

멀티모달 감정인식 모델

딥러닝팀

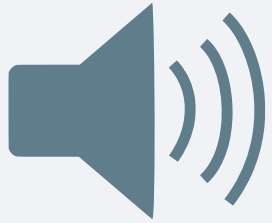
김예찬

윤지영

채소연

한지원

홍지우

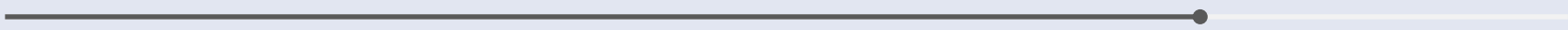


3주차 계획



트랜스포머 모델 학습

- 고안된 트랜스포머 모델을 학습시켜 음성 분석 모델과 텍스트 분석 모델과의 정확도 비교
- 우울증 데이터를 받아서 분석 - 우울증과 감정의 분포를 확인
- 음성인식 모델과 연동 / 음성을 텍스트화하여 트랜스포머 모델로 분석
- 플랫폼을 만들어서 직접 테스트할 수 있도록 만들 예정!!





트랜스포머 모델 학습

- 고안된 트랜스포머 모델을 학습시켜 음성 분석 모델과 텍스트 분석 모델과의 정확도 비교
- 우울증 데이터를 받아서 분석 - 우울증과 감정의 분포를 확인
- 음성인식 모델과 연동 / 음성을 텍스트화하여 트랜스포머 모델로 분석
- 플랫폼을 만들어서 직접 테스트할 수 있도록 만들 예정!!

더 이상.. 할 수 없게
되었습니다...



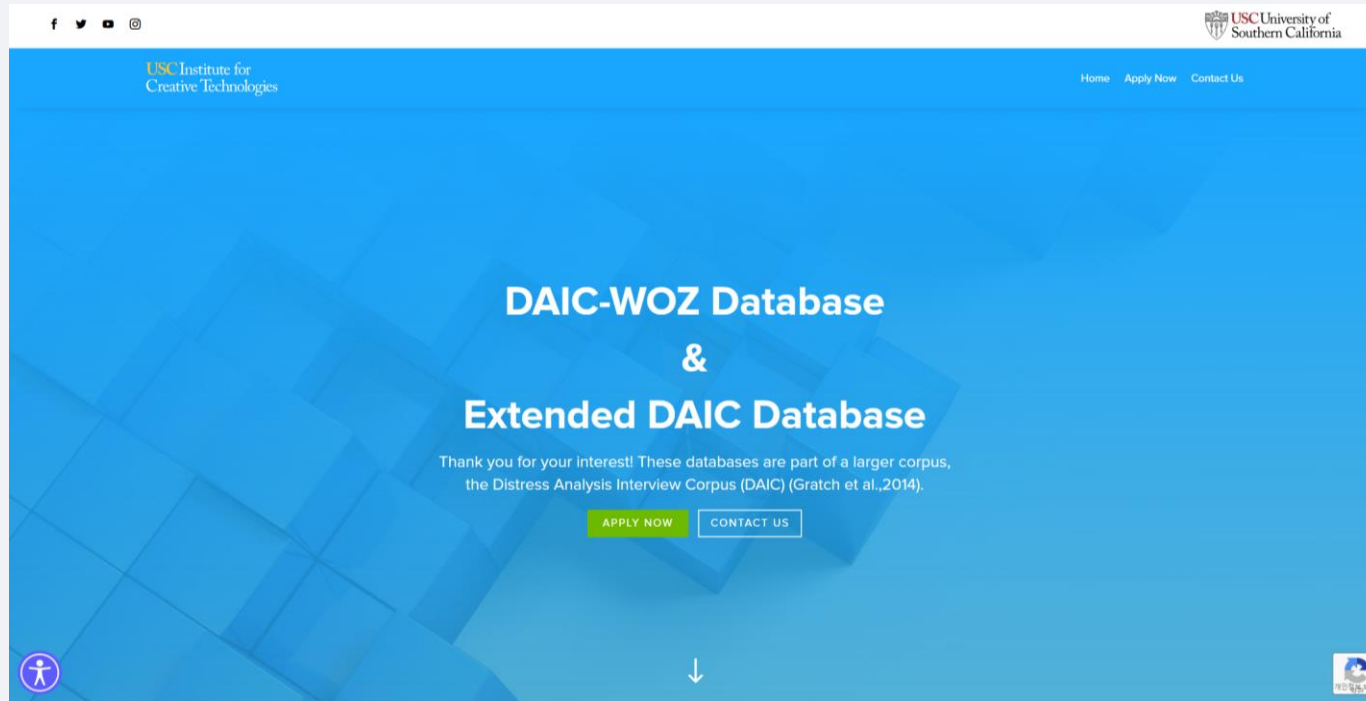


트랜스포머 모델 학습

- 고안된 트랜스포머 모델을 학습시켜 음성 분석 모델과 텍스트 분석 모델과의 정확도 비교
- 우울증 데이터를 받아서 분석 - 우울증과 감정의 분포를 확인
- 음성인식 모델과 연동 / 음성을 텍스트화하여 트랜스포머 모델로 분석
- 플랫폼을 만들어서 직접 테스트할 수 있도록 만들 예정!!



3주차 계획



우울증 데이터가 라벨링된
데이터를 제공하는 사이트
연구 목적/윤리에 맞게 사용
하겠다는 서약서 제출 필요

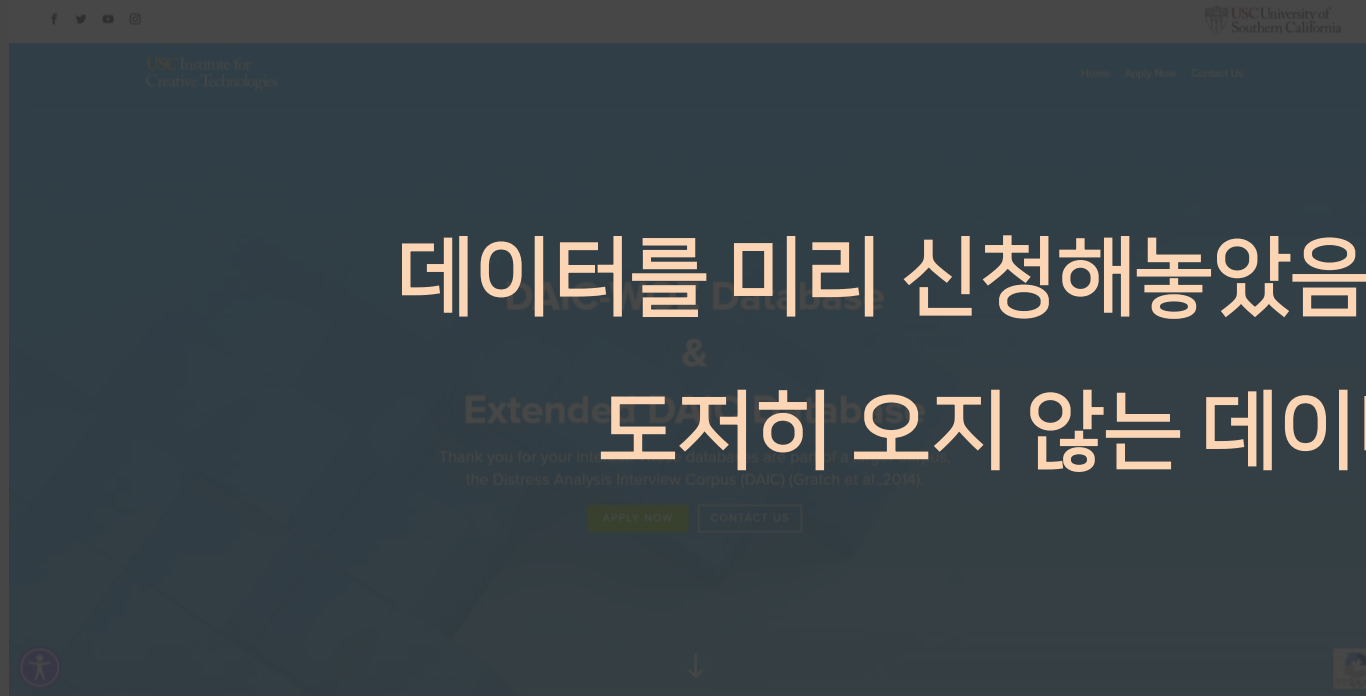


One Eternity Later

한 평생이 지난 후...



3주차 계획



데이터를 미리 신청해놓았음에도 불구하고
도저히 오지 않는 데이터파일...

우울증 데이터가 라벨링된

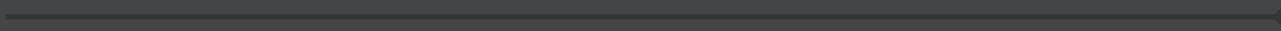
데이터를 제공하는 사이트

연구 목적/윤리에 맞게 사용

하겠다는 서약서 제출 필요

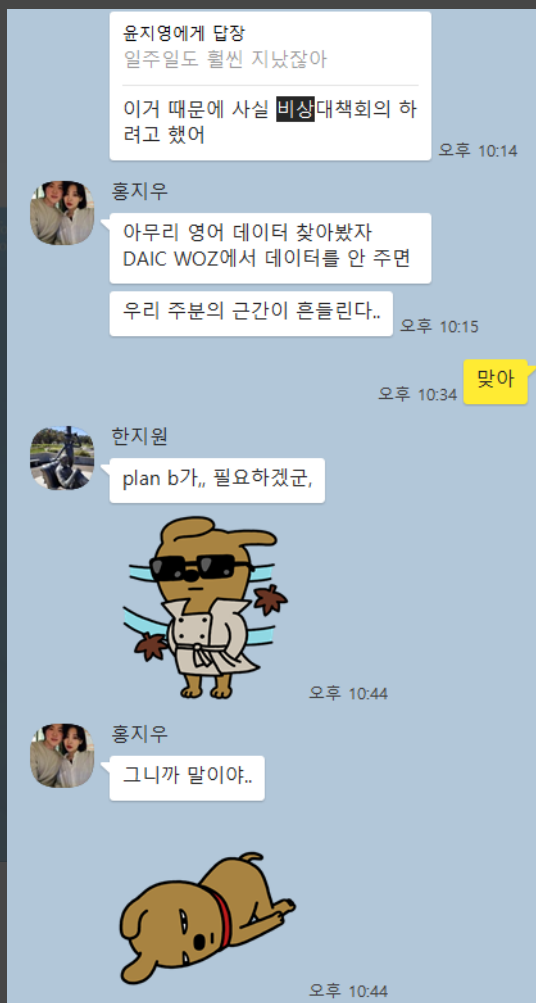


Deep Learning
-PSAT-





3주차 계획



결국...

