Journal of Next-generation Convergence Information Services Technology Vol.12, No.3, June (2023), pp.391-400
ISSN 2384-101X(p) 2672-1163(e) *http://dx.doi.org/10.29056/jncist.2023.06.11*

시니어 헬스케어를 위한 음성 기반 감정 분류 모델에 관한 연구

A Study on the Voice-Based Emotional Classification Model for Senior Healthcare

이은서[1], 최형선[2], 황보택근[3]*

Eun-Seo Lee[1], Hyoung-Sun Choi[2], Taeg-Keun Whangbo[3]*

요 약

본 논문은 감정 분류가 되어있는 한국어 음성 데이터셋을 기반으로 시니어 헬스케어를 위한 노인들의 감정 분석 모델을 제안한다. 제안한 모델은 주어진 데이터 세트에서 노인들의 경우 불분명한 발음과 느린 음성 속도와 같은 음성 특징들을 극복하기 위해 잡음 제거, 음성 정규화, 발음 정제와 같은 전처리 기술을 적용하였으며, log mel spectrogram 기술을 활용하였고 분류 모델은 컨볼루션 레이어와 리커런트 레이러를 합친 Convolution-Recurrent Neural Network(CRNN)를 조합하여 구성하였다. 개발 모델에 대하여 평가지표로 Area Under the Curve(AUC)를 사용하여 얻어낸 감정 분류 정확도는 85%가 나왔다. 이러한 결과는 노인들의 음성 감정 분류에 대하여 정확도가 높고 효과적인 접근 방법을 제시하며, 노인들의 감정 상태를 이해하여 헬스케어에 크게 기여할 수 있다. 또한, 이러한 연구 결과는 노인들의 음성 감정 분류 분야에 새로운 가능성을 열어주고, 관련 연구 및 실제 응용에 큰 도움이 될 것으로 기대된다.

핵심어 : 감정 분류, 음성, log mel spectrogram, Convolution-Recurrent Neural Network(CRNN), Area Under the Curve(AUC)

Abstract

In this paper proposes an emotional analysis model for the elderly for senior healthcare based on a Korean speech dataset that is emotionally classified. The proposed model used pre-processing techniques such as noise removal, speech normalization, and pronunciation purification to overcome speech characteristics such as unclear pronunciation and slow speech speed in the case of the elderly in a given

1 Department Computer Science, Gachon University, Gyeonggi-do, Korea [Graduate Student]
e-mail: dmstj5308@gachon.ac.kr
2 Department Computer Science, Gachon University, Gyeonggi-do, Korea [Graduate Student]
e-mail: hschoi@gachon.ac.kr
3 Department Computer Science, Gachon University, Gyeonggi-do, Korea [Professor]
e-mail: tkwhangbo@gachon.ac.kr (Corresponding author)

* 본 연구는 경기도의 경기도 지역협력연구센터 사업의 일환으로 수행하였음.[GRRC-가천2020(B03), AI기반 헬스케어 콘텐츠 개발]

Received(April 26, 2023), Review Result(1st: May 22, 2023), Accepted(June 12, 2023), Published(June 30, 2023)

dataset. For the development model, the accuracy of emotion classification obtained using Area Under the Curve (AUC) as an evaluation index was 85%. These results suggest an accurate and effective approach to the classification of speech emotions in the elderly, and can greatly contribute to healthcare by understanding the emotional state of the elderly. In addition, these research results are expected to open new possibilities in the field of voice emotion classification in the elderly and to be of great help in related research and practical applications.

Keyword : Emotional Classification, Voice, log mel spectrogram, Convolution-Recurrent Neural Network (CRNN), Area Under the Curve(AUC)

1. 서론

자연어 처리(Natural Language Processing, NLP)는 인간의 언어를 기계적으로 이해하고 처리하는 인공지능 분야입니다. 인간의 언어는 복잡하고 다양한 문법, 의미, 구조, 문맥을 가지고 있어 기계가 이를 이해하고 활용하는 것은 어렵습니다. 그러나 자연어 처리는 텍스트 문서를 분석하고 해석하며, 문장의 의미를 추출하고 자동으로 요약하거나 번역하는 등의 작업을 수행할 수 있습니다. 또한, 대화 시스템과 같은 인간과 기계 간의 자연어 인터페이스를 개발하여 자연스러운 대화를 할 수 있도록 합니다. 최근 몇 년 동안 자연어 처리는 급속한 성장을 이루었습니다. 특히 Chat GPT와 같은 혁신적인 기술이 나오면서 자연어 처리 기술에 대한 인식도 일반인들에게 퍼지게 되었습니다.

