestClassifier	62.5%
DecisionTreeClassifier	60%
LGBMClassifier	60%
KNeighborsClassifier	57.5%
AdaBoostClassifier	55%
LogisticClassifier	55%

다양한 분류 모델을 통해 합격자와 불합격자의 구분에 있어서 KoBERT가 67.5%의 정확도로 가장 높은 성능을 나타냈다. 이는 다른 기계학습 모델들과 비교할 때 유의미한 차이를 보였다.

## IV. Conclusions

이 논문은 최근의 AI 면접시스템의 도입과 그 효과에 논란이 있는 상황에서, AI 면접의 다면적 평가시스템이 과연 효과적인지에 대해 연구하기 위해, 시각, 음성, 자연어처리 분야에서의 평가방법론을 구현하고 검증했다. AI 면접에서 얼굴인식과 시선 처리를 통한 감정분석, 그리고 면접자의 음성 분석과 자연어 처리 결과의 통합적 해석은 지원자의 다면적인 평가를 가능하게 하는 중요한 요인이라 있다. 이는 면접 과정의 효율성과 정확성을 높이는 동시에, AI 면접의 실효성과 공정성에 대한 논란에 대해 중요한 시사점을 제공한다.

AI 기술의 발전은 인재선발 과정을 혁신적으로 변화시키고 있으며, 이러한 변화는 기업들이 더욱 객관적이고 비 편견적인 방법으로 인재를 선별하는 데 중요한 역할을 해야 할 것이기 때문이다.

논문의 분석 결과상, 현재 각 회사에서 활용되는 AI 면접시스템에서 시각화 부분은 상당히 높은 정확도를 가지는 것으로 평가되었다. 하지만, 음성분석과 자연어 처리 영역에서는 면접자의 태도 분석시 한계점을 지니고 있기 때문에, 음성인식을 통한 감정분석 모델을 더욱 세분화할 필요가 있으며, 특히 자연어처리의 경우, 단순히 감정이나 태도 분석을 넘어 다양한 평가척도를 마련할 필요가 있음을 파악했다. 앞으로 인공지능 기술의 발전과 인재선발 과정에서 이 기술이 어떻게 효과적으로 통합될 수 있는지에 대한 일견 중요한 지침을 제공한다고 할 수 있다. 더불어, 윤리적, 법적 측면 및 AI의 투명성과 관련된 중요한 고려사항들에 대해서도 논의가 필요하며, 향후 연구와 정책 개발이 필요함을 마지막으로 강조하고자 한다.

## REFERENCES

- [1] In-kyu Choi, Ha-eun Ahn, Jisang Yoo. (2018). Facial Expression Classification Using Deep Convolutional Neural Network. *Journal of Electrical Engineering & Technology*, 13(1), 485-492.
- [2] Ketan Sarvakar a, R. Senkamalavalli b, S. Raghavendra c, J. Santosh Kumar d, R. Manjunath e, Sushma Jaiswal f, Facial emotion recognition using convolutional neural networks, 2023
- [3] Soowon Park, & Tae-Young Choe. Specific sentiment analysis and sentiment trend analysis through natural language processing and machine learning. 2022
- [4] Min-A Lee1 ; Yeon-Ji Park2 ; Jun-Yeong Nal ; Chae-Bong Sohn3, Implementation of Review Sentiment Analysis Application Using KoBERT, KoGPT-2, and KoBART Optimized Hyperparameters, 2023
- [5] MJ Lim, P Kim, JH Shin, Sentiment Analysis Model with Semantic Topic Classification of Reviews, 2020
- [6] Alec Radford, Robust Speech Recognition via Large-Scale Weak Supervision, 2020
- [7] A-Hyeon Jo1, Keun-Chang Kwak2, "A Trend Analysis on Emotional Recognition Technology Using Speech Signals" The 2020 Annual Academic Conference
- [8] Choi Eun Bin, Kim Soo-Hyung, "Deep Learning-Based Emotion Recognition using Speech Features"
- [9] SangHyeuk Yoon, Dayun Jeon, Neungsoo Park, "Speech emotion recognition based on CNN - LSTM Model", ACK 2021 Academic Conference Proceedings (Volume 28, Issue 2)