

멀티태스크 러닝 심층신경망을 이용한 화자인증에서의 나이 정보 활용

Utilization of age information for speaker verification using multi-task learning deep neural networks

김주호,¹ 허희수,¹ 정지원,¹ 심혜진,¹ 김승빈,¹ 유하진[†]

(Ju-ho Kim,¹ Hee-Soo Heo,¹ Jee-weon Jung,¹ Hye-jin Shim,¹ Seung-Bin Kim,¹ and Ha-Jin Yu^{1†})

¹서울시립대학교 컴퓨터학과

(Received July 10, 2019; revised August 23, 2019; accepted September 11, 2019)

초 록: 화자 간 음색의 유사성은 화자 인증 시스템의 성능을 하락시킬 수 있는 요인이다. 본 논문은 화자 인증 시스템의 일반화 성능을 향상시키기 위해, 심층신경망에 멀티태스크 러닝 기법을 적용시켜 발화자의 화자 정보와 나이 정보를 함께 학습시키는 기법을 제안한다. 멀티태스크 러닝 기법은 은닉층들이 하나의 태스크에 과적합 되지 않도록 하여 심층신경망의 일반화 성능을 향상시킨다고 알려져 있다. 하지만 심층신경망을 멀티태스크 러닝 기법으로 학습시키는 과정에서, 나이 정보에 대한 학습이 효율적으로 수행되지 않는 것을 실험적으로 확인하였다. 이와 같은 현상을 방지하기 위해, 본 논문에서는 심층신경망의 학습 과정 중 화자 식별과 나이 추정 목적 함수의 가중치를 동적으로 변경하는 기법을 제안한다. 동일 오류율을 기준으로 RSR2015 평가 데이터셋에 대해 화자 인증 성능을 평가한 결과 나이 정보를 활용하지 않은 화자 인증 시스템의 경우 6.91 %, 나이 정보를 활용한 화자 인증 시스템의 경우 6.77 %, 나이 정보를 활용한 화자 인증 시스템에 가중치 변경 기법을 적용한 경우 4.73 %의 오류율을 확인하였다.

핵심용어: 화자 인증, 멀티태스크 러닝, 심층신경망, 나이추정

ABSTRACT: The similarity in tones between speakers can lower the performance of speaker verification. To improve the performance of speaker verification systems, we propose a multi-task learning technique using deep neural network to learn speaker information and age information. Multi-task learning can improve generalization performances, because it helps deep neural networks to prevent hidden layers from overfitting into one task. However, we found in experiments that learning of age information does not work well in the process of learning the deep neural network. In order to improve the learning, we propose a method to dynamically change the objective function weights of speaker identification and age estimation in the learning process. Results show the equal error rate based on RSR2015 evaluation data set, 6.91 % for the speaker verification system without using age information, 6.77 % using age information only, and 4.73 % using age information when weight change technique was applied.

Keywords: Speaker verification, Multi-task learning, Deep neural network, Age estimation

PACS numbers: 43.60.Bf, 43.72.Bs

1. 서 론

최근, 화자 인증 시스템에 심층신경망(Deep Neural Network, DNN)을 접목시키는 다양한 연구^[1,2]들이 보고되고 있다. 이 중 심층신경망에서 마지막 은닉층의 선형 활성화값을 화자 특징으로 사용하는 기법^[3,4]

[†]Corresponding author: Ha-Jin Yu (hjyu@uos.ac.kr)
School of Computer Science, College of Engineering, University of Seoul, 163 Siripdae-ro, Dongdaemun-gu, Seoul 02504, Republic of Korea
(Tel: 82-2-6490-5697, Fax: 82-2-6490-2444)

이 널리 활용되고 있다. 위와 같은 심층신경망은 학습 시 음성데이터로부터 화자 정보만을 학습한다. 여기서 화자 정보란 해당 음성이 어떤 화자의 음성 인지를 나타내는 정보를 의미한다. 해당 방식으로 학습된 심층신경망을 화자 인증 시스템에 활용할 경우, 음색이 유사한 화자 간의 발성에 대해 낮은 신뢰도의 화자 인증 결과가 나타날 수 있다. 이와 같은 문제점을 해결하여 화자 인증 성능을 향상시키기 위해, 본 논문에서는 화자 정보 이외의 추가 정보를 활용한다. 기존 연구 중에는 성별 정보^[5]나 문장 정보^[6]를 추가적으로 활용하여 성능 향상을 확인한 바 있다.

