주제분석

너의 목소리가 보여

1 딥러닝팀 이승우 안세현 이수정 전효림 홍지우

목차

01

복습 및 정리

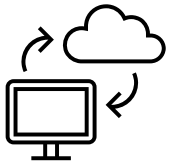
01 복습 및 정리

1주차 복습

딥러닝팀의 1주차

음성을 바탕으로 성별, 나이, 지역 유추하는 모델 구현

개발환경



서버에서 주피터 노트북을 실행해 서버를

연 후 포트 포워딩을 통해 로컬에서 직접 접속

데이터

여러 개의 JSON 파일 하나의 CSV로 변환

전처리를 거쳐 Feature engineering

모델링

MFCC / Mel Spectrogram을 이용한

CNN 모델과 RNN 모델 (4가지)

01 복습 및 정리

1주차 복습

딥러닝팀의 1주차

음성을 바탕으로 성별, 나이, 지역 유추하는 모델 구현

개발환경

서버에서 주피터 노트북을 실행해 서버를

연 후 포트 포워딩을 통해 로컬에서 직접 접속

데이터

여러 개의 JSON 파일 하나의 CSV로 변환

전처리를 거쳐 Feature engineering

모델링

MFCC / Mel Spectrogram을 이용한

CNN 모델과 RNN 모델 (4가지)

DeepVoice

01 복습 및 정리

1주차 복습

딥러닝팀의 1주차

음성을 바탕으로 성별, 나이, 지역 유추하는 모델 구현

개발환경

서버에서 주피터 노트북을 실행해 서버를

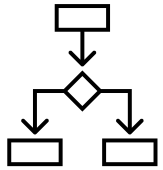
연 후 포트 포워딩을 통해 로컬에서 직접 접속

데이터

여러 개의 JSON 파일 하나의 CSV로 변환

전처리를 거쳐 Feature engineering

모델링



MFCC / Mel Spectrogram을 이용한

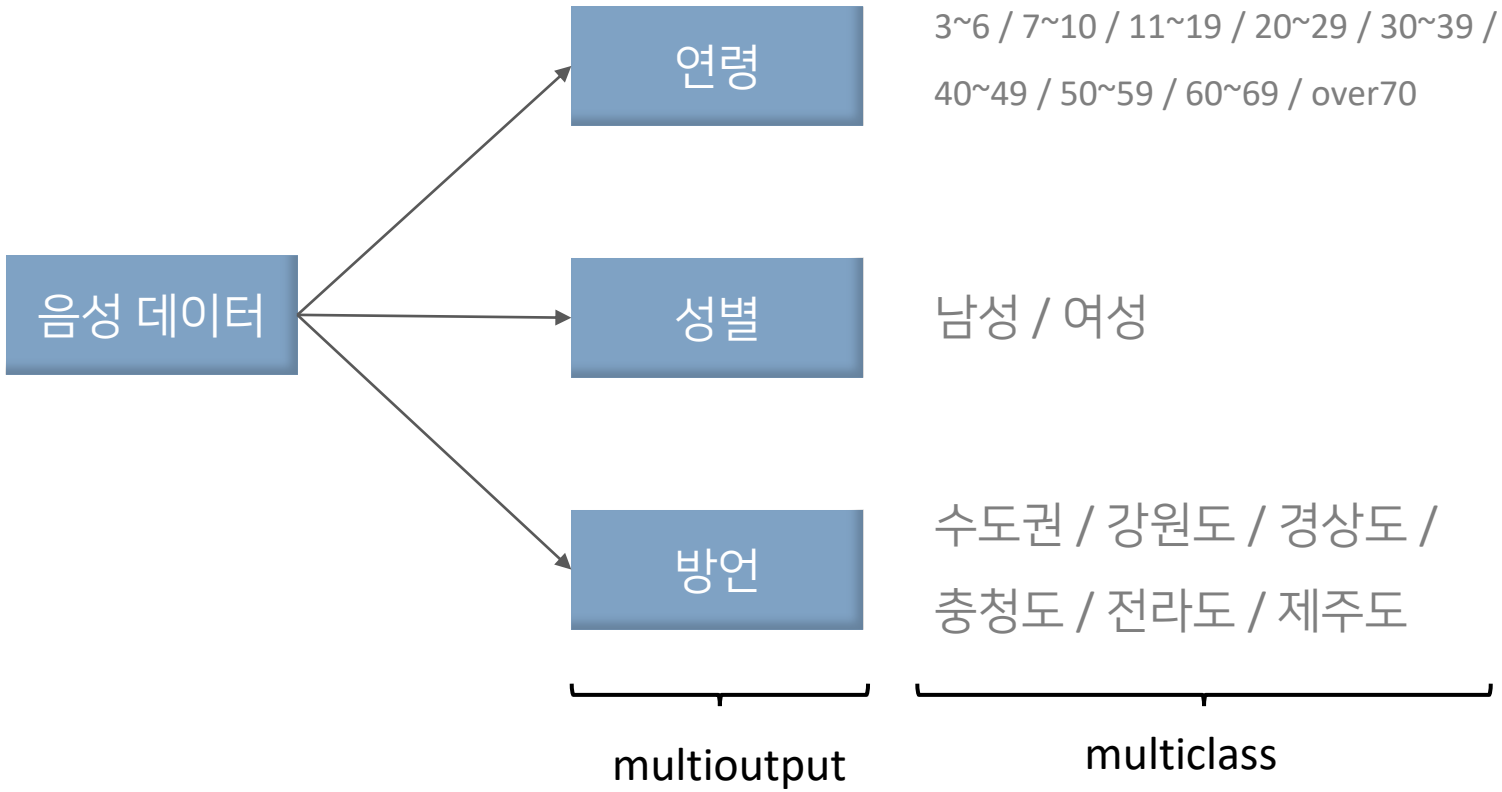
CNN 모델과 RNN 모델 (4가지)

DeepVoice

01 복습 및 정리

정리

multiclass-multioutput classification



01 복습 및 정리

정리

multiclass classification

	예측 label의 수	lable내 원소의 수
multiclass classification	1	>2
multilabel classification	>1	2 (0 or 1)
multiclass-multioutput classification	>1	>2
multioutput regression	>1	continuous

위 사진에 있는 동물의 종류를 선택해주세요

고양이 | 강아지 | 호랑이

01 복습 및 정리

정리

multilabel classification

	예측 label의 수	lable내 원소의 수
multiclass classification	1	>2
multilabel classification	>1	2 (0 or 1)
multiclass-multioutput classification	>1	>2
multioutput regression	>1	continuous

위 사진에 있는 강아지의 품종을 모두 선택해주세요

웰시코기 | 비글 | 요크셔테리어

01 복습 및 정리

정리

multiclass-multioutput classification

	예측 label의 수	lable내 원소의 수
multiclass classification	1	>2
multilabel classification	>1	2 (0 or 1)
multiclass-multioutput classification	>1	>2
multioutput regression	>1	continuous

위 사진에 있는 동물의 종류와 색은 ?

종류	고양이	강아지	호랑이
색상	회색	초록색	흰색

01 복습 및 정리

정리

multioutput regression

	예측 label의 수	lable내 원소의 수
multiclass classification	1	>2
multilabel classification	>1	2 (0 or 1)
multiclass-multioutput classification	>1	>2
multioutput regression	>1	continuous

위 인물의 신장과 체중은 ?

신장 — 185

체중 — 74

01 복습 및 정리

모델링 정리

모델 목록

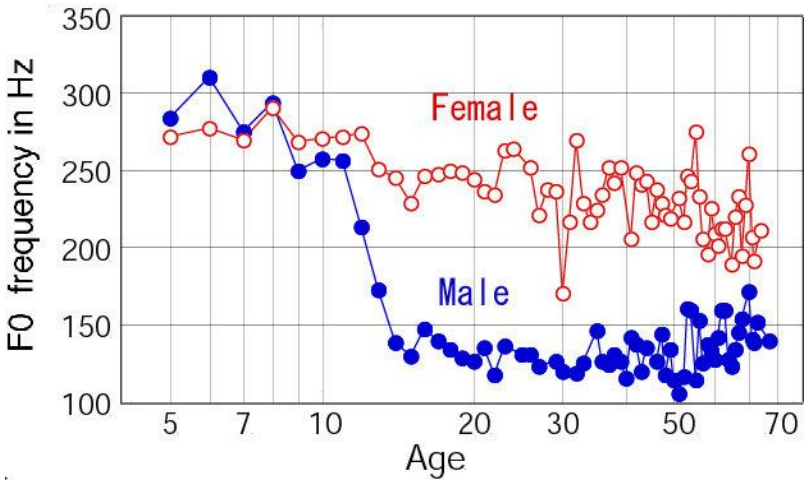
	성별	연령	방언
① Logistic Regression	○		
② CNN	○	○	○
③ RNN	○	○	○
④ CNN + LSTM	○	○	○
⑤ CNN for Raw Waveform	○	○	○

01 복습 및 정리

모델링 정리

Logistic Regression

연령, 성별에 따른 F0 주파수



Acoustic characteristics of Japanese vowels
- Tatsuya Hirahara, Reiko Akahane-Yamada (2004)

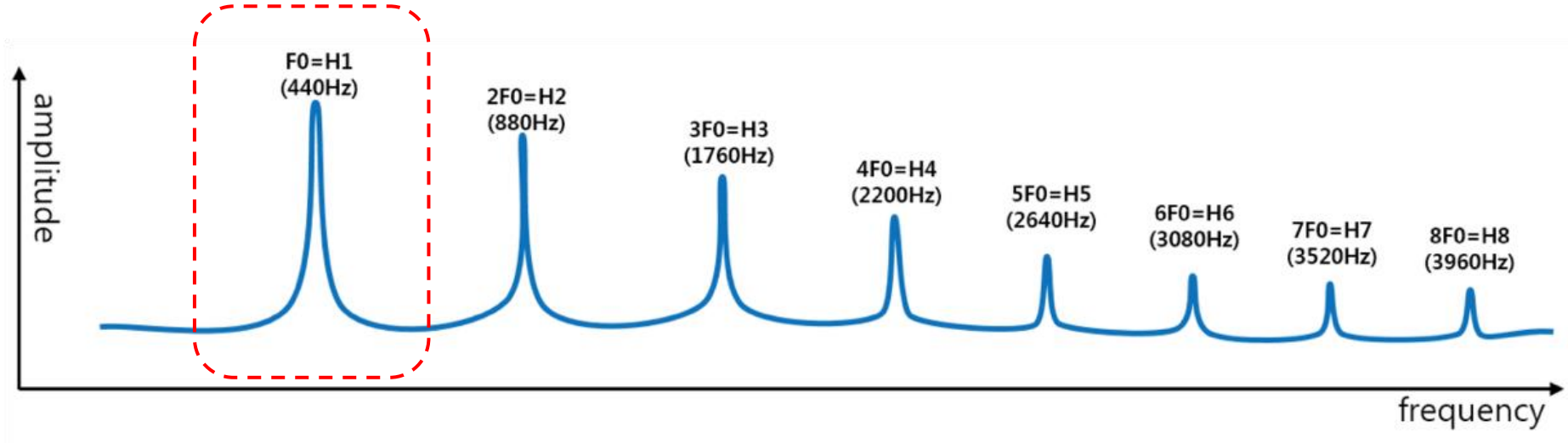
성별에 따라 F0 주파수에서 확연한 차이

성별 분류를 위한 Logistic Regression의
입력으로 F0 주파수 활용

01 복습 및 정리

모델링 정리

Logistic Regression



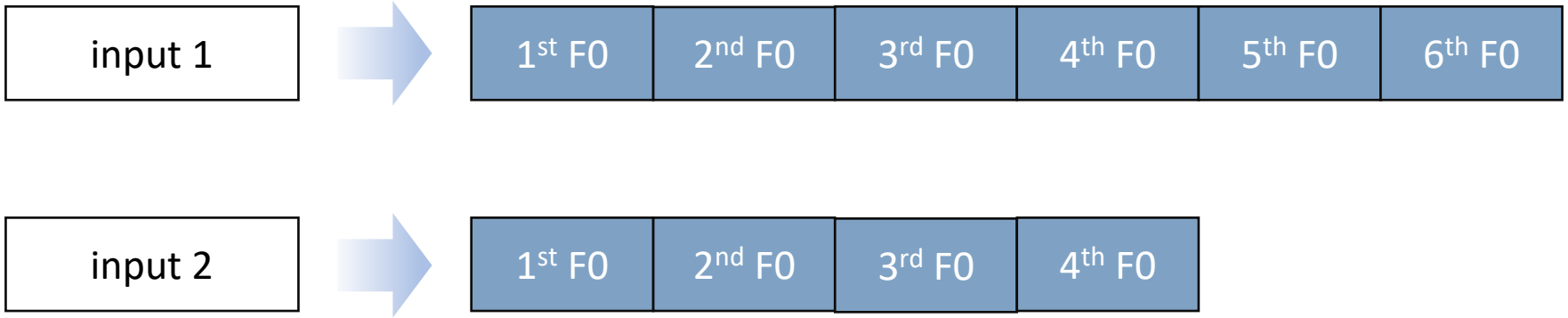
F0 주파수란 기본 주파수(Fundamental Frequency)로,

음성 신호를 구성하는 여러 주파수의 최대 공약수

01 복습 및 정리

모델링 정리

Logistic Regression

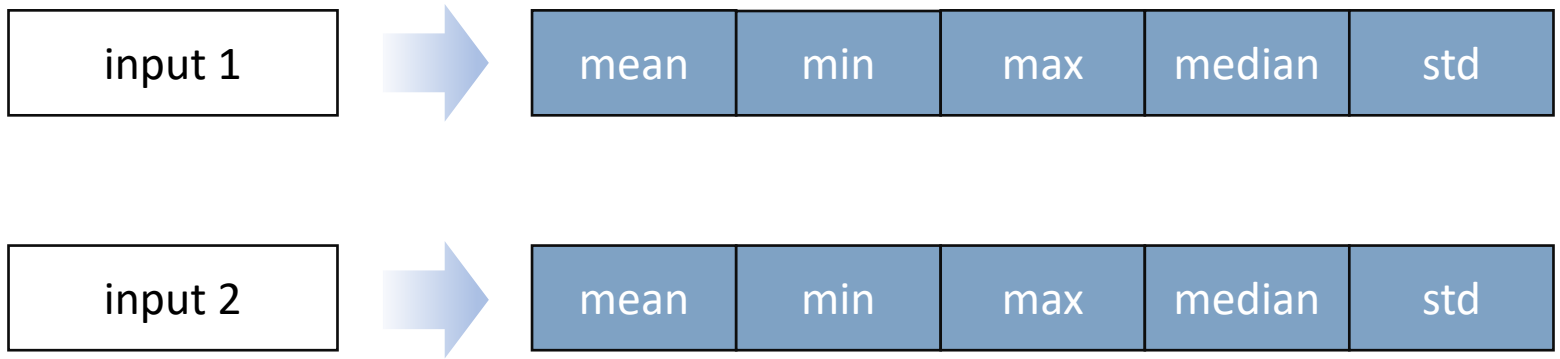


STFT를 거치며 F0 주파수를 찾기 때문에 오디오의 길이에 따라 F0 주파수의 길이가 제각각

01 복습 및 정리

모델링 정리

Logistic Regression

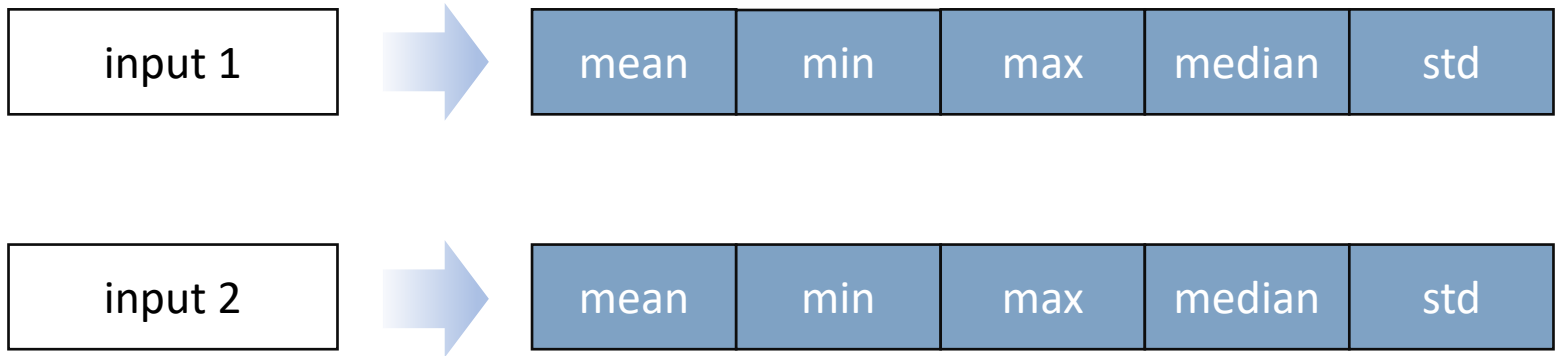


각 입력에서 F0 주파수의 통계량을 입력으로 이용

01 복습 및 정리

모델링 정리

Logistic Regression



성별 분류에서 0.84의 Validation Accuracy 기록

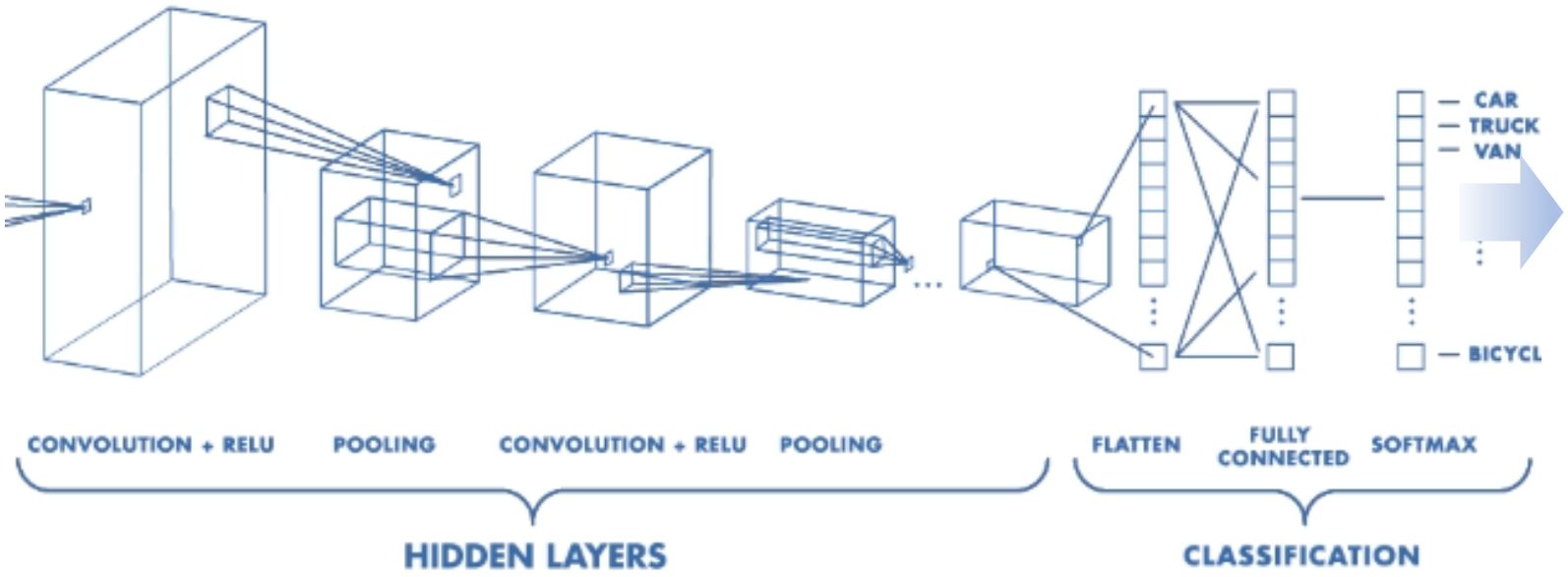
01 복습 및 정리

모델링 정리

CNN

[3, 13, 126000]의 tensor를 최종 input으로 사용

MFCC
+
1차 차분
+
2차 차분

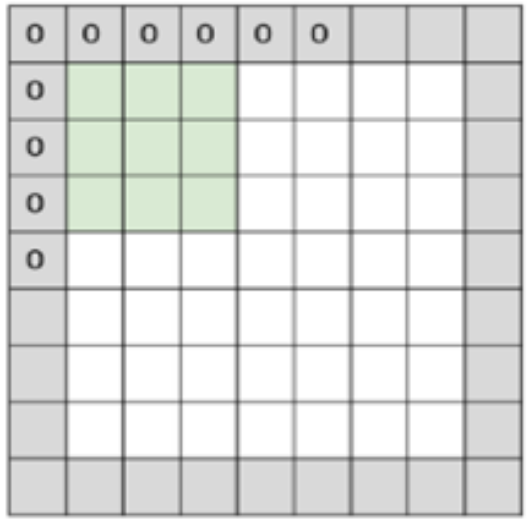


<예시>
① 강원도
② 경상도
③ 전라도
④ 제주도
⑤ 충청도

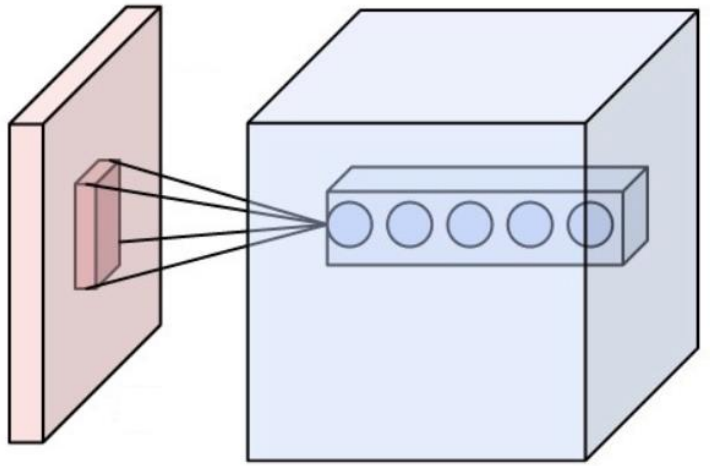
01 복습 및 정리

모델링 정리

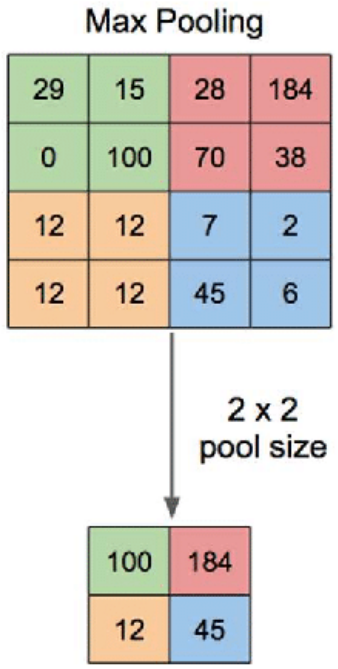
CNN – MFCC



Input 위, 아래
2칸씩 Zero Padding



Convolutional Layer



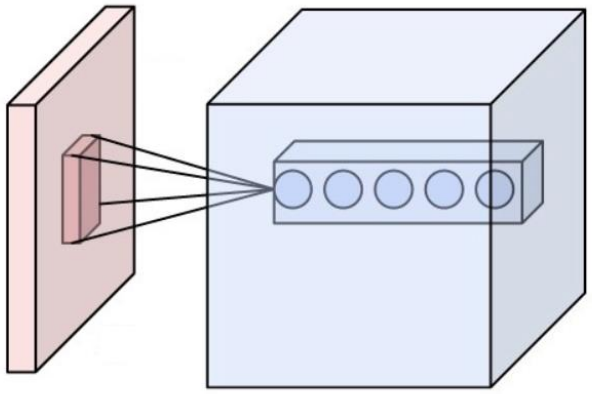
Max Pooling

DeepVoice

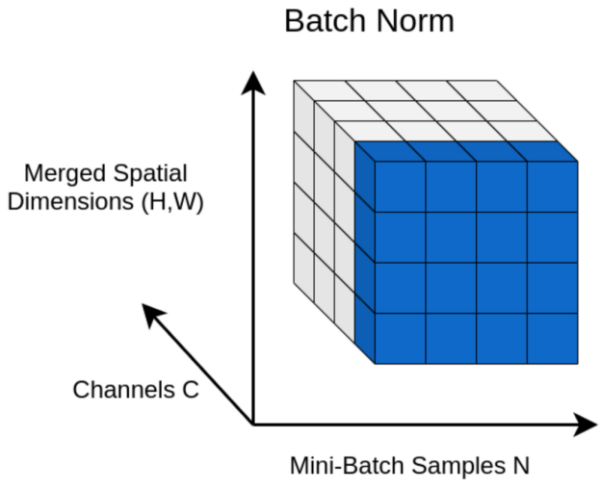
01 복습 및 정리

모델링 정리

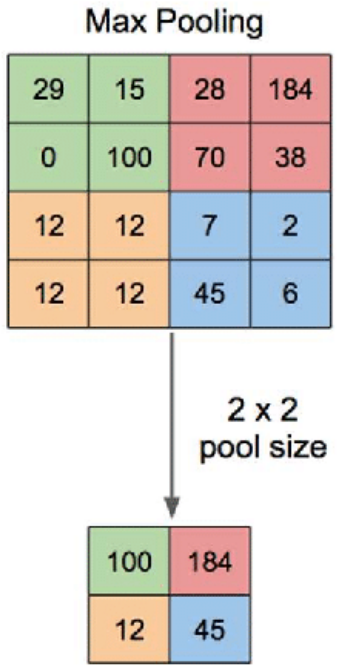
CNN – Mel Spectrogram



Convolutional Layer



Batch Normalization



Max Pooling

DeepVoice

01 복습 및 정리

모델링 정리

Vanilla RNN

분류 모델이므로 Many-to-One의 형태

∴ 최종 Hidden Layer의 Output만 사용!

