Journal of Biomedical Engineering Research 39: 63-68 (2018)
http://dx.doi.org/10.9718/JBER.2018.39.2.63

학 술 논 문

한국어 음성을 이용한 연령 분류 딥러닝 알고리즘 기술 개발

소순원[1] · 유승민[2] · 김주영[2] · 안현준[2] · 조백환[3] · 육순현[1] · 김인영[1]

[1]한양대학교 일반대학원 생체공학과, [2]한양대학교 의생명공학전문대학원 생체의공학과
[3]성균관대학교 삼성융합의과학원 의료기기산업학과

Development of Age Classification Deep Learning Algorithm Using Korean Speech

Soonwon So[1], Sung Min You[2], Joo Young Kim[2], Hyun Jun An[2], Baek Hwan Cho[3], Sunhyun Yook[1] and In Young Kim[1]

[1,2]Department of Biomedical Engineering, Hanyang University, Seoul, Republic of Korea
[3]Department of Medical Device Management and Research, Sungkyunkwan University
(Manuscript received 12 February 2018 ; revised 28 February 2018 ; accepted 1 March 2018)

Abstract: In modern society, speech recognition technology is emerging as an important technology for identification in electronic commerce, forensics, law enforcement, and other systems. In this study, we aim to develop an age classification algorithm for extracting only MFCC(Mel Frequency Cepstral Coefficient) expressing the characteristics of speech in Korean and applying it to deep learning technology. The algorithm for extracting the 13th order MFCC from Korean data and constructing a data set, and using the artificial intelligence algorithm, deep artificial neural network, to classify males in their 20s, 30s, and 50s, and females in their 20s, 40s, and 50s. finally, our model confirmed the classification accuracy of 78.6% and 71.9% for males and females, respectively.

Key words: Deep Neural Network, Artificial Intelligence, Speech Classification, Speech Recognition, Age Classification

I. 서 론

화자 인식 기술에 있어 음성 정보를 통한 화자 인식 기술은 영상 정보나 다양한 생체 정보를 통한 화자 인식 기술에 비해 측정 장비 및 측정 방법에 있어 비교적 간편하게 화자의 성별, 연령대 등의 정보를 획득할 수 있다. 이러한 장점을 기반으로 현대 사회에서 음성 화자 인식 기술은 전자 상거래, 법의학, 법 집행 등의 시스템에서 신원을 확인하는 데 있어 중요한 기술로 부상하고 있다[1]. 또한, 음성 기술의 발달과 오디오 컨텐츠 및 전자 상거래 시스템의 지속적인 확대로 인해 화자 인식의 중요성은 더욱 증가하고 있는 추세다[1].

전 세계적으로 음성을 통하여 화자의 연령이나 성별을 분류 할 수 있는 다양한 화자 인식 연구가 진행되고 있다. 특히, 국외에서는 'Interspeech 2010'에서 제공하는 795명의 화자의 전화 통화에서 발생하는 발화 영어 음성 데이터베이스 aGender[2]를 사용하여 연령을 분류하는 연구가 다양하게 진행되고 있다. 이 데이터베이스를 이용하여 Ming Li et al.는 13세 미만을 C(children) 클래스, 14-19세를 Y(young people) 클래스, 20-54세를 A(Adults) 클래스, 55세 초과인 경우를 S(senor) 클래스로 지정하여 데이터 셋을 구성하였고, 이 데이터 셋을 이용하여 52.03%의 정확도를 확인했다[3]. 또한, Nguyen et al.은 aGender 데이터베이스를 사용하여 제공되는 4개의 연령 클래스로 나눈 데이터 셋을 구성하고, 퍼지(Fuzzy) Support Vector Machine(SVM) 모델

Corresponding Author : In Young Kim
Department of Biomedical Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Korea
Tel : +82-2220-0690
E-mail : iykim@hanyang.ac.kr
본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신 방송 연구개발사업의 일환으로 수행하였음. [10045452, 사용자 의도 인지형 멀티모달 brain-machine 인터페이스 시스템 개발]