이러한 자연어 처리 기술 중에서도 감정 인식은 다양한 분야에서 중요한 역할을 합니다. 감정 인식은 텍스트나 음성 데이터에서 사용자의 감정을 파악하는 기술로, 상당한 관심을 받고 있습니다. 특히 헬스케어 분야에서는 감정 인식이 중요한 역할을 할 수 있습니다. 사용자, 특히 노인들의 감정을 이해하고 적절하게 대응하는 능력은 상호 작용의 정확도를 크게 향상하고 전반적인 사용자 경험을 개선할 수 있습니다. 최근에는 의료 분야의 노인들을 위한 대화 모델 개발에 관한 관심도 높아지고 있습니다. 노인들은 일반 성인과 달리 디스플레이 화면의 사용에 어려움을 겪는 경우가 많아 음성 기반의 의사소통이 가능한 매체를 선호합니다. 또한, 노인 세대는 다른 연령대와는 다른 감정 표현 방식을 가지고 있으며, 발음이 불분명하고 음성 빠르기가 느리기 때문에 일반적인 모델을 사용하는 것은 효과적이지 않습니다. 이에 따라 우리는 시니어 헬스케어를 위한 음성 기반 감정 분류 모델 개발을 제안합니다. 목소리로 표현된 감정을 분석함으로써, 우리는 노인들의 감정 상태에 따라 반응을 조정할 수 있는 목표를 가지고 있습니다. 이 모델은 인간 대 인간 대화와 유사한 경험을 제공하여 노인들과 공감적이고 매력적인 상호 작용을 끌어내는 것을 목표로 합니다.

이러한 기술의 발전은 시니어 헬스케어 분야를 촉진할 수 있습니다. 음성 기반 감정 분류 모델을 통해 노인들의 음성을 분석함으로써, 우리는 노인들의 감정 상태에 대한 귀중한 통찰력을 얻을 수 있습니다. 이를 토대로 노인들과의 상호 작용을 개선하고 헬스케어 서비스의 품질을 높일 수 있습니다. 그리고 노인들은 자연스러운 음성 기반 인터페이스를 통해 쉽게 의사소통할 수 있으며,

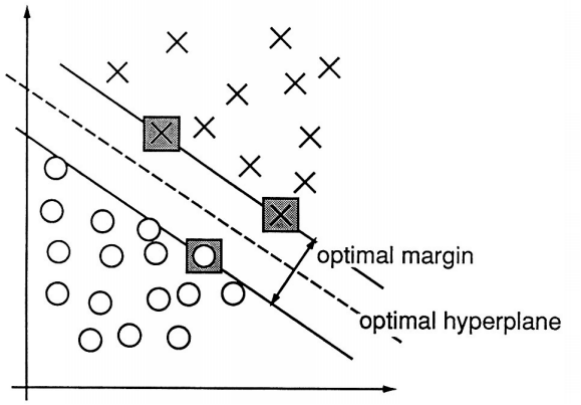
자신의 감정을 표현하고 공유할 수 있습니다. 따라서 음성 기반 감정 분류 모델은 노인들이 편안하게 사용할 수 있는 인터페이스를 제공하고, 노인의 감정을 이해하고 상황에 맞는 반응을 할 수 있는 기반이 될 수 있습니다.

2. 이론적 배경

자연어 처리 분야에서는 다양한 감정 분류 모델들이 개발됐습니다. 이 모델들은 텍스트 데이터에서 감정을 인식하고 분류하는 데 사용됩니다. 다음은 감정 분류의 몇 가지 모델들에 대하여 설명합니다.