MFCC

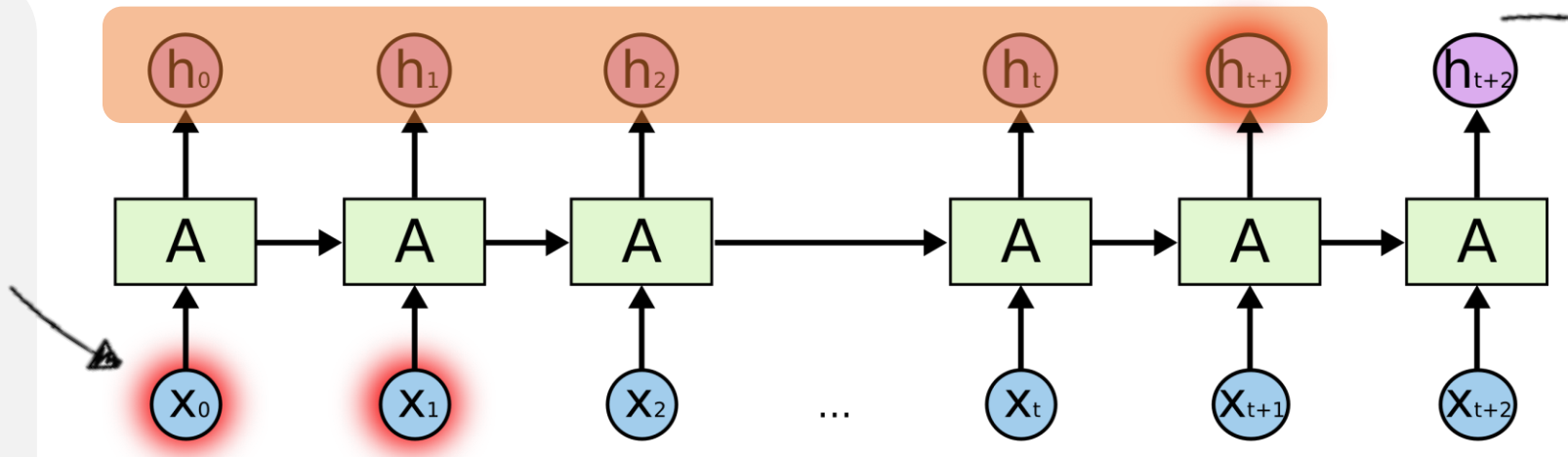
Mel Spectogram

+

1차 차분

+

2차 차분



<예시>

① 강원도

② 경상도

③ 전라도

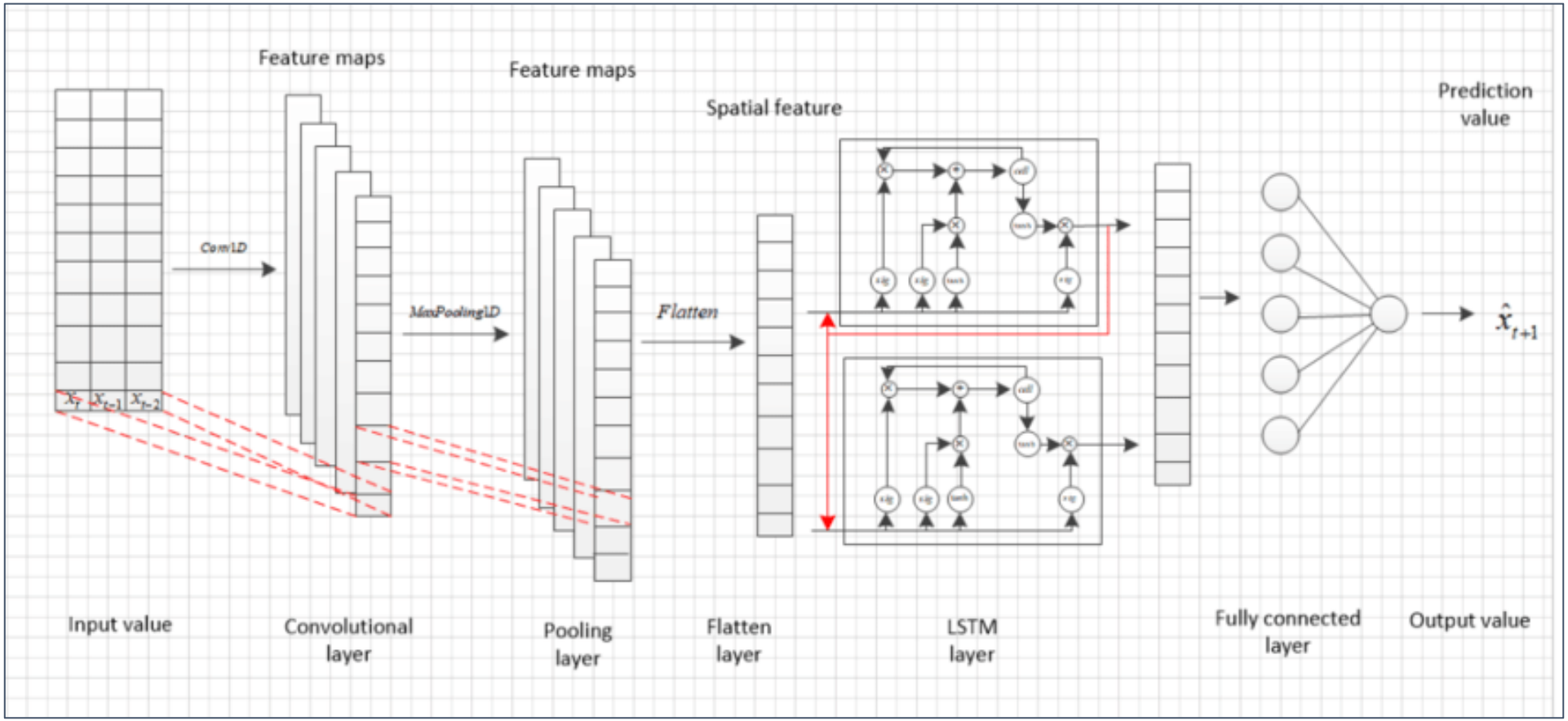
④ 제주도

⑤ 충청도

01 복습 및 정리

모델링 정리

CNN + LSTM

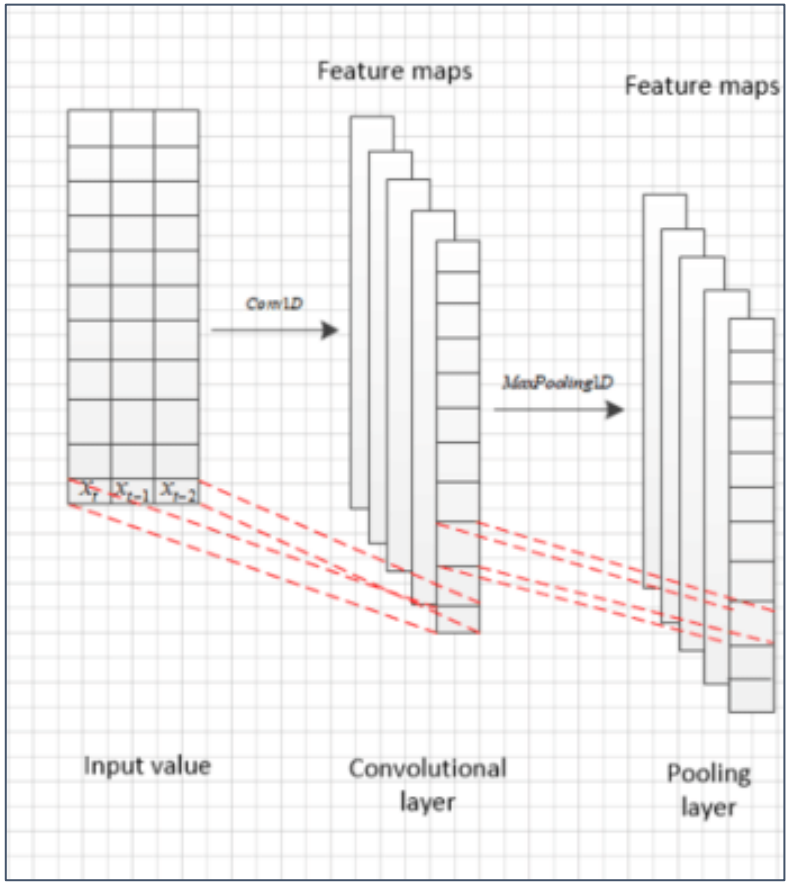


DeepVoice

01 복습 및 정리

모델링 정리

CNN + LSTM



Input

↓

CNN Model 통과

↓

Feature(CNN의 Output)

DeepVoice

01 복습 및 정리

모델링 정리

CNN + LSTM

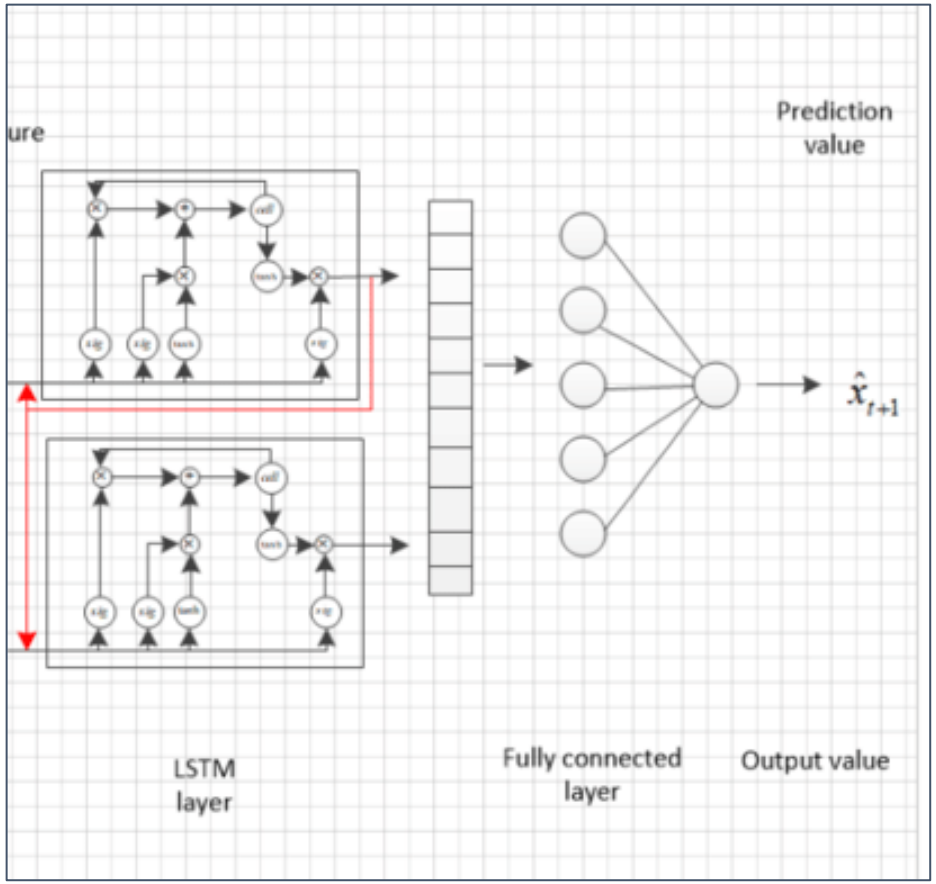
CNN의 Output = LSTM의 Input

↓

LSTM 셀 통과

↓

Hidden Layer 최종 Output → Softmax

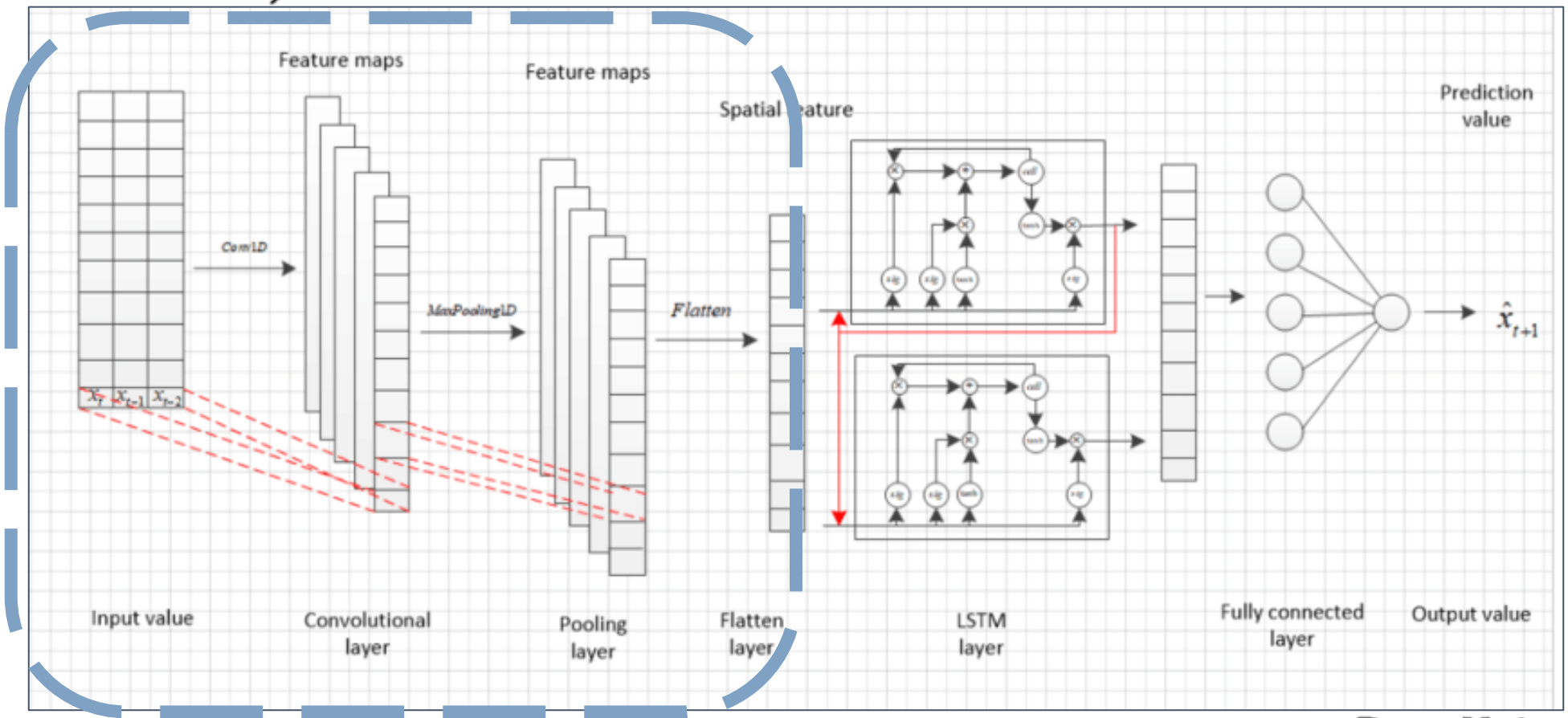


01 복습 및 정리

모델링 정리

CNN + LSTM

Convolutional Layer + Pooling Layer + Batch Normalization Layer = 데이터의 특징 추출

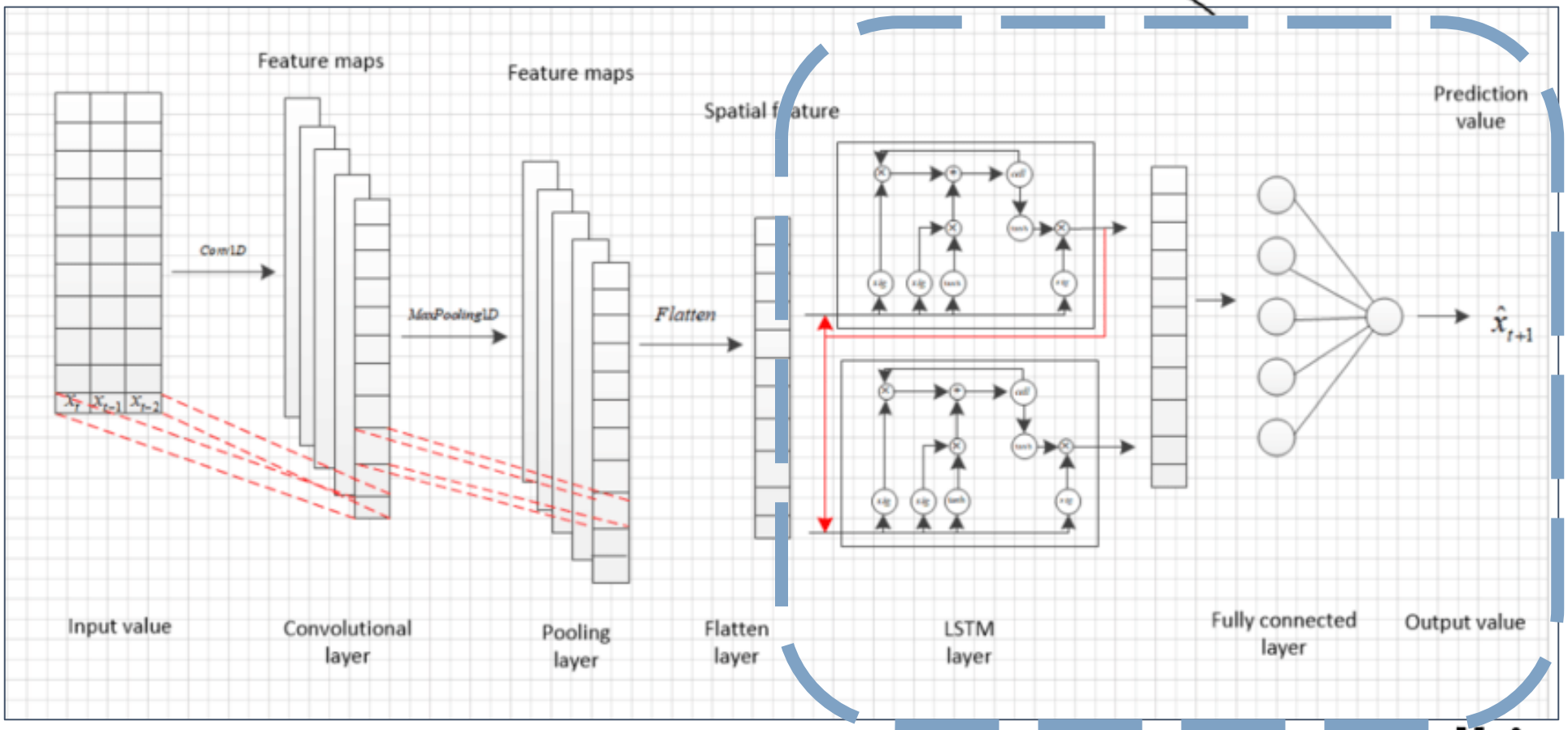


01 복습 및 정리

모델링 정리

CNN + LSTM

CNN에서 추출된 특징값을 바탕으로 Label 예측

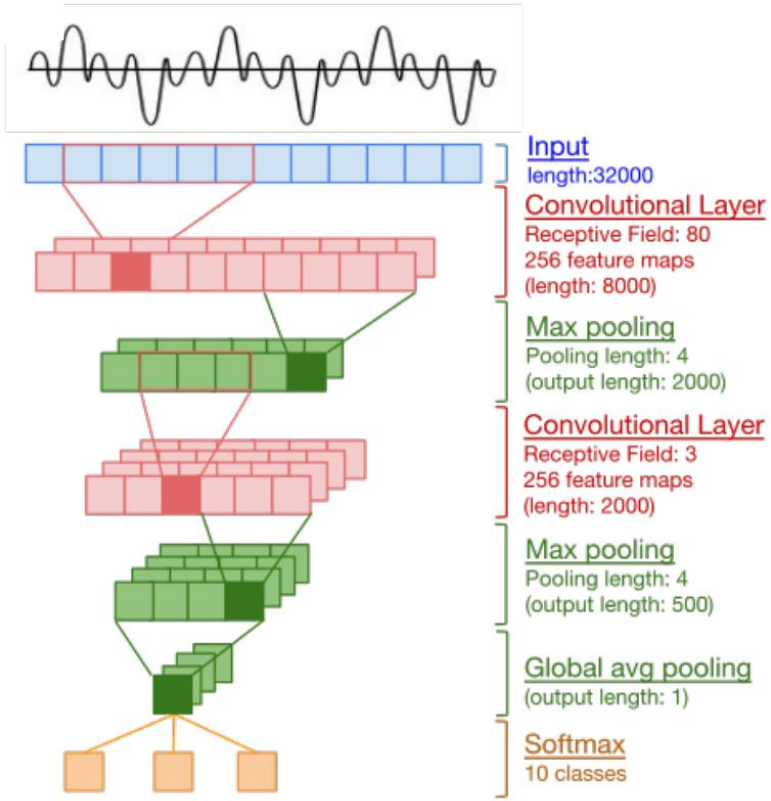


DeepVoice

01 복습 및 정리

모델링 정리

CNN for Raw Waveform



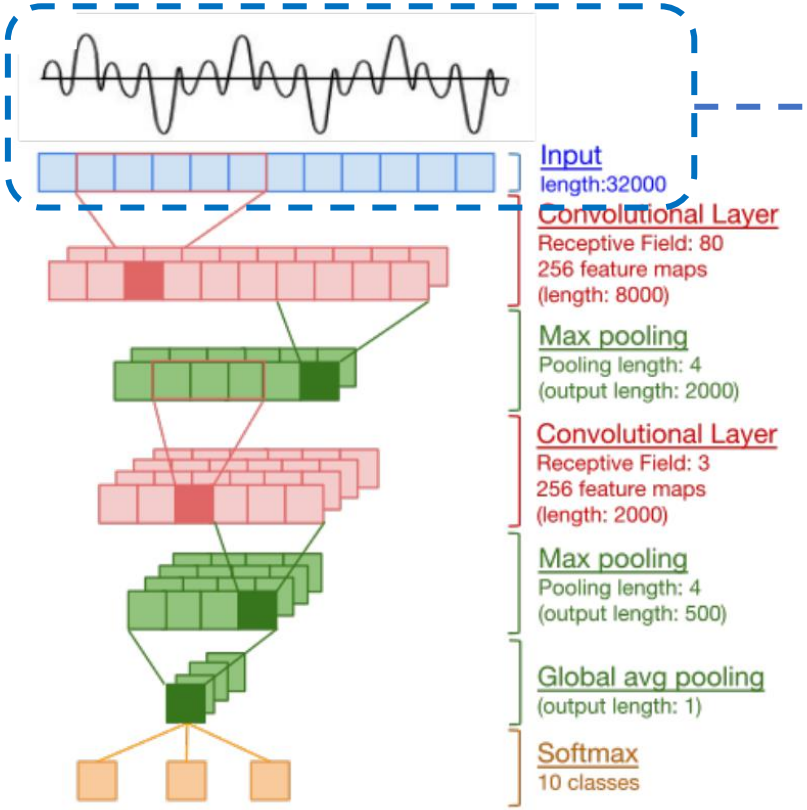
Feature Engineering 없이
입력으로 raw-audio를 사용하는
CNN 모델

VERY DEEP CONVOLUTIONAL NEURAL NETWORKS FOR RAW WAVEFORMS
- Wei Dai, et. al (2016)