비상대책회의를 열게되고...

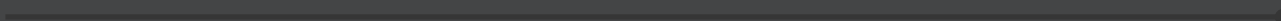
우울증 데이터가 라벨링된

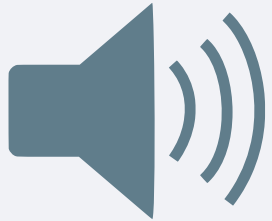
데이터를 제공하는 사이트

이 데이터를 윤리에 맞게 사용
하겠다는 서약서 제출 필요



Deep Learning
-PSAT-





0. 새로운 주제 소개



0. 새로운 주제 소개



2022 휴먼이해 인공지능 논문경진대회 참가



[대회안내] 2022 휴먼이해 인공지능 논문경진대회

본 대회는 한국전자통신연구원(ETRI)이 주최하고 과학기술정보통신부와 국가과학기술연구회(NST)가 후원합니다



주관/주최

에트리

보상

1,300만원

기간

22. 03. 16 ~ 22. 06. 30

참여인원

46



참여하기

2. 모집분야 II : 멀티모달 감정 데이터셋 활용 감정 인식 기술 분야

저번 주제 때 활용하던 멀티모달을 이용하여 대회에 참가하기로 결정



Deep Learning
-PSAT





0. 새로운 주제 소개



새로운 주제의 장단점

장점

대회이기 때문에 데이터셋이
기본적으로 다 주어져있다.
멀티모달 아이디어를 그대로
가져갈 수 있다.





0. 새로운 주제 소개



영어데이터의 특징

장점

대회이기 때문에 데이터셋이
기본적으로 다 주어져있다.
멀티모달 아이디어를 그대로
가져갈 수 있다.

단점

데이터의 수가 매우 불균형
주제를 새로 정한만큼 발표때까지
죽었다고 생각하고 해야함





0. 새로운 주제 소개



영어데이터의 특징

장점

대회이기 때문에 데이터셋이
기본적으로 다 주어져있다.
멀티모달 아이디어를 그대로
가져갈 수 있다.

단점

데이터의 수가 매우 불균형
주제를 새로 정한만큼 발표때까지
죽었다고 생각하고 해야함



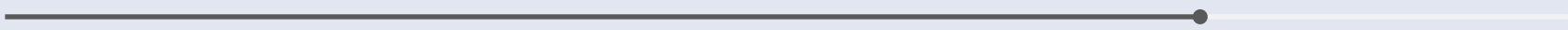


0. 새로운 주제 소개



최종 목표

- 데이터 분석, EDA를 통해 어떻게 진행할 지 확인
- 불균형한 데이터를 전처리
- 멀티모달을 이용하여 발화자의 감정을 예측하는 모델 만들기





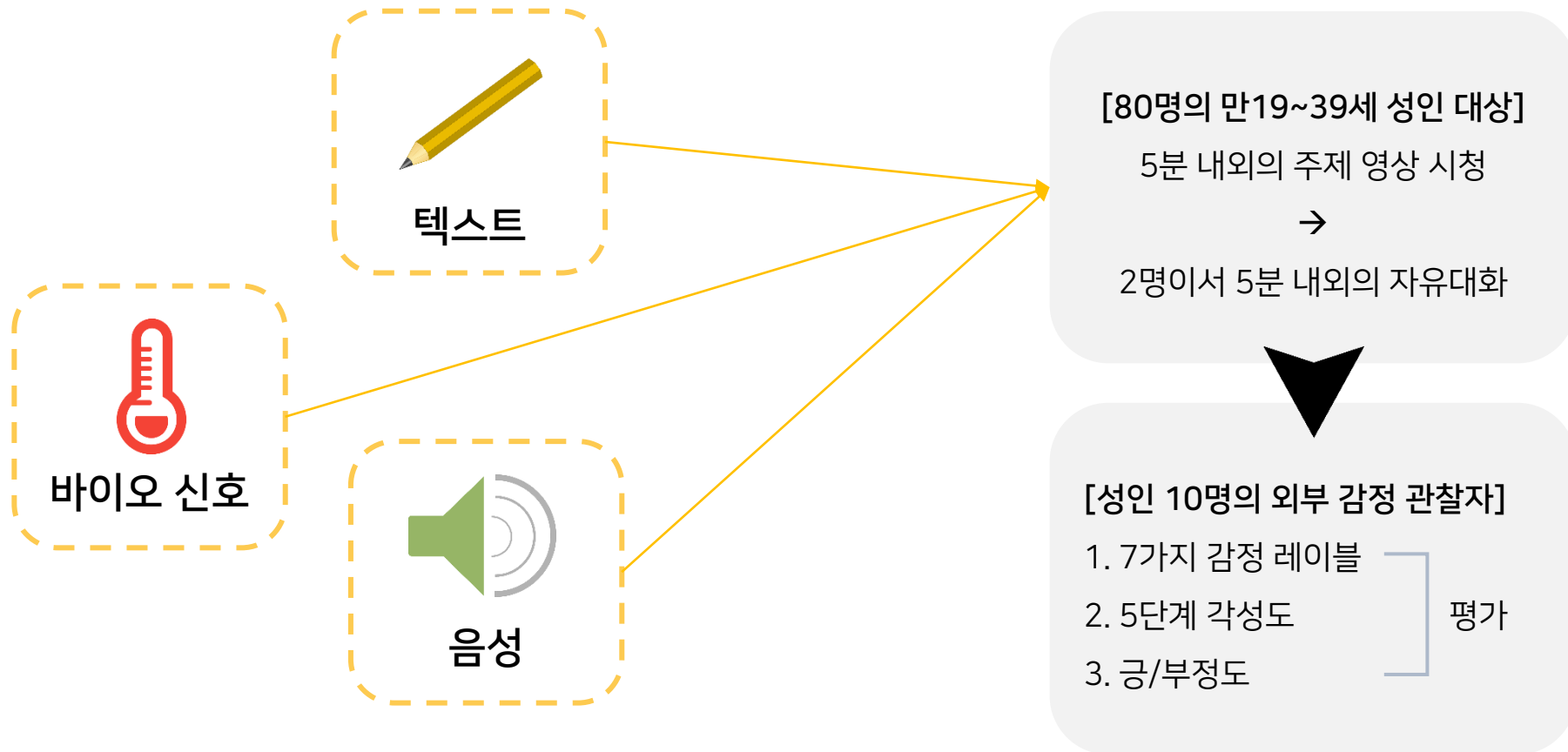
1. 데이터 소개 및 EDA



1. 데이터 소개 및 EDA



ETRI 한국어 감정 데이터셋 - KEMDy20(일반인 대상 자유발화)





1. 데이터 소개 및 EDA



KEMDy20 데이터셋 구성

디렉토리	구성	설명
annotation	.csv	발화에 대한 관찰자의 감정 레이블 평가
wav	.wav / .txt	자유대화 발화 음성/텍스트 파일
EDA	.csv	E4 장치로 수집한 피부전도도 데이터
IBI	.csv	E4 장치로 수집한 심장박동수 데이터
TEMP	.csv	E4 장치로 수집한 체온 데이터