2.1 SVM (Support Vector Machine)

SVM은 지도 학습 알고리즘 중 하나로, 주로 이진 분류 문제에 사용됩니다. SVM은 데이터 포인트들을 고차원 공간으로 매핑하여 서로 다른 클래스를 가장 잘 분리하는 결정 경계를 찾습니다. 이를 위해 SVM은 최대 마진(Maximum Margin) 분류를 수행합니다. 마진은 결정 경계와 클래스 사이의 거리로, SVM은 이를 최대화하여 분류의 안정성을 높입니다. [그림 1]은 최대 마진에 따른 분류 모습을 확인할 수 있습니다.



[그림 1] SVM 최대 마진에 따른 분류
[Fig. 1] Classification by SVM maximum margin

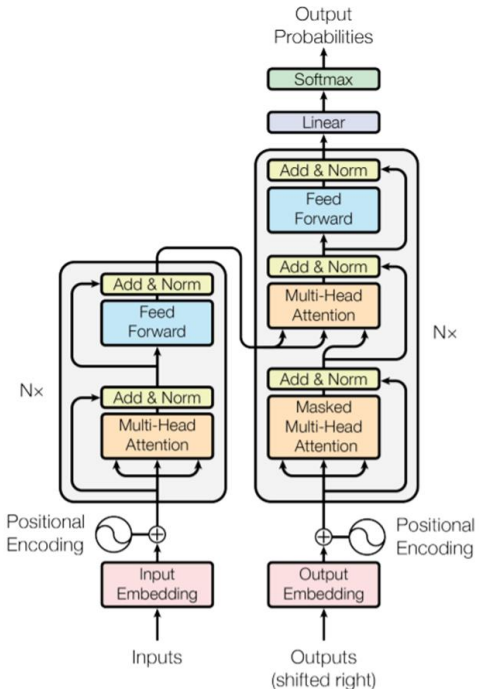
선형 SVM은 선형 분리 가능한 데이터를 처리하는 데 사용됩니다. 이 경우, SVM은 선형 결정 경계를 찾아 데이터를 분류합니다. 하지만 현실적인 데이터는 비선형일 수 있으므로 커널 SVM이 사용됩니다. 커널 SVM은 비선형 분리 가능한 데이터를 처리하기 위해 데이터를 고차원 공간으로 매핑하는 커널 함수를 사용합니다. 이렇게 고차원 공간에서 선형 결정 경계를 찾은 뒤, 다시 원래

공간으로 되돌려 비선형 결정 경계를 구합니다.

SVM은 분류 외에도 회귀 분석에도 사용될 수 있습니다. SVM 회귀는 데이터 포인트들이 결정 경계에 위치하도록 조정하는 회귀 모델을 만듭니다 [1].

2.2 BERT (Bidirectional Encoder Representations from Transformers)

BERT는 자연어 처리를 위한 사전 훈련된 언어 모델로, Transformer 아키텍처를 기반으로 합니다. BERT는 사전 훈련 단계와 미세 조정 단계로 구성되며 [그림 2]와 같은 구조를 가집니다.



[그림 2] BERT 구조도

[Fig. 2] BERT Structural Diagram

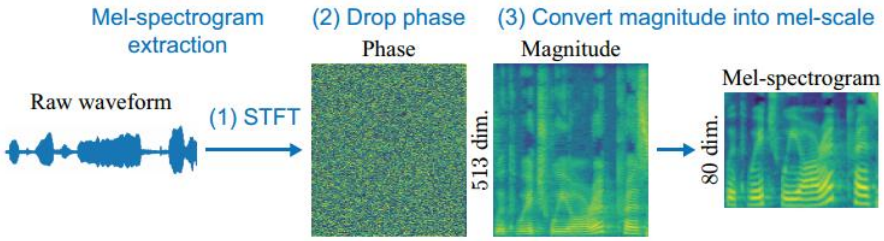
사전 훈련 단계에서 BERT는 대규모의 텍스트 데이터로 사전 훈련됩니다. 이 과정에서 양방향 언어 모델링(Objective)과 마스킹(Objective)이 사용됩니다. 양방향 언어 모델링은 문장 내의 단어를 예측하고, 마스킹은 문장 내의 일부 단어를 가려서 복원하는 작업을 수행합니다. 이를 통해 BERT는 문맥을 잘 이해하고 단어의 임베딩을 학습합니다. 미세 조정 단계에서는 BERT를 특정 자연어 처리 작업에 맞게 세부 조정합니다. 예를 들어, 문장 분류, 개체명 인식, 질의응답 등 다양한 작업에 BERT를 적용할 수 있습니다. 이 단계에서는 작은 데이터셋을 이용하여 BERT 모델을 미세 조정하고, 해당 작업에 대한 성능을 향상시킵니다 [2].