01 복습 및 정리

모델링 정리

CNN for Raw Waveform



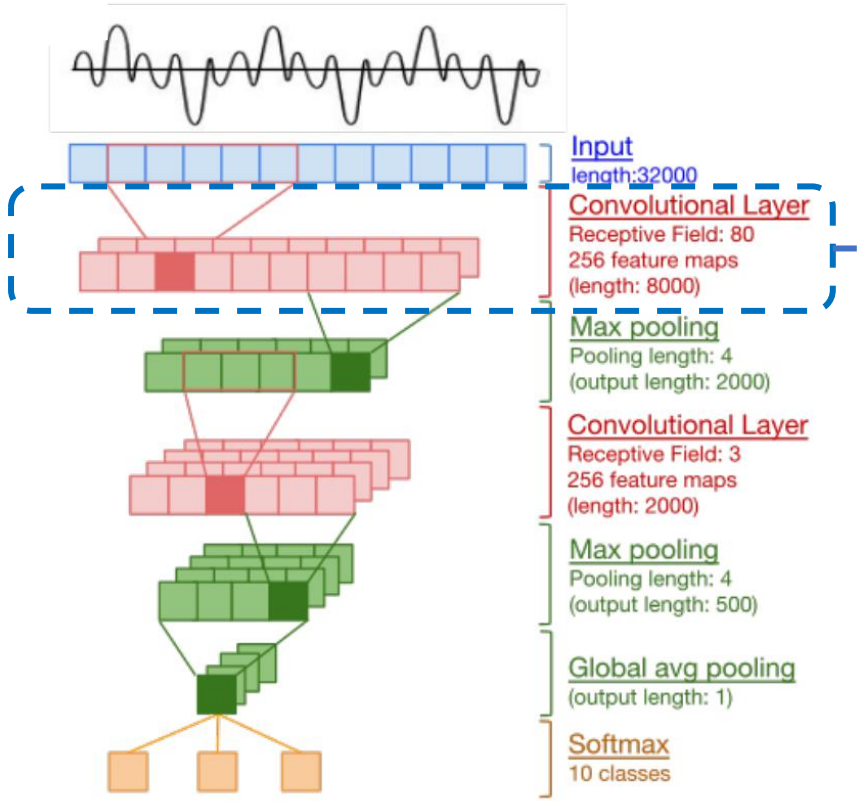
Feature Engineering 없이
입력으로 raw-audio를 사용하는
CNN 모델

VERY DEEP CONVOLUTIONAL NEURAL NETWORKS FOR RAW WAVEFORMS
- Wei Dai, et. al (2016)

01 복습 및 정리

모델링 정리

CNN for Raw Waveform



1-D CNN 의 filter size를 80으로 세팅

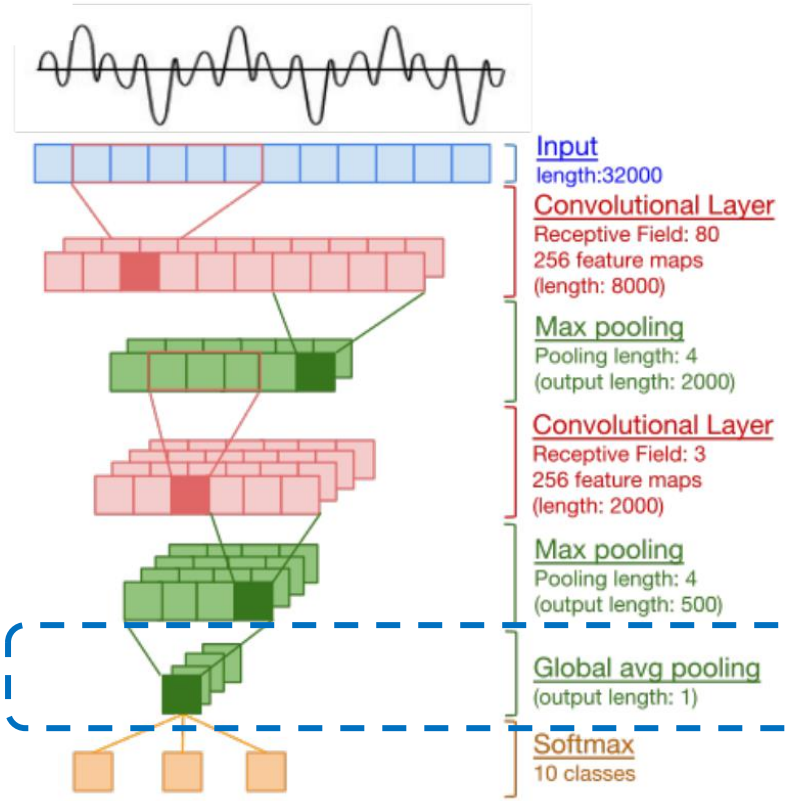
다운 샘플링 효과

VERY DEEP CONVOLUTIONAL NEURAL NETWORKS FOR RAW WAVEFORMS
- Wei Dai, et. al (2016)

01 복습 및 정리

모델링 정리

CNN for Raw Waveform



Global Average Pooling

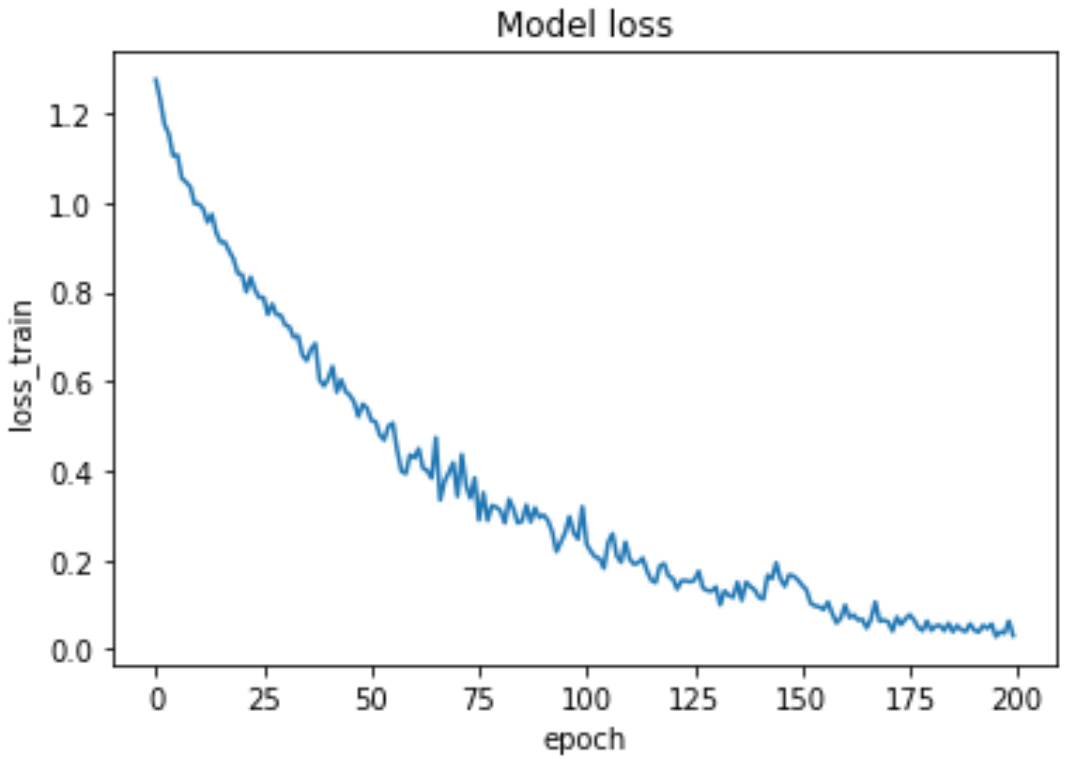
feature map을 1차원 feature로 변환

VERY DEEP CONVOLUTIONAL NEURAL NETWORKS FOR RAW WAVEFORMS
- Wei Dai, et. al (2016)

01 복습 및 정리

모델 성능

연령 분류를 위한 Baseline model – CNN

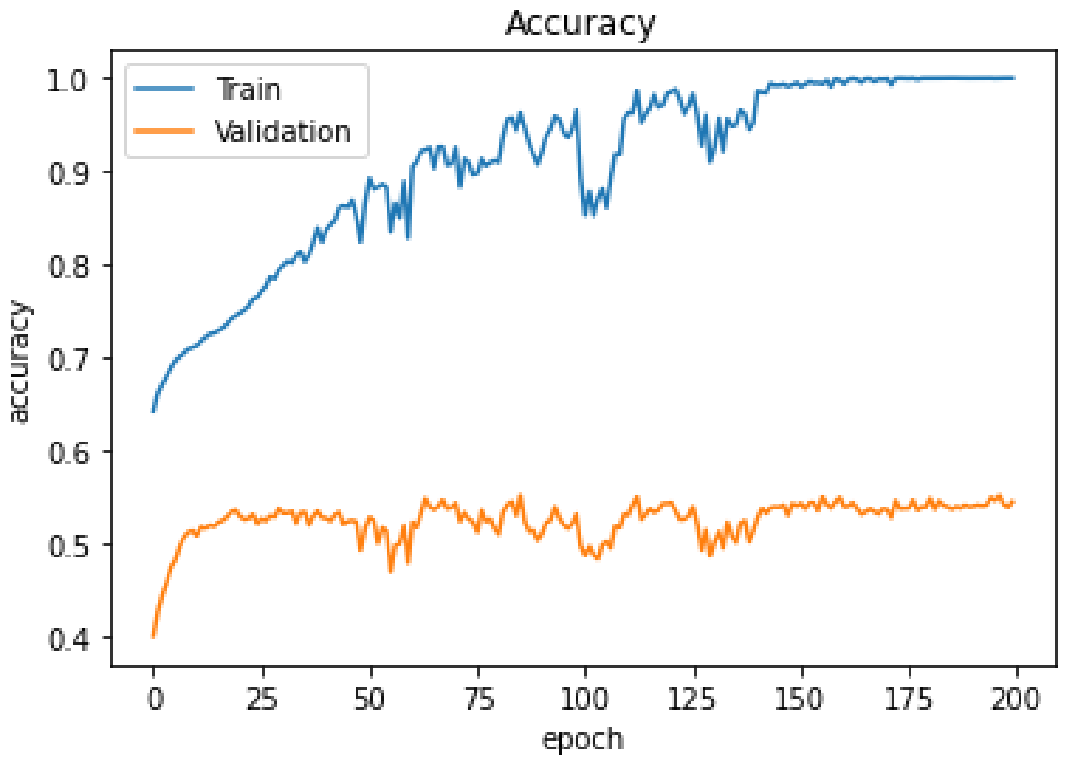


안정적으로 Train Loss 수렴

01 복습 및 정리

모델 성능

연령 분류를 위한 Baseline model – CNN



Train Accuracy 안정적 상승 → 학습 원활

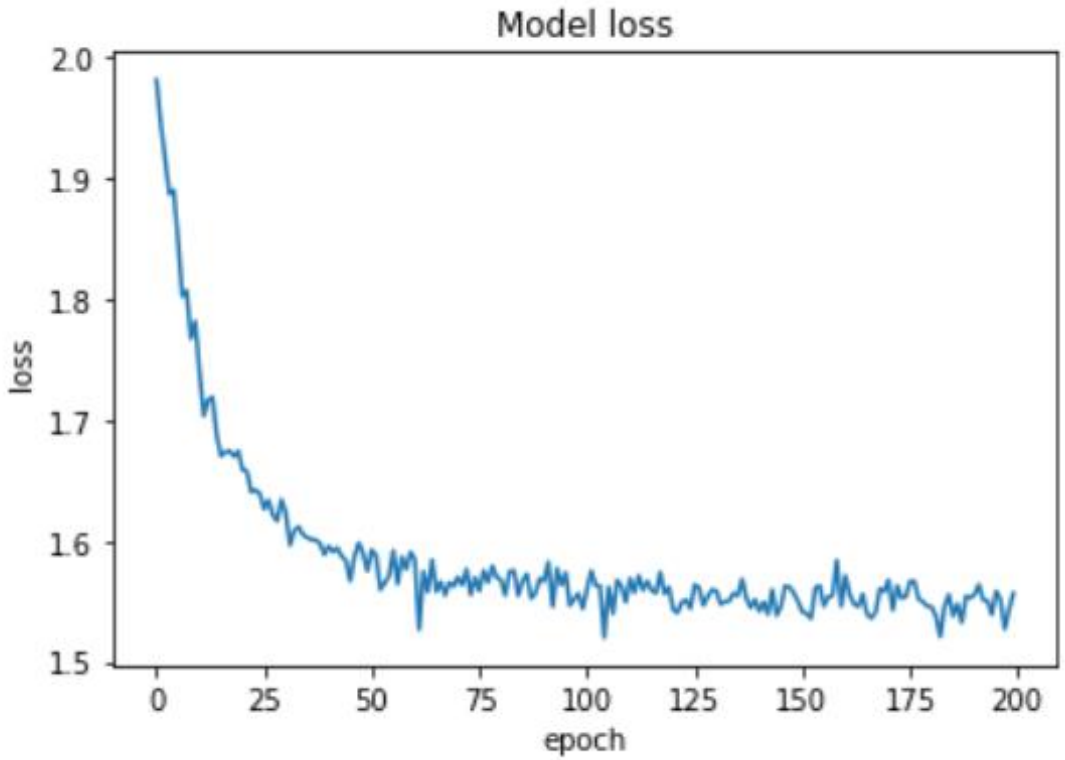
Validation Accuracy 약 55%에 수렴

과적합 존재?

01 복습 및 정리

모델 성능

연령 분류를 위한 Baseline model – RNN



초반 빠른 Loss 감소

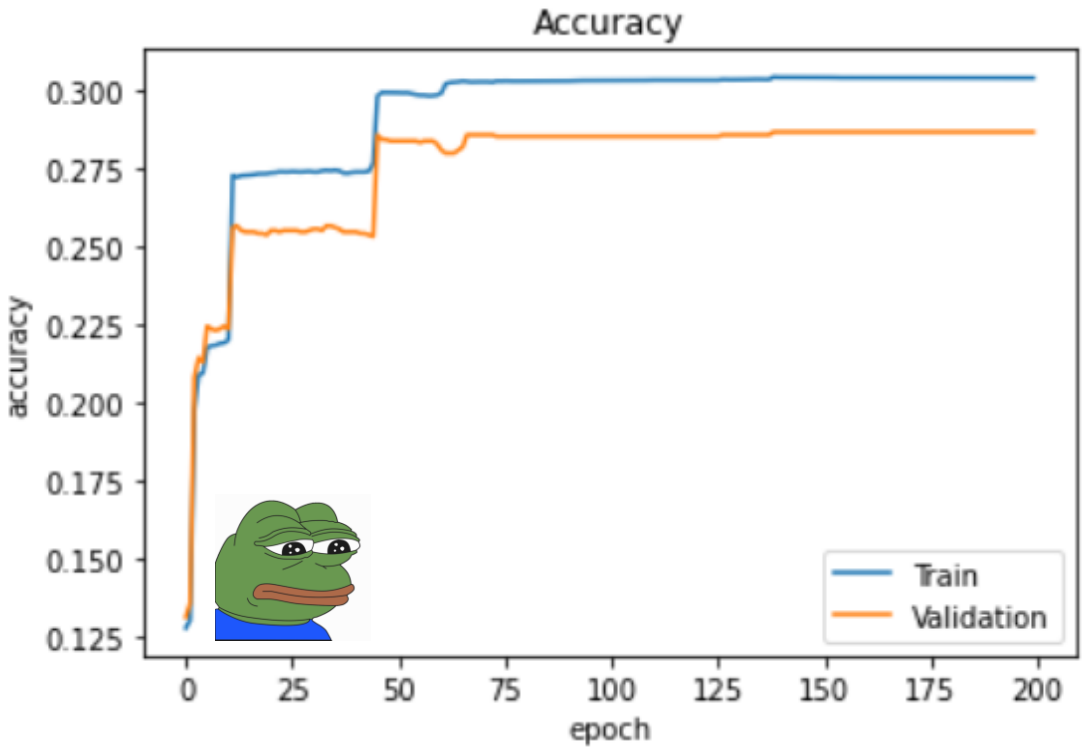
1.5 이하로 감소하지 못함

학습이 원활하게 이뤄지지 않음

01 복습 및 정리

모델 성능

연령 분류를 위한 Baseline model – RNN



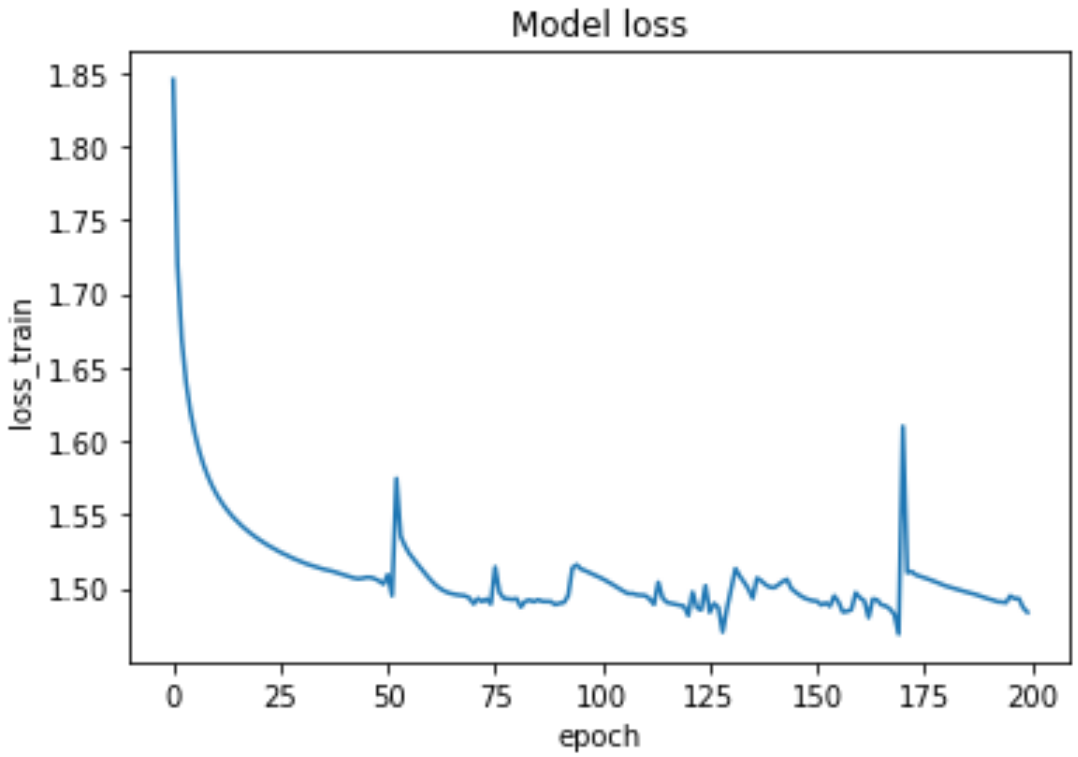
train accuracy와
validation accuracy의 움직임 유사
두 지표 모두 우상향 추세

성능 다소 떨어짐
validation accuracy=0.30

01 복습 및 정리

모델 성능

연령 분류를 위한 Baseline model – CNN + LSTM



초반 빠른 Loss 감소

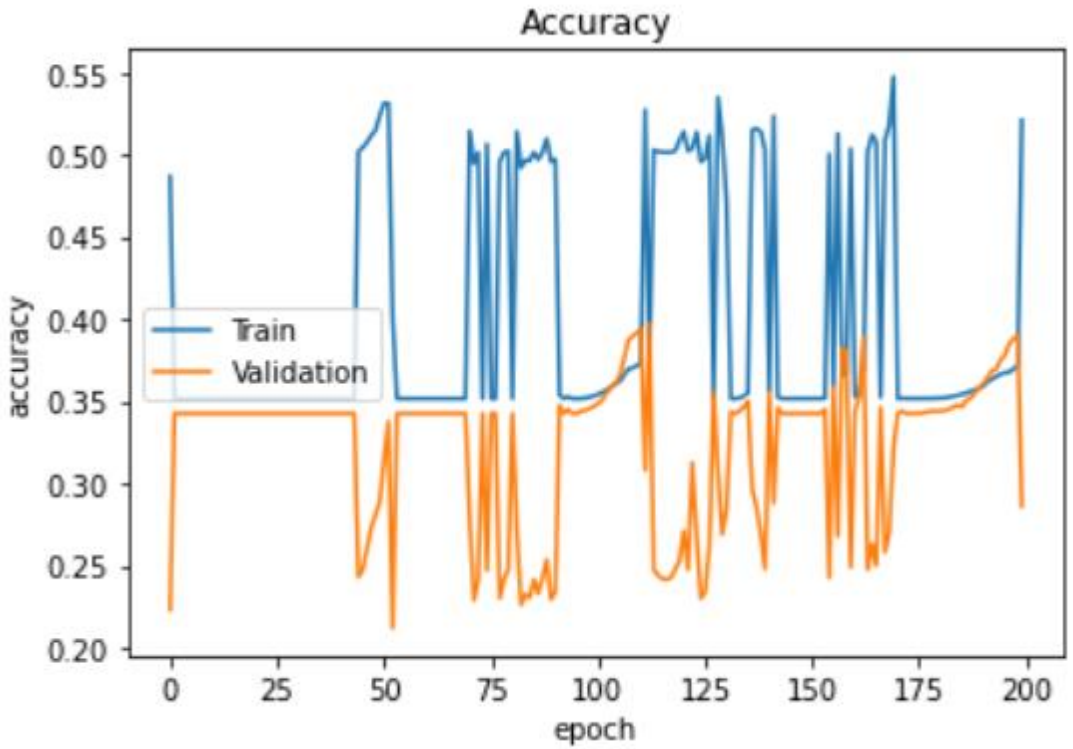
1.50 이하로 감소하지 못함

학습이 원활하지 않음

01 복습 및 정리

모델 성능

연령 분류를 위한 Baseline model – CNN + LSTM



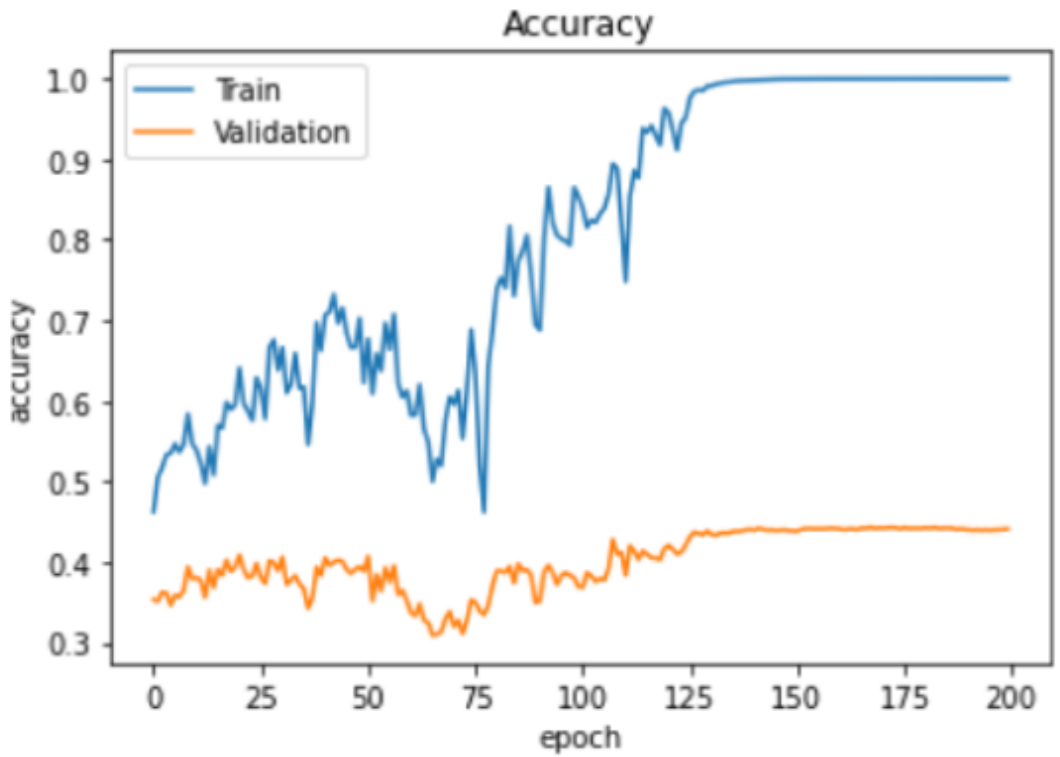
두 지표가 반대의 경향

train dataset에 모델 과적합(?)