과 다른 퍼지 멤버십 값을 적용한 퍼지 SVM 모델을 제안하여 48.61%와 48.8%의 정확도 결과를 보였다[4]. 음성을 통한 연구는 국외뿐만 아니라 국내에서도 진행되고 있다. 강우현 et al.은 음성정보기술산업지원센터(SiTEC)에서 구축한 한국어 문장 데이터셋(SiTECT Dict01 DB, SiTECT Dict02 DB)의 여성 음성 데이터를 사용하여 연령 분류 연구를 진행했다. 이 데이터셋을 이용하여 10대, 20대, 30대, 40대, 50대, 60대의 연령을 클래스(총 6개)를 나눈 데이터 셋과 10대, 20-30대, 40대 이상의 연령 클래스를 나눈 데이터 셋(총 3개)을 구성하여 인공 신경망 모델을 제안했다. 그 결과 6개의 클래스를 가진 데이터셋의 정확도는 40.29%, 3개의 클래스를 가진 데이터셋의 분류 정확도는 68.46%의 정확도를 보였다[5].

그러나 30-60% 정확도를 보이는 기존의 연구들은 전자상거래, 법의학 등 화자 식별에 있어 높은 정확도를 요구하는 실생활에 적용하기에는 무리가 있을 것으로 보이며, 또한, 한국어를 통한 높은 분류 정확도를 보이는 연령 분류 알고리즘이 제안되어 있지 못하다.

본 연구에서는 한국어 음성에서 음성의 특징을 표현하는 Mel Frequency Cepstral Coefficient(MFCC)만을 추출하여 딥러닝 기술에 적용한 연령 분류 알고리즘을 개발을 목표로 한다. Mel Frequency Cepstral Coefficient는 일정 구간에 대한 스펙트럼을 분석하여 특징을 추출하는 기법으로[6], 음성 관련 연구에서 많이 사용하는 특징점이다. 또한, 딥러닝 기술은 비선형 모듈을 구성하여 1차원적인 표현을 보다 추상적인 차원으로 표현하여 학습하는 방법으로써, 딥러닝 기술의 상위 계층은 중요한 입력 요소를 증폭시키고 관련성이 없는 변화를 억제시키는 강점이 있다[7]. 위와 같은 강점들을 이용한다면, 한국어 발화 음성 사운드를 통해 다양한 연령대 분류 정확도를 높일 수 있다.

본 연구에서는 한국어 음성 데이터에서 13차 MFCC를 추출하여 데이터셋을 구성하고, 심층 인공 신경망(DNN; Deep Neural Network)을 사용하여 20대, 30대, 50대의 남성을 분류하는 알고리즘과 20대, 40대, 50대의 여성을 분류하는 알고리즘을 제안하였다. 제안하는 모델의 성능을 비교하기 위해서 기존 연구에서 사용했던 SVM보다 성능이 좋은 랜덤 포레스트[8]의 분류 정확도와 비교하여 성능을 평가하였다.

II. 연구 방법

본 연구는 한국어 음성 파일인 '서울말 낭독체 발화 말뭉치'[9]를 이용하여 음성의 특징을 나타내는 MFCC만을 추출한 데이터셋을 구성하고 심층 인공 신경망의 하이퍼 파라미터를 변경시키며 화자의 연령을 분류하는 인공지능 모델을 학습시켰다(그림 1). 러닝 기술인 인공지능 모델은 Hold-out 학습 방법으로 학습을 진행하였고, 심층 인공 신경망의 경우, 학습을 진행하는 도중 기울기 소실(Vanishing Gradient) 문제가 야기되는 시그모이드, 탄젠트 함수 대신 비선형 함수인 ReLU(Rectified Linear Unit)을 활성화 함수(Activation Function)로 사용하여 인공지능의 구성 모듈을 변경시켜 학습시켰다[7].