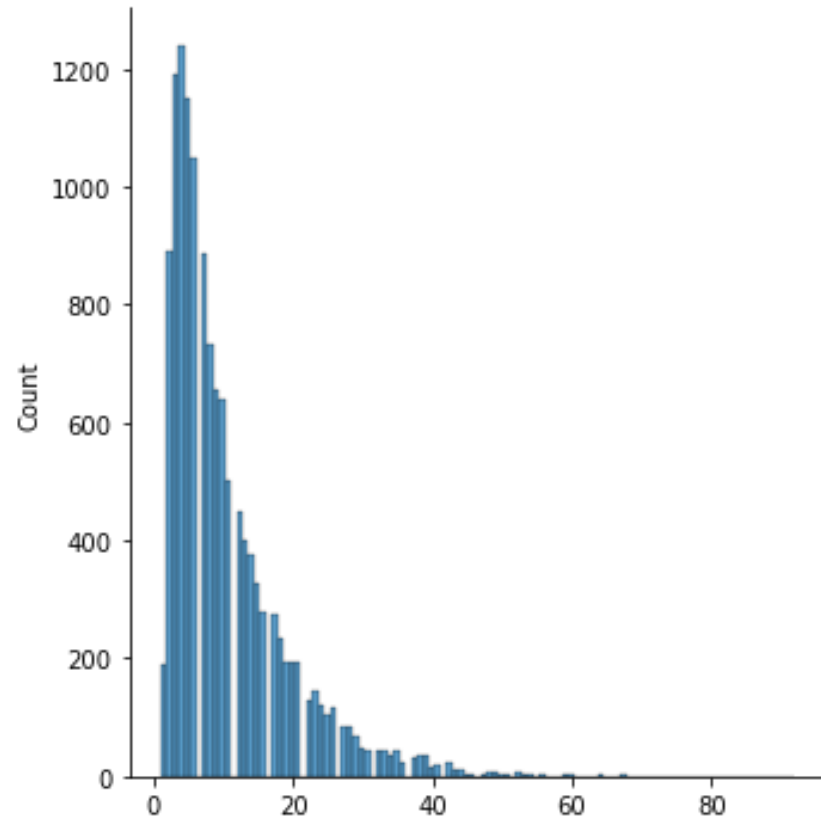




1. 데이터 소개 및 EDA



텍스트 데이터 길이



텍스트의 길이를 전처리한 결과
대략 5~12단어로 구성되어
있는 것을 확인할 수 있음

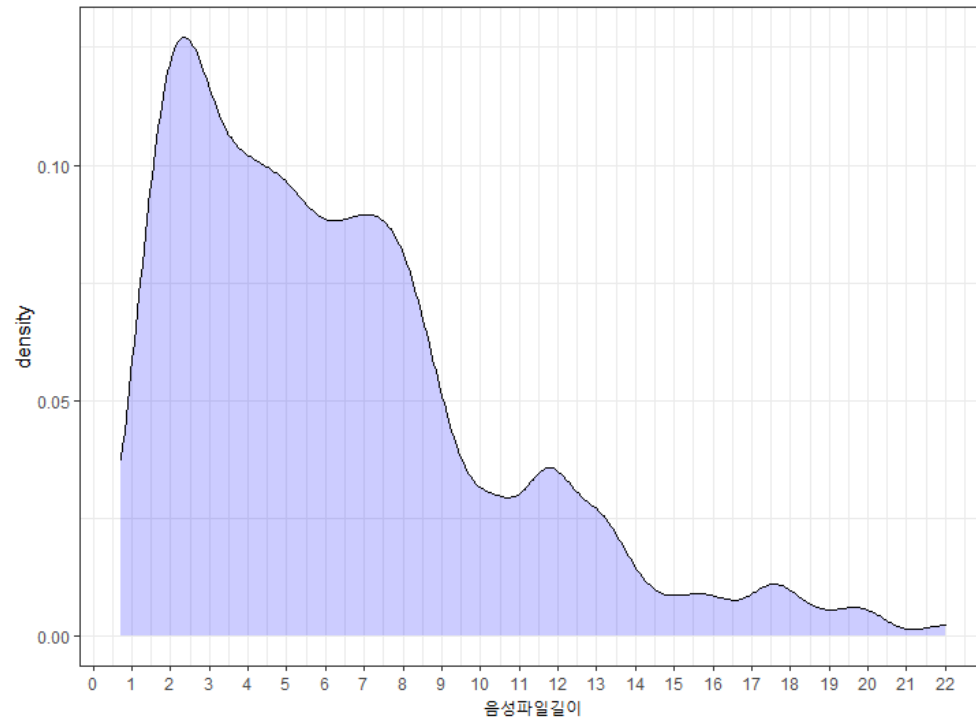




1. 데이터 소개 및 EDA



음성 데이터 길이



음성의 길이를 전처리한 결과
대략 3초~8초로 구성되어
있는 것을 확인할 수 있음

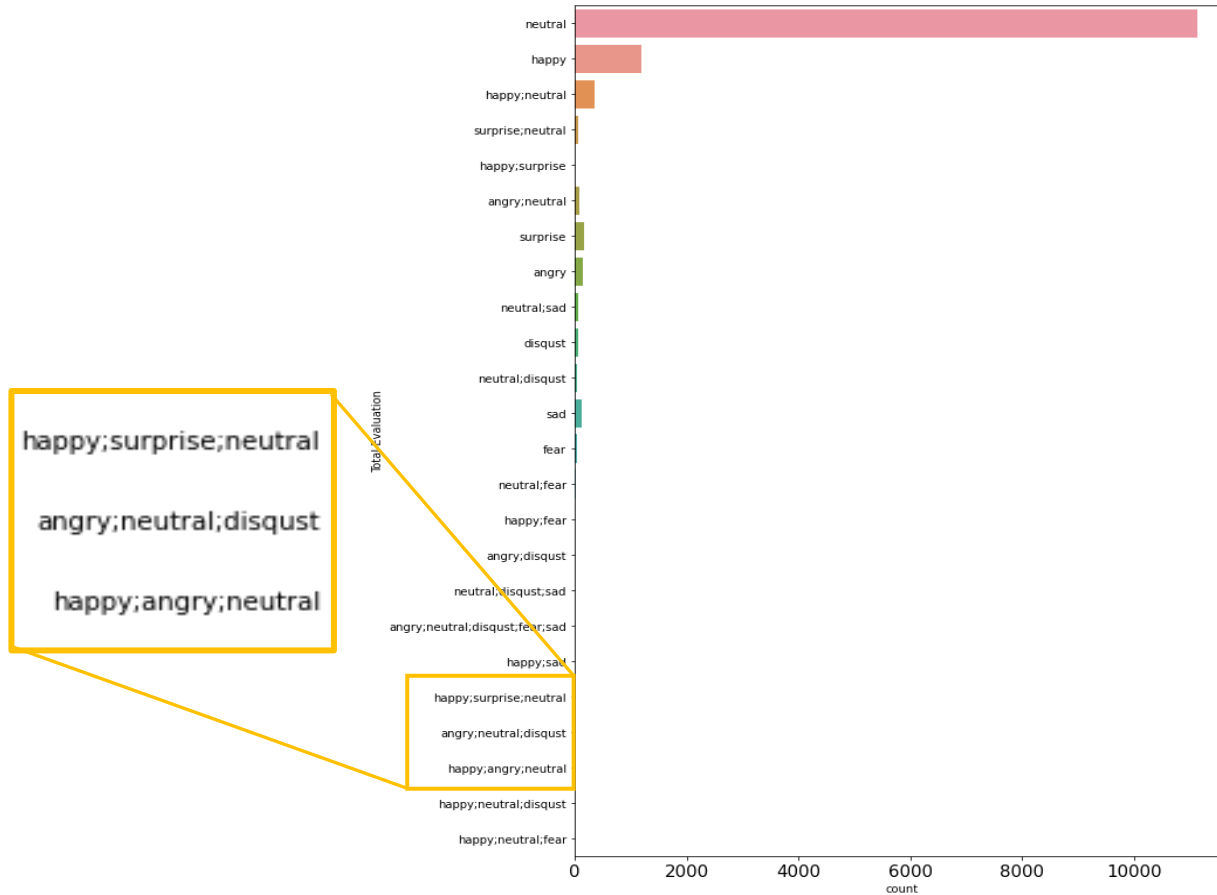




1. 데이터 소개 및 EDA



감정분포



← Neutral

Neutral(중립) 감정의 비율이

불균형적으로 큼

다수의 감정이 라벨링 된 데이터 존재

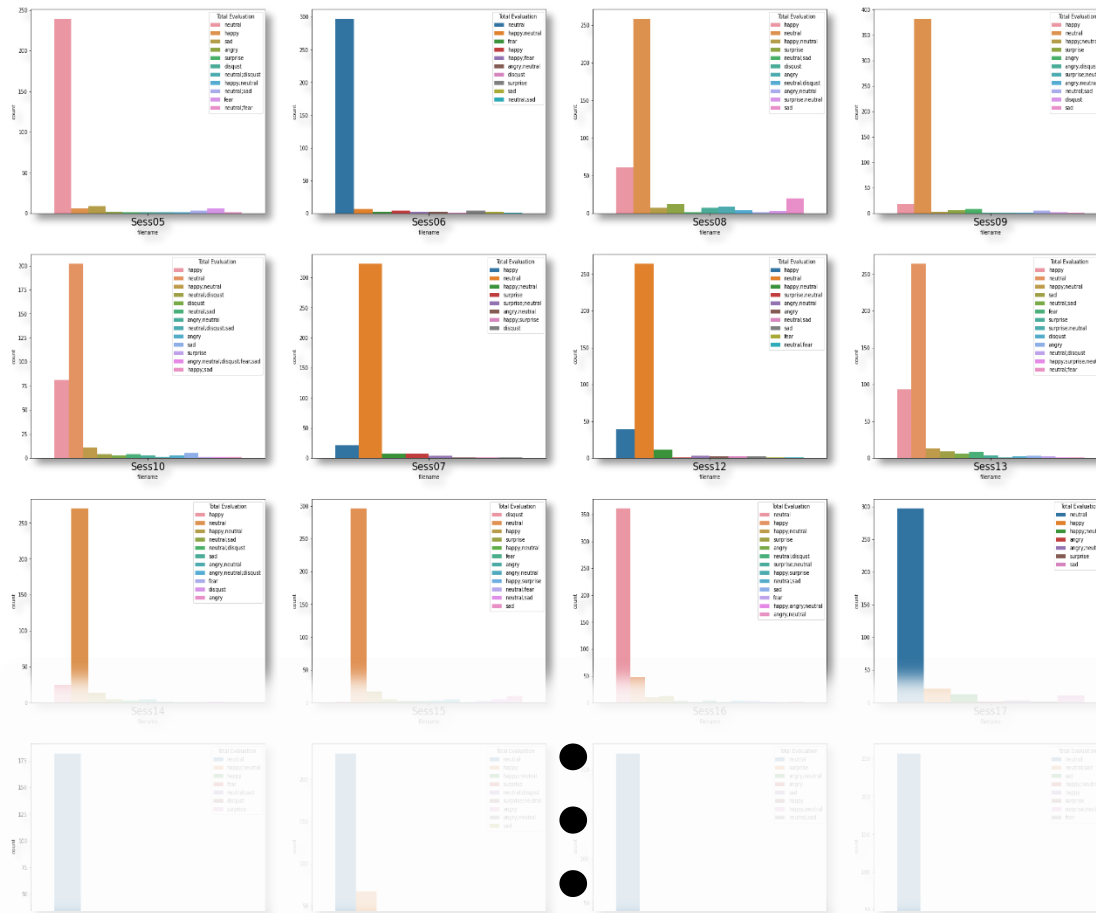




1. 데이터 소개 및 EDA



감정분포 - 세션별



감정분포 불균형 문제