01 복습 및 정리

모델 성능

연령 분류를 위한 Baseline model – CNN for Raw Waveform



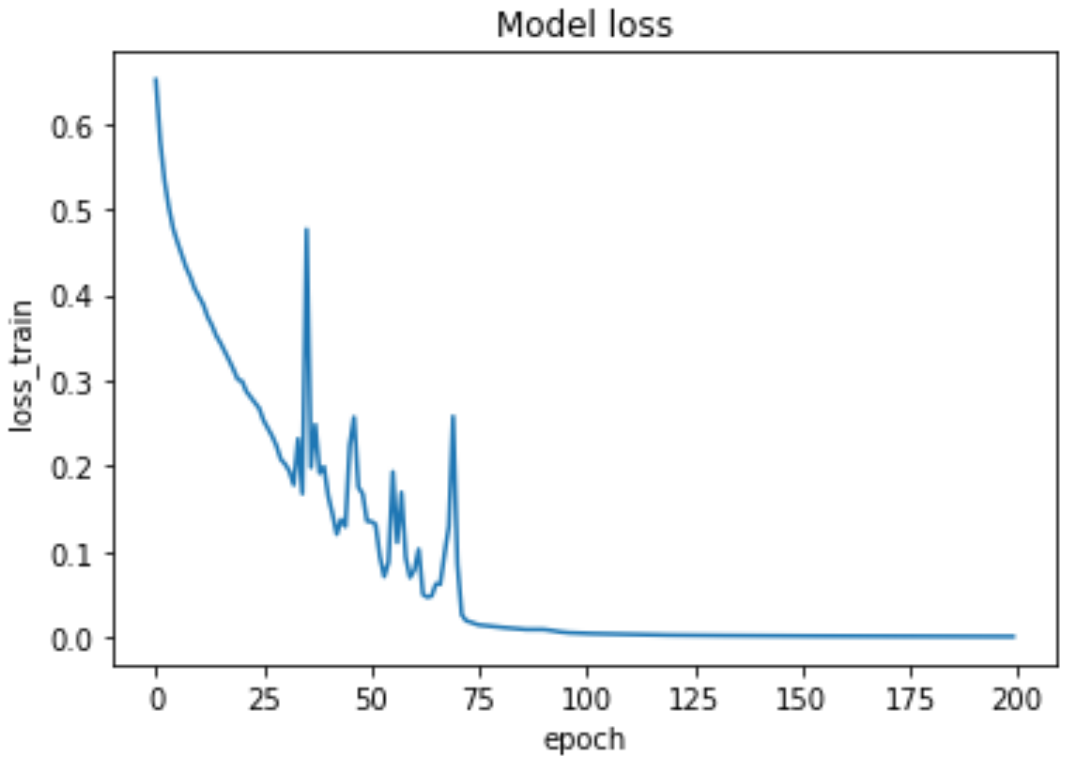
Validation accuracy가 0.4에서 수렴

원활한 학습 X

01 복습 및 정리

모델 성능

성별 분류를 위한 Baseline model – CNN



초반 빠른 Loss 감소

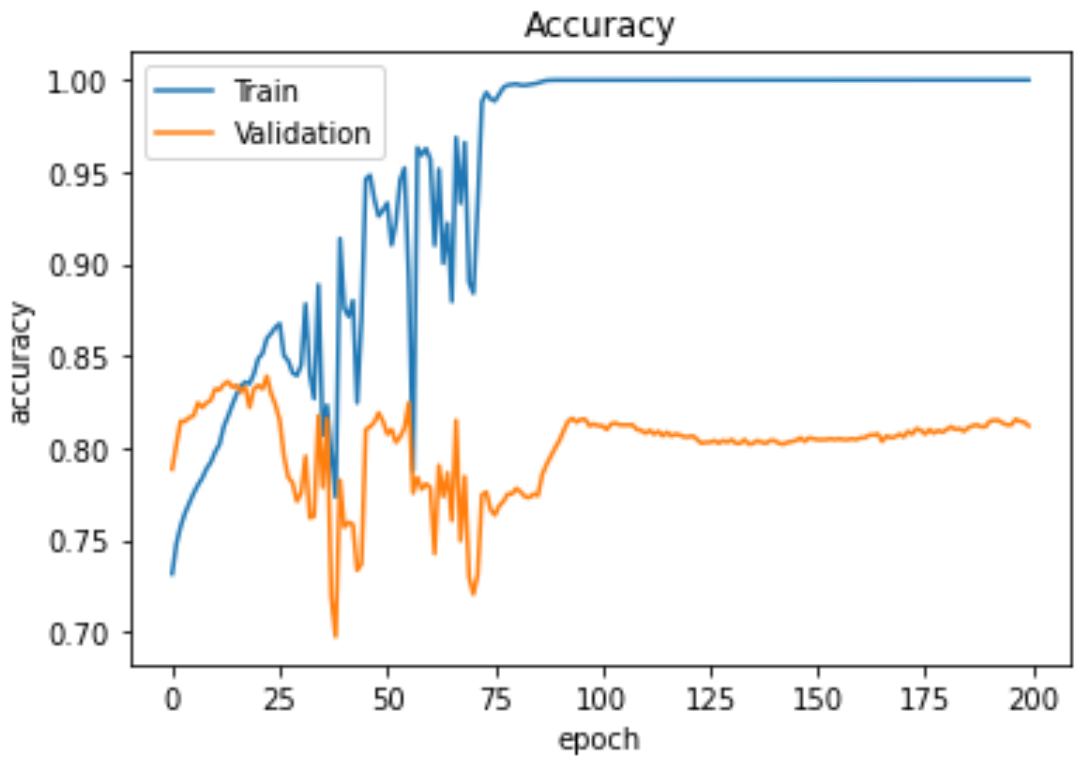
Loss가 0으로 수렴

Train Dataset에 대한 학습 원활

01 복습 및 정리

모델 성능

성별 분류를 위한 Baseline model – CNN



Train Accuracy : 변동폭이 크지만 1로 수렴

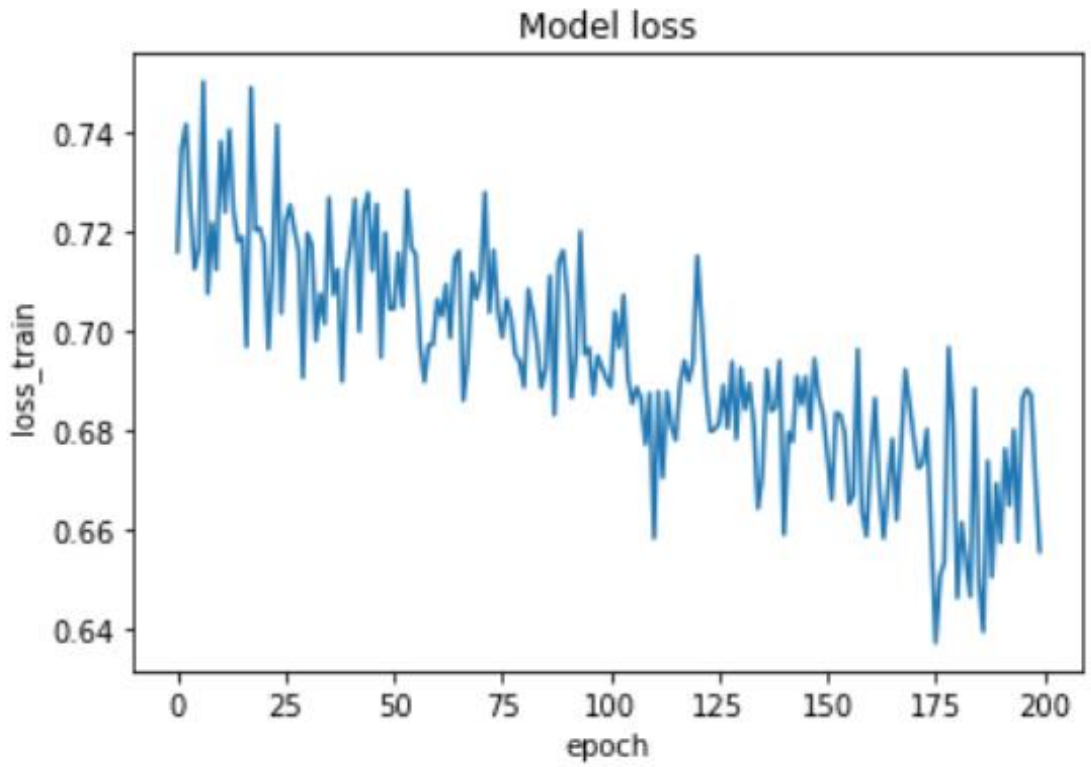
Validation Accuracy : 약 80%에서 진동

과적합 경향

01 복습 및 정리

모델 성능

성별 분류를 위한 Baseline model – RNN



Loss값이 안정적으로 수렴하지 못함

0.64 이하로 감소하지 않음

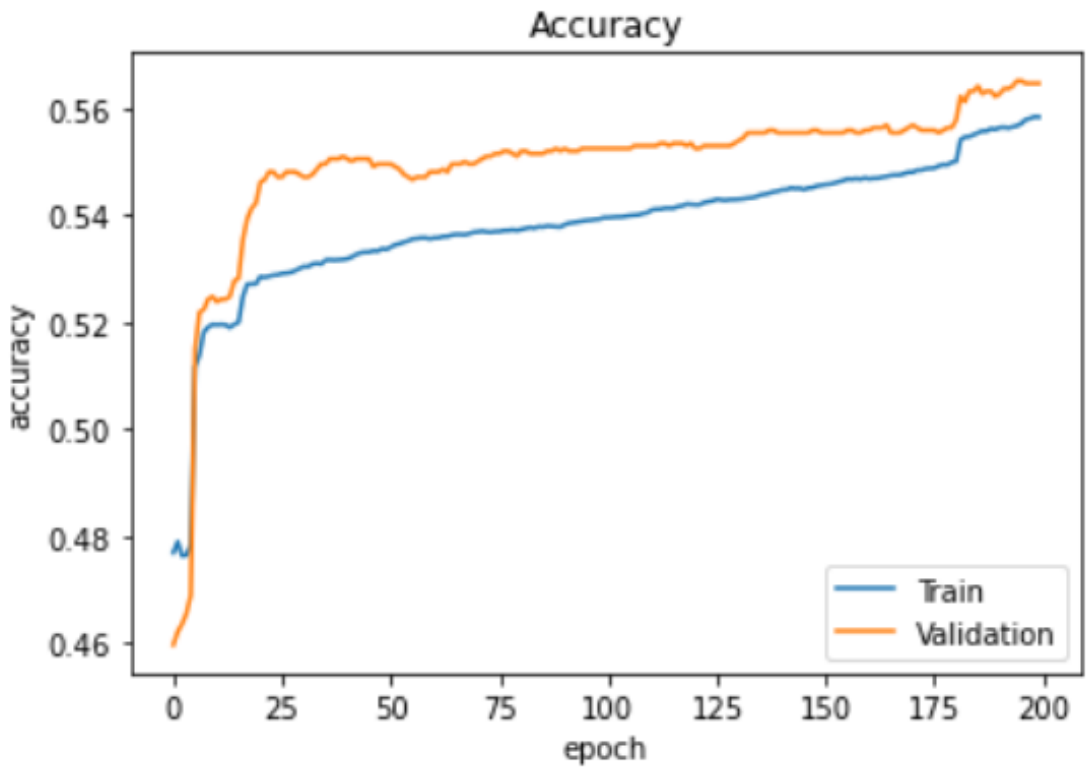
학습 원활하지 않음

DeepVoice

01 복습 및 정리

모델 성능

성별 분류를 위한 Baseline model – RNN



초반 급격한 Accuracy 증가

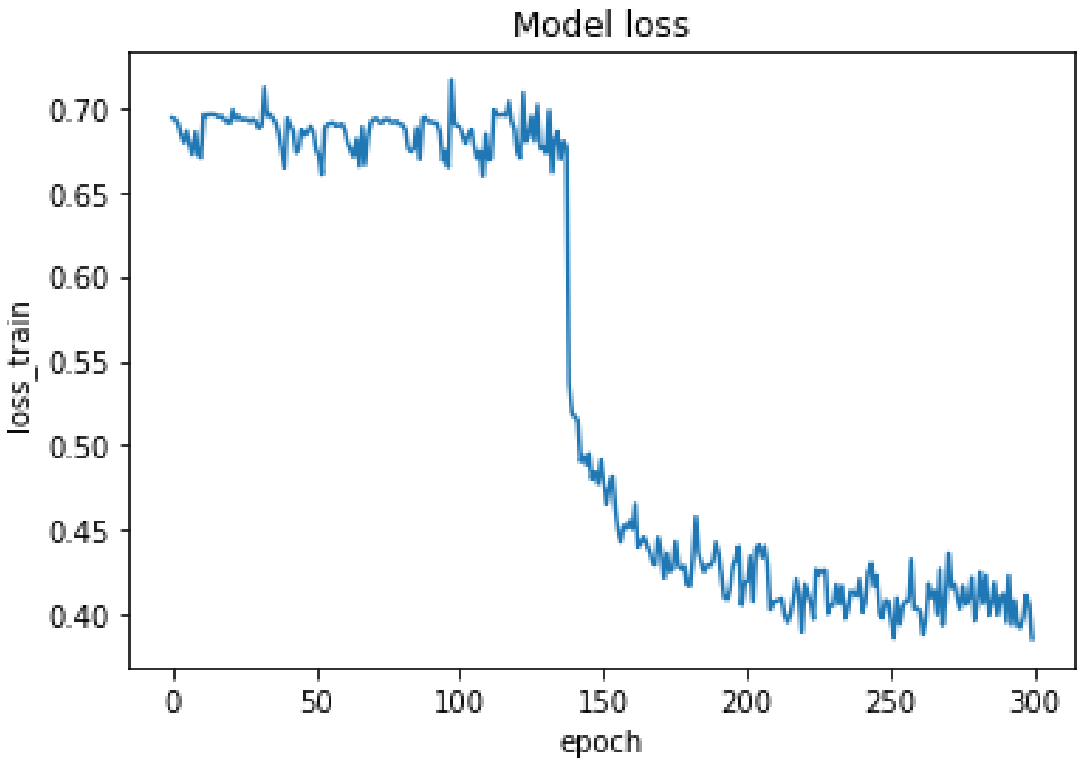
Validation accuracy 약 0.56

(CNN보다 성능 떨어짐)

01 복습 및 정리

모델 성능

성별 분류를 위한 Baseline model – CNN + LSTM



초반 Loss 0.70에서 진동

약 120 epoch 이후 Loss 급격한 감소

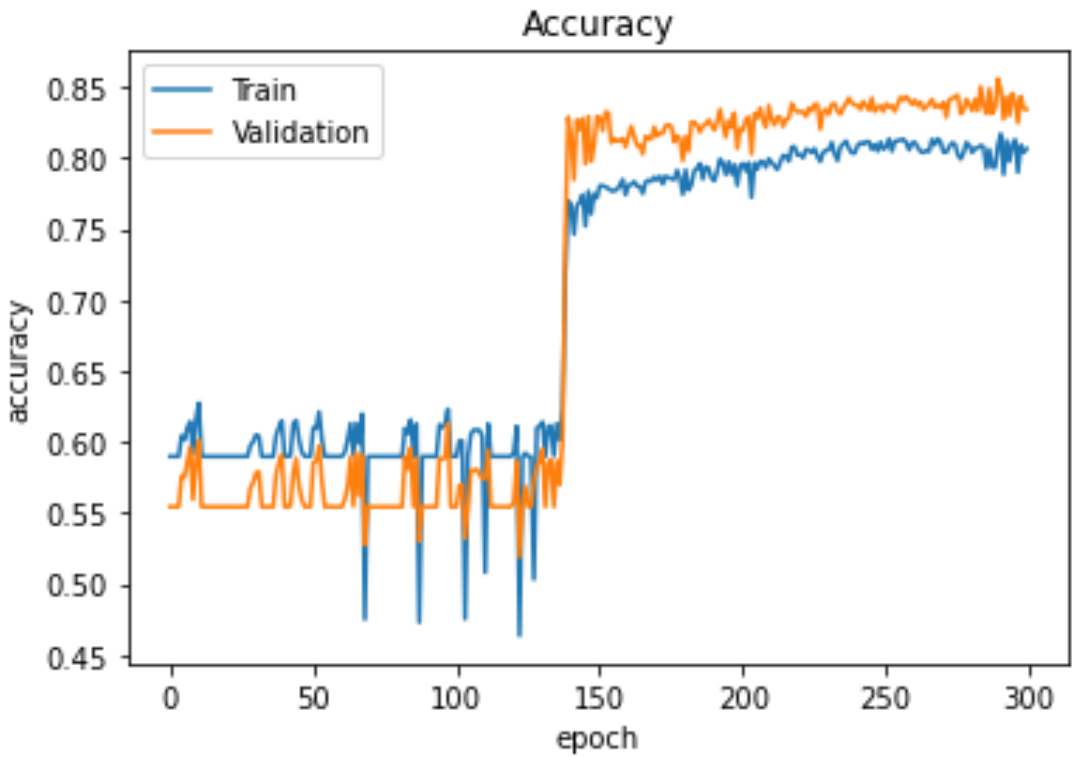
CNN에 비해 학습 속도가 느림

Loss가 완벽하게 0으로 수렴X

01 복습 및 정리

모델 성능

성별 분류를 위한 Baseline model – CNN + LSTM



약 120 epoch까지 Accuracy 진동

이후 급격한 Accuracy 증가

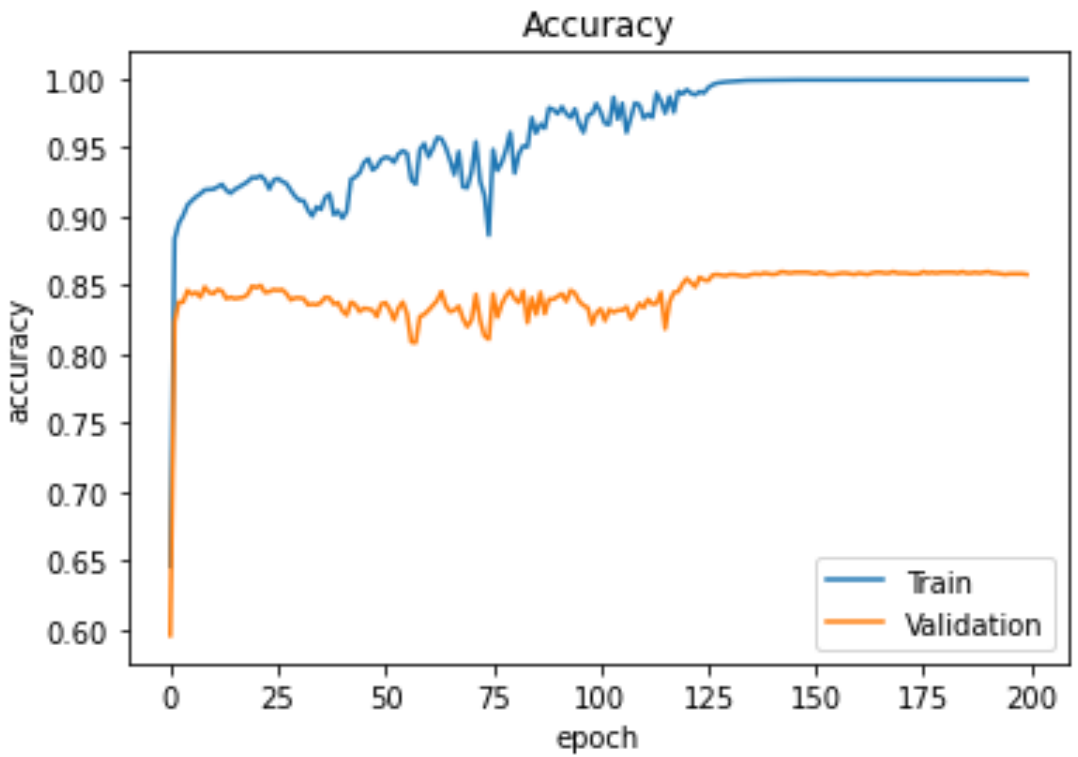
Validation Accuracy : 약 84%에 수렴

(CNN의 성능보다 뛰어남)

