조음 장애인의 음성 인식을 위한 인공 신경망 연구

전재민, 이혁, 황성수
한동대학교 전산전자공학부
e-mail : 21400611@handong.edu

Artificial Neural Network For Recognizing of Voice Of Person With Dysarthria

Jaemin Jeon, Hyuk Lee, Sung Soo Hwang
School of Computer Science and Electrical Engineering,
Handong Global University

1. 연구 필요성 및 문제점

인공지능 기술의 발전에 따라서 음성을 통해 프로그램을 실행하는 음성인식 기술도 함께 성장하고 있다. 음성을 활용한 명령어 입력이 타이핑을 통한 명령어 입력보다 빠르고 좋은 효율을 보임에 따라 '빅스비'나 'Siri' 같은 인공지능 비서 또는 '클로바', '카카오 미니'와 같은 AI 스피커 등의 다양한 상품들이 출시되고 있다. 반면 이런 시대의 흐름 속에서 조음장애인들은 발음의 부정확함으로 인한 기술적 소외를 경험할 수 밖에 없다. 이러한 기술적 소외가 현대 기술 발전에 따라 해결해야 할 문제이다.

기존의 음성인식 기술은 인간의 음성을 획득한 후 고유의 특징을 추출해 음소단위로 인식하고 미리 학습한 음소 사전의 도움으로 음소 순서에 가장 적합한 단어를 찾는다. 하지만 조음 장애인들의 경우 음성의 특성이 모두 다르기 때문에 기존의 학습된 일반적 모델로는 음성 인식을 수행하는 데에 큰 어려움을 가진다.

따라서 본 논문은 일반적인 음소단위 인식 방법이 아닌 classification을 통한 고립 단어 인식 방법을 제안한다. 제안 시스템에서는 인식할 단어들을 미리 지정하고, 시스템을 활용하려는 조음 장애인으로부터 지정한 단어들의 음성 데이터를 획득한다. 획득한 데이터들을 통해 신경망을 학습하며, 학습된 신경망을 통해서 입력된 음성에 대한 정답을 찾아내도록 한다.

2. 연구내용과 방법

먼저 진행할 단어들을 다음과 같이 선정하였다. 단어들은 기존의 AI 스피커에서 많이 사용되는 명령어들로 선정하였다.

'뉴스, 리모컨, 소리 작게, 소리 크게, 시간, 오늘 날씨, 오늘 일정, 지니야, 클로바'

한 명의 조음 장애인으로부터 선정된 9개의 명령어에 대한 녹음 데이터를 수집하였다. 수집된 데이터를 Training set으로 학습을 진행하였다. 음성 데이터들을 Neural network에 입력하기 전에 MFCC를 통해서 전처리를 진행하였다. MFCC 방식을 통해서 음성의 feature를 추출하고 이를 통해 학습하는 방식을 취했다.

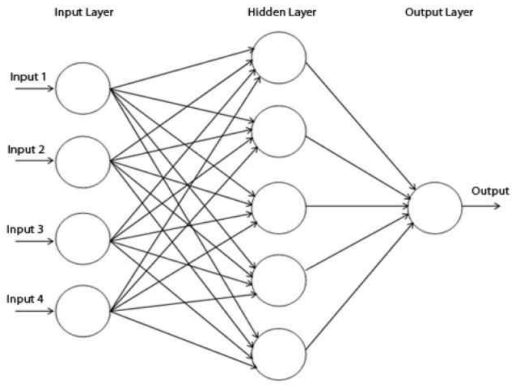

그림 1. MLP 구조

Neural network는 Multi-Layer Perceptron의 형식으로 구성하였고, output layer로 Softmax 함수를 사용하여서 각 class에 대한 결과값을 0~1 사이의 값으로 출력했다. 이를 통해서 출력된 output 중에서 가장 큰 값을 가진 class를 정답 class로 인식하고 출력하도록 구성하였다. 이를 통해서 한 명의 조음 장애인의 특성에 맞춘 음성인식 시스템을 구축하였다.

9 class classification을 위해 구축한 신경망은 비슷한 명령어에 대해서는 인식률이 낮은 모습을 보였다. (소리작게 - 소리크개, 오늘날씨 - 오늘일정의 경우) 이를 해결하기 위하여 각 pair를 따로 학습하여 two class classification을 진행하여 해결하였다. 9 class classification을 진행하여서 각각의 pari에 해당되는 경우에는 two class classification의 신경망을 한번 더 거치도록 구성하였다. 이를 통해서 비슷한 특성의 class간의 구분도 해결하였다.

3. 결론 및 향후 연구

본 논문에서는 고립 단어 인식 방법을 통한 조음 장애인의 음성 인식을 연구하였다. 기존의 음소 단위 음성 인식을 하는 음성 인식기와는 달리 조음 장애인의 음성을 단어 단위로 학습하여 인식하고 그에 대한 결과값을 표현할 수 있도록 하였다. 이를 통하여서 각각의 특성이 확연한 조음 장애인의 경우에도 자신의 목소리에 맞춘 음성인식기를 통해서 음성인식 인터페이스를 사용할 수 있도록 하였다. 아래는 기존의 AI 스피커 '클로바'와 인공지능 비서 '빅스비', '시리'와 본 논문에서 제안한 방법을 통해 개발한 프로그램의 인식률을 비교한 내용이다. 실험은 각 class별로 30번 씩을 들려주어서 일치하였는 지를 확인하였다. 아래의 일치율은 전체 test case에 대한 일치율이다.

	일치율
클로바	8.52%
빅스비	35.56%
시리	*
고립단어인식을 통한 음성인식	90.0%

(*시리는 녹음된 음성을 인식하지 못했다.)

향후 조음 장애인이 직접 본인의 목소리로 사용하고자 하는 명령어의 학습한 후 이를통해서 인공지능 비서나 AI 스피커 등의 음성 인식을 사용한다면 이전보다 일상생활에서의 편리함을 느낄 것이다.

4. 사사의 글

이 논문은 과학기술정보통신부의 소프트웨어중심대학 지원사업 (2017-0-00130)의 지원을 받아 수행하였음.

참고문헌

[1] 김영진, 김은주, 김명원. (2005). 고립단어 음성인식에서 신경망을 이용한 사용자 적응형 후처리. 한국정보과학회 학술발표논문집, 32(2), 736-738.

[2] 김연수, 김창석. (1992). Neural-HMM을 이용한 고립단어 인식 (Isolated-Word Recongnition Using Neural Network and Hidden Markov Models). 한국통신학회논문지, 17(11), 1199-1205.

[3] 이기희, 임인칠. (1995). 화자적응 신경망을 이용한 고립단어 인식 (Isolated Word Recognition Using a Speaker-Adaptive Neural Network). 전자공학회논문지-B, 32(5), 765-776.