+
다수 감정 라벨링





1. 데이터 소개 및 EDA



감정분포 - 세션별



감정분포 불균형 문제
+
다수 감정 라벨링

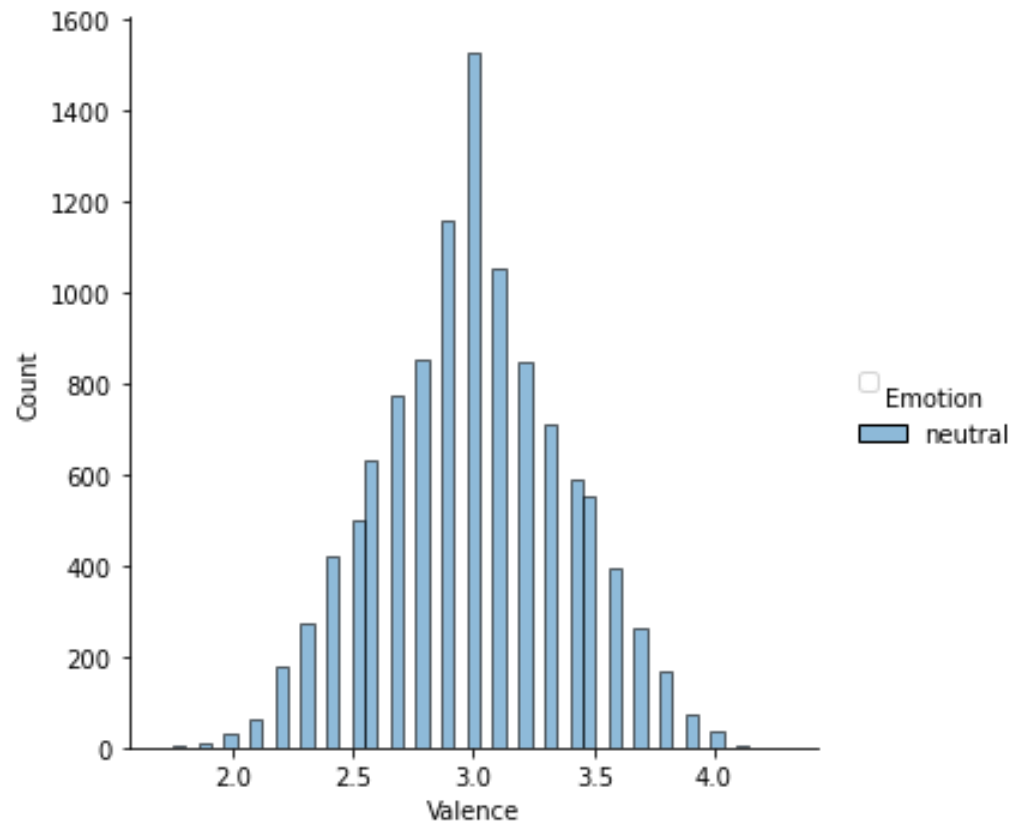




1. 데이터 소개 및 EDA



감정분포(Neutral) - 긍/부정도



(부정) 1-2-3-4-5 (긍정)

Neutral은

3에 가장 많이 분포되어 있으며

정규분포를 따름

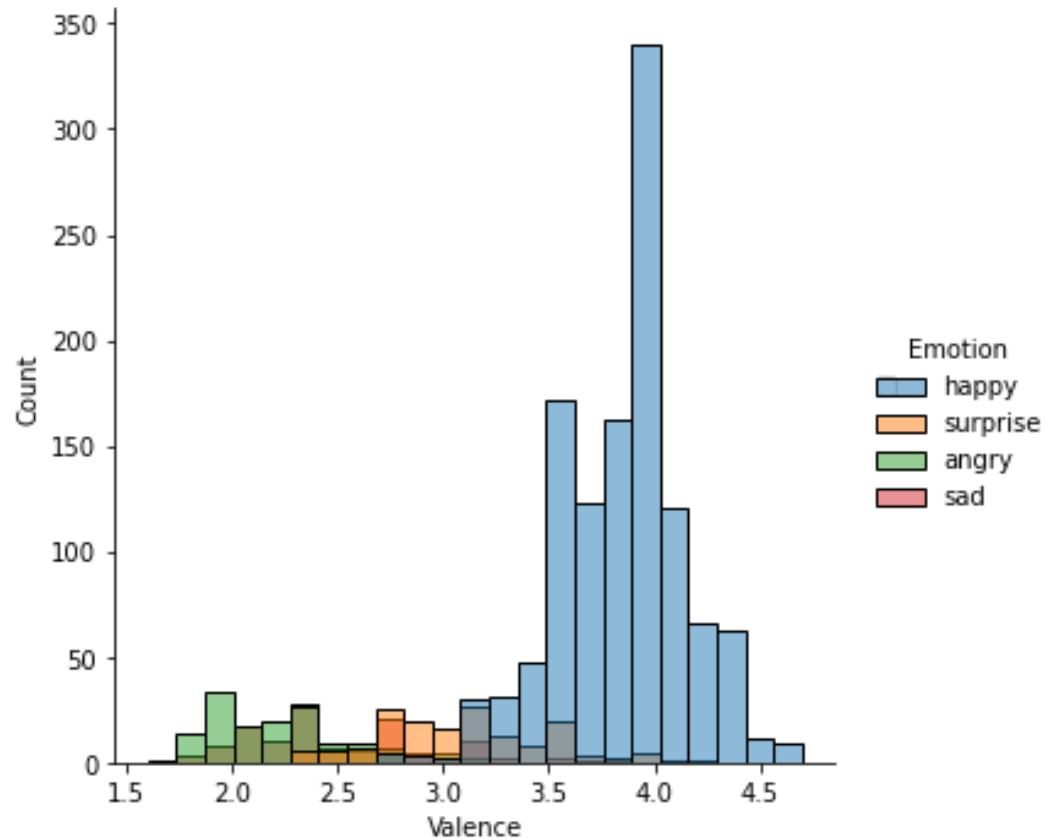




1. 데이터 소개 및 EDA



감정분포 - 긍/부정도



(부정) 1-2-3-4-5 (긍정)

Angry는 부정, Happy는 긍정에
대체로 분포되어 있음

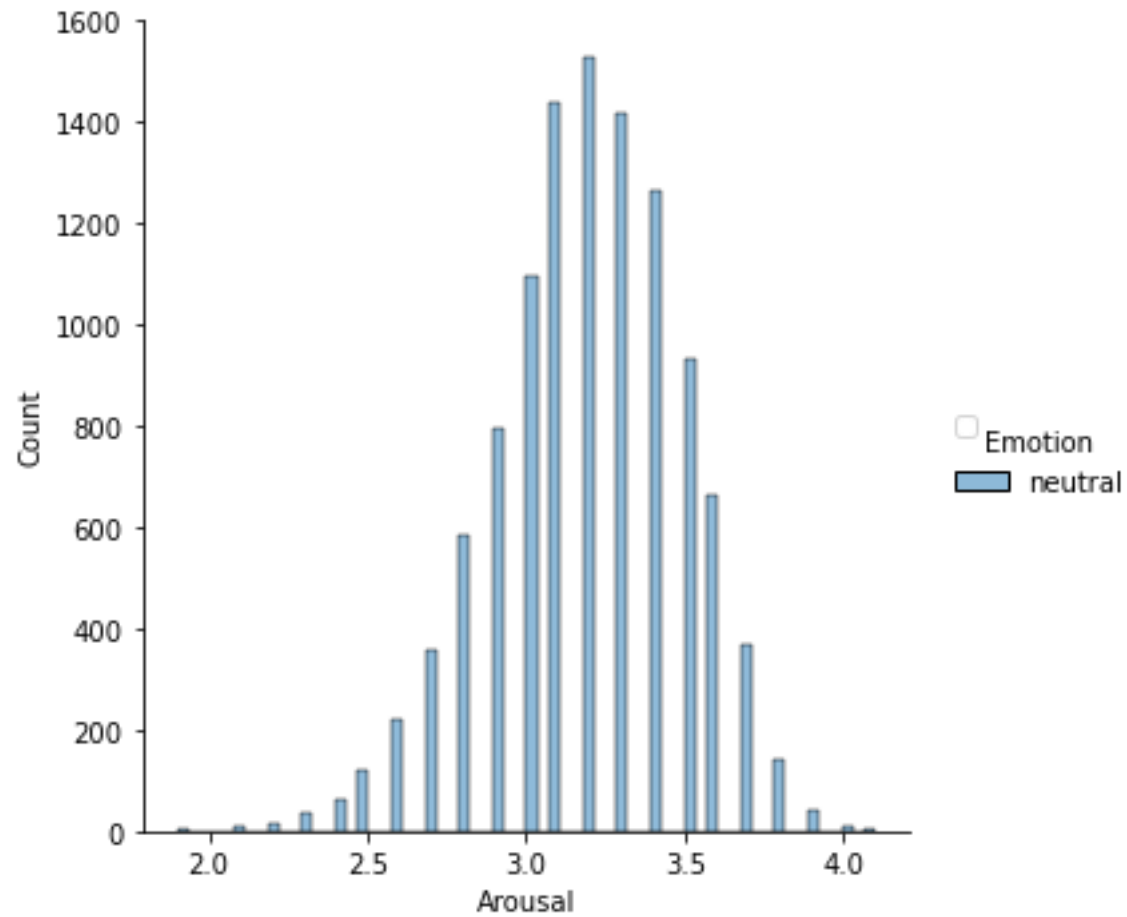




1. 데이터 소개 및 EDA



감정분포(Neutral) - 각성도



(이완) 1-2-3-4-5 (각성)

