딥러닝심화 프로젝트

음성인식을 통한 휴대용 장비 제어 시스템

박길순
충북대학교 산업인공지능연구센터

INDUSTRIAL AI RESEARCH CENTER

CONTENTS

프로젝트 주제 및 소개

INDUSTRIAL AI RESEARCH CENTER

■ 프로젝트 주제

간단한 음성 명령어를 통한 휴대용 장비 제어 시스템 개발

■ 프로젝트 배경 및 필요성

1) 프로젝트 배경

- 산업현장이나 위험상황에서 사용되는 휴대용 측정장비들의 경우 대부분 버튼 조작식으로 장비를 제어하고 있음
- 현장에 따라서 장갑을 끼거나 양손으로 다른 장비를 조작하는 일이 있음
- 딥러닝 음성 인식 모델이 소형 장비에서도 활용 가능한 수준으로 경량화됨

2) 프로젝트 필요성

- PPE 착용 상황이나 현장 제약으로 손을 자유롭게 사용하지 못하는 상황에서 장비 제어 필요
- 현장 작업자의 편의성 증대
- 휴대용 장비에 스마트 제어 기능을 추가하여 차세대 경쟁력 확보

학습내용과의 연계성

■ 머신러닝 이론 적용

√ 딥러닝 모델 설계 : CNN기반 음성인식 모델 아키텍처 구현

√ 특징추출 : MFCC(Mel-Frequency Cepstral Coefficients) 기법을 통한 음성 신호 전처리

√ 모델 훈련: TensorFlow/Keras를 이용한 지도학습 기반 분류 모델 학습

■ 신호처리 기법

√ 필터링 기법: Mel 필터뱅크를 통한 인간 청각 특성 반영

√ FFT를 통한 시간/주파수 분석

■ 임베디드 시스템 최적화

√ 메모리 관리: Arduino Nano RP2040의 SRAM 제약내 효율적 메모리 사용

√ 학습모델 경량화

방법 및 구현 (Methodology & Implementation)

INDUSTRIAL AI RESEARCH CENTER

■ 시스템 구성

1) 하드웨어

- Arduino Nano RP2040 CONNECT

2) 데이터셋

- Google Speech Commands Dataset v0.02

3) 명령어

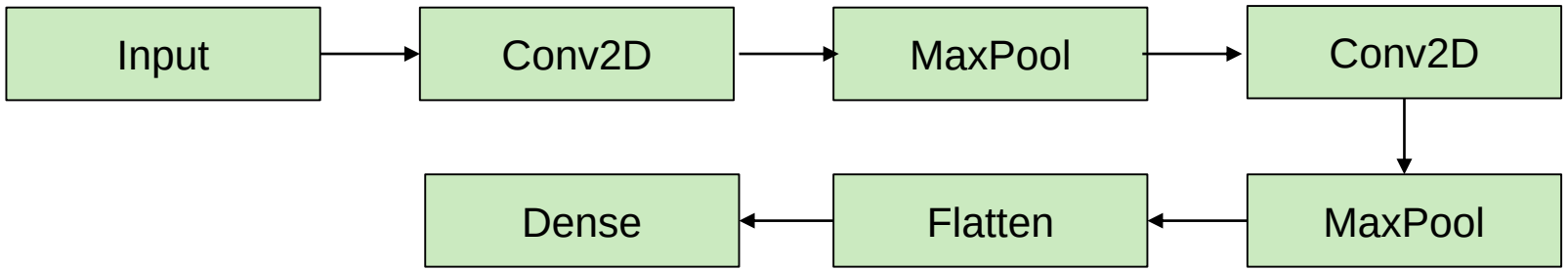
- "marvin" : 내장LED 0.2초간격 5회 깜빡임, LED on, 대기상태
- "on" : LED 1초간격 1회 깜빡임
- "go" : LED 1초간격 2회 깜빡임
- "stop" : LED 1초간격 3회 깜빡임
- "down" : LED 1초간격 4회 깜빡임
- "happy" : LED off, 슬립모드
- 명령어당 약 2000개 샘플

■ 전처리 과정

√ 오디오 정규화, MFCC 특징 추출, 데이터 증강(노이즈, 시간추가)

실험 구성 (Experiment Settings)

