

다운 샘플링 기법을 활용한 뉴로모픽 아키텍처 기반 음성인식 추론모델의 성능 분석

조진성¹, 김봉재²¹충북대학교 전기·전자·정보·컴퓨터공학부 ²충북대학교 컴퓨터공학과

e-mail : {cjs97111, bjkim}@chungbuk.ac.kr

Performance analysis of speech recognition model based on neuromorphic architecture using down-sampling technique

Jinsung Cho¹, Bongjae Kim²¹Department of Electrical, Electronics, Information and Computer
Engineering, Chungbuk National University, Korea²Department of Computer Engineering, Chungbuk National
University, Korea

요약

뉴로모픽 아키텍처 기반 추론 모델은 딥러닝 기반 추론 모델에 비해 매우 낮은 전력으로 동작 가능하며, 추론 및 학습 속도가 빠르다. 때문에 뉴로모픽 아키텍처 기반 인공지능 모델을 활용하려는 여러 연구가 활발하다. 본 논문에서는 음성인식 분야에 뉴로모픽 아키텍처를 활용할 수 있는 기반을 마련하기 위하여 뉴로모픽 아키텍처 기반 음성인식 모델의 성능을 분석하였다. 전처리 기법으로 다운샘플링 기법을 사용하였으며, 실험에는 짧은 음성 데이터를 사용하였다. 실험 결과 최대 64.2 퍼센트의 정확도를 보였다. 음성 데이터에 푸리에 변환을 이용한 전처리 기법 등이 추가 적용된다면 음성인식 성능의 향상이 가능할 것으로 기대된다.

1. 서 론

뉴로모픽 아키텍처는 딥러닝을 대체할 차세대 인공지능 아키텍처로 인공지능 연구자들에게 많은 관심을 받고 있다[1]. 뉴로모픽 아키텍처는 GPU 기반의 딥러닝에 비해 매우 낮은 전력으로 동작 가능하며, GPU 기반 연산을 하는 딥러닝보다 추론 및 학습 속도가 빠르다. 뉴로모픽 아키텍처에서 동작 가능한 3세대 인공신경망 모델인 SNN(Spiking Neural Network)은 딥러닝에서 사용하는 2세대 인공지능 신경망과 달리 실제 인간의 생물학적인 뇌의 뉴런과 유사하게 동작한다[2].

본 논문에서는 뉴로모픽 아키텍처 기반 인공신경망 모델이 다양한 분야에 적용 가능함을 확인하기 위하여 뉴로모픽 아키텍처 기반 음성인식 추론모델의 성능을 측정하고 분석하였다. 실험에는 짧은 단어 음성 데이터(위, 아래, 가운데, 왼쪽, 오른쪽)를 사용하였다. 또한 뉴로모픽 아키텍처에서 사용 가능하도록 데이터를 전처리하기 위하여 다운 샘플링 기법을 사용하였다.

실험 결과 두 개의 단어를 분류할 때 최대 64.2 퍼센트의 정확도를 보였으며, 다섯 개의 단어를 분류할 때 28.6 퍼센트의 정확도를 보였다. 대체로 낮은 정확도를 보였으나 다른 데이터 전처리 방법을 적용하거나 SNN 모델을 확장하는 방법 등으로 모델 성능의 향상이 가능할 것으로 예상된다.

이후 본 논문의 구성은 다음과 같다. 2장에서 실험에 사

용한 데이터와 전처리 기법을 서술한다. 3장에서 실험 및 결과를 서술한 후 4장 결론을 끝으로 논문을 마친다.

2. 데이터셋 및 전처리 기법

2.1 데이터셋

본 논문에서는 뉴로모픽 아키텍처 기반 음성인식 모델의 성능을 분석하기 위하여 직접 녹음한 다섯 가지 한국어 단어의 음성 데이터를 사용하였다. 단어의 종류는 “위”, “아래”, “가운데”, “왼쪽”, “오른쪽”으로 총 다섯 종류이며 아래 표1은 각 단어의 학습 데이터 개수와 검증 데이터 개수를 보여준다. 학습 데이터와 검증 데이터는 4명의 남자, 1명의 여자의 목소리를 녹음하였다.

(표 1) 음성 데이터 종류 및 개수

단어	학습 데이터 개수	검증 데이터 개수
위	105	7
아래	106	7
가운데	110	7
왼쪽	107	7
오른쪽	108	7

데이터셋을 구성하는 각 음성 데이터는 WAV 파일 포맷이며 1채널로 구성되어 있다. 또한 각 음성 데이터의 샘플링 레이트(Sampling Rate)는 44,100Hz 이다.