Neutral은

3~3.5에 가장 많이 분포되어 있으며

정규분포를 따름

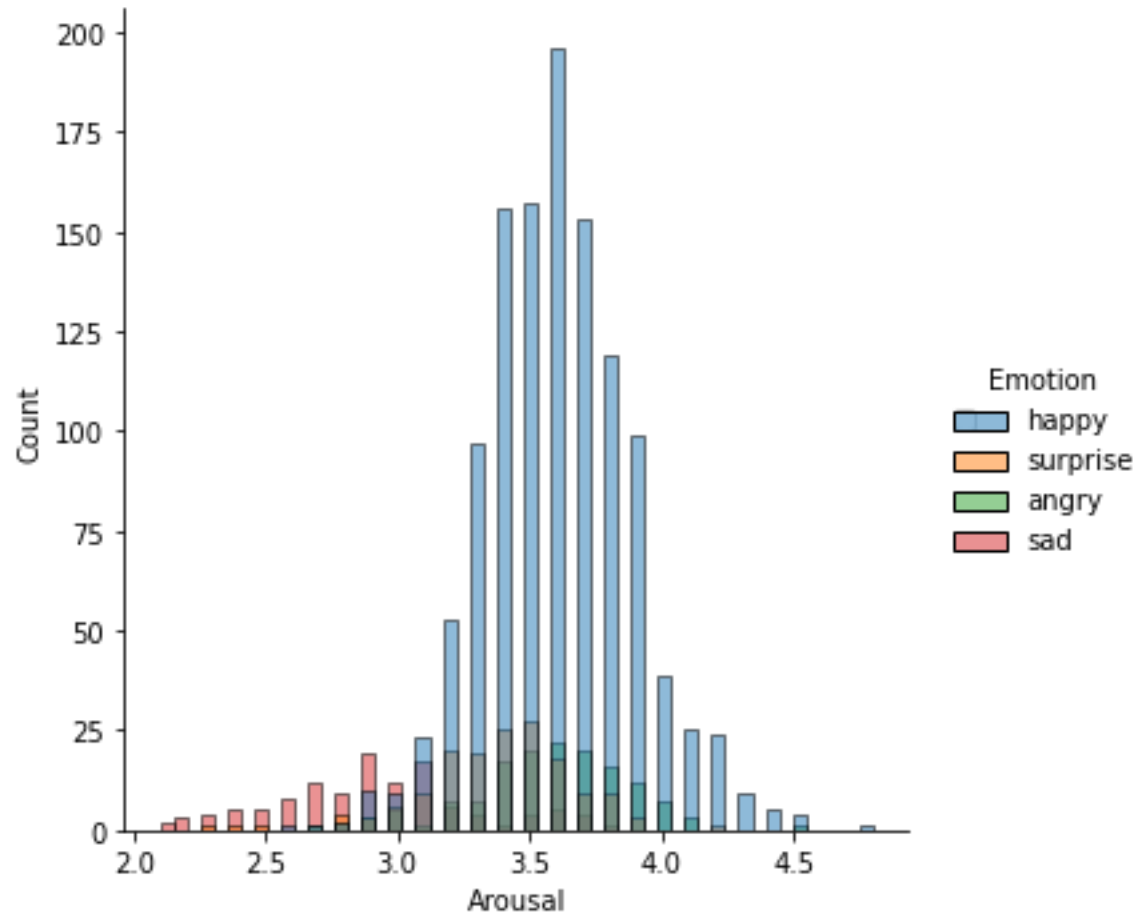




1. 데이터 소개 및 EDA



감정분포 - 각성도



(이완) 1-2-3-4-5 (각성)

Sad는 이완,

Happy는 중립부터 각성까지

분포되어 있음





2. 분석 흐름



2. 분석흐름



분석 흐름 - 음성 전처리

KEMDy20 - wav
Sess01_script01_User001F_001.wav
Sess01_script01_User001F_002.wav
:
Sess01_script03_User001F_012.wav
:
Sess01_script06_User002M_043.wav

Mel
Spectrogram

Augmentation

감정 불균형 해소





4-1. 음성 감정 분류



분석 흐름 - 텍스트 전처리

./wav/~/.txt

: 발화세그먼트의 음성 **이외 사운드상황에 대한 태깅**

- c/ : 휴지구간이 확보되지 않은 연속발성(0.3초 미만)
- n/ : 발성 이외의 단발적인 소음
- N/ : 음성 구간의 50% 이상 잡음이 포함된 상황
- u/ : 단어의 내용을 알아 들을 수 없는 상황
- l/ : 발성중 음음 소리가 포함된 상황 (small 'L')
- b/ : 발성 중 숨소리, 김침 소리가 포함된 상황
- * : 단어 중 일부만 알아 듣거나 알아들었으나 애매한 상황
- + : 발성 중 말을 반복적으로 더듬는 상황
- / : 간투사

토큰화
&
Augmentation

감정 불균형 해소

Index base Encoding





2. 분석흐름



분석 흐름 - 바이오 전처리

KEMDy20 - 바이오

- EDA : 피부전도도
- IBI : 심장박동수
- TEMP : 체온

시계열 군집화

사용 여부 결정

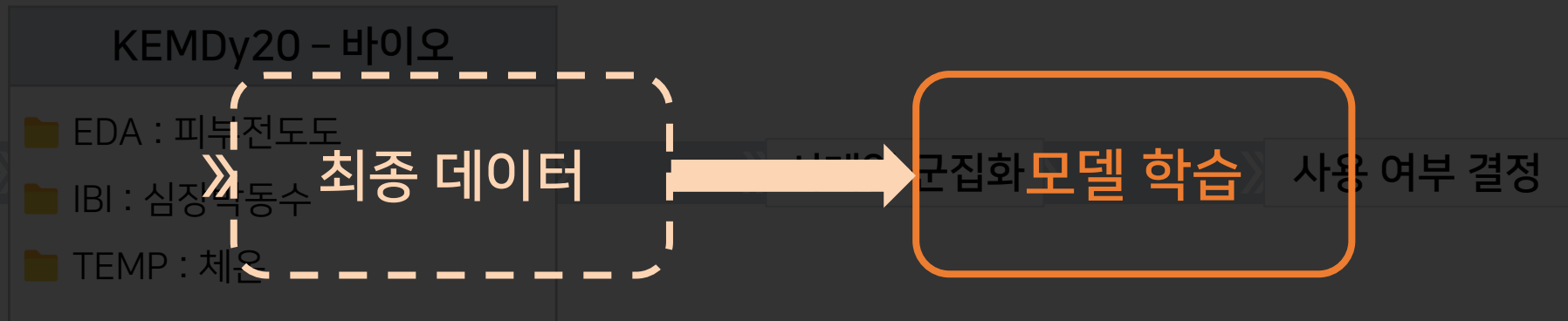




2. 분석흐름



분석 흐름 - 바이오 전처리





3. 음성 데이터 전처리



3. 음성 데이터 전처리

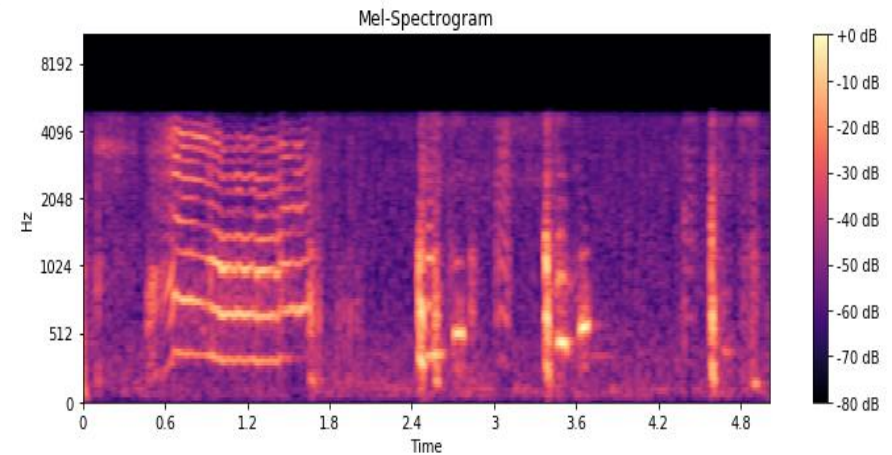


Mel Spectrogram

.wav 파일



[Librosa] 패키지
멜(mel) 필터 수 = 128
자르는 간격 = 512



Mel Spectrogram

음성 전처리에 대한 설명은
[딥러닝팀 주제분석 1주차]
에서!