1. 한국어 문장 데이터베이스

본 연구에 사용된 데이터셋은 '서울말 낭독체 발화 말뭉치'로써[9], 2002년에 개발하여 2005년도에 공개 및 분배한 음성 데이터베이스이다. 총 19개 소설에 대한 문장이 녹음되어 있으며, 20대의 남녀, 30대의 남성, 40대의 여성, 50대의 남녀, 60대 이상의 남녀 화자의 발화가 문장별로 구성되어 있고, 16bit 양자화 및 16kHZ로 표본 추출되어 있다. 이 중 파일이 손실된 부분을 제외시키고, 남성과 여성의 데

표 1. 인공지능 모델 학습에 사용된 서울말 낭독체 발화 말뭉치 화자 수; 남성은 20대, 30대, 50대, 4개 연령대, 여성은 20대, 40대, 50대 4개 연령대로 구성.
Table 1. Number of Speech Corpus Speakers used in Artificial Intelligence Model Learning; Males in their 20s, 30s, 50s, and females in their 20s, 40s, 50s.

(Unit: Person)	20s	30s	40s	50s
Male	17	20	-	11
Female	19	-	20	18

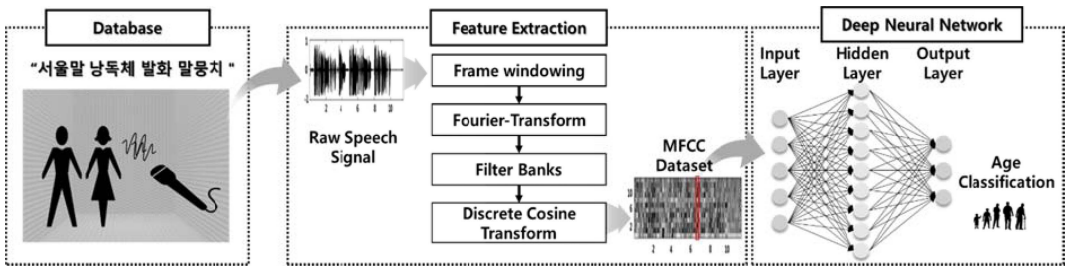

그림 1. 딥러닝 기술 기반 한국어 음성을 이용한 연령 분류 알고리즘 개발.
Fig. 1. Development of Age Classification Algorithm Using Korean Speech Based on Deep Learning.

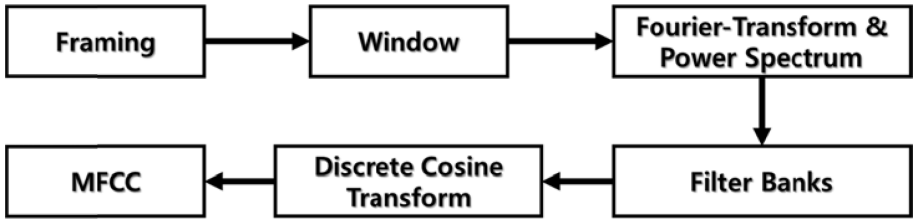

그림 2. MFCC 특징점 추출 단계.
Fig. 2. The Step of extraction of MFCC.

이터셋을 분리하여 데이터셋을 구성하고 인공지능 모델 학습에 사용되었다(표 1).

2. 특징점 추출

데이터셋은 화자의 포만트(Formant)와 음색의 특징을 나타낼 수 있는 MFCC만을 추출하여 구성하였다. MFCC를 추출하는 단계는 다음과 같다(그림 2).

먼저, 단구간 음성에 대해 고속 푸리에 변환을 하기 위해서 frame 크기를 25ms, 중복사이즈를 10ms로 설정하여 프레임을 구분한 후, 식(1)과 같이 해밍 윈도우(Hamming Window)를 적용했다. 여기서, $0 \le n \le N-1$ 이며 N은 프레임의 크기와 샘플링 레이트의 곱인 윈도우의 길이이다.

$$\omega[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{n-1}\right) \quad \text{식 (1)}$$

그리고 512개의 샘플을 갖는 고속 푸리에 변환(FFT: Fast Fourier Transform) 후, 파워 스펙트럼을 추출했다. 이렇게 추출된 파워 스펙트럼은 멜 스캐일의 삼각 필터를 적용하여 필터 뱅크를 구한다. 멜 스케일이란 사람이 소리를 인식할 수 있는 피치와 실제 주파수 사이의 맵핑으로, 본 연구에서는 식(2)를 이용하여 샘플링 레이트 16kHz의 최대 고주파 주파수의 멜 주파수를 구한 후[10],

$$m = 2595 log_{10}\left(1 + \frac{sampling\ rate}{700}\right) \quad \text{식 (2)}$$