01 복습 및 정리

모델 성능

성별 분류를 위한 Baseline model – CNN for Raw Waveform



train accuracy와

validation accuracy의 움직임 유사

두 지표 모두 빠른 수렴

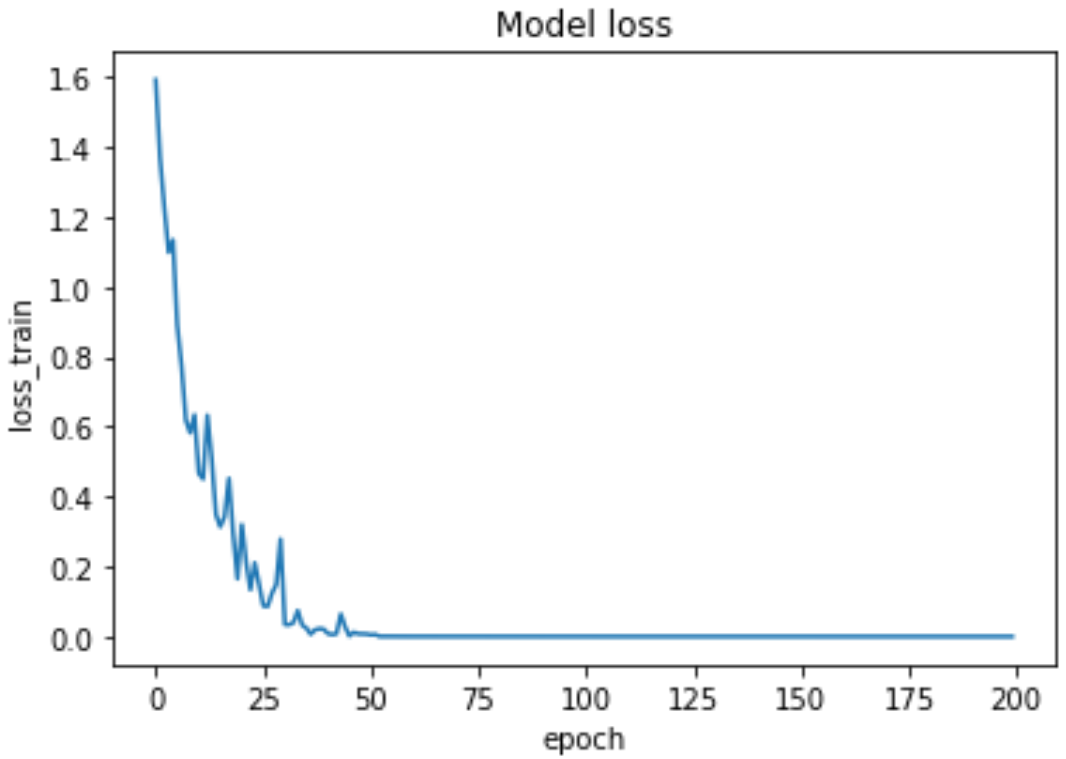
원활한 학습 상태

validation accuracy=0.85

01 복습 및 정리

모델 성능

방언 분류를 위한 Baseline model – CNN



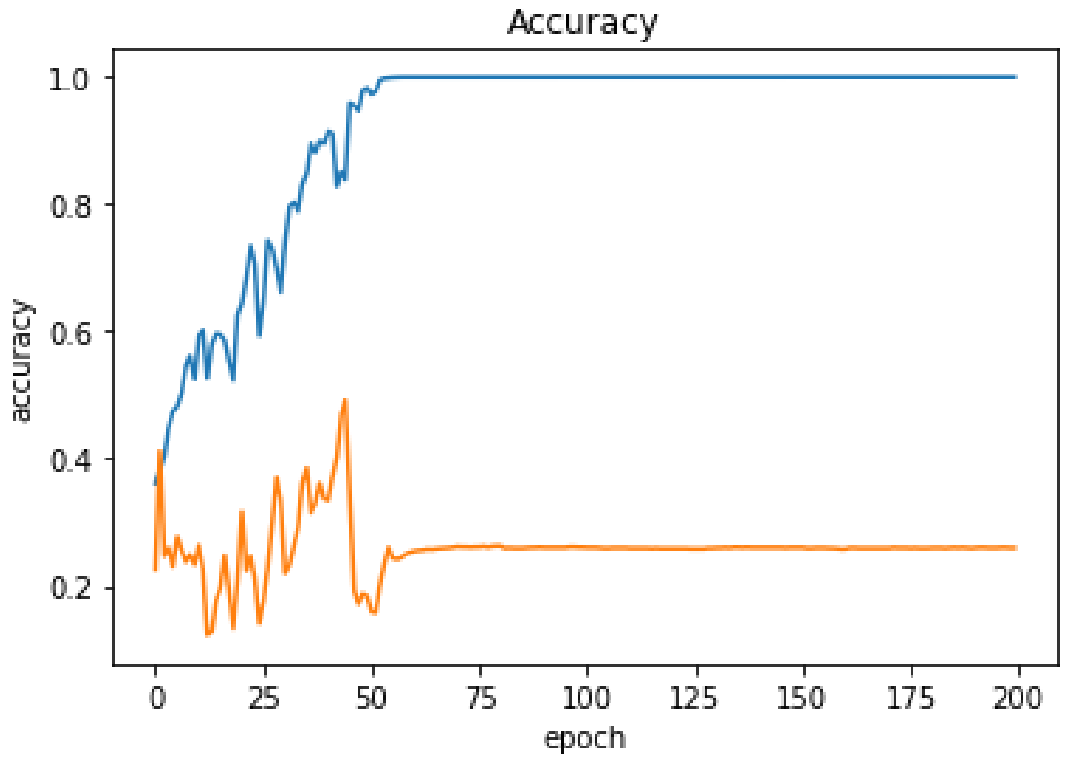
Loss 빠른 속도로 0으로 수렴

안정적인 학습

01 복습 및 정리

모델 성능

방언 분류를 위한 Baseline model – CNN



Train Accuracy 1.0으로 수렴

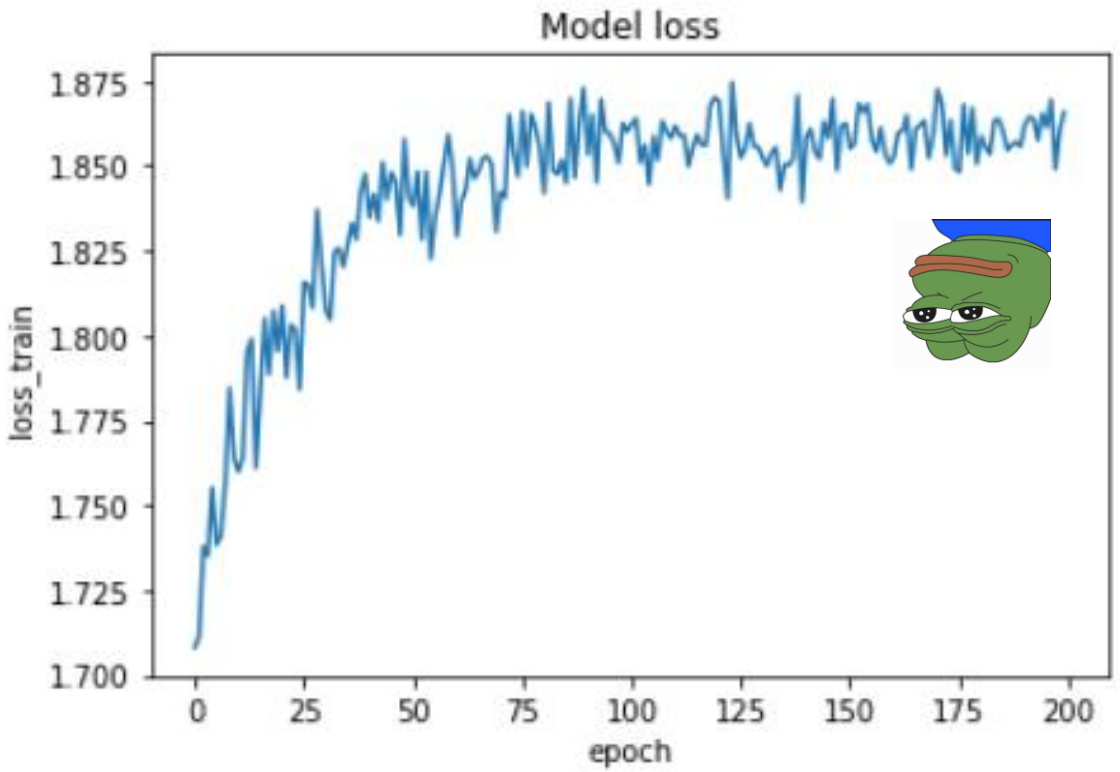
Validation Accuracy 약 30%에서 진동

과적합 발생? → 데이터 수 증가 필요

01 복습 및 정리

모델 성능

방언 분류를 위한 Baseline model – RNN



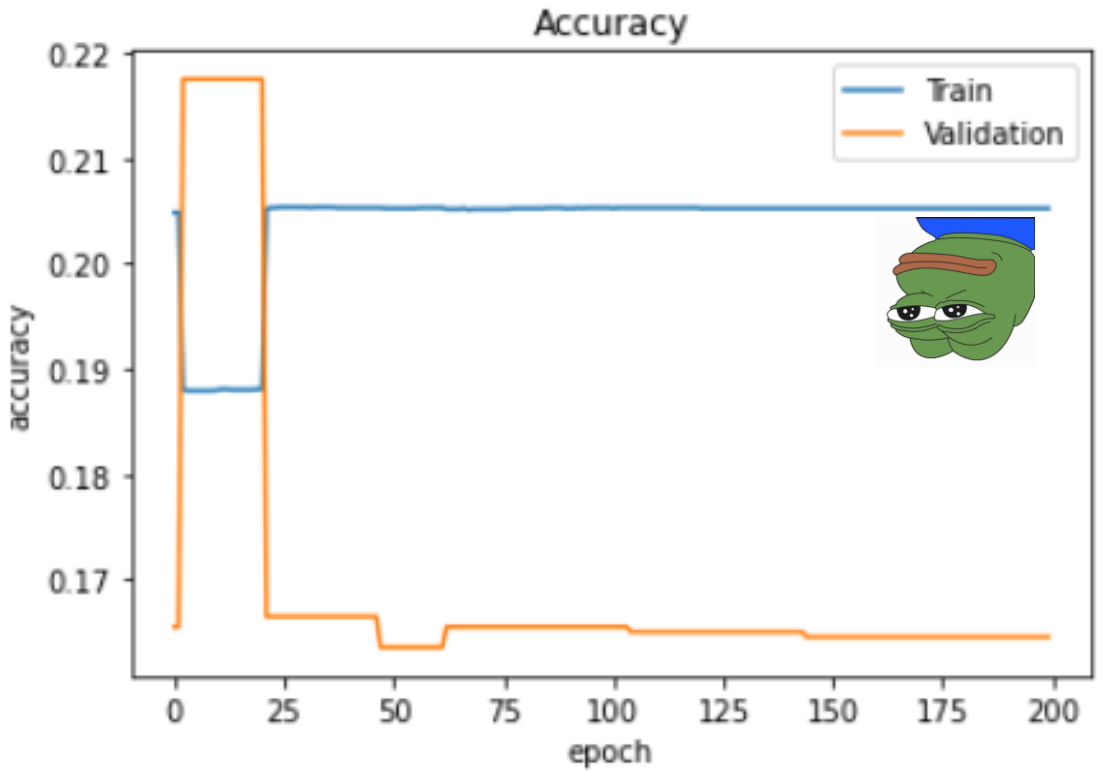
Loss 수렴하지 못하고 우상향

학습 X

01 복습 및 정리

모델 성능

방언 분류를 위한 Baseline model – RNN



Train accuracy = 0.25

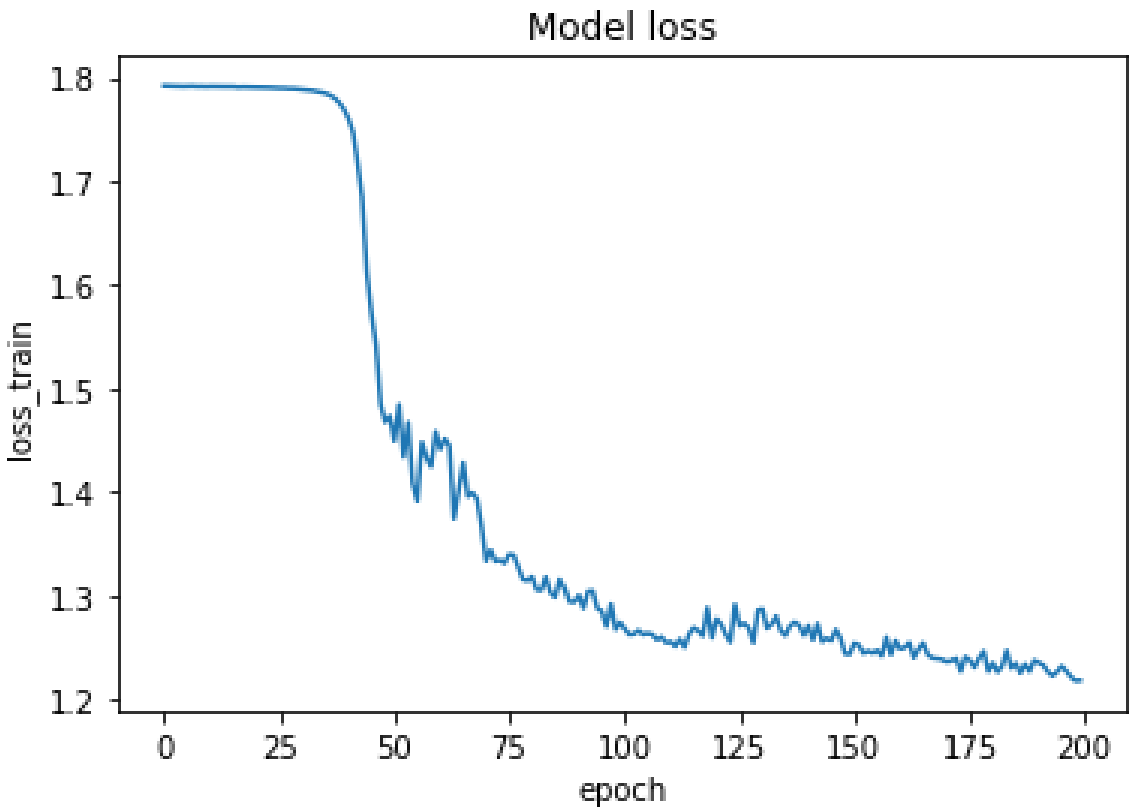
Validation accuracy = 0.16으로 수렴

성능 매우 떨어짐

01 복습 및 정리

모델 성능

방언 분류를 위한 Baseline model – CNN + LSTM



약 40epoch 내외부터 Loss 급격히 감소

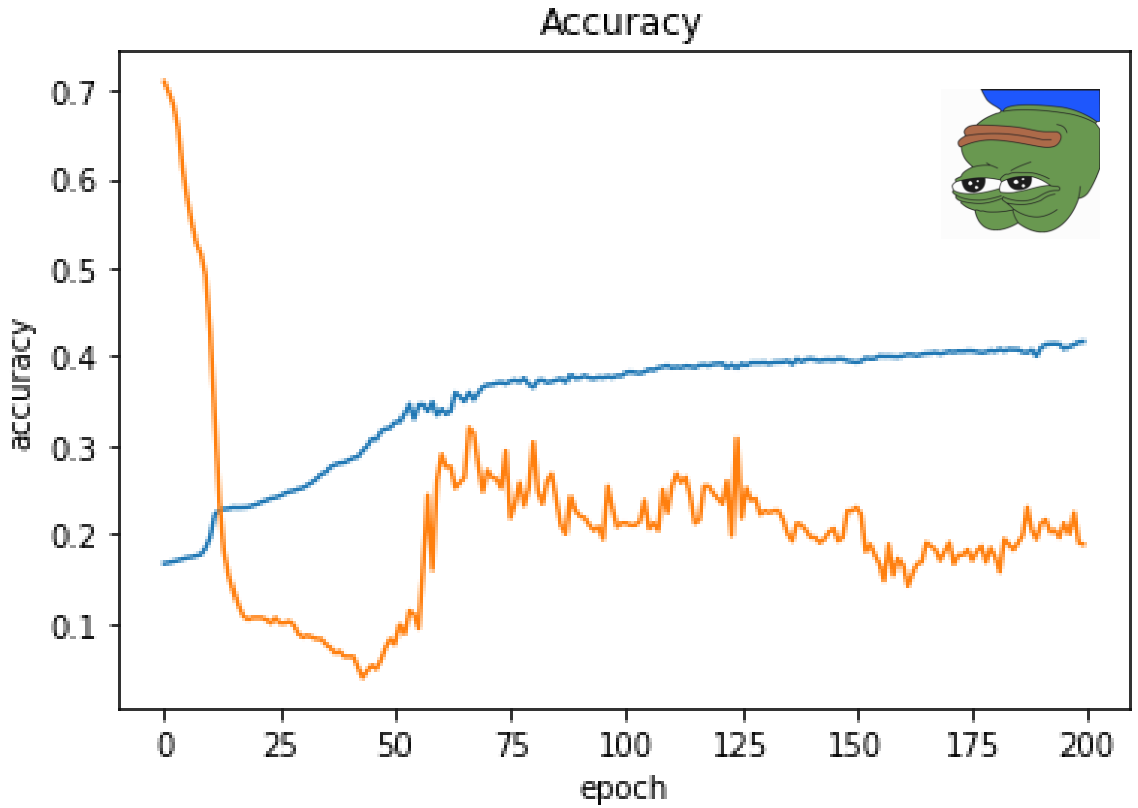
Loss 약 1.2로 수렴

학습이 원활하지 않음

01 복습 및 정리

모델 성능

방언 분류를 위한 Baseline model – CNN + LSTM



Train Accuracy 약 45%에 수렴

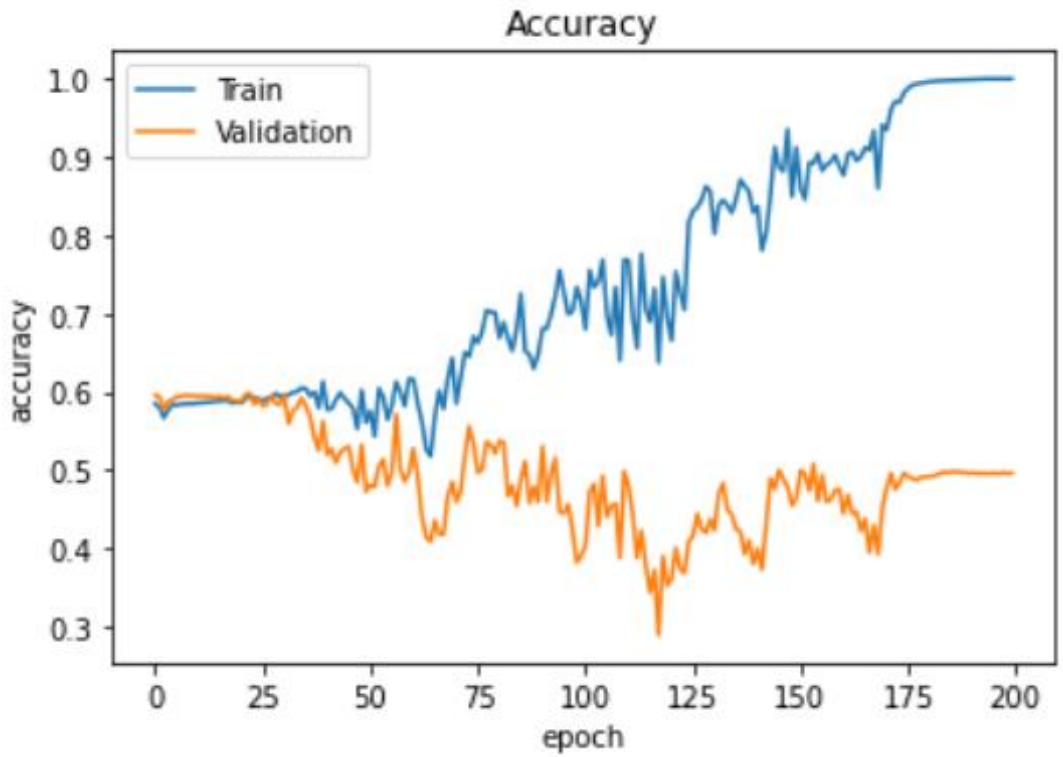
Validation Accuracy 약 20%에 수렴

학습이 원활하지 않음

01 복습 및 정리

모델 성능

방언 분류를 위한 Baseline model – CNN for Raw Waveform



train accuracy 우상향

validation accuracy 우하향

train dataset에 모델 과적합

02

시행착오 및 최종 input

02 시행착오 및 최종 input

시행착오

모델 성능 문제

모델 유형	연령	성별	방언
Vanilla RNN (Recurrent NN)	train acc : 0.3 val acc : 0.28	train acc : 0.56 val acc : 0.55	train acc : 0.25 val acc : 0.16
CNN	train acc : 0.9 val acc : 0.55	train acc : 0.99 val acc : 0.82	train acc : 0.99 val acc : 0.28
CNN + LSTM	train acc : 0.52 val acc : 0.28	train acc : 0.79 val acc : 0.84	train acc : 0.45 Val acc : 0.2
CNN for raw waveform	train acc : 0.99 val acc : 0.5	train acc : 0.99 val acc : 0.86	train acc : 0.99 val acc : 0.51

02 시행착오 및 최종 input

시행착오

모델 성능 문제

모델 유형	연령	성별	방언
Vanilla RNN (Recurrent NN)	train acc : 0.3 val acc : 0.28	train acc : 0.56 val acc : 0.55	train acc : 0.25 val acc : 0.16
CNN	train acc : 0.9 val acc : 0.5	train acc : 0.99 val acc : 0.82	train acc : 0.99 val acc : 0.28
CNN + LSTM	train acc : 0.5 val acc : 0.28	train acc : 0.99 val acc : 0.84	train acc : 0.45 Val acc : 0.2
CNN for raw waveform	train acc : 0.99 val acc : 0.5	train acc : 0.99 val acc : 0.86	train acc : 0.99 val acc : 0.51

모델을 성능을 높이고자…

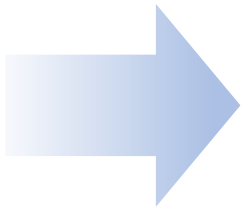
1. 연령 재범주화
2. 훈련 데이터 균등 샘플링
3. 1초 미만 데이터 제거
4. 노이즈 제거

02 시행착오 및 최종 input

시행착오

연령 재범주화

Age
3-10
11-19
20-29
30-39
40-49
50-59
60-69
Over 70

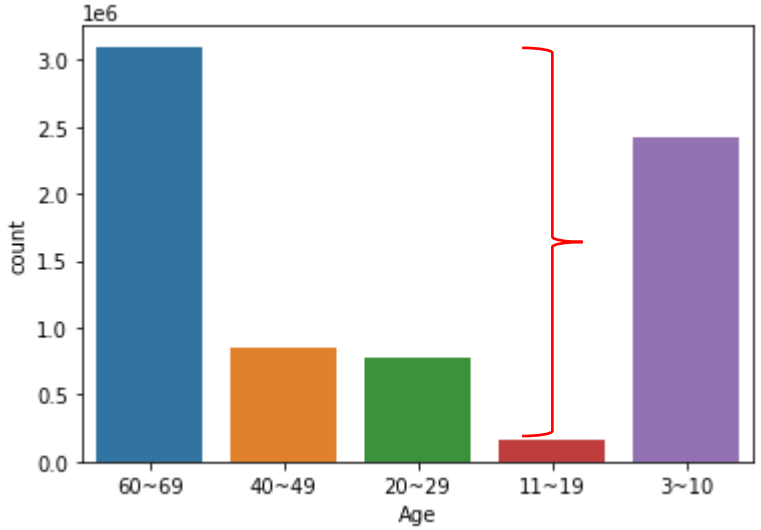


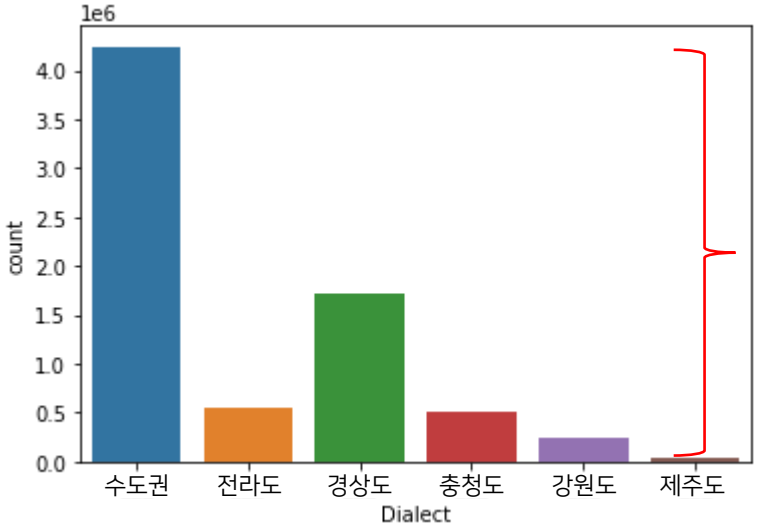
Age	
3-10	유년
11-19	소년
20-29	청년
30-39	
40-49	중장년
50-59	
60-69	노년
Over 70	

02 시행착오 및 최종 input

시행착오

훈련 데이터 균등 샘플링





연령, 방언 label에 불균형이 심해 모델의 학습에 지장이 있을 것이라 판단

02 시행착오 및 최종 input

시행착오

훈련 데이터 균등 샘플링

data → groupby →

			title	File_Path
Gender	Age	Dialect		
Female	11-19	수도권	n_0879-12001-02-01	Training/[원천]1.AI비서_원천...
Female	11-19	수도권	...	...
Female	20-39	수도권	zzpw4036_1-870081(1)	Training/[원천]3.비정형_원...
Female	20-39	수도권	...	...
...	...	...	...	...

→ sampling → train data

데이터를 각 label마다 균등하게 샘플링하여 훈련 데이터로 이용

02 시행착오 및 최종 input

시행착오

1초 미만 데이터 제거

데이터의 길이가 너무 짧은 경우
MFCC와 F0이 원활하게 추출 불가

+

사전 연구

“ 0.5초 ~ 1.0초 구간은 잡음이 섞인 데이터도 많기 때문에 본 연구의 실험 데이터로 쓰기 적절하지 않다.
따라서, 본 연구 데이터는 1.0초 이상의 데이터를 대상으로 실험을 진행한다. … ”

→ 1초 미만 데이터 제거

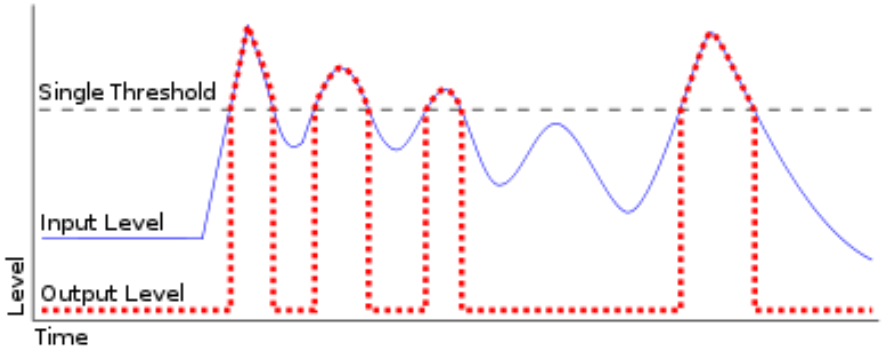
기계학습 기반의 연령분류
-서울대학교 대학원, 언어학과 언어학 전공 이주영

02 시행착오 및 최종 input

시행착오

Noise 제거 : noise reduce

고정 소음 감소(Stationary Noise Reduction)