2.3 Log Mel

Log Mel은 음성 신호 처리에서 주로 사용되는 특성 추출 방법입니다. Log Mel은 Mel Spectrogram에 로그 변환을 적용한 것으로 음성 신호의 주파수 스펙트럼을 멜 척도(Mel scale)에 따라 변환하여 표현한 것입니다. 멜 척도는 인간의 청각 특성을 모델링 한 척도로, 인간의 청각 시스템이 주파수를 인식하는 방식에 기반을 둡니다. 일반적으로 Mel Spectrogram에 로그 변환을 적용하는 이유는 주파수의 크기 스케일을 조정하여 인간의 청각 특성을 더 잘 모델링하기 위함입니다. 로그 변환은 주파수의 크기 차이에 대한 인식 능력을 개선하고, 데이터의 동적 범위를 조정하여 특성 추출의 효과를 높입니다. Log Mel은 음성 인식, 음성 감정 인식, 음성 처리와 같은 음성 관련 작업에서 특성 추출에 널리 사용됩니다.

2.4 Log Mel Spectrogram

Log Mel Spectrogram은 음성 신호에서 특징을 추출하기 위해 사용되는 변환 기법입니다. [그림 3]은 음성 신호를 시간 및 주파수 정보를 가진 Mel Spectrogram으로 변환하는 과정입니다.



[그림 3] Mel Spectrogram 추출의 처리 흐름

[Fig. 3] Processing flow of Mel Spectrogram extraction

먼저, 오디오 신호를 작은 프레임으로 분할합니다. 각 프레임에 대해 짧은 시간 동안의 주파수 콘텐츠를 분석하기 위해 Fourier 변환을 적용합니다. 이를 통해 각 프레임의 주파수 성분을 얻을 수 있습니다. 이후, mel-scale을 적용하여 주파수를 인간의 청각에 더 잘 맞도록 변환합니다. mel-scale은 주파수 스펙트럼을 인간의 청각 특성에 맞게 변환하는 척도입니다. mel-scale 변환을 통해 주파수 영역을 선형적으로 나타내는 대신, 인간의 청각에 더 잘 부합하는 비선형적인 척도로 변환합니다. 마지막으로, 변환된 주파수 스펙트럼에 로그를 적용하여 스케일을 조정합니다. 로그를 취함으로써 주파수의 다양성을 강조하고, 작은 값과 큰 값 사이의 차이를 줄입니다. 이를 통해 더 좋은 동적 범위를 얻을 수 있습니다. 결과적으로, Log Mel Spectrogram은 시간에 따른 주파수의 변화를 고려한 음성 신호의 주파수 특성을 표현하는 이미지입니다. 이를 입력으로 사용하여 감정 분

류 모델을 학습하고 예측할 수 있습니다 [3].

3. 관련 연구

Rojas Víctor et al [4]의 연구는 노인들의 기분을 음성 처리를 통해 모니터링하는 시스템을 제안합니다. 이 시스템은 특히 슬픔을 감지하는 데 초점을 맞추었으며, 가족 구성원들이 필요한 사람을 적시에 지원할 수 있도록 돕습니다. Circumflex 감정 모델을 기반으로 하여 감정을 그룹으로 분류하여 슬픔을 인식합니다. 다양한 감정 데이터베이스를 사용하여 시스템을 평가한 결과, 남성의 경우는 94%의 사례를 인식하고 여성의 경우는 79%를 인식할 수 있는 성과를 얻었습니다. 또한, 이 솔루션은 모바일 시스템에서 사람들의 기분을 모니터링하기 위해 사용될 수 있습니다.