식(3), 식(4)를 이용하여 멜 스케일을 주파수 영역으로 변경하고, 주파수 분할 구간(Bin)의 크기를 구하였다.

$$f = 700\left(10^{\frac{m}{2595}} - 1\right) \quad \text{식 (3)}$$

$$Bin = floor\left(\frac{(sample\ of\ FFT+1)*f}{sample\ rate}\right) \quad \text{식 (4)}$$

위와 같이 구한 주파수 분할 구간을 통해서 식(5)와 같이 멜 스캐일의 삼각필터를 적용하게 되면, 각 필터는 중심 주파수에서 1의 응답을 갖는 삼각형이며, 응답이 0인 두 개의 인접한 필터의 중심 주파수에 도달할 때까지 선형으로 감소하는 필터 뱅크($H_m(k)$)를 구할 수 있다. 여기서 k는 중앙점을 기준으로 좌우의 길이를 말하며, m은 삼각 필터의 개수를 말한다.

$$H_m(k) = \begin{cases} 0, & k < bin(m-1) \\ \dfrac{k - bin(m-1)}{bin(m) - bin(m-1)}, & bin(m-1) < k < bin(m) \\ 1, & k = bin(m) \\ \dfrac{bin(m+1) - k}{bin(m+1) - bin(m)}, & bin(m) < k < bin(m+1) \\ 0, & k = bin(m-1) \end{cases} \quad \text{식 (5)}$$

최종적 특징점인 MFCC를 구하기 위해선, 산출된 필터 뱅크에 이산 코사인 변환(DCT: Discreate Cosine Transform)을 적용하여 구할 수 있으며, 그림 2과 같은 단계를 거쳐 13차 MFCC를 추출하여 데이터셋[11,12]을 구성하였다.

3. 인공 신경망 모델

제안하는 인공지능 모델은 각각 남성 연령과 여성 연령을 구분하는 개별 알고리즘으로 구성되었다. 우선 남성 연령 분류 모델은 입력층과 출력층 사이에 2개의 은닉층을 가지는 심층 인공 신경망 기반의 알고리즘으로써, 각각 은닉층은 256개의 뉴런으로 구성하였다. 각 은닉층의 활성 함수로 ReLU 함수를 적용함으로써 기울기 소실에 의한 학습 지연을 최소화하였다. 그리고 2번째 은닉층 이후에 인공 신경망의 과적합을 방지하고 보편적 성능을 향상 시키기 위하여 0.5의 확률을 가지는 Dropout층을 추가하였다(그림 3(a)).

여성 연령 분류 모델의 경우, 입력층과 출력층 사이에 3개의 은닉층을 지니는 심층 인공 신경망 기반의 알고리즘으로 구성하였으며, 각각의 은닉층은 128개의 뉴런으로 구성되었다. 남성 연령분류 모델과 동일하게 각 은닉층의 활성 함수로는 ReLU 함수를 사용하였으며, 최종 은닉층 이후에 Dropout층을 추가하였다(그림 3(b)).

두 가지 모델에 공통적으로 학습에 사용된 손실함수(Loss Function)는 분류 모델에 일반적으로 사용되는 손실 함수인 범주형 교차엔트로피(Categorical Cross-entropy)를 이용하였다. 범주형 교차엔트로피는 다음 식(6)과 같이 정의되며, 여기서 $y_k^*(t)$와 $y_k(t)$는 k번째 클래스에 속한 n번째 학