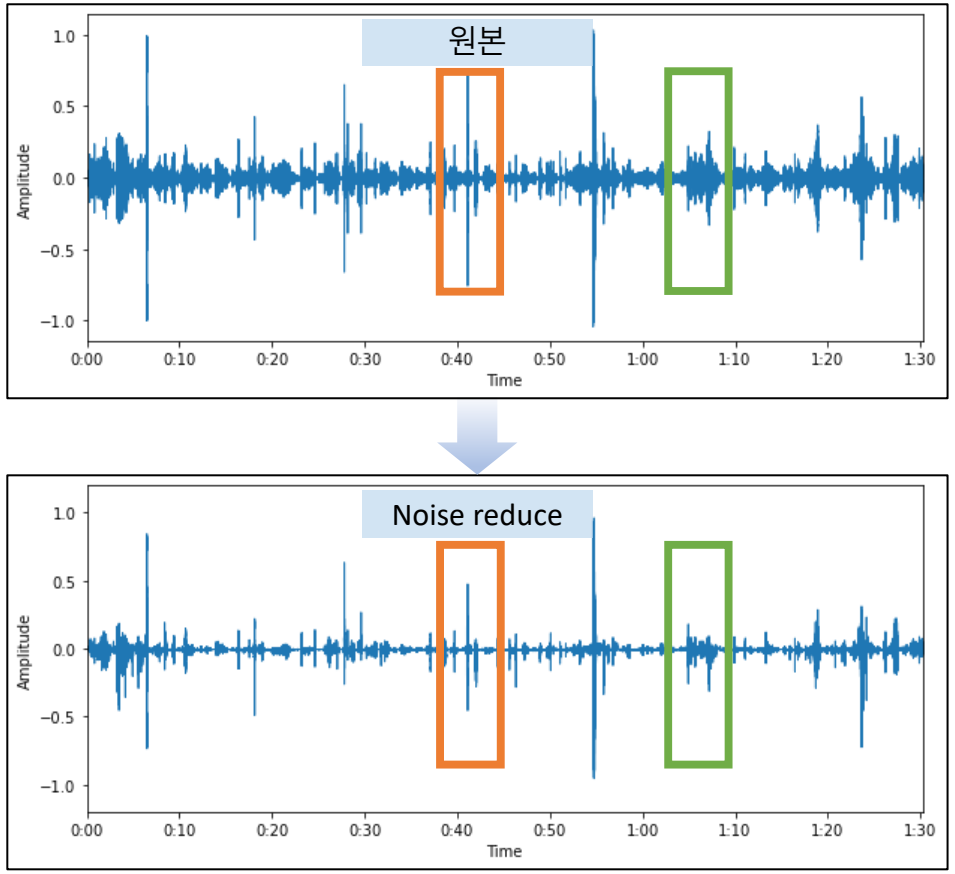


임계값을 넘지 못하는 신호는 차단하는 방식

02 시행착오 및 최종 input

시행착오

Noise 제거 : noise reduce



Noise도 작아졌지만 음성이 더 많이 감소

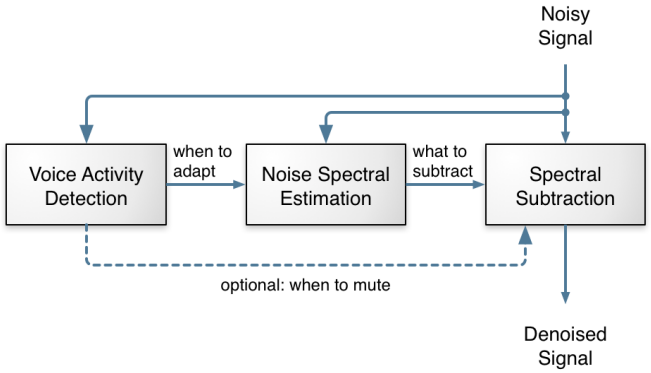
Noise reduce 사용 X

02 시행착오 및 최종 input

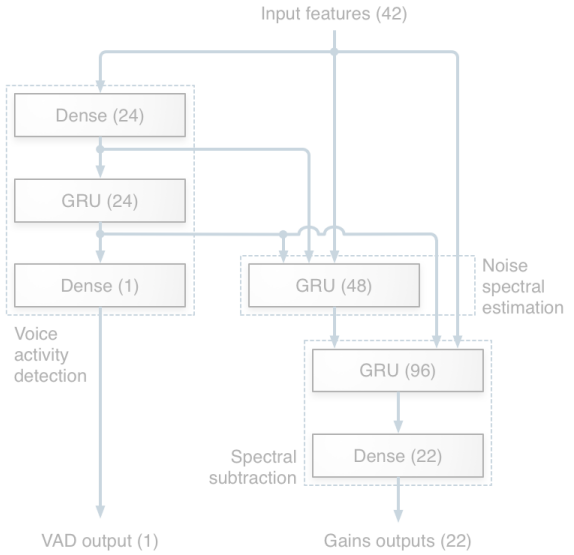
시행착오

Noise 제거 : RNNoise

conventional noise suppression algorithm



neural network for noise suppression



신호를 받아 관심 있는 음성에 최소한의 왜곡을 일으키며 잡음을 제거하는 알고리즘

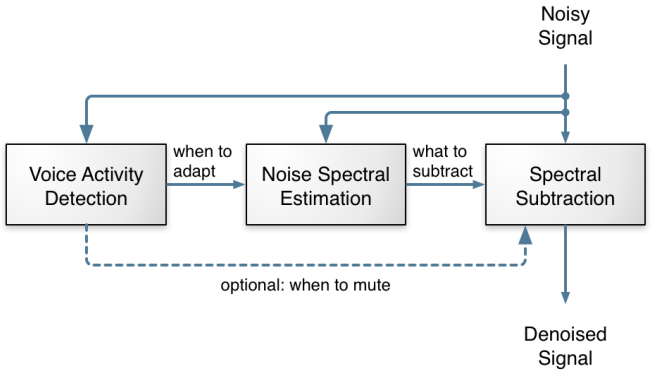
DeepVoice

02 시행착오 및 최종 input

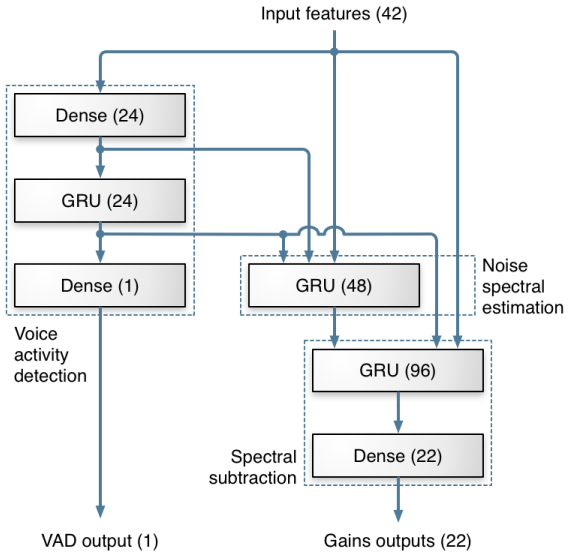
시행착오

Noise 제거 : RNNoise

conventional noise suppression algorithm



neural network for noise suppression



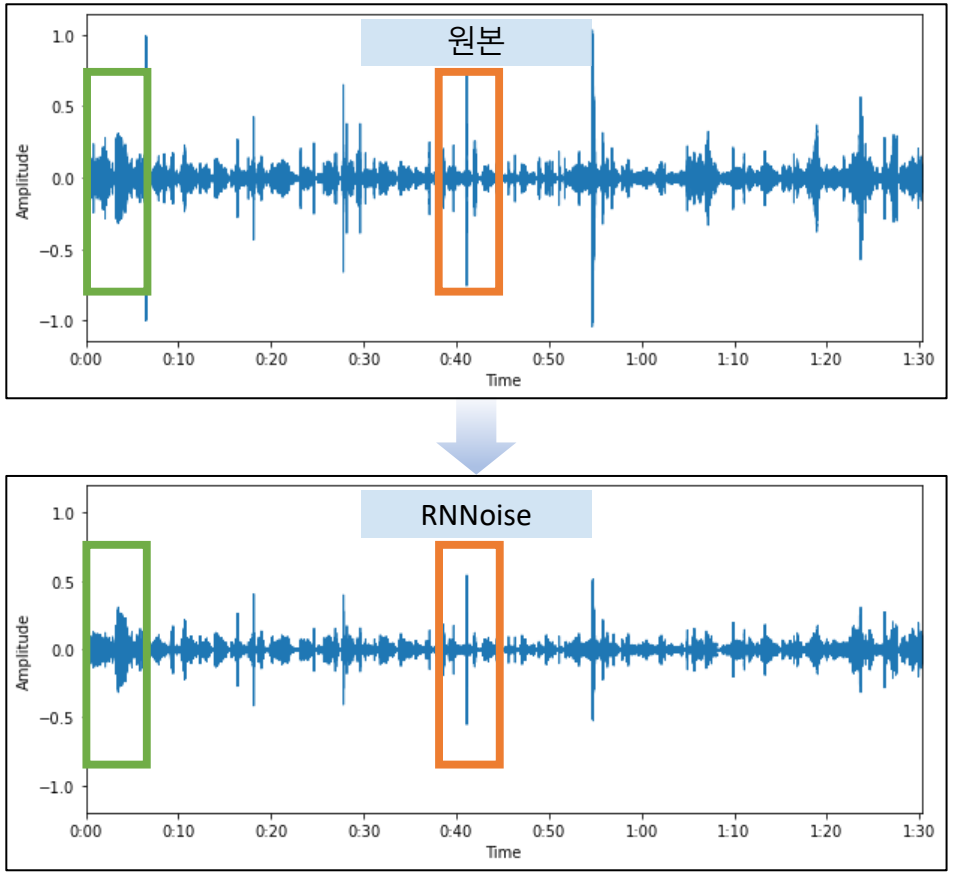
noise suppress algorithm에 딥러닝 적용

DeepVoice

02 시행착오 및 최종 input

시행착오

Noise 제거 : RNNoise



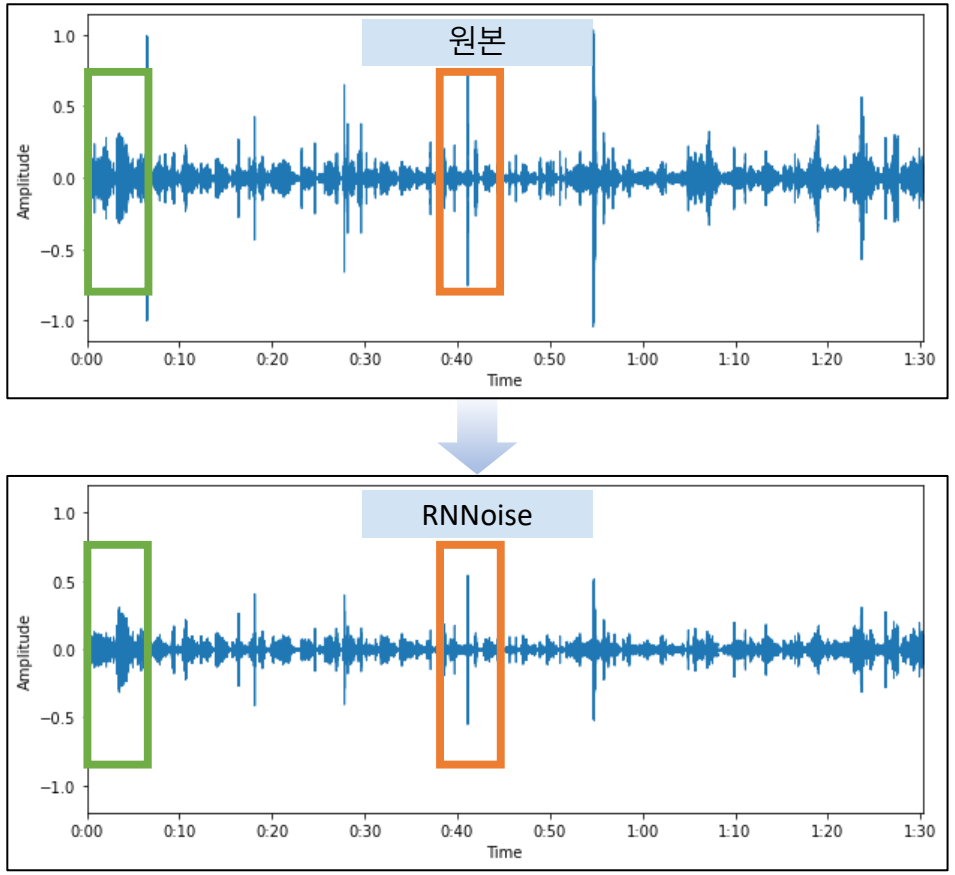
음성은 거의 그대로이지만 Noise는 작아짐

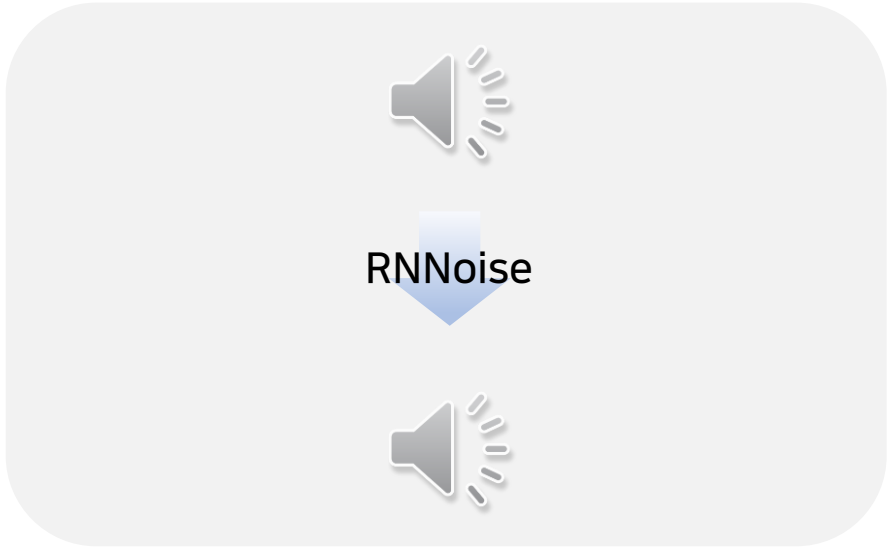
RNNoise 사용

02 시행착오 및 최종 input

시행착오

Noise 제거 : RNNoise





DeepVoice

02 시행착오 및 최종 input

시행착오

Noise 제거 : RNNoise

validation accuracy

	연령 분류 딥러닝	성별 분류 딥러닝	방언 분류 딥러닝	Logistic Regression
노이즈 미제거	40.72	81.07	50.34	0.85
노이즈 제거	43.24	82.13	50.27	0.65

약간의 성능 향상

큰 성능 하락

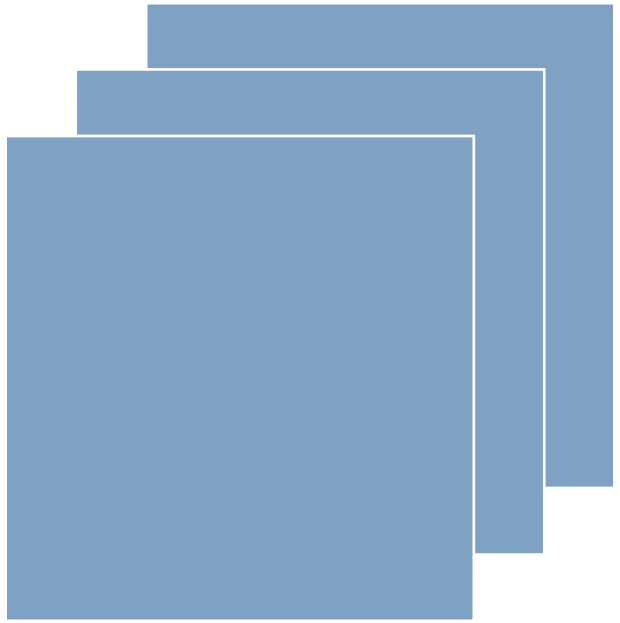
노이즈 제거 + 딥러닝 모델을 사용

성별 분류에 유용한 F0 주파수를 딥러닝 모델의 입력에 추가

02 시행착오 및 최종 input

최종 input

최종 Input



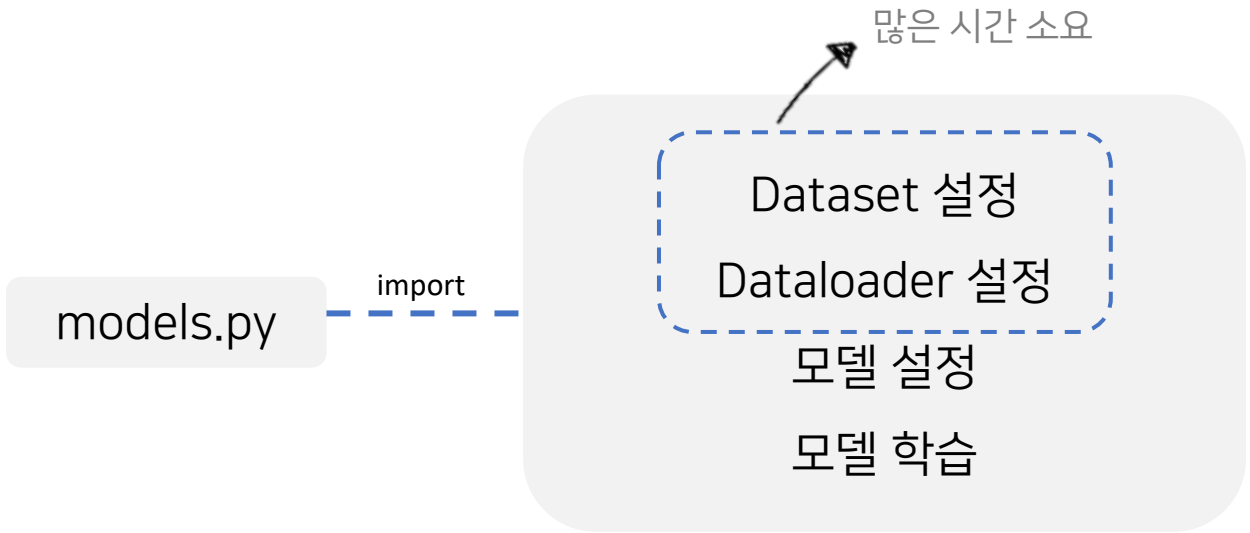
[3, 14, 400]

MFCC + F0 / MFCC + F0의 1차 차분 / MFCC + F0의 2차 차분

02 시행착오 및 최종 input

시행착오

모듈 reload



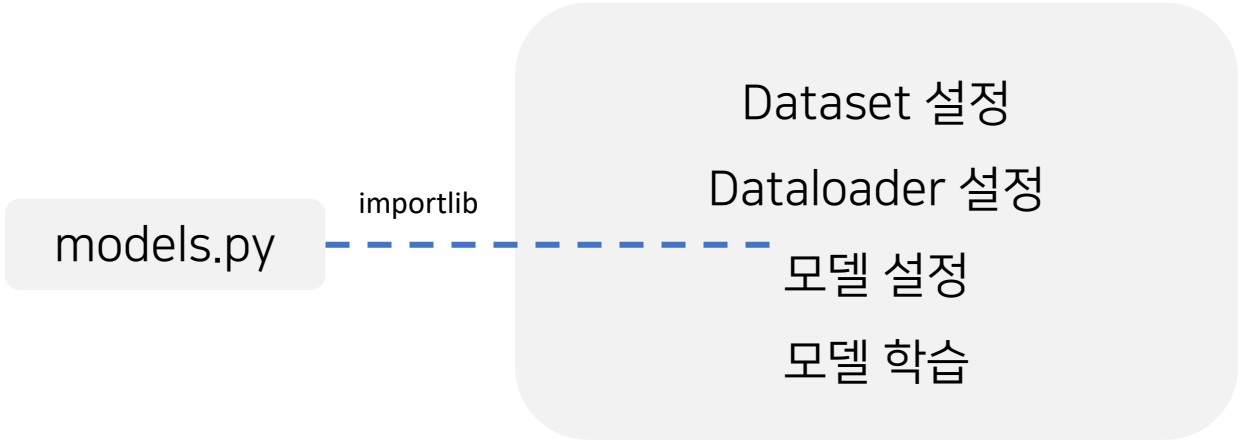
모델 수정 시 다시 import 해주기 위해 kernel을 재시작 필요

데이터를 다시 불러오며 많은 시간 소요

02 시행착오 및 최종 input

시행착오

모듈 reload



모델 설정 이전에 importlib을 이용해 모듈 reload

한 번 불러온 데이터로 여러 차례 모델을 수정하며 실험 가능

03

최종 모델링

03 최종 모델링

모델 연결

필요성

우리가 예측하려 하는 label이

다른 label의 예측에 도움을 줄 수 있지 않을까?

음성 외에 아무 정보도 주어지지 않은 상태의 방언 분류

vs

음성과 함께 60대 남성이라는 정보가 주어진 상태의 방언 분류

03 최종 모델링

모델 연결

방법

연령 분류 모델

- 유년 성별 분류 모델
  - 유년 남성 방언 분류 모델
  - 유년 여성 방언 분류 모델
- 소년 성별 분류 모델
  - 소년 남성 방언 분류 모델
  - 소년 여성 방언 분류 모델
- 청년 성별 분류 모델
  - 청년 남성 방언 분류 모델
  - 청년 여성 방언 분류 모델
- 중장년 성별 분류 모델
  - 중장년 남성 방언 분류 모델
  - 중장년 여성 방언 분류 모델
- 노년 성별 분류 모델
  - 노년 남성 방언 분류 모델
  - 노년 여성 방언 분류 모델

03 최종 모델링

모델 연결

방법

연령 분류 모델

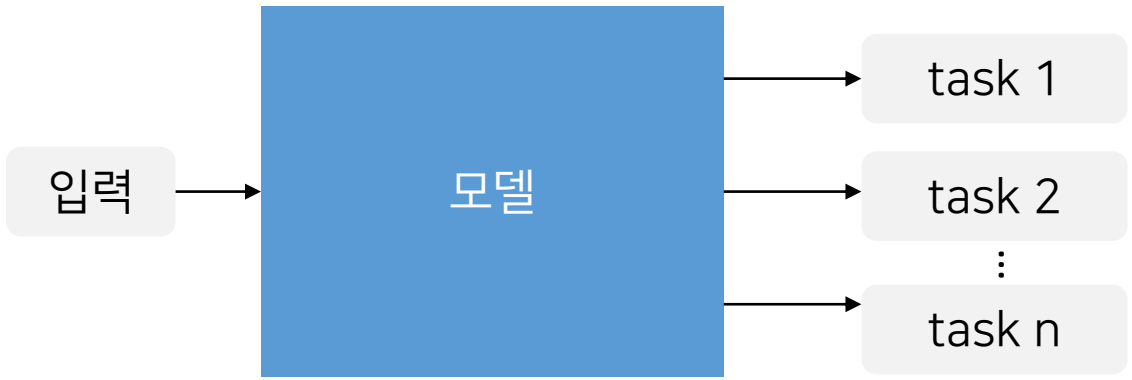
유년 성별 분류 모델
유년 남성 방언 분류 모델
유년 여성 방언 분류 모델
소년 성별 분류 모델
소년 남성 방언 분류 모델
소년 여성 방언 분류 모델
청년 성별 분류 모델
청년 남성 방언 분류 모델
청년 여성 방언 분류 모델
중장년 성별 분류 모델
중장년 남성 방언 분류 모델
중장년 여성 방언 분류 모델
노년 성별 분류 모델
노년 남성 방언 분류 모델
노년 여성 방언 분류 모델