■ 네트워크 아키텍처



■ 훈련설정

1) Optimizer : Adam(learning rate : 0.001)
2) Loss Function : Categorical Crossentropy
3) Batch Size : 32
4) Epochs : 30

결과 및 분석 (Results & Analysis)

INDUSTRIAL AI RESEARCH CENTER

■ 모델학습

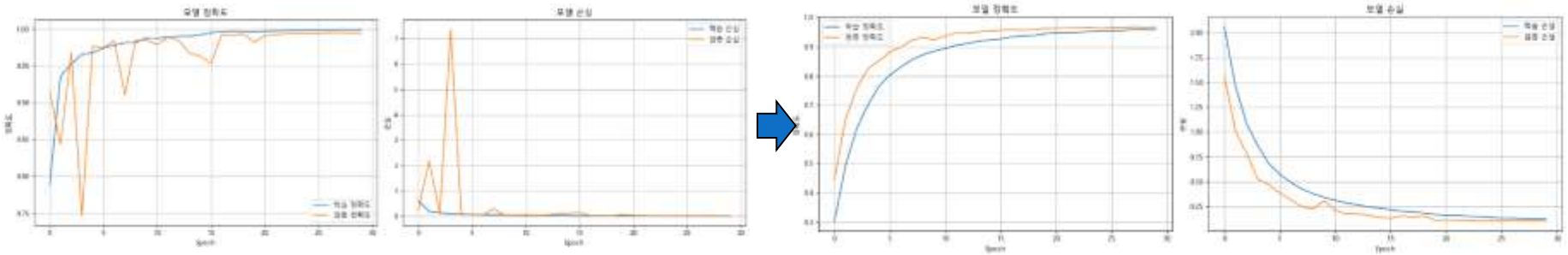
```
Epoch 29/30
1157/1157 ━━━━━━━━━━━━━━━━━━━━ 0s 1s/step - accuracy: 0.9990 - loss: 0.0033
Epoch 29: val_accuracy did not improve from 0.99507
1157/1157 ━━━━━━━━━━━━━━━━━━━━ 1349s 1s/step - accuracy: 0.9990 - loss: 0.0033 - val_accuracy: 0.9949 - val_loss: 0.0161 - learning_rate: 1.2500e-04
Epoch 30/30
1157/1157 ━━━━━━━━━━━━━━━━━━━━ 0s 1s/step - accuracy: 0.9994 - loss: 0.0023
Epoch 30: val_accuracy did not improve from 0.99507

Epoch 30: ReduceLROnPlateau reducing learning rate to 6.25000029685907e-05.
1157/1157 ━━━━━━━━━━━━━━━━━━━━ 1295s 1s/step - accuracy: 0.9994 - loss: 0.0023 - val_accuracy: 0.9946 - val_loss: 0.0172 - learning_rate: 1.2500e-04

학습 결과가 'training_history.png'로 저장되었습니다.

학습 결과 분석:
- 전체 학습 epoch 수: 30
- 최적의 epoch: 24
- 최고 검증 정확도: 0.9951
- 최종 검증 정확도: 0.9946
* Early stopping이 작동하여 과적합을 방지했습니다.
248/248 ━━━━━━━━━━━━━━━━━━━━ 40s 160ms/step - accuracy: 0.9913 - loss: 0.0334

테스트 정확도: 0.9928
모델이 'speech_recognition_model.keras'로 저장되었습니다.
모델 정보가 'model_info.json'에 저장되었습니다.
```



결과 및 분석 (Results & Analysis)

INDUSTRIAL AI RESEARCH CENTER

■ 모델검증(핸드폰녹음파일 사용)