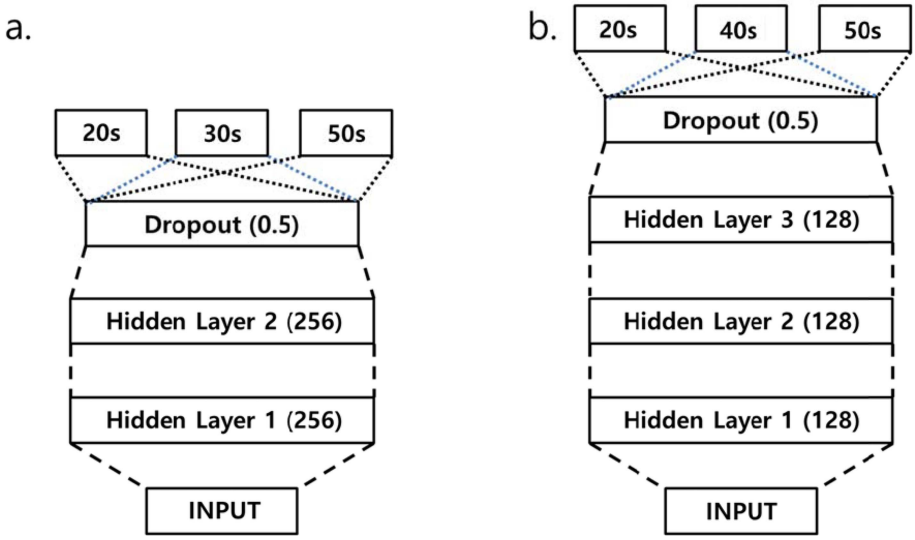

그림 3. 제안하는 연령 분류 심층 인공 신경망 구조; a. 남성 연령 분류 모델, b. 여성 연령 분류 모델.
Fig. 3. Proposed structure of age classification deep neural network; a. Male Age Classification Model, b. Female Age Classification Model.

습 데이터에 대한 네트워크의 목표(정답)과 산출된 결과(예측 값)을 의미한다.

$$L = -\sum_n \sum_k y_k^*(t)\log(y_k(t)) \quad \text{식 (6)}$$

두 가지 모델의 학습은 Mini-batch를 이용한 확률적 기울기 하강법(Stochastic Gradient Descent)를 이용하여 진행하였다. 이러한 확률적 기울기 하강법의 경우 매번 학습에서 전체 학습 데이터에 대하여 손실함수를 산출하지 않고, 학습 데이터에서 무작위로 추출된 데이터에 대해서 손실함수를 계산하고 이에 따른 오차를 업데이트를 하는 방식으로 학습이 진행된다. 이러한 확률적 기울기 하강법의 장점은 학습 데이터의 무작위 추출을 통한 난수성을 통하여, 알고리즘 손실함수가 국소 최저치(Local Minima)에 도달하여 학습이 원활이 진행되지 못하는 문제를 해결하고 종국에는 전역 최저치(Global Minima)까지 도달할 수 있다는 장점을 지닌다. 본 연구에서는 512 샘플의 크기를 가지는 Mini-batch를 구성하여 학습을 진행하였으며, 확률적 기울기 하강법을 기반으로 하는 Adam 최적화기(Optimizer)[13]을 적용하여 네트워크의 학습을 수행하였다.

남성 연령 분류 모델과 여성 연령 분류 모델은은 각각 20에폭, 40에폭을 진행하였으며, 네트워크의 효과적인 학습을 위하여 초기 10 에폭의 학습은 러닝레이트를 0.0001로 설정하여 진행하였고, 이후 이어진 10 에폭에는 초기치보다 10% 감소된 러닝레이트를 적용하였다.

본 연구에서 제안하는 모델 성능은 전체 데이터를 학습 데이터와 테스트 데이터로 구분하여 평가하는 Hold-out 학습방법을 통해서 평가하였다. 성능을 평가하기 위한 인덱스는 정확히 분류되는 정도를 나타내는 정확도로 나타냈다. 또한, 이 모델의 성능을 비교하기 위해 동일한 특징점을 사용한 랜덤 포레스트의 분류 정확도와 비교하였다.

III. 연구 결과 및 고찰

1. 결과

표2는 본 연구에서 제안하는 심층 인공 신경망 기반의 연령 분류 시스템과 전통적인 기계학습인 랜덤 포레스트의 연령 분류 시스템의 결과를 비교하고 있다.

제안하는 심층 인공 신경망 기반의 남성 연령 분류 모델은 78.6%의 분류 정확도를 나타냈고, 전통적인 기계학습인 랜덤 포레스트 모델보다 26.8% 높은 결과를 확인했다. 심층 인공 신경망 기반의 여성 연령 분류 모델의 경우 71.9%의 분류 정확도를 보였으며 랜덤 포레스트 모델보다 27.28% 높은 결과를 보였다.