본 논문은 기존 심층신경망 기반 화자 인증 시스템에서 나이 정보를 추가로 활용하는 기법을 제안한다. 입력된 발성으로부터 나이 정보를 활용하면 나이가 많이 나는 두 화자의 발성을 구별하기 용이하다는 장점이 있다. 구체적인 예시로서, 한 가정 내의 부자간 혹은 모녀간의 발성처럼 같은 성별이면서 음색이 유사하여 분류가 어려운 경우, 나이 정보 분석을 통해 다른 화자의 발성으로 분류가 가능하다. 그러나 음성으로 사람의 나이를 정확히 추정하는 것은 어려운 일이고, 본 연구의 목적은 화자인증 성능을 향상시키는 것이므로, 멀티태스크 심층신경망에서 화자 정보와 나이 정보를 동시에 학습시키는 방법을 사용하였다. 이때 두 목적 함수의 가중치를 동적으로 변경하여 최종적으로 화자인증의 성능을 향상시키는 기법을 제안하고 실험으로 우수성을 확인하였다.

본 논문의 II장에서는 연구에서 활용한 기존 기술의 개념을 소개한다. III장에서는 본 논문에서 제안한 기법들을 소개한다. 구체적으로, 나이 정보를 활용한 화자 인증 시스템의 구조와 인증 과정을 설명한 뒤, 실험 결과 분석을 통해 발견한 가중치 변경 기법을 소개한다. IV장은 제안한 기법들을 활용한 화자 인증 실험의 설계 및 실험 결과의 분석을 다루며, 마지막으로 V장에서 결론 및 향후 연구 계획을 보인다.

II. 기존 기술

2.1 심층신경망 기반 화자 인증 시스템

화자인증은 한 사용자가 본인의 아이디와 음성을

입력하였을 때, 음성을 비밀번호 대신 사용하여 본인 인증을 하는 것을 말한다. 이 때 한 사람의 음성을 충분히 수집하여 시스템을 학습시키는 것은 어려우므로, 일반적으로 다수의 사용자 중에 누구의 음성 인지를 구분하는 화자 식별 시스템을 기반으로 하여 화자 인증 시스템을 개발한다. 본 논문에서 활용하는 심층신경망 기반 화자 인증 시스템^[3]의 학습 및 평가 과정은 다음과 같다. 먼저 학습 데이터셋에 포함된 화자를 식별하도록 심층신경망을 학습시킨다. 학습이 완료된 심층신경망은 학습 데이터셋의 화자를 식별하도록 학습 되어 있으므로, 출력층의 노드 수는 학습 데이터셋에 포함된 화자 수와 동일하다. 즉, 학습 과정에서 각 발성의 화자 정보를 원-핫 벡터(one-hot vector)의 형태로 표현해 화자 식별이 가능하도록 한다.

하지만 이와 같이 학습된 심층신경망은 학습 데이터에 포함되지 않은 화자의 발성을 비교할 수 없다. 알려져 있지 않은 화자에 대한 발성을 비교하기 위해, 심층신경망에서 화자 식별을 수행하는 출력층을 제거한 후, 심층신경망에 발성이 입력되었을 때 마지막 은닉층의 선형 활성화 값을 화자 특징으로 사용한다. 화자 인증 수행을 위해서는, 먼저 복수의 등록 발성을 심층신경망에 입력하여 화자 특징들을 추출하고, 이를 활용하여 화자 모델을 구성한다. 이후 평가 발성이 입력될 경우, 심층신경망에 입력하여 화자 특징을 추출한 뒤, 대상 화자의 화자 모델과 코사인 유사도를 계산하여 사전에 정의된 임계값을 넘을 경우 동일 화자인 것으로 판별한다. 대상 화자들의 평가 발성들로 구성된 평가 데이터셋에 대한 코사인 유사도 점수를 기준으로 동일 오류율을 계산하고 화자 인증 시스템을 평가한다. Fig. 1은 본 논문에서 베이스라인으로 사용한 화자 인증 시스템의 학습 및 평가 과정을 나타낸다.

