임베디드 환경에서의 다중소리 식별 모델을 위한 경량화 기법 비교 연구

하옥균[O], 이태민[*], 성병준[*], 이창헌[*], 김성수[*]
[O]경운대학교 소프트웨어학부,
[*]경운대학교 소프트웨어학부
e-mail: okha@ikw.ac.kr, {xoals4354, wnsdl700440, amckdgjs, tjdtn1031}@naver.com

A Comparative Study of Lightweight Techniques for Multi-sound Recognition Models in Embedded Environments

Ok-kyoon Ha[O], Tae-min Lee[*], Byung-jun Sung[*], Chang-heon Lee[*], Seong-soo Kim[*]
[O]School of Software, Kyungwoon University,
[*]School of Software, Kyungwoon University

● 요 약 ●

본 논문은 딥러닝 기반의 소리 인식 모델을 기반으로 실내에서 발생하는 다양한 소리를 시각적인 정보로 제공하는 시스템을 위해 경량화된 CNN ResNet 구조의 인공지능 모델을 제시한다. 적용하는 경량화 기법은 모델의 크기와 연산량을 최적화하여 자원이 제한된 장치에서도 효율적으로 동작할 수 있도록 한다. 이를 위해 마이크로 컴퓨터나 휴대용 기기와 같은 임베디드 장치에서도 원활한 인공지능 추론을 가능하게 하는 모델을 양자화 기법을 적용한 경량화 방법들을 실험적으로 비교한다.

키워드: 딥러닝(Deep learning), 소리 인식(Sound recognition), 임베디드 환경(Embedded environments)

I. Introduction

코로나의 장기화로 줄었던 장애인 고용 비율이 점점 회복됨에 따라 기업에서도 장애인 고용 비율이 상승하고 있다. 그러나, 여전히 안전이 보장되지 않는 곳에서 근무하는 청각장애인의 경우 안전사고나 재난이 발생했을 때 비장애인에 비해 정보 습득이 느리고, 시각에 의존하기 때문에 사이렌과 같은 소리를 통한 즉각적인 상황 파악에 어려움이 있다. 이에 청각장애인의 일상생활을 보조하고 On-device 환경에서 활용할 수 있는 인공지능 기반 다중소리 인식 기술이 필요하다.

II. Background

본 논문에서는 2023년 5월 AIHub에서 개방한 극한 소음 환경 소리 데이터셋을 활용하여 청각장애인을 위한 소리 인식 모델을 개발하고, 스마트 폰 및 마이크로 컴퓨터와 같은 임베디드 환경에서도 활용 가능하도록 양자화 기법을 적용하여 경량화하는 것을 목표로 한다. 소리 식별 학습을 위해 사용한 데이터셋은 566GB의 용량과 363개의 클래스로 구분되는 대용량 데이터이며, 이를 기반으로 학습된 소리 식별 모델 역시 용량 및 처리의 한계를 보여 임베디드 환경에서 적절한 성능을 발휘하기 어렵다. 따라서, 양자화 방법 등을 고려하여 모델의 경량화가 반드시 필요하다.

III. Design and Development

1. Development Environment

인공지능 학습은 우분투 환경에서 두 개의 가상환경을 활용하였고, 활용 라이브러리는 Table 1과 같다.

Table 1. Development Environment of Sound Recognition Model

Env 1 : upgrade38		Env 2 : tf_gpu_38	
Python	3.8.16	Python	3.8.16
Numpy	1.24.3	Numpy	1.19.5
Pandas	2.0.2	tensorflow-gpu	2.4.0
Sounddevice	0.4.6	scikit-learn	1.2.2
librosa	0.10.0	matplotlib	3.6.0
matplotlib	3.7.1	spicy	1.10.1

2. Data Pre-processing

CNN 학습에서 Input Shape를 일정하게 만들기 위해 184,931개의 데이터를 2초 사이즈, Stride는 1.6초로 Frame Processing을 적용해 주었고, 소리 데이터에서 특징을 추출하기 위해 Mel-Spectrogram으로 변환해주었다. 또한, 363가지의 클래스별로 데이터들을 섞은 후 무작위로 언더샘플링(Under -sampling)을 진행하였고 학습 시 클래

스 별로 가중치를 다르게 부여하여 데이터 불균형 문제를 해결하였다.