Jian Qijian et al [5]의 연구는 노인의 음성 특성을 통합하고 주의 메커니즘을 포함한 감정 인식 방법을 제안하며, 노인 음성 감정 인식의 낮은 정확도에 대응하는 방법을 연구합니다. 먼저 노인의 음성 특성을 추출하고 통합합니다. 그다음, 통합된 특성은 양방향 장·단기 기억망 (BLSTM)의 입력으로 사용하여 음성 각 프레임의 깊은 감정적 특징을 학습합니다. 주의 메커니즘은 각 프레임 특성의 감정 분류 가중치를 계산하기 위해 사용합니다. 마지막으로, 각 프레임의 특성은 해당 가중계수와 곱해져 완전 연결층의 입력으로 사용되어 음성 감정 인식을 완료합니다. 노인 음성 감정 데이터베이스에서의 실험 결과는 전통적인 BLSTM과 비교하여 연구에서 제안하는 방법이 노인 음성 감정 인식의 정확도를 효과적으로 향상시킬 수 있음을 보여줍니다.

Shafran Izhak et al [6]의 연구는 화자의 음성에서 음성 서명을 자동으로 정확하게 추출하는 문제를 연구합니다. 화자 특성을 추출하기 위해 두 가지 접근 방식으로 접근하였는데, 첫 번째 접근 방식은 일반적인 음향 및 운율적 특징에 초점을 맞추는 방법이고, 두 번째 접근 방식은 화자가 사용한 단어 선택에 초점을 맞추는 방법입니다. 첫 번째 접근 방식에서는 화자 특성에 의존한 HMM(숨겨진 마르코프 모델)을 사용하여 특성 및 음성 특징을 평가하여 모든 조사된 특성에 대해 우연 수준 이상의 정확도를 달성합니다. 두 번째 접근 방식은 음성 인식 라티스에 적용된 합리적 커널을 가진 지지 벡터 머신을 사용하여 감정의 이진 분류 과제에서 약 8.1%의 정확도를 달성하였습니다.

Wu Lei et al [7]의 연구는 노인을 대상으로 한 지능형 음성 비서의 이미지 선호도에 관해 연구합니다. 연구는 PAD (Pleasure, Arousal, Dominance) 감정 모델을 기반으로 진행되었습니다. 노인들이 어떤 음성 비서의 이미지를 선호하는지 조사하기 위해 실험을 진행하였습니다. 실험에서는 노인 참가자들에게 다양한 음성 비서의 이미지를 제시하고, 각 이미지에 대한 선호도를 평가하도록 요청하였습니다. 연구 결과는 노인들의 특성에 따라 선호도가 다르다는 것을 보여주었습니다. 선호도는 비서 이미지의 Pleasure, Excitement, Control에 영향을 받았으며, 노인들의 이미지 선호도는 각각의 노인 자신의 감정 차원에 따라 다양하게 변화하는 것으로 나타났습니다.

Han Kun et al [8]의 연구는 심층 신경망(DNN)을 활용하여 원시 데이터로부터 고수준 특징을 추출하고, 이러한 특징이 음성 감정 인식에 효과적임을 보여줍니다. DNN을 사용하여 각 음성 세그먼트에 대한 감정 상태 확률 분포를 생성합니다. 그런 다음 세그먼트 수준의 확률 분포로부터 문장 수준의 특징을 생성합니다. 이러한 문장 수준의 특징은 음성 감정을 식별하기 위해 특수한 단일 은닉층 신경망인 extreme learning machine (ELM)에 입력됩니다. 실험 결과는 제안된 접근 방법이 저수준 특징에서 감정 정보를 효과적으로 학습하며, 최신 기법과 비교하여 상대적인 정확도가 20% 향상되는 것을 확인하였습니다.