2.2 멀티태스크 러닝

멀티태스크 러닝 기법(Multi-Task Learning, MTL)은 하나의 심층신경망에 다수의 출력층을 사용하여 여러 태스크를 동시에 학습시키는 방법^[7]으로, 학습 과정에서 각 태스크간의 연관성을 활용하는 방법이다. 다수의 태스크가 하나의 심층신경망을 공유하

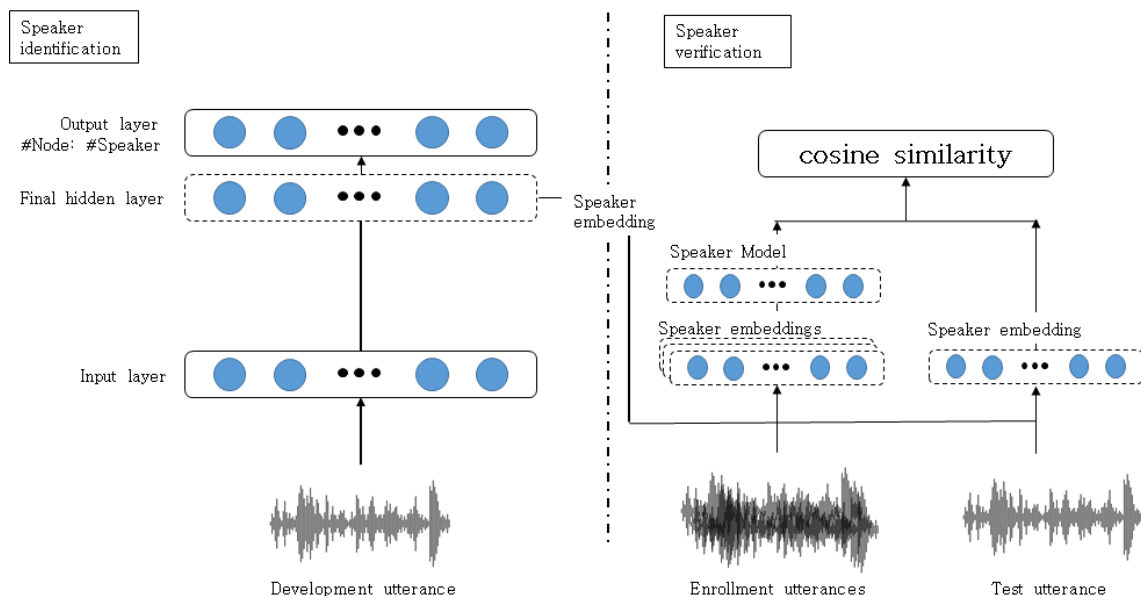


Fig. 1. Learning and evaluation process of conventional speaker verification system using DNN.

로 학습 효율이 증대 될 수 있으며, 은닉층들이 하나의 태스크에 과적합 되지 않도록 하여 심층신경망의 일반화 성능 향상을 기대할 수 있다.

III. 제안한 기법

3.1 나이 정보를 활용한 화자 인증 시스템

본 논문에서는 화자의 나이 정보를 함께 고려할 수 있는 화자 인증 시스템을 제안한다. 제안한 시스템은 화자 특징에 화자 정보뿐만 아니라, 나이 정보를 추가로 포함한다. 나이 정보를 동시에 활용하는 과정을 통해, 세대 간의 구분 능력이 생기고, 화자인증 시스템의 성능 향상을 기대하였다. 제안한 시스템의 학습 및 평가 과정은 다음과 같다. 먼저 학습 데이터셋을 이용하여 심층신경망이 화자 식별과 나이 추정을 동시에 수행하도록 학습시킨다. 그리고 제안한 방법으로 학습된 심층신경망으로부터 화자 정보와 나이 정보가 함께 포함된 특징을 추출하기 위해, 마지막 공유 은닉층의 출력을 화자 특징으로 사용한다. 마지막 공유 은닉층이란, 각 태스크별로 분리된 은닉층의 이전 은닉층을 지칭한다. 해당 방식을 이용해 등록 발성과 평가 발성으로부터 화자 특징을 추출한다. 하나의 화자마다 다수의 등록 발

성이 존재하는 경우, 화자별로 계산한 화자 특징의 평균값을 화자 모델로 구성한다. 화자 모델과 평가 발성의 화자 특징 간의 코사인 유사도를 계산한 뒤, 이를 기준으로 화자 인증을 수행한다. Fig. 2는 본 논문에서 제안한 나이 정보를 활용한 화자 인증 시스템의 학습 및 평가 과정을 보여준다. 멀티태스크 러닝 기법을 적용한 심층신경망의 목적 함수는 아래와 같이 정의된다.