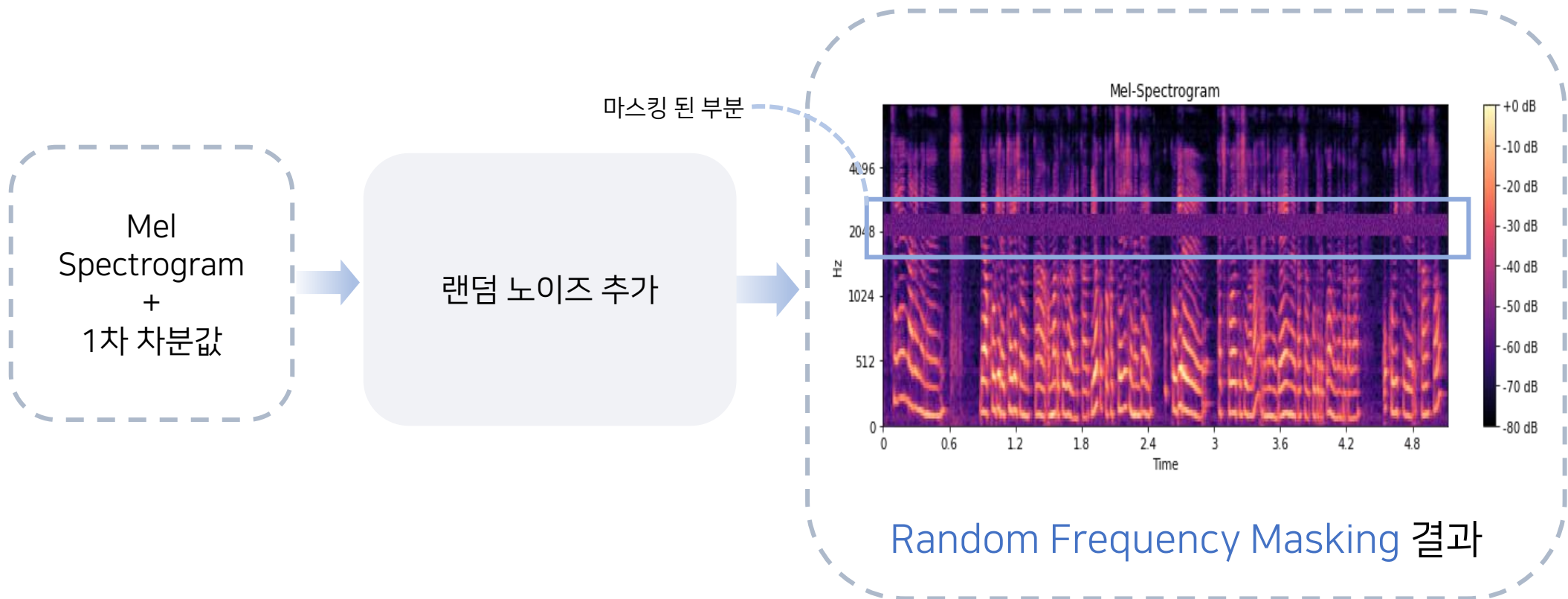




3. 음성 데이터 전처리



데이터 증강(augmentation) – Random Frequency Masking





4. 텍스트 데이터 전처리



4. 텍스트 데이터 전처리



텍스트 데이터

Audio	Text	Emotion	Arousal	Valence	Segment
[[[-46.271706 - 56.406227	n/ 아 친구들도? l/	[0.0.0.1.0.0.0.]	2.8	3.8	Sess01_script01_User001F_001
[[[-54.416702 ... -75...	l/ 나는 생일?	[0.0.0.1.0.0.0.]	3.1	3.3	Sess01_script01_User001F_002
[[[-5.5659237e+01 - 5.0546646e+0...	생일날이면은 b/ 내가 고기를 되...	[0.0.0.1.0.0.0.]	3.1	3.3	Sess01_script01_User001F_003
[[[-63.03308 - 58.875835 - 57.511215	b/ 선물이라 이 보통 돈으로 주...	[1.0.0.0.0.0.0.]	3.1	3.6	Sess01_script01_User001F_004
		:			
[[[-5.3404884e+01 - 5.4611393e+0...	u/ 그걸 점점 발전을 시키잖아, 사람처럼 느껴지게.	[0.0.0.1.0.0.0.]	3.7	2.4	Sess40_script06_User080F_024
[[[-5.7383400e+01 - 5.5686134e+0...	뭐 진짜 막 어떤 욕심이 가득 찬 사람은 내가...	[0.0.0.1.0.0.0.]	3.5	2.7	Sess40_script06_User080F_025





4. 텍스트 데이터 전처리



불용어 지정

텍스트 분석에 도움이 되지 않는 음성 관련 효과음들에 대한 표시들을 모두 불용어로 지정

불용어 리스트	
C/	휴지구간이 확보되지 않은 연속발성
n/	발성 이외의 단발적 소음
N/	음성 구간의 50% 이상 잡음이 포함
u/	단어의 내용을 알아들을 수 없음
l/	발성 중 음음 소리가 포함
b/	발성 중 숨소리, 기침 소리가 포함
*	단어 중 일부만 알아듣거나 애매한 상황
+	발성 중 말을 반복적으로 더듬는 상황
/	간투사





4. 텍스트 데이터 전처리



불용어 제거

Text
n/ 아 친구들도? l/
l/ 나는 생일?
생일날이면은 b/ 내가 고기를 되...
b/ 선물이라 이 보통 돈으로 주...
:
u/ 그걸 점점 발전을 시키잖아, 사람처럼 느껴지게.
뭐 진짜 막 어떤 욕심이 가득 찬 사람은 내가...



Text
아 친구들도?
나는 생일?
생일날이면은 내가 고기를 되...
선물이라 이 보통 돈으로 주...
:
그걸 점점 발전을 시키잖아, 사람처럼 느껴지게.
뭐 진짜 막 어떤 욕심이 가득 찬 사람은 내가...



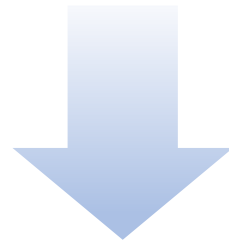


4. 텍스트 데이터 전처리



텍스트 데이터 토큰화 (koBERT)

너는 대선 때 투표할 수 있어?



사전 훈련된 KoBERT 토크나이저

_너	는	_내년	_대선	_때	_투표	할	_수	_있어	?
----	---	-----	-----	----	-----	---	----	-----	---





4. 텍스트 데이터 전처리



최종 텍스트 데이터

Text
어 나는 워낙 많아 가지고 고등 고등학교 3학년 3학년 한 중반쯤에...
근데 이제 진지 이거 진짜예요 진짜 꾸민 게 아니라.
그리고 그걸 또 그걸 또 빌려달라고 하는 남자애들이 있다는 거랑.
아 뭐였지 그.
:
그때도 아마 믿었었을 때였으니까 엄마 아빠가 했겠지 엄마가...
그냥 오히려 군대 둘 다 갔다 왔잖아...



Text
[_어, _나는, _워낙, _많아, _가지고, _고, 등, _, 고등학교, _3, ...
[_근, 데, _이제, _진, 지, _이, 거, _진짜, 예, 요, _진짜, _꾸,...
[_그리고, _그, 걸, _또, _그, 걸, _또, _빌려, 달라, 고, _하는, ...
[_아, _뭐, 였, 지, _그, .]
:
[_그때, 도, _아, 마, _, 믿, 었, 었, 을, _때, 였, 으니, 까,
[_그냥, _오히려, _군, 대, _둘, _다, _, 갔다, _, 왔, 잤, 아, ...