4. 연구 방법

본 연구에서는 노인들의 음성 감정 분류를 위해 log mel spectrogram과 연결된 모델을 개발하는 데 초점을 두었습니다.

음성 데이터 세트는 [그림 4]와 같이 AI Hub의 7가지 감정(happiness, angry, disgust, fear, neutral, sadness, surprise)을 다루는 한국어 대화 음성으로 구성되어 있습니다.

[그림 4] 감정 분류를 위한 대화 음성 데이터셋

[Fig.4] Conversational voice dataset for emotion classification

음성 데이터는 16bit, 48kHz wav 파일로 구성되어 있으며 [그림 5]와 같이 CSV 파일에 대화의 상황, 음성 인식 결과, 감정 라벨링 정보, 사용자의 성별 나이 정보가 담겨 있습니다.

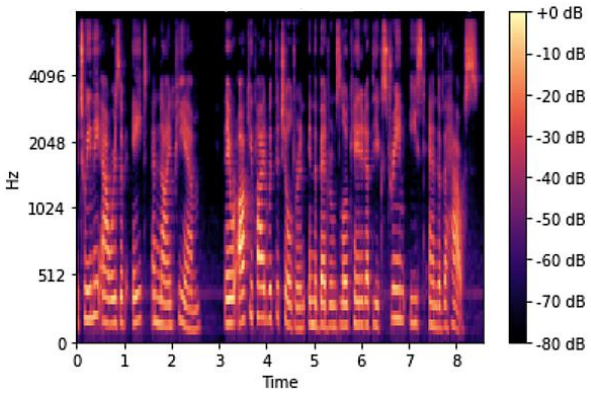
wav_id	발화문	상황	1번 감정	1번 감정세기	2번 감정	2번 감정세기	3번 감정	3번 감정세기	4번 감정	4번감정세기	5번 감정	5번 감정세기
5ee1e7939aa8ea0eec53fac7	나 어제 헤어졌어	sad	Sadness	1	Sadness	1	Sadness	1	Sadness	1	Sadness	1
5ee1e7a379bf120ed2b81ba2	어쩌다 보니까 그렇게 됐네	sad	Sadness	1	Sadness	1	Neutral	0	Sadness	1	Sadness	1
5ed9793b9aa8ea0eec53f7f1	유기견 다큐멘터리를 봤는데 무책임한 사람들 때문에 너무 화가 나	disgust	Sadness	2	Sadness	2	Angry	1	Angry	1	Angry	1
5ed979582880d70f286128c3	버려지는 유기견들의 생활을 다룬 다큐멘터리 없어	disgust	Sadness	1	Sadness	2	Sadness	1	Angry	1	Sadness	2
5ed9797b1dcf350eeded50b8	작년보다 5배는 더 늘어나는 것 같대 최근 들어 더 늘어나고 있고	disgust	Fear	1	Sadness	1	Sadness	1	Angry	1	Sadness	2
5ed9799ac90a530ee56b5f6b	정부보조금이랑 사람들의 기부금으로 운영되고 있어	disgust	Neutral	0	Neutral	0	Sadness	1	Sadness	1	Sadness	2
5ed979dc7e21a10eee253e90	맞아 벌어지는 대부분의 아이들이 다 큰 강아지 들였어 아직도 상처	disgust	Sadness	2	Sadness	2	Sadness	2	Sadness	2	Sadness	2
5ed97a22c90a530ee56b5f6c	어제저녁에 진짜 무서웠어	fear	Fear	2	Fear	2	Fear	1	Fear	1	Fear	2
5ed97a381dcf350eeded50bc	친구랑 약속 있어 나갔는데 밥 먹고 이따가 지진이 발생하는 거야 얼	fear	Fear	2	Fear	2	Fear	2	Fear	1	Fear	1
5ed97a5cc90a530ee56b5f6d	큰 지진은 지나갔는데 여진이 조금씩 있는 거 같기도 해	fear	Fear	2	Fear	2	Fear	1	Fear	1	Fear	1
5ed97a779aa8ea0eec53f7f5	다른 테이블에서 밥 먹던 사람들이 도망치다가 넘어지고 그 와중에 니	fear	Sadness	1	Fear	2	Sadness	1	Fear	1	Fear	1
5ed97a977e21a10eee253e92	그렇지 않아도 오늘 친구 데리고 병원 가기로 했어	fear	Neutral	0	Sadness	1	Sadness	1	Sadness	1	Fear	2
5ed97ac29aa8ea0eec53f7f8	룸메이트와 너무 자주 싸우게 돼	anger	Disgust	2	Angry	2	Angry	1	Angry	1	Fear	1