No	녹음파일	녹음명령어	예측명령어	신뢰도	marvin	happy	on	down	stop	go
1	음성 250611_205559m.m4a	marvin	marvin	0.999	0.999	0.001	0	0	0	0
2	음성 250611_205637m.m4a	marvin	marvin	1	1	0	0	0	0	0
3	음성 250611_205649o.m4a	on	on	0.9355	0.0484	0.0002	0.9355	0.0083	0.0055	0.0022
4	음성 250611_205700o.m4a	on	on	0.9858	0.007	0.0002	0.9858	0.0009	0.0005	0.0056
5	음성 250611_205709g.m4a	go	go	0.9989	0.0001	0.0003	0.0003	0.0001	0.0004	0.9989
6	음성 250611_205740s.m4a	stop	stop	0.9962	0	0.0038	0	0	0.9962	0
7	음성 250611_205749s.m4a	stop	stop	0.9178	0	0.0821	0.0001	0	0.9178	0
8	음성 250611_205806d.m4a	down	down	0.8684	0.0011	0.0002	0.0039	0.8684	0.0329	0.0936
9	음성 250611_205816d.m4a	down	down	0.912	0	0	0	0.912	0.0871	0.0008
10	음성 250611_205825h.m4a	happy	happy	0.9992	0.0001	0.9992	0	0	0.0002	0.0005
11	음성 250611_205841h.m4a	happy	happy	0.9998	0	0.9998	0	0	0.0001	0.0001
12	음성 250611_215846g.m4a	go	go	0.999	0	0.0001	0.0003	0.0003	0.0003	0.999

CHUNGBUK NATIONAL UNIVERSITY

결과 및 분석 (Results & Analysis)

■ 모델 최적화

√ TensorFlow Lite 모델로 변환

√ 양자화(Quantization) 적용

√ 가중치 압축: Arduino 메모리 제약 고려

```
모델 최적화 완료:
- 원본 모델 크기: 24621912 bytes
- 최적화된 모델 크기: 4077656 bytes
- 저장 위치: model_optimized
```

```
예측된 명령어 분포:
  - marvin: 3개 (25.0%)
  - on: 0개 (0.0%)
  - go: 7개 (58.3%)
  - stop: 2개 (16.7%)
  - down: 0개 (0.0%)
  - happy: 0개 (0.0%)

신뢰도 분석:
  - 평균: 0.2398
  - 최대: 0.2697
  - 최소: 0.2107
```

```
예측된 명령어 분포:
  - marvin: 2개 (16.7%)
  - on: 2개 (16.7%)
  - go: 3개 (25.0%)
  - stop: 1개 (8.3%)
  - down: 0개 (0.0%)
  - happy: 4개 (33.3%)

신뢰도 분석:
  - 평균: 0.7871
  - 최대: 0.9961
  - 최소: 0.3203
```

결과 및 분석 (Results & Analysis)

INDUSTRIAL AI RESEARCH CENTER

■ 모델 경량화

```
예측된 명령어 분포:
  - marvin: 3개 (25.0%)
  - on: 0개 (0.0%)
  - go: 7개 (58.3%)
  - stop: 2개 (16.7%)
  - down: 0개 (0.0%)
  - happy: 0개 (0.0%)

신뢰도 분석:
  - 평균: 0.2398
  - 최대: 0.2697
  - 최소: 0.2107
```

```
예측된 명령어 분포:
  - marvin: 2개 (16.7%)
  - on: 2개 (16.7%)
  - go: 3개 (25.0%)
  - stop: 1개 (8.3%)
  - down: 0개 (0.0%)
  - happy: 4개 (33.3%)

신뢰도 분석:
  - 평균: 0.7871
  - 최대: 0.9961
  - 최소: 0.3203
```

No	녹음파일	녹음명령어	예측명령어	신뢰도	marvin	happy	on	down	stop	go
1	음성 250611_205559m.m4a	marvin	marvin	0.9609	0.9609	0.0352	0.0000	0.0000	0.0000	0.0000
2	음성 250611_205637m.m4a	marvin	marvin	0.9961	0.9961	0.0000	0.0000	0.0000	0.0000	0.0000
3	음성 250611_205649o.m4a	on	on	0.8789	0.0742	0.0352	0.8789	0.0078	0.0039	0.0039
4	음성 250611_205700o.m4a	on	on	0.3828	0.0742	0.1680	0.3828	0.1680	0.0352	0.1680
5	음성 250611_205709g.m4a	go	go	0.9570	0.0000	0.0078	0.0156	0.0156	0.0039	0.9570
6	음성 250611_205740s.m4a	stop	happy	0.8359	0.0000	0.8359	0.0000	0.0000	0.1641	0.0000
7	음성 250611_205749s.m4a	stop	happy	0.6914	0.0000	0.6914	0.0000	0.0000	0.3086	0.0000
8	음성 250611_205806d.m4a	down	go	0.3203	0.0039	0.0039	0.0273	0.3203	0.3203	0.3203
9	음성 250611_205816d.m4a	down	stop	0.8477	0.0000	0.0742	0.0000	0.0742	0.8477	0.0039
10	음성 250611_205825h.m4a	happy	happy	0.9961	0.0000	0.9961	0.0000	0.0000	0.0000	0.0000
11	음성 250611_205841h.m4a	happy	happy	0.9961	0.0000	0.9961	0.0000	0.0000	0.0000	0.0000
12	음성 250611_215846g.m4a	go	go	0.5820	0.0234	0.0234	0.2578	0.1133	0.0039	0.5820