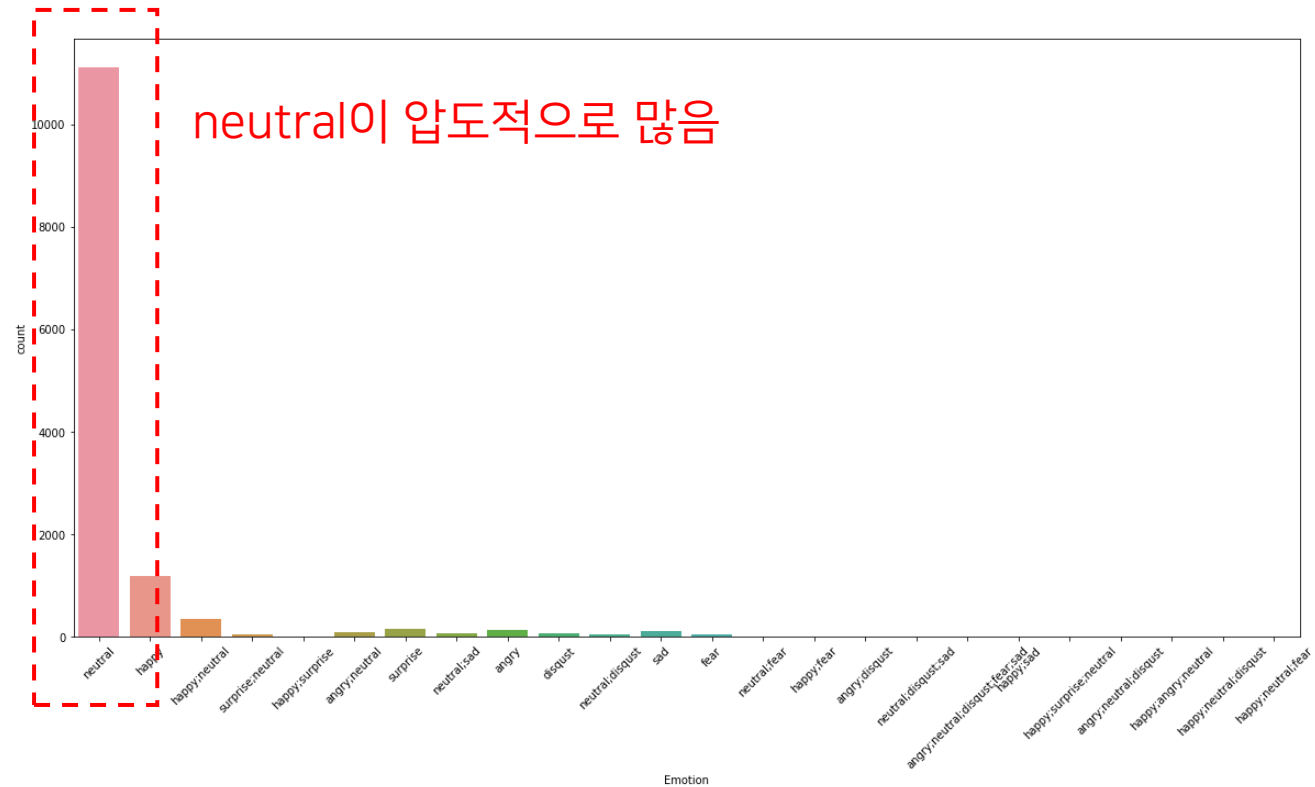




4. 텍스트 데이터 전처리



불균형한 텍스트 데이터



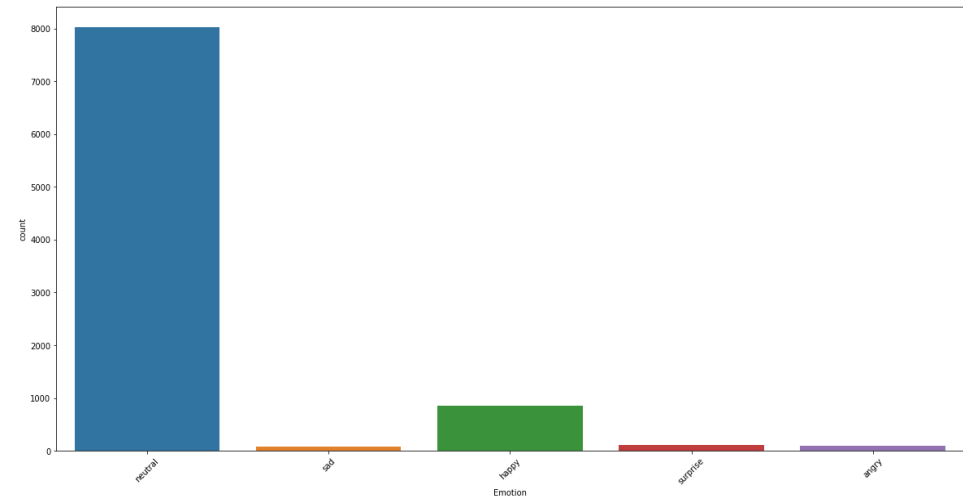
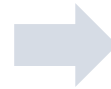
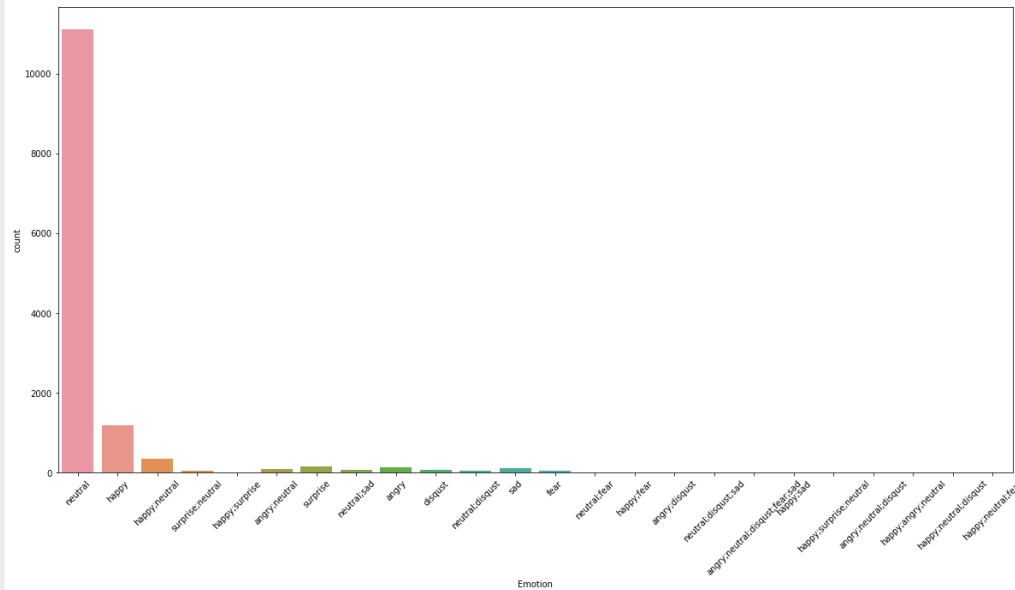


4. 텍스트 데이터 전처리



불균형한 텍스트 데이터

상위 5개 감정을 제외한 감정들 제거



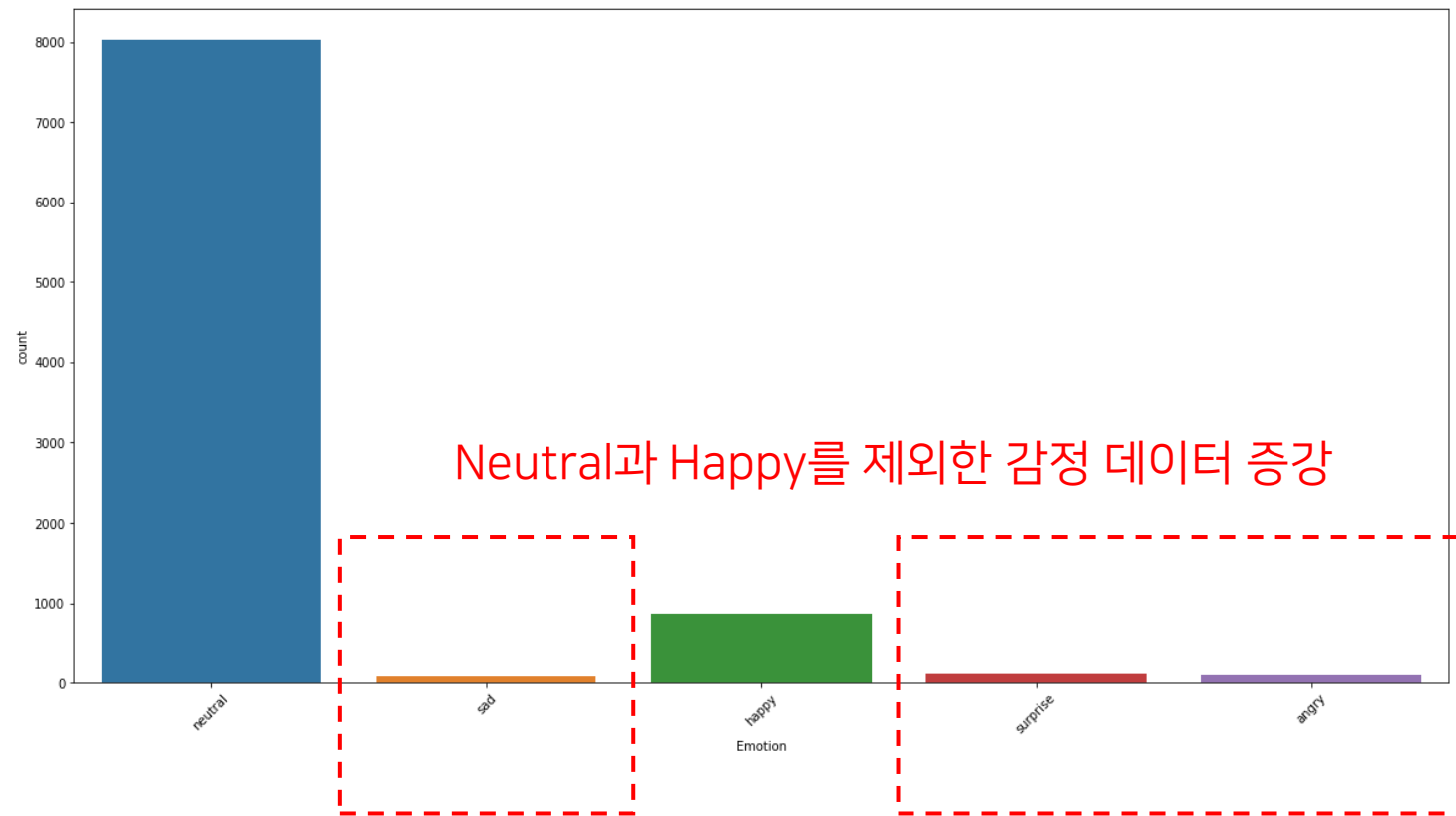


4. 텍스트 데이터 전처리



불균형한 텍스트 데이터

상위 5개 감정을 제외한 감정들 제거





4. 텍스트 데이터 전처리



불균형한 텍스트 데이터

텍스트 데이터 증강방법

방법	문장
원래 문장	중심을 잃고 목소리도 잃고 비난받고 사람들과 멀어지는 착각 속에
SR (동의어 대체)	핵심을 잃고 목소리도 잃고 비판받고 인간들과 멀어지는 혼란 속에
RI (단어 임의 대체)	돈을 잃고 목소리도 잃고 비난받고 사람들과 멀어지는 착각 속에
RS (단어 위치 변경)	중심을 잃고 사람들과 잃고 비난받고 목소리도 멀어지는 착각 속에
RD (단어 임의 삭제)	중심도 목소리도 사람들과 멀어지는 착각





4. 텍스트 데이터 전처리



불균형한 텍스트 데이터

텍스트 데이터 증강방법

방법	문장
원래 문장	중심을 잃고 목소리도 잃고 비난받고 사람들과 멀어지는 착각 속에
SR (동의어 대체)	중심을 잃고 목소리도 잃고 비난받고 인간들과 멀어지는 혼란 속에
RI (단어 임의 대체)	돈을 잃고 목소리도 잃고 비난받고 사람들과 멀어지는 착각 속에
RS (단어 위치 변경)	중심을 잃고 사람들과 잃고 비난받고 목소리도 멀어지는 착각 속에
RD (단어 임의 삭제)	중심도 목소리도 사람들과 멀어지는 착각

SR, RI와 같은 단어 대체 방법은 문장의 의미를 왜곡할 수 있다고 판단!