시간적, 공간적 비용이 크기 때문에 탈락

03 최종 모델링

모델 연결

multitask learning



장점

- Knowledge Transfer

task 1을 학습하며 얻은 정보가

다른 task를 해결하는 데 도움

- Overfitting 감소

여러 task를 동시에 해결해야 하기 때문에

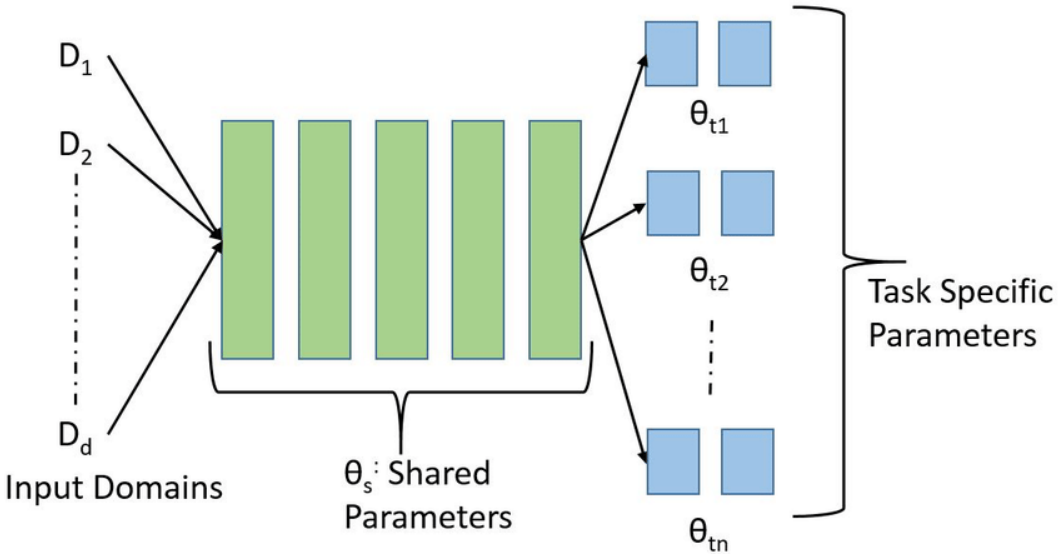
보다 일반화 된 feature를 추출하도록 학습

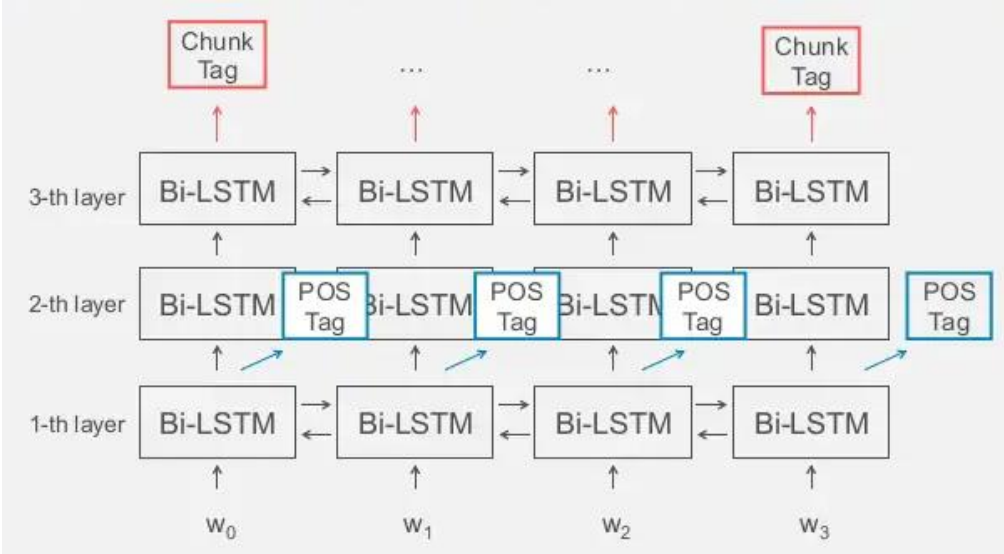
03 최종 모델링

모델 연결

개요

Hard Parameter Sharing	LSTM Base MTL



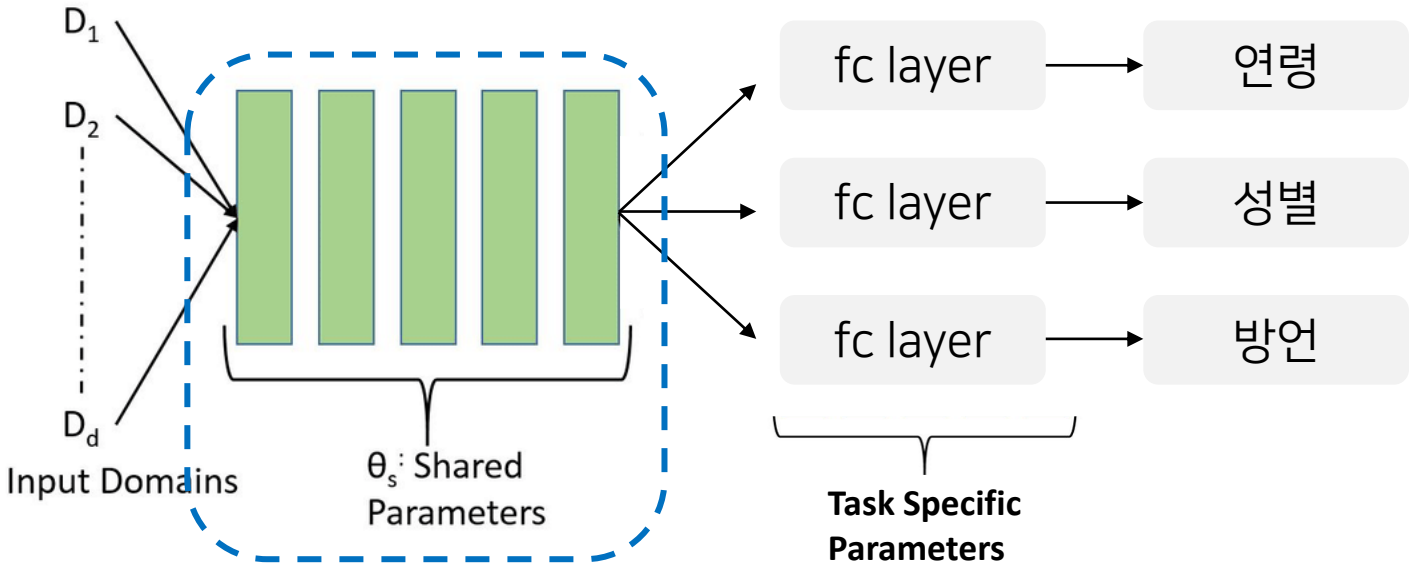


DeepVoice

03 최종 모델링

모델 연결

Hard Parameter Sharing



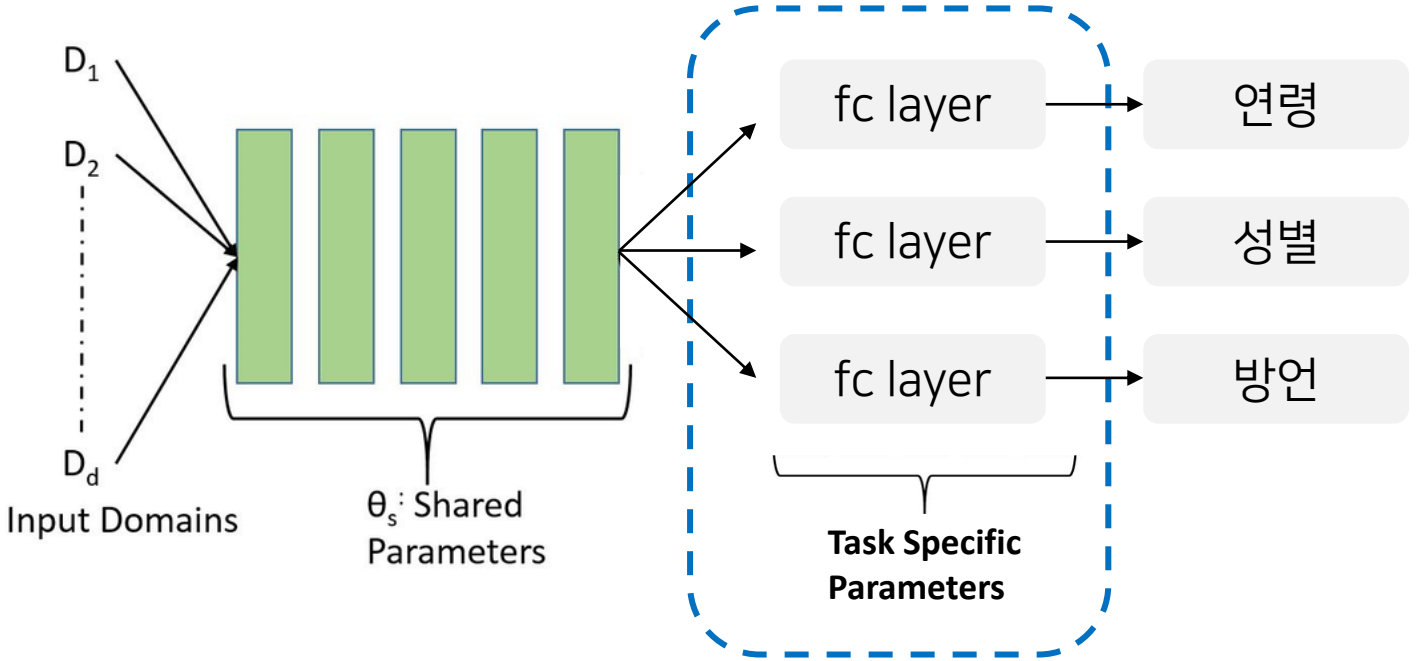
각 task에 대해 입력에서 공통된 feature 추출

Multitask Learning – Caruana, R. (1997)

03 최종 모델링

모델 연결

Hard Parameter Sharing



뽑아낸 feature에서 각 task 해결에 적합한 feature 추출

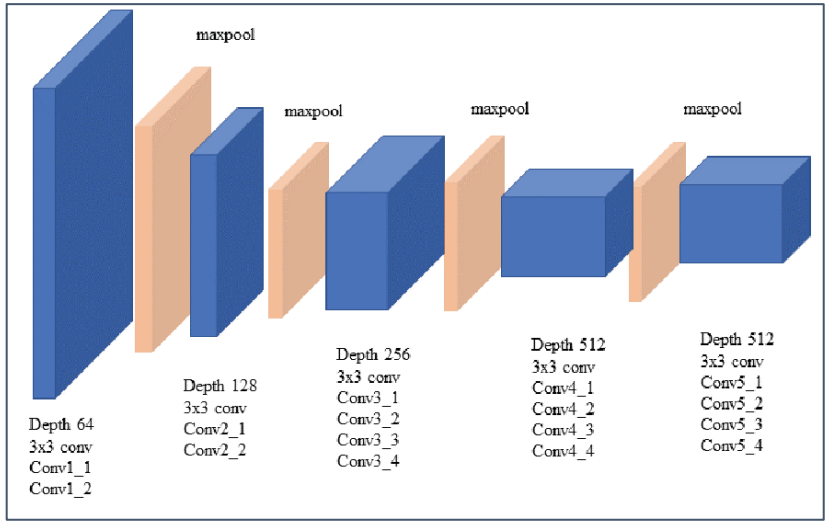
Multitask Learning – Caruana, R. (1997)

03 최종 모델링

모델 연결

CLSTM

① CNN 모델 → 특징값 추출

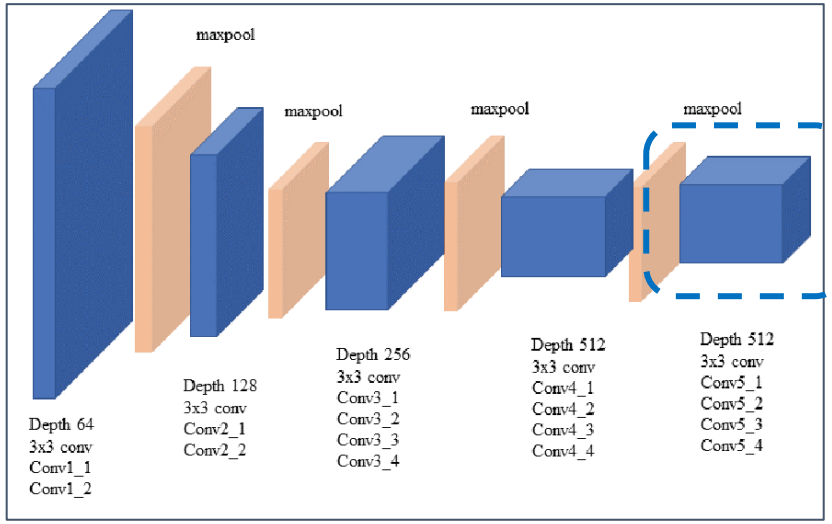


03 최종 모델링

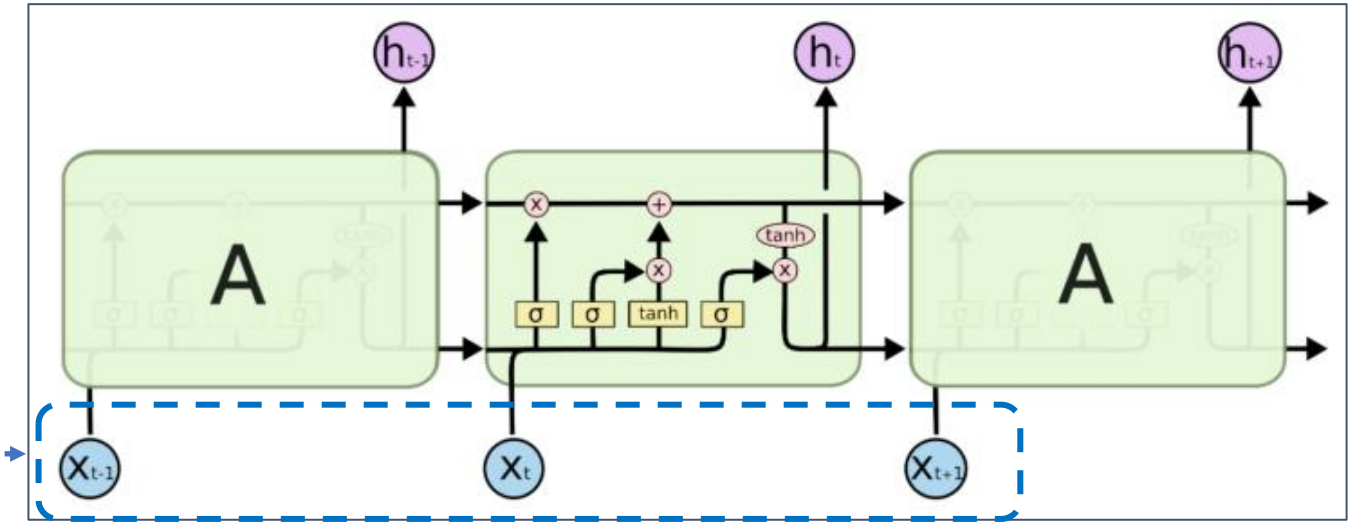
모델 연결

CLSTM

① CNN 모델 → 특징값 추출



② CNN 모델 Output → 1st LSTM의 Input

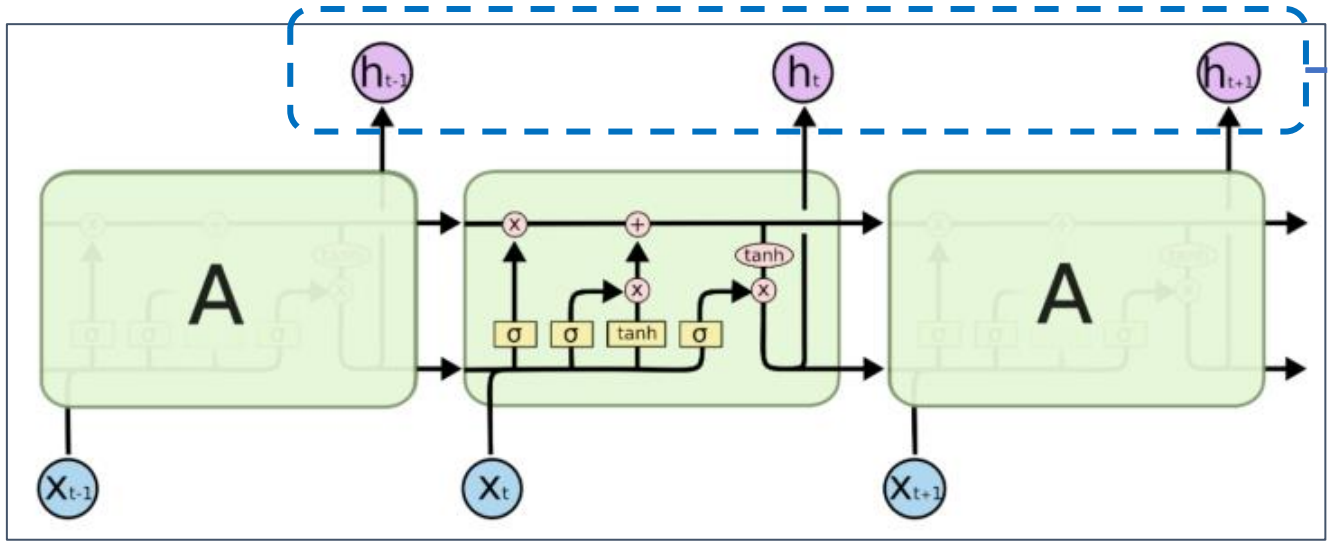


03 최종 모델링

모델 연결

CLSTM

② CNN 모델 Output → 1st LSTM의 Input



Loss 저장

연령 분류

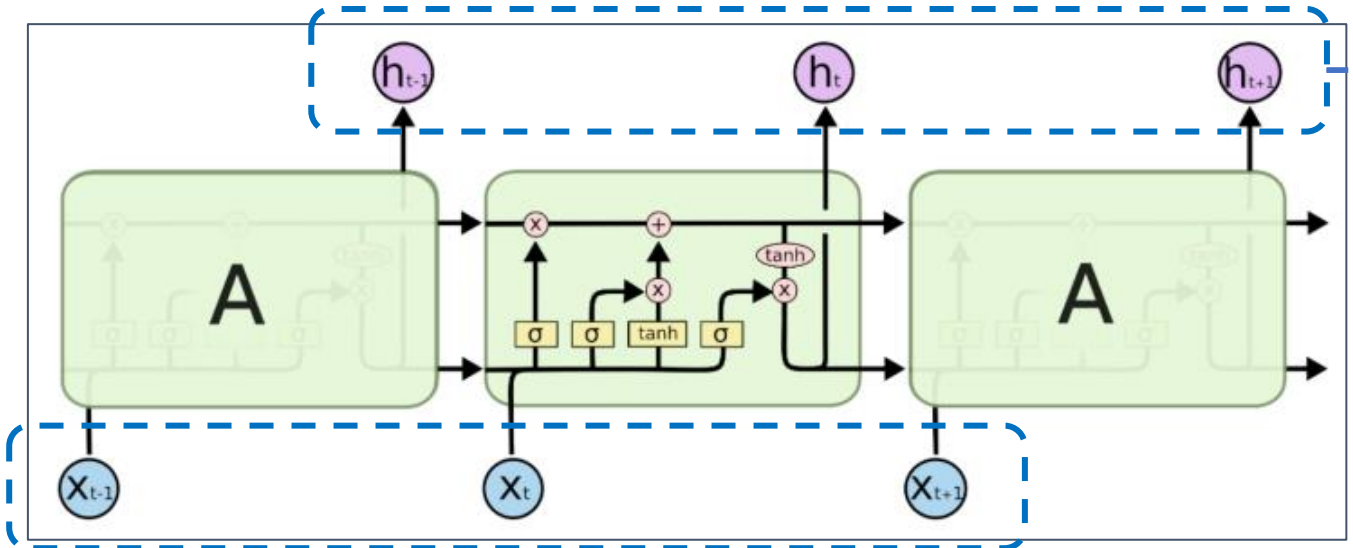
Hidden Layer의 Output
다음 LSTM으로 전달

03 최종 모델링

모델 연결

CLSTM

③ 1st LSTM의 Output → 2nd LSTM의 Input



Loss 저장

성별 분류

Hidden Layer의 Output
다음 LSTM으로 전달

이전 LSTM의 Hidden Output

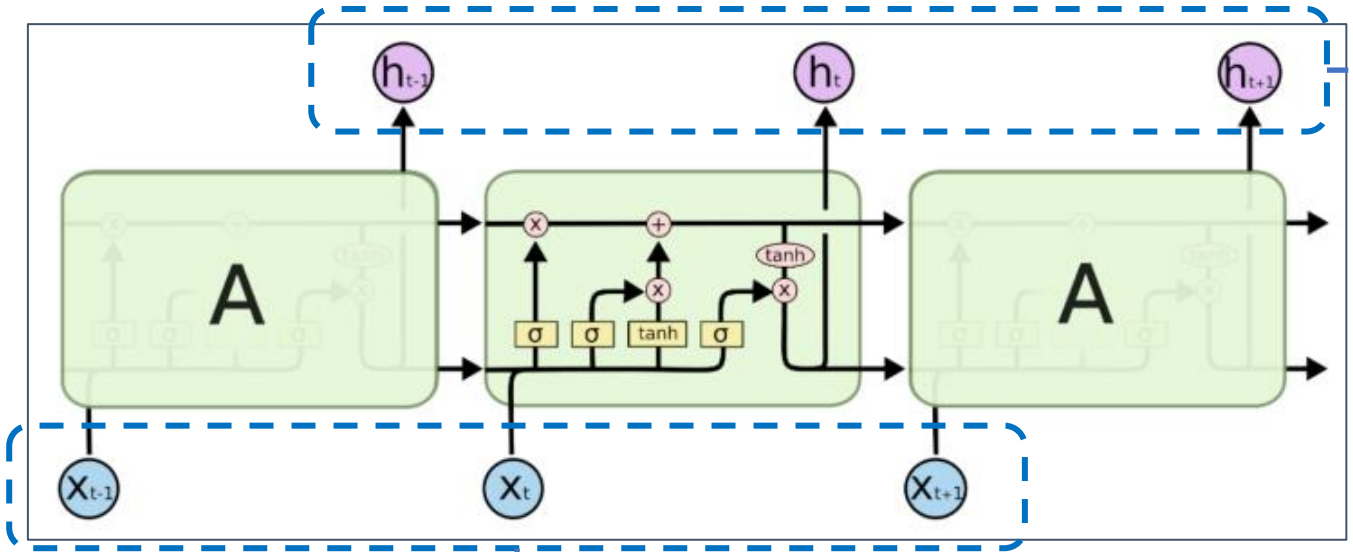
DeepVoice

03 최종 모델링

모델 연결

CLSTM

④ 2nd LSTM의 Output → 마지막 LSTM의 Input



Loss 저장

방언 분류

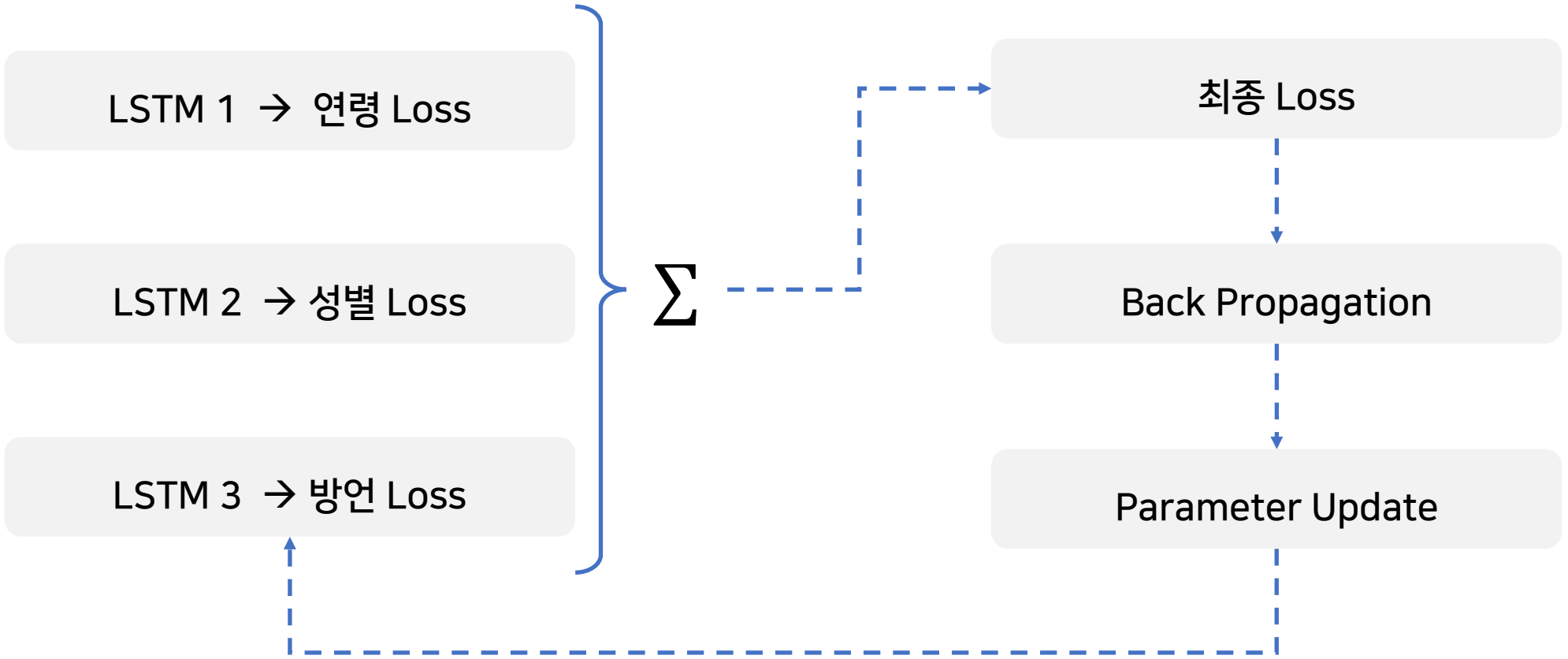
이전 LSTM의 Hidden Output

DeepVoice

03 최종 모델링

모델 연결

CLSTM



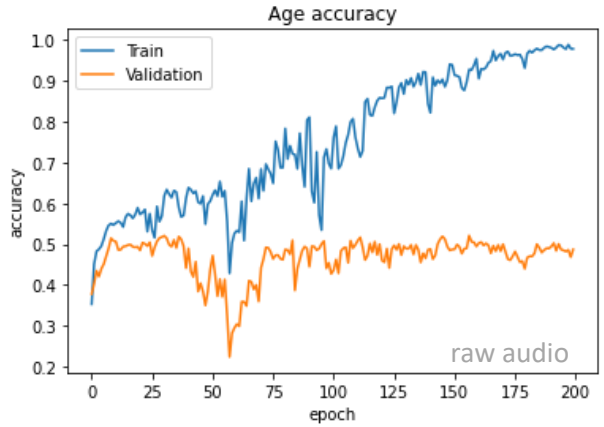
04

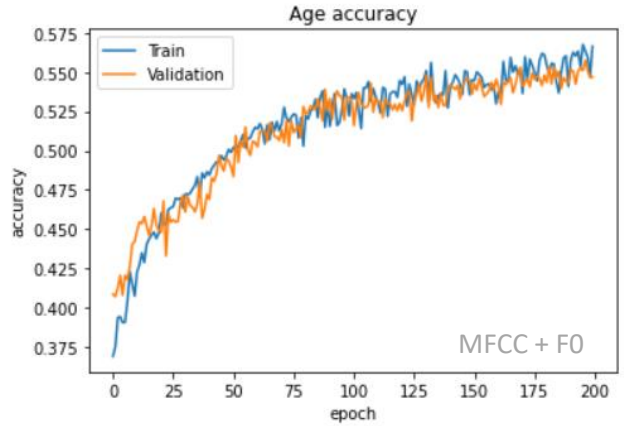
결과 및 의의

04 결과 및 의의

Hard Parameter Sharing

주제별 Accuracy - 연령





raw audio 이용 시 train acc만 우상향

MFCC + F0 이용 시 두 acc 모두 우상향

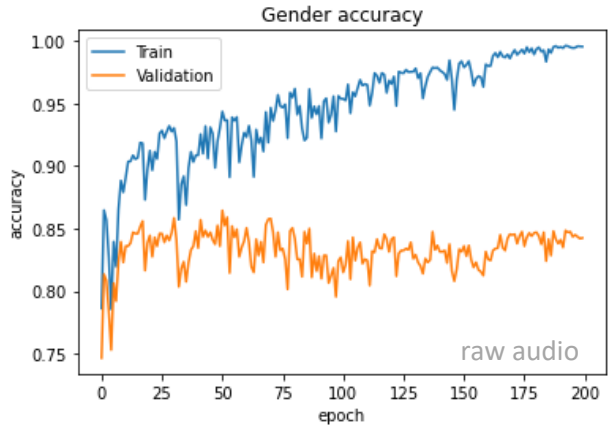
raw audio에 비해 MFCC + F0 데이터 이용 시 더욱 원활한 학습 양상