[그림 5] 라벨링 파일

[Fig. 5] Labeling file

수집한 데이터를 전처리하는 과정에서는 주어진 데이터셋에서 노인들의 경우, 불분명한 발음과 느린 음성 속도와 같은 문제를 극복하기 위해 잡음 제거, 음성 정규화, 발음 정제와 같은 전처리 기술을 적용하였습니다.

전처리를 마친 음성 데이터셋을 학습데이터와 테스트 데이터를 나누고 학습 데이터셋을 log mel spectrogram으로 변환합니다. Mel Spectrogram은 [그림 6]과 같이 시각화할 수 있습니다.



[그림 6] mel spectrogram 시각화

[Fig. 6] Visualization of mel spectrogram

이는 음성의 주파수 성분을 추출하는 방법의 하나입니다. log scale을 적용하여 음성 데이터의 다양한 주파수 영역을 더욱 잘 표현되도록 하며 데이터를 학습 및 평가를 위해 훈련 세트와 테스트 세트로 분할합니다. 각 세트는 다양한 음성 감정을 포함하도록 랜덤하게 선택되도록 하였습니다.

다음으로, log mel spectrogram을 추출하기 위해 오디오 데이터에서 주파수 영역을 시각적으로 표현하는 기술을 사용했습니다. 이를 위해 음성 신호를 프레임 단위로 분할하고, Mel 스케일 필터를 적용하여 주파수 영역을 변환합니다. 변환된 주파수 영역에 로그를 취하여 로그 스케일로 변환한 Log Mel spectrogram을 얻습니다. log mel spectrogram을 입력으로 받아 감정 분류를 수행하기 위해, 모델은 일련의 Convolution Layer로 시작하여 주파수적인 특징을 추출하고, 이후에 Recurrent Layer로 시계열적인 특징을 모델링합니다. 최종적으로 감정 클래스에 대한 출력 레이어가 추가하여 모델을 설계하였습니다. 모델은 softmax 활성화 함수를 사용하여 각각의 감정 클래스에 대한 확률 분포를 출력합니다. log mel spectrogram을 입력으로 사용하여 모델을 학습하며, 각각의 입력 spectrogram은 해당하는 감정 클래스에 대한 레이블과 함께 사용됩니다. Cross Entropy Loss를 사용하여 모델의 출력과 실제 감정 클래스 간의 차이를 최소화하고자 하였으며 학습에는 Adam Optimizer와 적절한 학습률을 사용하였습니다.

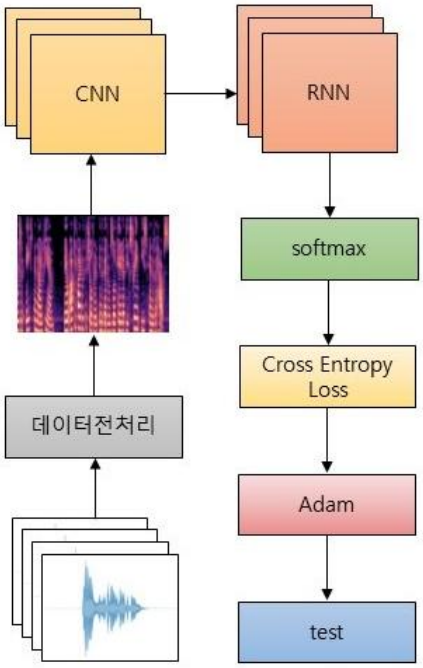
[표 1] AUC를 이용한 모델 정확도 평가

[Table 1] Model accuracy evaluation using AUC

파라미터	평균(%)
AUC	0.85

테스트 세트를 이용하여 학습된 모델을 평가합니다. 모델의 예측 결과와 실제 감정 클래스를 비교하여 Area Under the Curve(AUC) 평가 지표를 사용하여 모델의 성능을 평가하였습니다. 이를 통해 [표 1]과 같이 85%라는 높은 정확도를 얻을 수 있었습니다.