$$L_{total} = W_{spk, T} L_{spk} + W_{age, T} L_{age}, \quad (1)$$

$$W_{spk, T} = \alpha, \quad W_{age, T} = \beta, \quad (2)$$

여기서 L_{total} 은 제안한 목적 함수를 지칭하며, L_{spk} 와 L_{age} 의 가중치 합으로 구성된다. L_{spk} 는 화자 식별의 목적 함수로, CCE(Categorical Cross Entropy)를 활용하였다. L_{age} 는 나이 추정의 목적 함수로, MSE(Mean Squared Error)를 활용하였다.

$W_{spk, T}$ 와 $W_{age, T}$ 는 각 태스크 별 목적 함수의 T번째 epoch에 해당하는 가중치를 의미한다. 해당 가중치들은 3.2장에서 소개하는 기법을 적용하는 경우에만 변경 되고, 그 이외의 경우는 고정되어 학습된다.

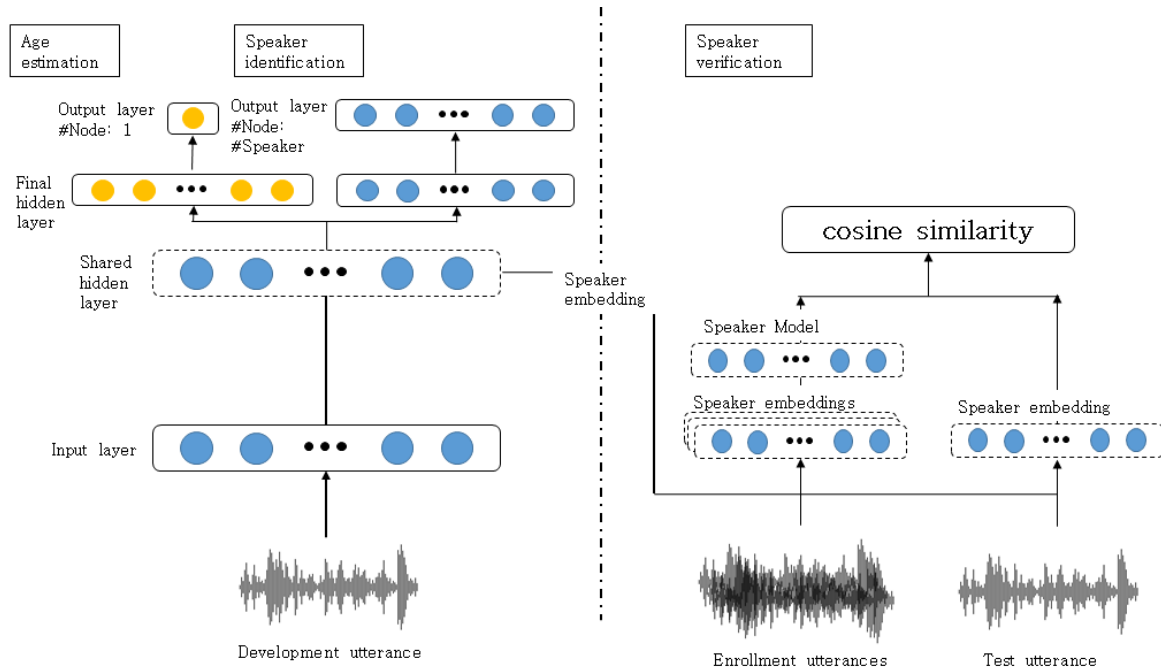


Fig. 2. Learning and evaluation process of speaker verification system using age information.

3.2 가중치 변경 기법

멀티태스크 러닝 기법을 적용해 모델을 학습시킬 때, 두 개 이상의 목적 함수를 가중합하여 새로운 목적 함수를 정의하게 된다. 본 연구를 수행하는 과정에서 두 개 이상의 목적 함수를 효과적으로 가중합하기 위해 고려해야 하는 두 가지 요소를 발견하였다.