표 2. 서울말 낭독체 발화 말뭉치에서의 연령 분류 모델 성능 비교.
Table 2. Performance comparison of age classification model in reading speech corpus of Seoul words.

Dataset	Accuracy (Unit :%)	
	Random forest	Deep neural network
Male (20s, 30s, 50s)	51.8	78.6
Female (20s, 40s, 50s)	44.62	71.9

표 3. 심층 인공 신경만 기반 연령 분류 모델의 오차 행렬.
Table 3. Confusion matrix of age classification model based on proposed deep neural network.

		Predicted class					Predicted class		
		20s	30s	50s			20s	40s	50s
Male actual class	20s	0.802	0.178	0.02	Female actual class	20s	0.972	0.027	0.001
	30s	0	1	0		40s	0.281	0.669	0.05
	50s	0.375	0.558	0.067		50s	0.229	0.703	0.068

표 3의 심층 인공 신경망 기반 연령 분류 모델의 오차 행렬(confusion matrix)을 보면 남성의 경우 30대의 분류 정확도가 100%로 가장 높았으며 여성의 경우 20대의 분류 정확도가 97%로 가장 높은 정확도 성능을 보였다. 반면 남성, 여성 모두 50대의 데이터가 각각 남성은 30대로 56%, 여성은 40대로 70% 분류되는 오류를 보였다.

2. 고찰

딥러닝 기술은 데이터 분류 기수에 있어 비약적인 발전을 하는데 큰 공헌을 하였다. 이러한 딥러닝 기술은 특정 분야의 지식(Domain Knowledge)에 기반하는 연구자가 설계한 다양한 알고리즘들을 이용한 특징(Handcrafted Feature)에 의존하기보다, 다수의 학습 데이터를 통해 미가공 데이터(Raw Data)가 표현하고자 하는 바를 직접적으로 학습하는 표현학습(Representative Learning)을 수행하고자 개발되었다. 다만 음성 연구에서는 미가공 데이터인 시간영역의 1차원 신호를 그대로 학습에 사용하지는 않고, 음성의 다양한 주파수적, 시간적 특징을 포함하고 있는 MFCC와 음향 에너지가 집중된 주파수 대역인 포만트(Formant)를 학습 데이터로 이용한다. 본 연구에서도 기존 음성 분류 연구에서 많이 쓰이는 13차 MFCC를 학습 데이터셋의 특징점으로 사용하였다[11,12]. 그러나 추후에는 피치나 주기 등의 운율적인 특징점[3]과 음성 에너지적인 특징점[14], 그리고 이를 통해 얻은 데이터 특징점들의 통계적인 특징점을 추출하여 데이터셋을 구성하여 인공지능 알고리즘 모델을 학습시킬 수도 있을 것이다.

제안하는 심층 인공 신경망 기반 한국어 음성을 통한 연령 분류 모델은 남성, 여성 각각 78.6%, 71.9%의 정확도로 40-60%의 기존 분류 모델에 비해 높은 정확도를 확인할 수 있었다. 단지 남성, 여성 모두 50대의 데이터가 분류가 잘 될지 않는 것을 확인 하였고 이 때문에 전체 분류 정확도를 낮추는 현상을 확인 하였다. 심층 인공 신경망은 실제 라벨과 예측된 라벨의 차이를 통한 비용을 계산하며, 코스트가 낮아지는 방향으로 학습한다. 이는 상대적으로 낮은 분포를 갖는 클래스가 코스트의 영향을 적게 끼치므로, 본 연구의 결과와 같이 나타났다고 해석되어 진다.

또한, 여성 연령 분류 모델은 남성 연령 분류 모델과 구조적 차이를 가진다. 이는 심층인공 신경망 모델이 남성과 여성 데이터셋을 학습할 때, 데이터셋을 잘 표현하기 위한 모델 최적화에 따라 발생하는 차이라 해석되어지며, 남성과 여성의 구조에 따른 실험이 추가로 진행되어 진다면, 높은 성능의 모델 제시가 가능할 것으로 예상되어 진다.