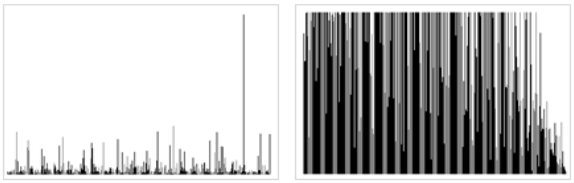

Fig. 1. Dataset distribution: the original(left), after undersampling(right)

3. Model Structure

모델은 잔여 연결과 입력과 출력 사이에 스킵 연결을 추가함으로써, 그래디언트 소실 문제를 완화하고 더 깊은 네트워크에서도 성능을 유지할 수 있는 ResNet layer-50[1] 모델을 사용하여 학습을 진행하였다.

4. Learning results

모델의 학습 결과는 Test data에 대한 예측으로 측정되었으며, 정밀도 88.8%, 재현율 87.9%, F1 Score 88.3%, 정확도 87.9%를 기록하였고, Fig. 2의 결과와 같이 이상적인 학습이 진행되었음을 보였다.

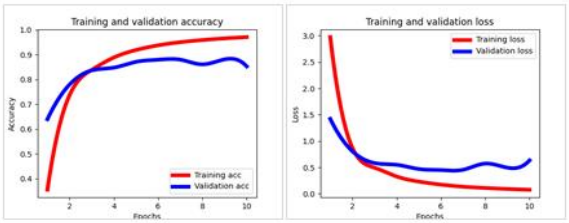

Fig. 2. Training results: Accuracy(left) and Loss(right)

5. Model quantization

학습된 인공지능 HDF5 모델을 TFLite형식으로 각각 Int, Full Int, Float16, Dynamic Range 양자화 기법을 적용하여 변환하고, 각 모델의 추론 속도와 F1-Score를 도출하여 각각의 결과를 비교하였다. Table 2는 측정된 각 경량화 방법들의 비교 결과를 보인다.

Table 2. Comparison of artificial intelligence model performance

Model Name (File format)	Quantized	Size	Inference Time(s)	F1-Score
Original Model (HDF5)	X	285.34MB	1.1509	0.890
Basic Model (TFLite)	X	94.75MB	0.5975	0.890
Int quantized (TFLite)	O	24.59MB	0.6287	0.2831
Full int quantized (TFLite)	O	24.59MB	0.6303	0.0001
Float16 quantized (TFLite)	O	47.42MB	0.5851	0.8902
Dynamic Range quantized (TFLite)	O	24.40MB	0.6308	0.8499

속도와 F1-Score 수치를 기준으로 비교한 결과, Float16과 Dynamic Range 양자화 기법은 원본 모델의 성능을 크게 훼손하지 않으면서도 모델의 크기를 획기적으로 줄여 디바이스에서의 운영 효율성을 높일 수 있을 정도의 효과가 있었다. 결과적으로 양자화가 진행되면 정밀도와 연산의 변환 과정에서 다양한 정보가 손실될 확률이 높은 데 반해 Float 16과 Dynamic Range 양자화 모델은 적은 정보의 손실로 원본 모델 성능의 유지가 가능함을 알 수 있다.

IV. Conclusion

본 논문은 청각장애인의 일상생활을 보조하기 위한 인공지능 기반 다중소리 인식 모델 개발 시 마이크로 컴퓨터 및 스마트 폰을 고려하여 소리 식별 인공지능 모델의 경량화 기법의 적용과 실험적인 비교 결과를 제시하였다. 결과에서 양자화 기법으로 경량화한 모델 중 Float16과 Dynamic Range 양자화 방법을 적용한 모델이 원본 성능을 유지하면서도 약 10배 정도 저용량으로 경량화되어 임베디드 환경에서도 사용 가능함을 보였다.

REFERENCES

[1] Chang-Hui Bae, Won-Young Cho, Hyeong-Jun Kim, Ok-Kyoon Ha, “An Experimental Comparison of CNN-based Deep Learning Algorithms for Recognition of Beauty-related Skin Disease,” Journal of the Korea Society of Computer and Information, Vol. 25, No. 12, pp. 25-34, 2020.