첫 번째 요소는 다른 방식으로 정의된 목적 함수의 손실 값 크기를 고려하는 것이다. 목적 함수의 손실 값은 네트워크를 학습시키는 정도와 관련 있기 때문에, 목적 함수의 손실 값 크기를 조정하지 않는다면 의도하지 않은 방식으로 두 태스크가 학습될 수 있다. 예를 들어, 특정 목적 함수의 손실 값이 다른 목적 함수에 비해 지나치게 큰 경우, 하나의 태스크 위주로 네트워크가 학습될 수 있다. 본 연구에서는 멀티태스킹 러닝 기법을 통해, CCE로 정의된 화자 식별 목적 함수와 MSE로 정의된 나이 추정 목적 함수를 동시에 활용하였다. 두 목적 함수는 정의되는 방식이 다르기 때문에, 동일한 입력에 대해 계산된 결과의 손실 값이 크게 차이날 수 있다. Fig. 3의 그래프는 가중치 1:1로 동일하게 고정하여 모델을 학습시킬 때, CCE와 MSE의 손실 값을 나타내고 있다. MSE 손실 값이 CCE 손실 값에 비해 크기 때문에, 모

델이 나이 정보 학습에 가중 되는 것을 실험적으로 발견하였다. 따라서 3.1에서 제안한 나이 정보를 활용한 화자 인증 시스템은 화자 식별과 나이 추정의 목적 함수의 손실 값 크기를 고려하여 가중치를 다양한 값으로 고정시켜 학습을 수행하였다.

두 번째 요소는 서로 다른 태스크의 학습 속도를 고려하는 것이다. 서로 다른 태스크를 하나의 네트워크에 동시에 학습시키는 경우, 데이터의 구성이나 태스크의 특성에 의해 충분한 학습에 소요되는 시간이 다를 수 있다. 상대적으로 태스크의 난이도가 어렵고 데이터가 많은 경우, 학습에 소요되는 시간이 길다. 반면에 태스크의 난이도가 쉽고 데이터가 적은 경우, 학습에 소요되는 시간이 짧을 것이다. 따라서 본 연구에서는 멀티태스킹 러닝을 적용해 서로 다른 태스크를 동시에 학습시키는 과정에서 하나의 태스크가 먼저 학습이 될 경우, 학습이 완료된 태스크를 지속적으로 학습시키는 것이 다른 태스크의 학습에 방해할 수 있다고 가정하였다. Fig. 4의 그래프는 3.1에서 제안한 화자인증 시스템의 나이 정보 학습 동향을 나타내고 있다. 해당 그래프는 매 epoch마다 학습 데이터세트에 대한 나이 정보 학습 결과를 의미하는 MSE의 손실 값과 검증 데이터세트에

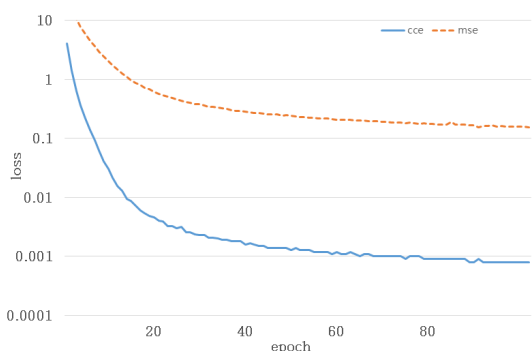


Fig. 3. Loss of speaker identification objective function and age estimation objective function. The solid line shows the loss of the speaker identification for the training data set. The dotted line shows the loss of the age estimation.

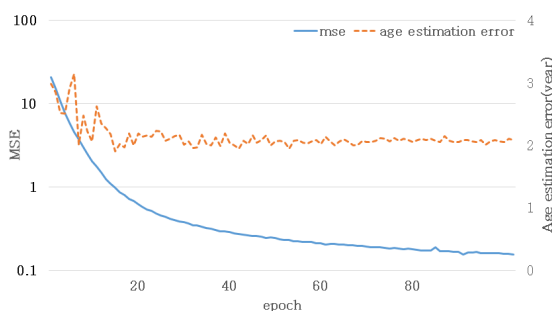


Fig. 4. Age information learning trend of speaker verification system using age information. The solid line shows the loss value of the MSE for the training data set. The dotted line shows the age estimation error using the trained DNN for the verification data set.

대한 나이 추정 결과를 의미하는 나이추정오차를 보여주고 있다. 학습초기에는 나이추정오차가 감소하였지만, 이후 일정한 값으로 수렴함을 확인할 수 있다. 반면 MSE의 손실 값은 지속적으로 감소하는 동향을 파악할 수 있다.