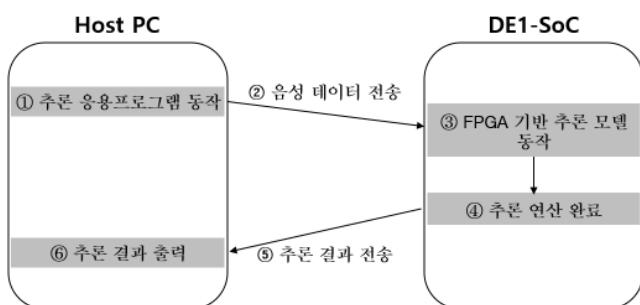
2.2 전처리 기법

본 실험에서 사용한 뉴로모픽 아키텍처 기반 추론모델은 196의 입력 크기를 사용한다. 하지만 녹음된 데이터의 샘플링 레이트는 44,100Hz로 뉴로모픽 아키텍처에서 사용하기에 적합하지 않다. 때문에 데이터를 추론모델에 적합한 형태로 전처리할 필요가 있다. 본 논문에서는 다운 샘플링기법을 활용하여 샘플링 주기를 낮추어 데이터를 축소하는 형태로 전처리하였다. 이후 뉴로모픽 아키텍처 기반 추론 연산에 용이하도록 각각의 다운 샘플링된 음성 데이터 값을 0에서 1 사이로 정규화를 진행하였다.

3. 실험 및 결과

3.1 실험 환경

본 논문에서는 뉴로모픽 아키텍처로 DE1-SoC를 사용하였다. DE1-SoC는 FPGA 기반 뉴로모픽 아키텍처 보드이며, FPGA 칩에 음성인식 추론모델이 프로그래밍되어 동작한다. DE1-SoC는 Host PC와 이더넷 네트워크를 기반으로 통신하며, Host PC에서 추론할 전처리된 음성 데이터를 전송하고, DE1-SoC에서 추론 결과를 다시 반환하는 방식으로 동작한다. 아래 그림은 뉴로모픽 아키텍처 기반 음성인식 추론모델의 동작 과정을 도식화한 것이다.



(그림 1) 뉴로모픽 아키텍처 기반 음성인식 추론모델 동작 과정

아래 표2와 표3은 실험에 사용한 Host PC의 세부 사양과 DE1-SoC의 하드웨어 사양을 보여준다.

(표 2) Host PC 세부 사양

항목	내용
OS	Ubuntu 20.04 LTS
CPU	AMD 라이젠 스레드리퍼 3960X
Memory	DDR4 64GB
GPU	RTX 2080Ti (4EA)

(표 3) DE1-SoC 세부 사양

항목	내용
CPU	Dual-core ARM Cortex-A9
Memory	DDR3 1GB
Network Interface	10/100/1000 Megabyte Ethernet
Number of Neurons	16k

3.2 실험 결과

본 논문의 실험에서는 인식하는 단어 종류 및 조합에 따른 정확도를 측정하기 위하여 다양한 실험을 진행하였다. 두 종류의 단어를 분류하는 경우, 세 종류의 단어를 분류하는 경우, 그리고 다섯 종류의 단어를 분류하는 경우의 정확도를 각각 측정하였다. 두 종류의 단어를 분류하는 경우 글자 수가 모델의 성능에 미치는 영향을 확인하기 위하여 다양한 단어 조합을 사용하여 모델의 성능을 측정하였다. 표4는 인식 단어 조합에 따른 정확도를 보여준다.

(표 4) 인식 단어 조합에 따른 모델 성능

인식 단어	정확도 (%)
위, 아래	50.0
위, 가운데	42.9
왼쪽, 오른쪽	64.2
위, 오른쪽	50.0
위, 아래, 가운데	33.3
위, 아래, 가운데, 왼쪽, 오른쪽	28.6

두 종류의 단어를 인식하는 경우 “왼쪽”, “오른쪽”을 분류할 때 가장 64.2 퍼센트로 가장 높은 성능을 보였다. 세 종류와 다섯 종류의 단어를 분류하는 경우 각각 33.3 퍼센트 28.6 퍼센트로 낮은 성능을 보였다. 인식하는 단어의 종류가 많아질수록 모델의 성능이 낮아지는 경향을 보였으며, 인식 단어의 글자 수는 모델의 성능에 큰 영향을 미치지 않았다.

4. 결론

본 논문의 실험 결과 두 종류 단어 분류에서 64.2 퍼센트, 세 종류 단어 분류에서 33.3 퍼센트, 다섯 종류 단어 분류에서 28.6 퍼센트의 정확도를 보였다. 실험 결과 대체로 낮은 정확도를 보였다.

향후 연구에서는 다양한 음성 데이터 전처리 방법을 사용하여 모델의 성능을 향상시킬 방안을 마련할 것이다. 다양한 실험을 통하여 뉴로모픽 아키텍처 기반 음성인식 모델의 성능을 향상시킬 수 있을 것으로 기대된다.

Acknowledgment

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2019-0-00708, 뉴로모픽 아키텍처 기반 자율형 IoT 응용 통합개발환경).

참고 문헌

- [1] Oh, K. I., et al., "Trend of AI Neuromorphic Semiconductor Technology", Electronics and Telecommunications Trends. Vol.35 No.3, 2020. pp.76–84.
- [2] Park, Sangmin, Junyoung Heo. "Conversion Tools of Spiking Deep Neural Network based on ONNX", The Journal of The Institute of Internet, Broadcasting and Communication. Vol.20 No.2, 2020. pp.165–170.