본 논문에서 제안한 모델의 전체 그림은 [그림 7]과 같습니다.



[그림 7] 제안한 모델 구조도

[Fig. 7] Schematic diagram of the proposed model

5. 결론

본 연구는 노인들이 표현한 감정을 분류하는 데 85%라는 높은 정확도를 얻을 수 있었습니다. 본 연구의 방법을 통해 개발한 모델은 노인들의 음성 특징 문제에 대하여 잡음 제거, 정규화 등의 데이터 전처리를 통하여 전체 연령대의 데이터로 학습된 모델에 테스트가 용이하도록 하였으며,

모델 또한 CNN과 RNN 및 softmax를 이용하여 정확하고 신뢰성 있는 음성 감정 분류를 수행할 수 있었습니다. 이러한 모델은 노인들의 정서적 요구를 이해하고 대응하는 데 있어 노인을 위한 헬스케어에 많은 기여를 할 것입니다. 결론적으로, 본 연구는 음성 기반 감정 인식을 활용하여 노인과의 원활한 상호 작용을 가능하게 하고 전반적인 사용자 경험을 향상하는 것을 목표로 합니다.

더 나아가 향후에 더 깊고 넓은 모델 구조를 사용 및 드롭아웃 또는 배치 정규화와 같은 정규화 기법을 이용하거나 하이퍼 파라미터 조정을 통해 더욱 성능을 향상시킬 수 있을 것으로 기대합니다.

References

[1] C. Corinna, V. Vapnik, “Support-vector networks.”, Machine learning, vol. 20, September 1995, pp. 273-297, doi: 10.1007/BF00994018.

[2] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding”, arXiv preprint, October 2018, doi: 10.48550/arXiv.1810.04805.

[3] K. Takuhiro, T. Kou, K. Hirokazu, S. Shogo, “iSTFTNet: Fast and lightweight mel-spectrogram vocoder incorporating inverse short-time Fourier transform”, ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 23-27, 2022, Singapore, pp. 6207-6211, doi: 10.1109/ICASSP43922.2022.9746713.

[4] R. Víctor, S. F. Ochoa, R. Hervás, “Monitoring moods in elderly people through voice processing”, Ambient Assisted Living and Daily Activities: 6th International Work-Conference, IWAAL 2014, December 2-5, 2014, Belfast, UK, pp. 139-146, doi: 10.1007/978-3-319-13105-4_22.

[5] J. Qijian, M. Xiang, W. Huang, “A speech emotion recognition method for the elderly based on feature fusion and attention mechanism”, Third International Conference on Electronics and Communication; Network and Computer Technology (ECNCT 2021), December 3-5, 2021, Harbin, China, pp. 398-403, doi: 10.1117/12.2628643.

[6] S. Izhak, M. Riley, M. Mohri, “Voice signatures.”, 2003 IEEE workshop on automatic speech recognition and understanding, November 30-December 4, 2003, VI, USA, pp. 31-36, doi: 10.1109/ASRU.2003.1318399.

[7] W. Lei, M. Chen, “Image Preference of Intelligent Voice Assistant for the Elderly Based on PAD Emotion Model”, International Conference on Human-Computer Interaction(HCII 2022), June 26-July 1, 2022, Virtual Event, pp. 425-435, doi: 10.1007/978-3-031-06050-2_30.

[8] H. Kun, D. Yu, I. Tashev, “Speech emotion recognition using deep neural network and extreme learning machine”, Interspeech 2014 : 15th Annual Conference of the International Speech Communication Association, September 14-18, 2014, Singapore, pp. 223-227, doi: 10.21437/Interspeech.2014-57.