이는 심층신경망이 검증 데이터셋에 대한 나이 추정 성능 향상을 보이지 않음에도 불구하고 나이 정보를 지속적으로 학습하는 것으로 해석할 수 있다. 즉, 제한한 화자 인증 시스템이 지속적으로 나이 정보를 학습 하는 것은 화자 인증 성능 향상에 비효율적일 수 있을 것이라 판단하였다. 뿐만 아니라 최종적으로 수행하고자 하는 태스크는 화자 인증이기 때문에 나이 정보보다는 화자 식별에 더 큰 가중치를 부여하여 학습하는 것이 화자 인증 시스템 성능

을 향상시킬 수 있을 것이라 기대하였다. 그리하여 본 논문에서는 각 태스크의 손실 값 크기와 모델의 학습 목적을 고려하여 가중치 변경 기법을 도입하였다.

따라서 학습이 진행됨에 따라 화자 정보의 학습 비중을 증가시켜 심층신경망이 화자 정보와 나이 정보를 효과적으로 학습할 수 있도록 설계하였다. Eq. (2)를 다음과 같이 변경하여 가중치 변경 기법을 적용하였다.

$$W_{spk, T} = W_{spk, T-1} * 1.1, (W_{spk, 0} = \alpha), \quad (3)$$

$$W_{age, T} = W_{age, T-1} / 1.1, (W_{age, 0} = \beta), \quad (4)$$

IV. 실험 설계 및 결과

4.1 데이터셋

화자 인증 실험은 RSR2015 데이터셋^[8]를 사용하여 문장 중속 환경에서 수행하였다. RSR2015 데이터셋은 300명의 화자로 구성되어 있으며, 총 8시간 분량의 발성으로 구성 되어있다. 화자의 연령대는 17세에서 42세까지 분포 되어있다. Fig. 5는 성별에 따른 나이 분포를 보여 주고 있다. 실험을 위해 데이터셋을 다음과 같이 학습 및 검증, 평가 세트로 구분하여 사용하였다. 총 300명 화자 중 194명 화자의 발성을 화자식별기의 학습 데이터셋으로 사용하였고, 53명 화자의 발성을 화자 인증 시스템의 검증 데이터셋으로 사용하였다. 나머지 53명 화자의 발성은 화자 인증 시스템의 최종 평가를 위해 평가 데이터셋으로 사용하였다. 학습 시 mini-batch 구성을 위해 발성의 길이는 약 4.86s로 지정하여, 발성이 지정 길이보다 짧을 경우 동일한 발성을 중복하여 늘려서 사용하였고, 지정 길이보다 길 경우 지정 길이만큼 잘라서 사용하였다.

4.2 화자 인증 실험 설계

본 논문에서 실험에 사용한 심층신경망은 멜-필터뱅크 에너지 특징을 사용하여 원 음성으로부터 음향 특징이 추출된 벡터를 입력 값으로 사용한다. 멜-

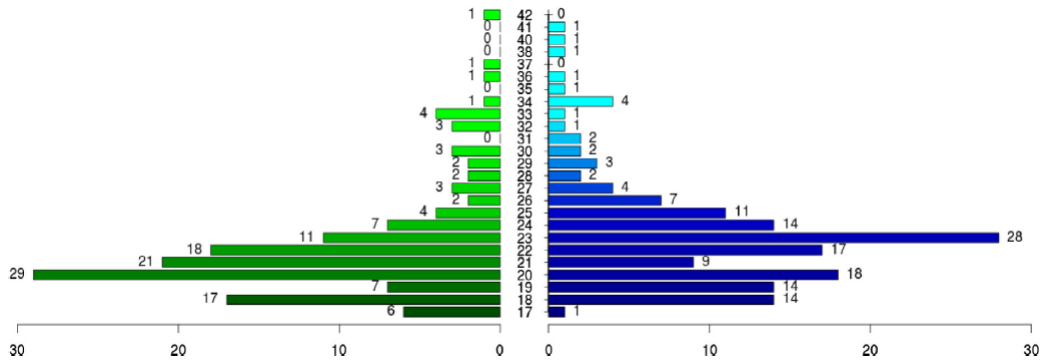


Fig. 5. Age pyramid of both male (left) and female (right) speakers of the RSR 2015 dataset.