본 연구에서 사용한 서울말 낭독체 발화 말뭉치는 각 연령별 남녀 화자 20명의 음성데이터를 사용한 것으로, 이는 국외 데이터베이스인 aGender가 가지는 화자의 수보다 적으며 남성은 40대, 여성은 30대 데이터가 부재하다. 그래서 충분한 데이터 확보를 하지 못한 한계로 분류 정확도의 결과에도 영향을 끼쳤을 것이다. 그러나 추후에 전 연령대에 있어 더 많은 양의 화자 발화 음성데이터를 확보한다면 보다 높은 성능의 모델 제시가 가능할 것으로 기대되어 진다.

여러 가지 한계에도 불구하고 본 연구에서는 딥러닝 기술을 이용하여 한국어 음성만을 통한 연령 구분 모델을 제시하였고, 추후에 데이터의 추가 확보 및 모델의 고도화를 진행한다면 범죄 수사나 전자상거래 등 실생활에 응용될 수 있는 가능성을 확인하였다.

IV. 결 론

본 연구에서는 한국어 음성데이터를 이용하여 딥러닝 기반 연령 분류 알고리즘을 개발하였고, 78.6%의 남성 연령 분류 정확도와 71.9%의 여성 연령 분류 정확도를 얻었다. 기술의 성능을 비교하기 위하여 동일한 데이터셋을 이용하여 랜덤 포레스트에 적용한 결과, 본 연구에서 개발한 알고리즘이 랜덤 포레스트의 성능보다 남녀 각각 26.8%, 27.28%의 높은 정확도를 보였다.

참고문헌

[1] J.H.L. Hansen and T. Hasan, “Speaker recognition by machines and humans: A tutorial review,” IEEE Signal Proc. Mag., vol. 32, no. 6, pp. 74-99, 2015.[2] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., M¨uller, C., Narayanan, S, “The INTERSPEECH 2010 Paralinguistic

Challenge," In: Proc. INTERSPEECH 2010, Makuhari, Japan, 2010, pp. 2794-2797.
[3] M. Li, K. J. Han, and S. Narayanan, "Automatic speaker age and gender recognition using acoustic and prosodic level information fusion," Computer Speech & Language, vol. 27, no. 1, pp. 151-167, 2013.
[4] Phuoc Nguyen, Trung Le, Dat Tran, Xu Huang, and Dharmendra Sharma. "Fuzzy support vector machines for age and gender classification," In INTERSPEECH 2010, Makuhari, Japan, 2010, pp. 2806-2809.
[5] 강우현, 이강현, 강태균, 김남수. "I-벡터 특징을 이용하는 NN 기반의 화자 연령 분류,"한국통신학회 학술대회논문집, 2015, pp. 589-590.
[6] Logan, Beth. "Mel Frequency Cepstral Coefficients for Music Modeling," ISMIR, vol. 270, 2000.
[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, 2015.
[8]Katerenchuk, Denys. "Age Group Classification with Speech and Metadata Multimodality Fusion." Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers," vol. 2, 2017.
[9] 윤태진, 강윤정, "한국어 대용량발화말뭉치의 단모음분석," 말소리와 음성과학, 제6권, 제3호, 2014, pp. 139-145.
[10] Muda, L., M. Begam and I. Elamvazuthi (2010). "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques," arXiv preprint arXiv:1003.4083.
[11] D. Mahmoodi, H. Marvi, M. Taghizadeh, A. Soleimani, F. Razzazi, and M. Mahmoodi, "Age estimation based on speech features and support vector machine," in Proceedings of the 3rd Computer Science and Electronic Engineering Conference (CEEC '11), July. 2011, pp. 60-64.
[12] A. Kumar, P. Agarwal, P. Dighe, S. S. Bhiksha Raj, and K. Prahallad, "Speech Emotion Recognition by AdaBoost Algorithm and Feature Selection for Support Vector Machines," http://home.iitk.ac.in/?subhali/reports/report iptse.pdf.
[13] KINGMA, Diederik P.; BA, Jimmy. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
[14] B. D. Barkana and J. Zhou, "A new pitch-range based feature set for a speaker's age and gender classification," Appl. Acoust., vol. 98, pp. 52-61, 2015.