결과 및 분석 (Results & Analysis)

INDUSTRIAL AI RESEARCH CENTER

■ 모델 경량화 개선

√ 클래스별 균형 조정

- stop과 down 명령어에 2배 가중치 적용
- 각 클래스별로 균등한 대표 샘플 생성

√ 혼합 정밀도 양자화

- 중요한 특징 추출 레이어는 FP16 유지
- 분류 레이어만 INT8로 양자화

```
경량화 모델 테스트 결과:
총 테스트 파일: 12

예측된 명령어 분포:
  - marvin: 3개 (25.0%)
  - on: 1개 (8.3%)
  - go: 2개 (16.7%)
  - stop: 2개 (16.7%)
  - down: 2개 (16.7%)
  - happy: 2개 (16.7%)

신뢰도 분석:
  - 평균: 0.9251
  - 최대: 0.9961
  - 최소: 0.6484
```

No	녹음파일	녹음명령어	예측명령어	신뢰도	marvin	happy	on	down	stop	go
1	음성 250611_205559m.m4a	marvin	marvin	0.9609 0.9961	0.9609 0.9961	0.0352 0.0000	0.0000	0.0000	0.0000	0.0000
2	음성 250611_205637m.m4a	marvin	marvin	0.9609 0.9961	0.9609 0.9961	0.0352 0.0000	0.0000	0.0000	0.0000	0.0000
3	음성 250611_205649o.m4a	on	marvin	0.8789 0.6484	0.0742 0.6484	0.0352 0.0117	0.8789 0.2891	0.0078 0.0273	0.0039	0.0039 0.0273
4	음성 250611_205700o.m4a	on	on	0.3828 0.9375	0.0742 0.0391	0.1680 0.0039	0.3828 0.9375	0.1680 0.0039	0.0352 0.0000	0.1680 0.0156
5	음성 250611_205709g.m4a	go	go	0.9570 0.9796	0.0000	0.0078	0.0156	0.0156	0.0039	0.9570 0.9796
6	음성 250611_205740s.m4a	stop	stop	0.8359 0.9844	0.0000	0.8359 0.0156	0.0000	0.0000	0.1641 0.9844	0.0000
7	음성 250611_205749s.m4a	stop	stop	0.6914 0.8320	0.0000	0.6914 0.1680	0.0000	0.0000	0.3086 0.8320	0.0000
8	음성 250611_205806d.m4a	down	down	0.3203 0.8281	0.0039	0.0039	0.0273 0.0156	0.3203 0.8281	0.3203 0.0742	0.3203 0.0742
9	음성 250611_205816d.m4a	down	down	0.8477 0.9922	0.0000	0.0742	0.0000	0.0742 0.9922	0.8477 0.0078	0.0039
10	음성 250611_205825h.m4a	happy	happy	0.9961	0.0000	0.9961	0.0000	0.0000	0.0000	0.0000
11	음성 250611_205841h.m4a	happy	happy	0.9961	0.0000	0.9961	0.0000	0.0000	0.0000	0.0000
12	음성 250611_215846g.m4a	go	go	0.5820 0.9180	0.0234 0.0156	0.0234	0.2578 0.0352	0.1133	0.0039	0.5820 0.9180

결과 및 분석 (Results & Analysis)

INDUSTRIAL AI RESEARCH CENTER

■ 아두이노 모델 경량화

```
PS D:\CursorAI\Sounddetect\Sounddetect> python model_extreme_compression.py
원본 모델 크기: 2057232 bytes (2009.02 KB)
INFO: Created TensorFlow Lite XNNPACK delegate for CPU.
입력 형태: [  1 124 129   1]
출력 형태: [1 6]
Arduino 헤더 파일 생성: ultra_compressed_model.h
압축된 모델 크기: 514308 bytes
```

감사합니다

Q&A