Table 1. DNN architecture ('l' refers the length of input sequence, 'Conv' refers convolution layer, 'Res' refers residual block, 'A_pool' refers average pooling layer, 'M_pool' refers max pooling layer, 'Con' refers concatenating layer for 'A_pool' and 'M_pool'.

Layer	Output shape	Kernel size	Stride
Conv1	$l \times 40 \times 32$	7×7	1×1
Res x 3	$l \times 40 \times 32$	3×3	1×1
Res x 4	$(l/2) \times 20 \times 64$	3×3	3×2
Res x 4	$(l/4) \times 10 \times 128$	3×3	3×2
Res x 3	$(l/8) \times 5 \times 256$	3×3	3×2
A_Pool	256	Global	Global
M_Pool	256	Global	Global
Con	512		

Residual block	BatchNormalization
	LeakyReLU
	Convolution layer
	BatchNormalization
	LeakyReLU
	Convolution layer

필터 बैं크 에너지 특징을 추출하는 과정에서 pre-emphasis를 0.97, window length를 25 ms, shift size를 10 ms, 필터 개수를 40개로 구성하였다. 심층신경망은 Adam 알고리즘을 활용해 학습하였다. 이때의 학습률은 0.001로 초기화 한 뒤, 매 반복마다 0.0001만큼 감쇄시켰다. 심층신경망의 구조는 ResNet34^{9,10}을 변형시켜 사용하였다. Table 1은 본 논문에서 활용한 심층신경망의 구조를 나타낸다. 제안한 화자 인증 시스템은 베이스라인과 동일한 구조의 심층신경망을 활용하였고, 마지막 은닉층과 모델의 학습 방식을 변경해 가며 다양한 화자 인증 시스템의 성능

Table 2. Performances of the baseline and the proposed speaker verification system using the evaluation set in terms of EER (%). 'MTL' refers to the speaker verification system using age information, 'WC' refers to the proposed weight change technique. ' α ' refers to the initial loss weight of the speaker identification function. ' β ' refer to the initial loss weight of the age estimation objective function.

$\alpha : \beta$ \ Model	MTL	MTL+WC
0.5 : 0.5	7.86	6.21
0.909 : 0.09	5.57	5.1
1 : 1	7.71	5.77
10 : 1	6.77	4.73

을 동일 오류율(Equal Error Rate, EER)을 기준으로 평가하였다. 동일 화자의 3개 발성으로부터 각각 화자 특징을 추출한 뒤, 이를 평균 내어 화자 모델을 구성하였다. 한 개의 발성을 사용하는 각 trial에 대해 대상화자의 화자 모델과 평가발성으로부터 추출한 화자 특징간의 코사인 유사도를 계산하였다.

4.3 실험 결과

본 논문의 실험 결과들은 Table 2에 나타내었다. Table 2의 각 행은 학습 방식에 따른 화자 인증 성능을 동일 오류율로 나타낸다. 베이스라인은 기존 심층신경망을 활용한 화자 인증 실험을 가리킨다. MTL과 WC는 각각 나이 정보를 활용한 화자 인증 시스템과 제안한 가중치 변경 기법을 지칭한다. 제안한 시스템의 초기 손실 값 가중치를 변경해가며 실험을 수행하였다. 베이스라인과 MTL의 비교를 통해, 화자 식별과 나이 추정 목적 함수의 손실 값의 크기 차

Table 3. Comparison of the baseline and the proposed model. 'WC' refers to the proposed weight change technique.

Model	EER (%)
Baseline	6.91
MTL	6.77
MTL+WC	4.73

이를 고려하여 학습시킨 제안한 화자 인증 시스템의 성능 향상을 확인하였다.

MTL과 MTL + WC 시스템의 비교를 통해 나이 정보를 활용한 화자 인증 시스템을 학습 시킬 때 가중치 변경 기법을 적용시키는 것이 화자 인증 시스템의 성능 향상에 효과적임을 확인하였다.