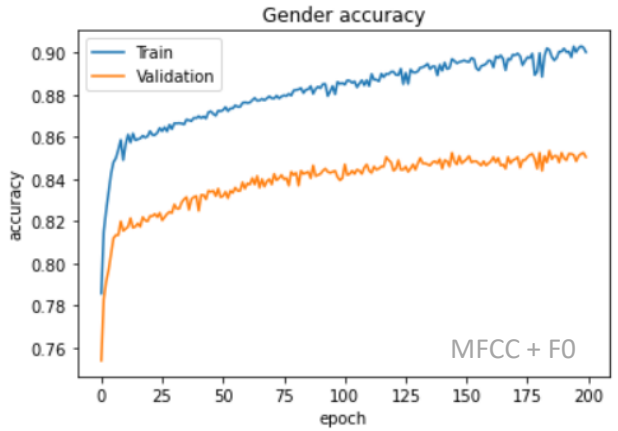
DeepVoice

04 결과 및 의의

Hard Parameter Sharing

주제별 Accuracy - 성별





raw audio 이용 시 train acc만 우상향

MFCC + F0 이용 시 두 acc 모두 우상향

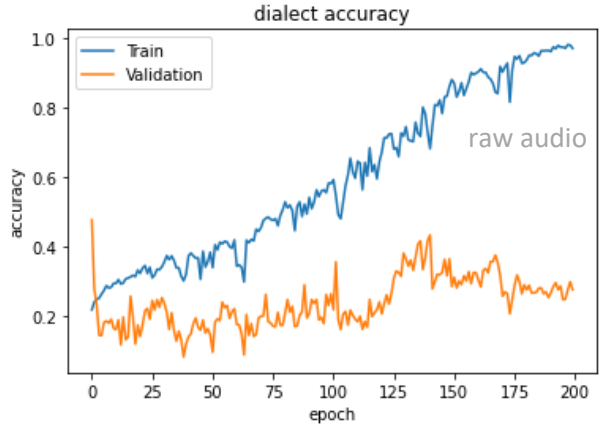
raw audio에 비해 MFCC + F0 데이터 이용 시 더욱 원활한 학습 양상

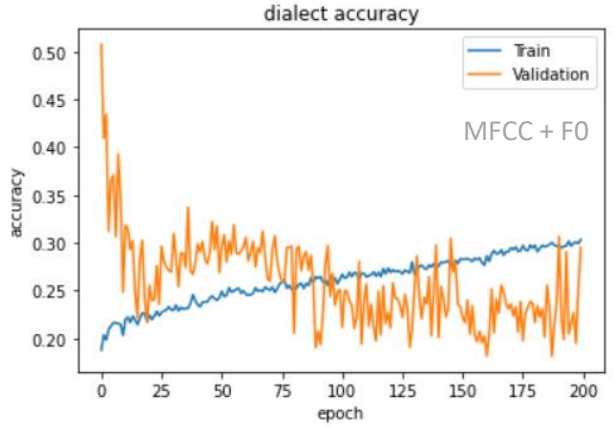
DeepVoice

04 결과 및 의의

Hard Parameter Sharing

주제별 Accuracy – 방언





어떤 데이터를 사용하는 지와 관계 없이
train acc는 우상향, validation acc는 수렴

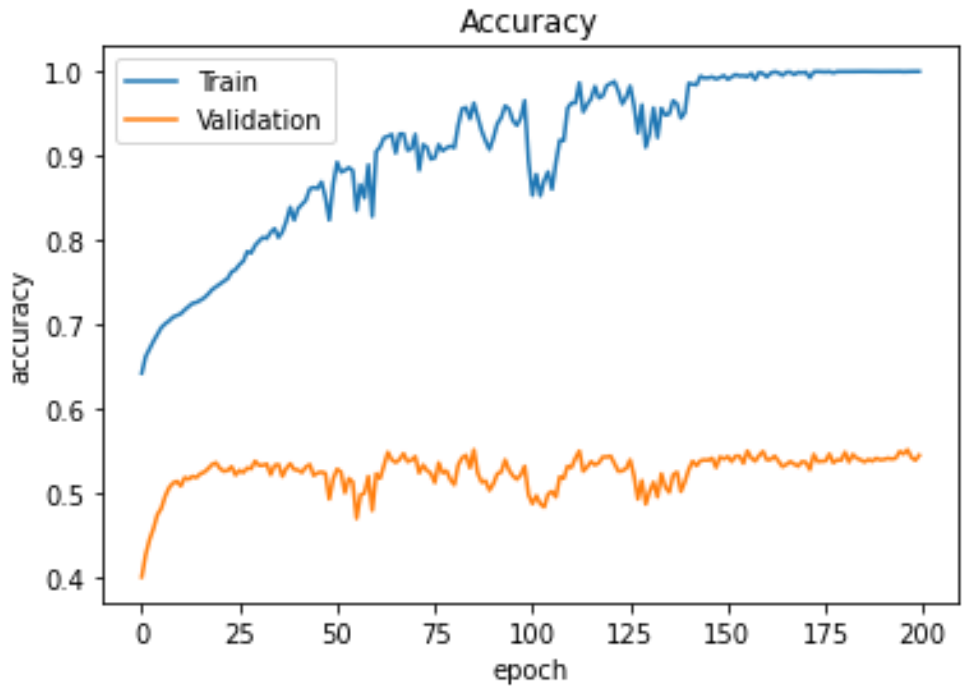
raw audio에 비해 MFCC + F0 데이터 이용 시
더욱 원활한 학습 양상

DeepVoice

04 결과 및 의의

CLSTM

주제별 Accuracy - 연령



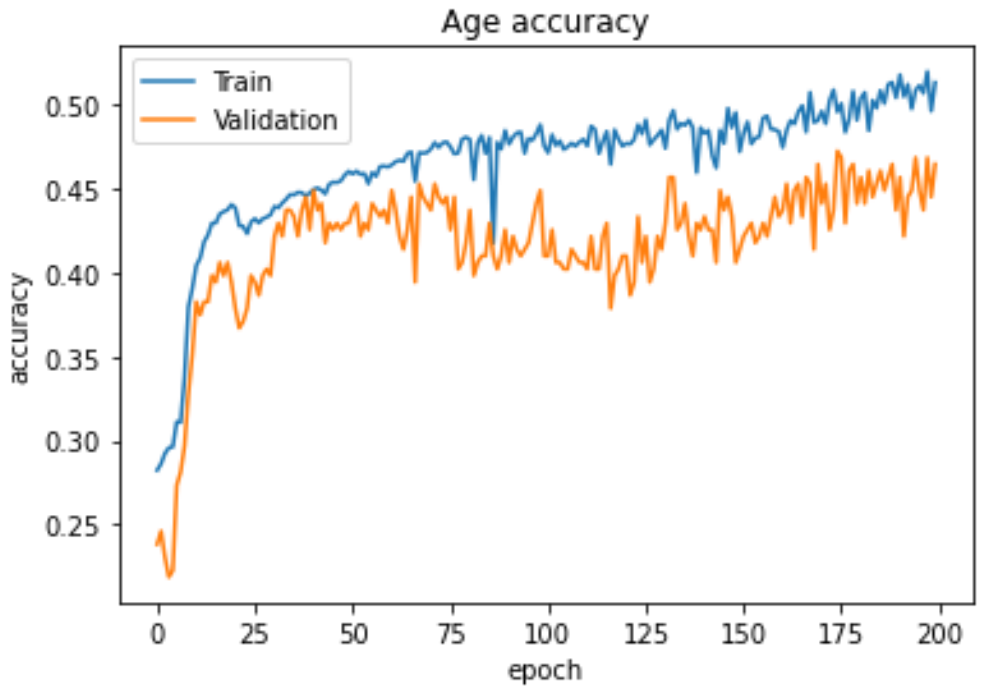
CNN 기반 연령 베이스 라인 모델

Validation Accuracy : 약 55%

04 결과 및 의의

CLSTM

주제별 Accuracy - 연령



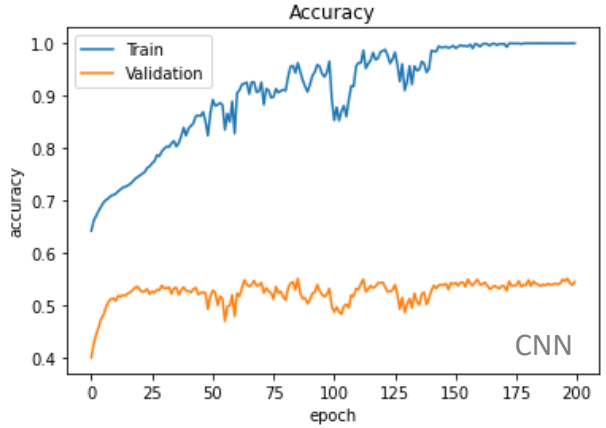
CLSTM 모델의 연령 분류

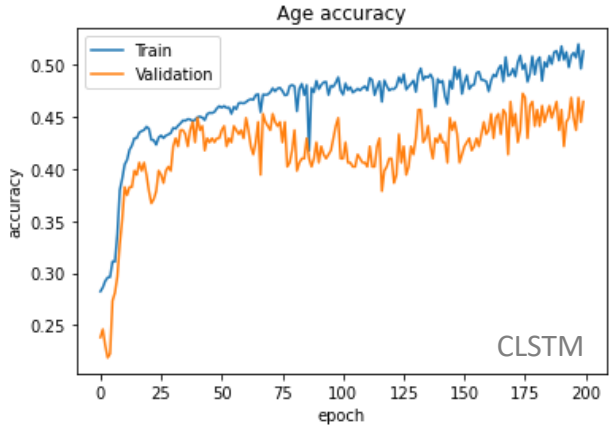
Validation Accuracy : 약 45%

04 결과 및 의의

CLSTM

주제별 Accuracy - 연령





① 연령 분류 성능 ↓

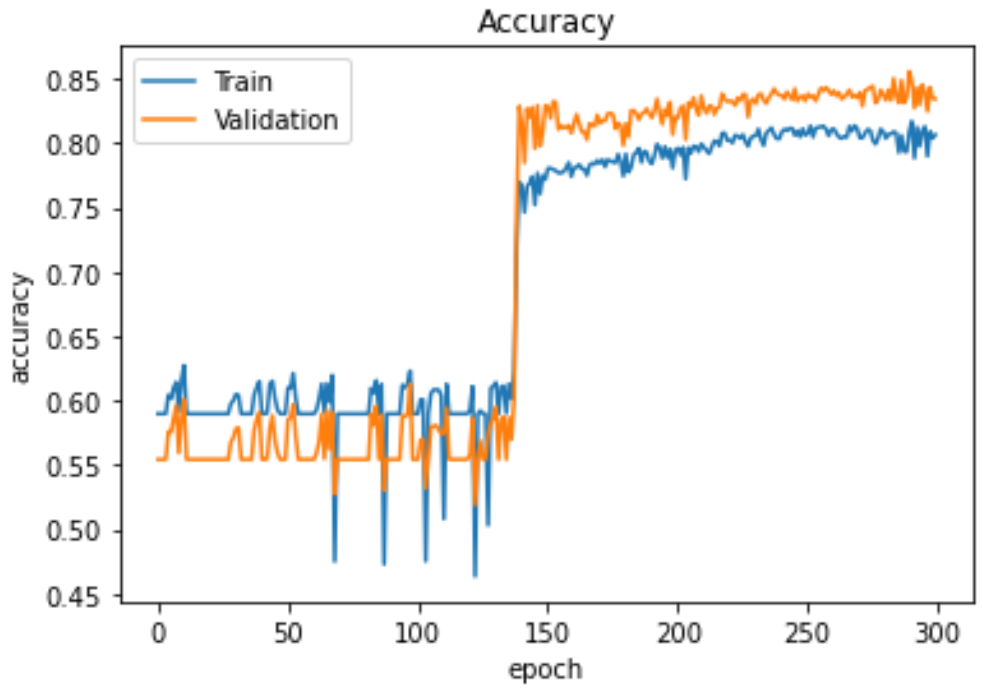
② Validation Accuracy 상승폭 안정화

→ Train Accuracy와 유사한 속도로 향상

04 결과 및 의의

CLSTM

주제별 Accuracy - 성별



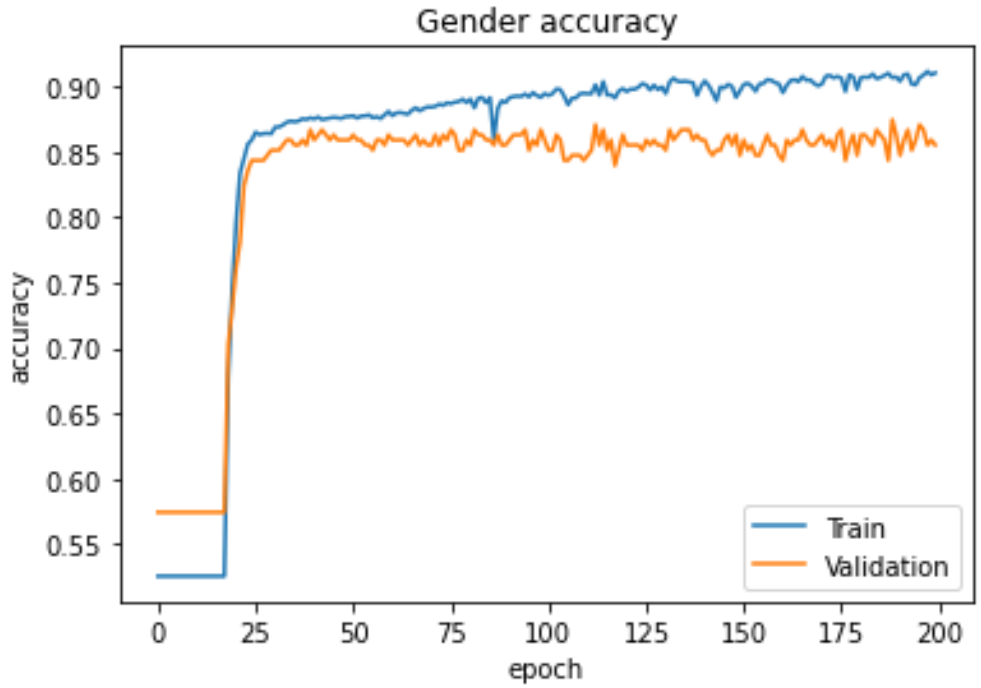
CNN + LSTM 모델의 성별 분류

Validation Accuracy : 약 84%

04 결과 및 의의

CLSTM

주제별 Accuracy - 성별



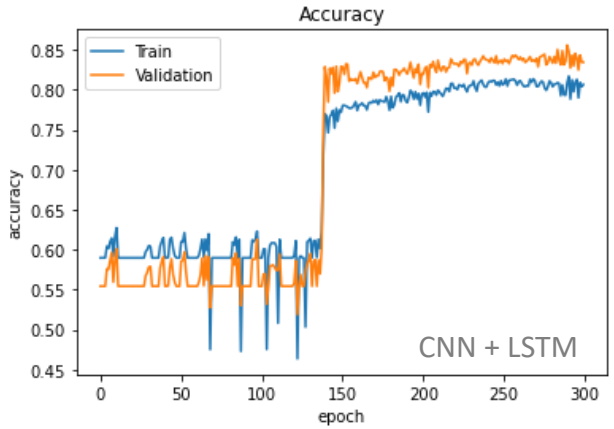
CLSTM 모델의 연령 분류

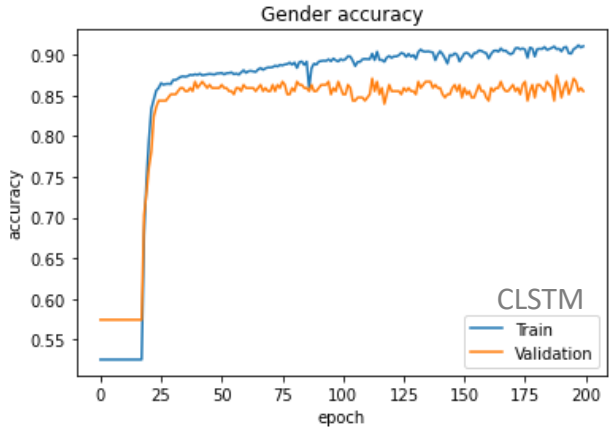
Validation Accuracy : 약 86%

04 결과 및 의의

CLSTM

주제별 Accuracy - 성별





① 성별 분류 성능 ↑

② Validation Accuracy 상승폭 안정화

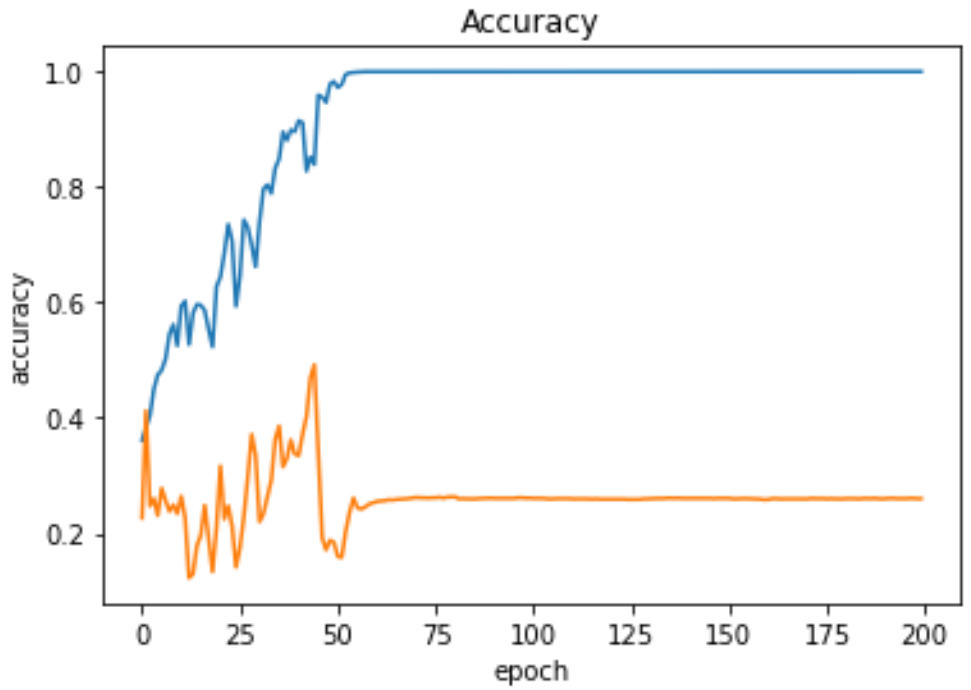
→ 빠른 속도로 수렴

DeepVoice

04 결과 및 의의

CLSTM

주제별 Accuracy - 방언



CNN 모델의 방언 분류

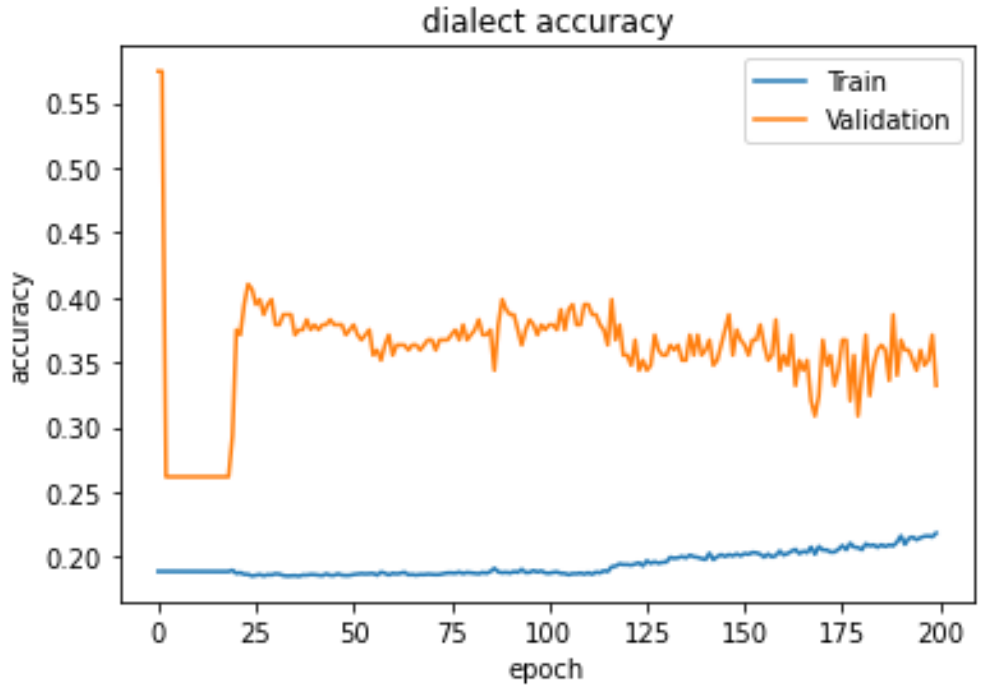
Validation Accuracy : 약 30% 내외

(초반 변동폭 ↑)

04 결과 및 의의

CLSTM

주제별 Accuracy - 방언



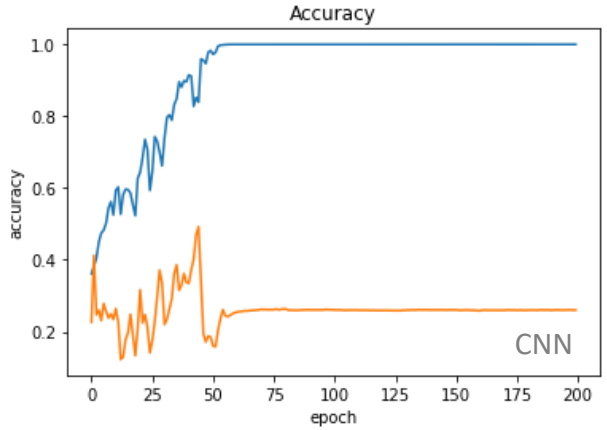
CLSTM모델의 방언 분류

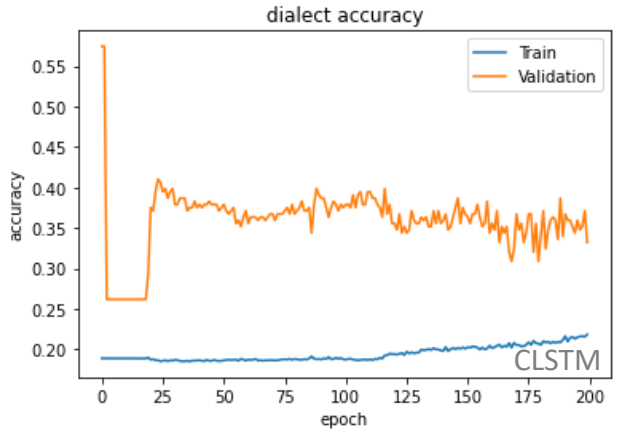
Validation Accuracy : 약 35%

04 결과 및 의의

CLSTM

주제별 Accuracy - 방언





① 방언 분류 성능 ↑

② Train Accuracy의 상승 저조

→ 학습이 적절하게 이루어지지 않음

DeepVoice

04 결과 및 의의

최종 결과

최종 결과 정리

모델 유형	연령	성별	방언
MFCC + F0 Hard Parameter Sharing	train acc : 0.57 val acc : 0.55	train acc : 0.9 val acc : 0.84	train acc : 0.3 val acc : 0.29
CNN For raw-waveform Hard Parameter Sharing	train acc : 0.99 val acc : 0.5	train acc : 0.99 val acc : 0.85	train acc : 0.99 val acc : 0.28
CLSTM	train acc : 0.51 val acc : 0.45	train acc : 0.9 val acc : 0.85	train acc : 0.23 val acc : 0.35

04 결과 및 의의

한계 및 의의

한계

방언 데이터의 경우 직접 들어본 결과
사투리가 느껴지지 않음

시간 제약으로 인해 모든 데이터 활용X

방언, 연령 분류 모델링 성능 부족

웹 서비스 구현에 난항

04 결과 및 의의

한계 및 의의

의의

음성 데이터를 다루는 여러 방법 습득
(MFCC, Mel Spectrogram, F0 주파수)

모델 성능 향상을 위한 방법을
다방면으로 모색

원격 개발 환경에서 작업 진행
(docker, ubuntu, tmux)

교수님과의 면담을 통한
전문성 향상

04 결과 및 의의

한계와 의의

의의

그러나, 가장 큰 의의는

음성 데이터를 다루는 여러 방법 습득
(MFCC, Mel Spectrogram, F0 주파수)

원격 개발 환경에서 작업 진행
(docker, ubuntu, tmux)

모델 성능 향상을 위한 방법을
다방면으로 모색

교수님과의 면담을 통한
전문성 향상

딥러닝 팀의 성능 향상을 위한 고군분투는

아직 끝나지 않았다는 사실!이다

DeepVoice

감사합니다