V. 결론

본 논문에서는 나이 정보를 활용하여 화자 인증 시스템의 성능을 향상시키는 방안을 제안하였다. 구체적으로, 심층신경망에 멀티태스크 러닝 기법을 적용하여 화자 식별과 나이 추정을 수행하도록 학습시킨 후, 학습된 심층신경망을 활용하여 화자 인증을 수행하도록 구성하였다. 그리고 나이 정보를 더욱 효과적으로 활용하기 위해 가중치 변경 기법을 제안하였다. RSR2015 데이터셋을 이용하여 제안한 기법들의 유효성을 검증하였다. Table 3은 각각의 모델의 성능 비교를 나타내고 있다. 기존 심층신경망을 이용한 화자 인증 시스템에 손실 값 가중치를 10 : 1로 지정하여 나이 정보를 추가로 활용한 결과 동일오류율이 6.91%에서 6.77%로 감소함을 확인하였다. 추가적으로 제안한 화자 인증 시스템에 가중치 변경 기법을 적용한 결과, 동일오류율이 6.77%에서 4.73%로 더욱 감소하는 것을 확인하였다. 위 두 실험 결과를 통해 나이 정보를 추가로 활용하는 제안한 기법의 유효성을 확인하였고, 추가로 제안한 가중치 조절 기법을 통해 더욱 효과적으로 나이 정보를 활용할 수 있음을 확인하였다.

감사의 글

이 논문은 2017년도 정부(미래창조과학부)의 재

원으로 한국연구재단의 지원을 받아 수행된 기초 연구사업임(2017R1A2B4011609).

References

1. G. Heigold, I. Morenno, S. Bengio, and N. Shazeer, "End-to-end text-dependent speaker verification," Proc. IEEE ICASSP, 2-3 (2016).
2. D. Snyder, D. G. -Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust DNN embeddings for speaker recognition," Proc. IEEE ICASSP, 1-2 (2018).
3. J. W. Jung, H. S. Heo, I. H. Yang, H. J. Shim, and H. J. Yu, "A complete end-to-end speaker verification system using deep neural networks: From raw signals to verification result," Proc. IEEE ICASSP, 1-3, (2018).
4. E. Variani, X. Lei, E. McDermott, I. L. Moreno, and J. G. Dominguez, "Deep neural networks for small footprint text-dependent speaker verification," Proc. IEEE ICASSP, 1-2 (2014).
5. A. Kanervisto, V. Vestman, M. Sahidullah, V. Hautamaki, and T. Kinnunen, "Effects of gender information in text-independent and text-dependent speaker verification," Proc. IEEE ICASSP, 1-3 (2017).
6. N. Chen, Y. Qian, and K. Yu. "Multi-task learning for text-dependent speaker verification," Proc. Interspeech, 185-189 (2015).
7. R. Caruana, "Multitask learning," Machine learning, **28**, 41-75 (1997).
8. A. Larcher, K. A. Lee, B. Ma, and H. Li, "Text-dependent speaker verification: Classifiers, databases and RSR2015," Speech Communication, **60** 56-77 (2014).
9. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. the IEEE Conf. computer vision and pattern recognition, 770-778 (2016).
10. K. He, X. Zhang, S. Ren, and J. Sun, *Identity Mappings in Deep Residual Networks* (Springer, Amsterdam, 2016), pp. 1-15.

저자 약력

▶ 김 주 호 (Ju-ho Kim)

2014년 ~ 현재: 서울시립대학교 컴퓨터과 학부 학사



▶ 허 희 수 (Hee-Soo Heo)



2013년 2월 : 서울시립대학교 컴퓨터과학
부 학사
2013년~현재 : 서울시립대학교 컴퓨터과
학부 석·박사통합과정

▶ 정 지 원 (Jee-weon Jung)



2017년 2월 : 서울시립대학교 컴퓨터과학
부 학사
2017년~현재 : 서울시립대학교 컴퓨터과
학과 석·박사통합과정

▶ 심 혜 진 (Hye-jin Shim)



2017년 2월 : 광운대학교 동북아통상학부
학사
2019년 2월 : 서울시립대학교 컴퓨터과학
부 석사
2019년~현재 : 서울시립대학교 산업기술
연구소 전임연구원

▶ 김 승 빈 (Seung-Bin Kim)



2014년~현재 : 서울시립대학교 컴퓨터과
학부 학사

▶ 유 하 진 (Ha-Jin Yu)



1990년 2월 : KAIST 전산학과 학사
1992년 2월 : KAIST 전산학과 석사
1997년 2월 : KAIST 전산학과 박사
1997년~2000년 : LG 전자 전자기술원 선
임연구원
2000년~2002년 : SL2(주) 연구소장
2002년~현재 : 서울시립대학교 컴퓨터과
학부 교수