4. 텍스트 데이터 전처리

불균형한 텍스트 데이터

텍스트 데이터 증강방법

방법	문장
원래 문장	중심을 잃고 목소리도 잃고 비난받고 사람들과 멀어지는 착각 속에
SR (동의어 대체)	핵심을 잃고 목소리도 잃고 비판받고 인간들과 멀어지는 혼란 속에
RI (단어 임의 대체)	돈을 잃고 목소리도 잃고 비난받고 사람들과 멀어지는 착각 속에
RS (단어 위치 변경)	중심을 잃고 사람들과 잃고 비난받고 목소리도 멀어지는 착각 속에
RD (단어 임의 삭제)	중심도 목소리도 사람들과 멀어지는 착각

의미 왜곡 위험이 적은 RS, RD 선택



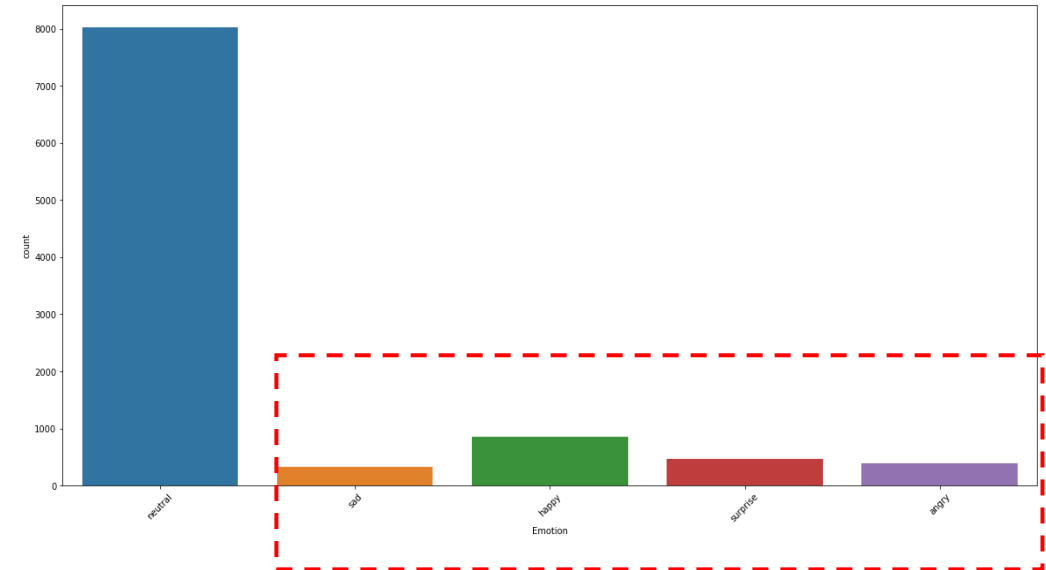
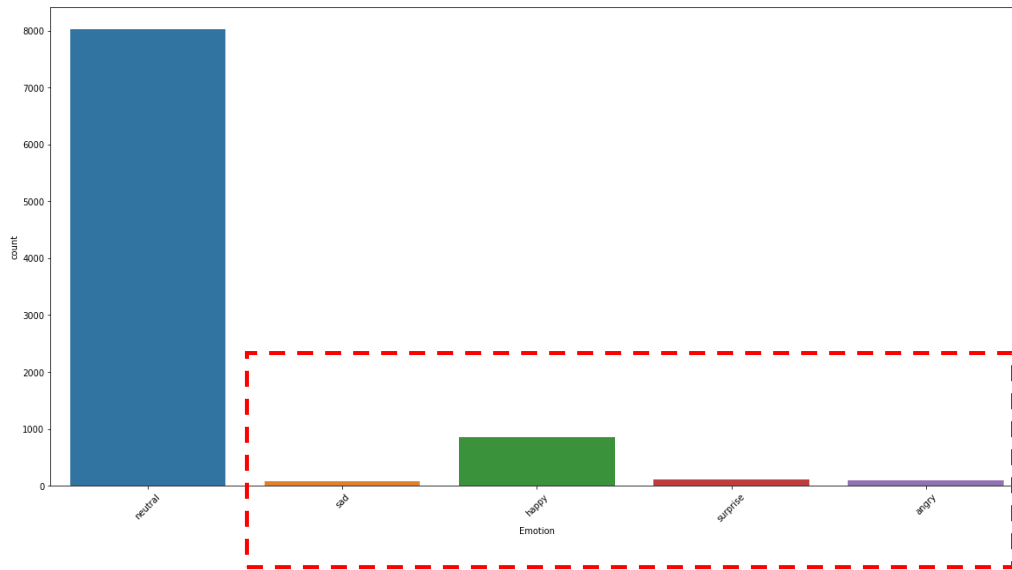


4. 텍스트 데이터 전처리



불균형한 텍스트 데이터

상위 5개 감정을 제외한 감정들 제거



라벨 별 250 % 증강!





5. 바이오 데이터 분석



5. 바이오 데이터 분석



바이오 데이터

Emotion	TEMP	IBI	EDA
[0. 0. 0. 1. 0. 0. 0.]	[34.55 34.55 34.55 34.55 34.55 34.55 34.55 34....	[0.640625, 0.640625, 0.578125]	[4.412055, 4.536246, 4.661...
[1. 0. 0. 0. 0. 0. 0.]	[34.59 34.59 34.59 34.59 34.59 34.59 34.59 34....	[]	[4.842703, 4.808134, 4.7940...
[1. 0. 0. 1. 0. 0. 0.]	[34.61 34.61 34.61 34.61 34.61 34.61 34.61 34....	[0.609375, 0.687500, 0.625000]	[5.233203, 4.979698, 4.932...
[0. 1. 0. 1. 0. 0. 0.]	[34.63 34.63 34.63 34.63 34.63 34.63 34.63 34....	[]	[5.17896, 5.085496, 5.0752...
[1. 1. 0. 0. 0. 0. 0.]	[35.16 35.16 35.18 35.18 35.18 35.18 35.18 35....	[0.687500, 0.593750, 0.609...	[5.37357, 5.41198, 5.60915...
:			
[0. 0. 1. 1. 0. 0. 0.]	[35.21 35.21 35.18 35.18 35.18 35.18 35.21 35....	[0.703125, 0.687500, 0.718...	[2.717516, 2.961985, 3.0426...
[0. 0. 0. 0. 0. 1. 0.]	[35.18 35.18 35.18 35.18 35.18 35.18 35.18 35....	[0.640625]	[2.720076, 2.694477, 2.737...
[0. 0. 0. 1. 0. 1. 0.]	[35.18 35.18 35.16 35.16 35.16 35.16 35.18 35....	[0.671875]	[3.011903, 3.247412, 3.4778...





5. 바이오 데이터 분석



바이오 데이터 - EDA(피부전도도)

EDA CSV 파일 형태 - 📁 Sess01_script03_User002M.csv		
0	2020-0821-1212-52-000	
2.820783	2020-0821-1213-19-250	
2.55749	2020-0821-1213-21-500	Sess01_script03_User002M_001
2.413964	2020-0821-1214-18-000	Sess01_script03_User002M_011
:		
2.493416	2020-0821-1215-19-750	Sess01_script03_User002M_019
4.127432	2020-0821-1219-26-250	
샘플링 주기	데이터 측정 시간	해당 EDA값이 속하는 Segment ID

대화 음성을 기반으로 측정



말을 하지 않을 때도
EDA 값이 찍힘





5. 바이오 데이터 분석



바이오 데이터 - EDA(피부전도도)

EDA CSV 파일 형태 - 📁 Sess01_script03_User002M.csv

0	2020-0821-1212-52-000	
2.820783	2020-0821-1213-19-250	
2.55749	2020-0821-1213-21-500	Sess01_script03_User002M_001
2.413964	2020-0821-1214-18-000	Sess01_script03_User002M_011
:		
2.493416	2020-0821-1215-19-750	Sess01_script03_User002M_019
4.127432	2020-0821-1219-26-250	
샘플링 주기	데이터 측정 시간	해당 EDA값이 속하는 Segment ID

EDA CSV 파일의
Segment ID

해당 Segment ID에서
감정 라벨 추출

대화 음성을 기반으로 측정



말을 하지 않을 때도
EDA 값이 찍힘



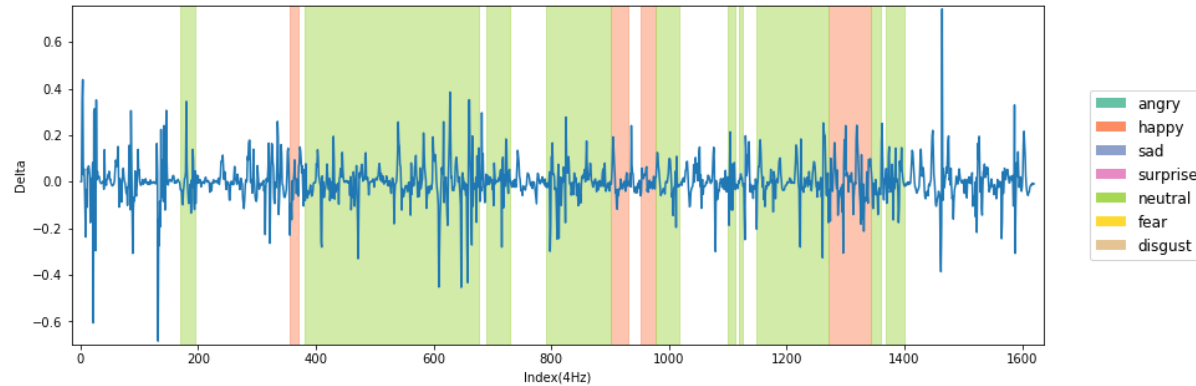


5. 바이오 데이터 분석

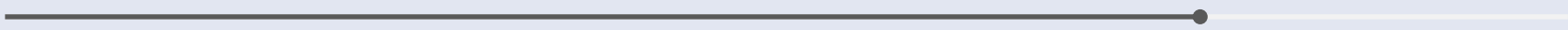
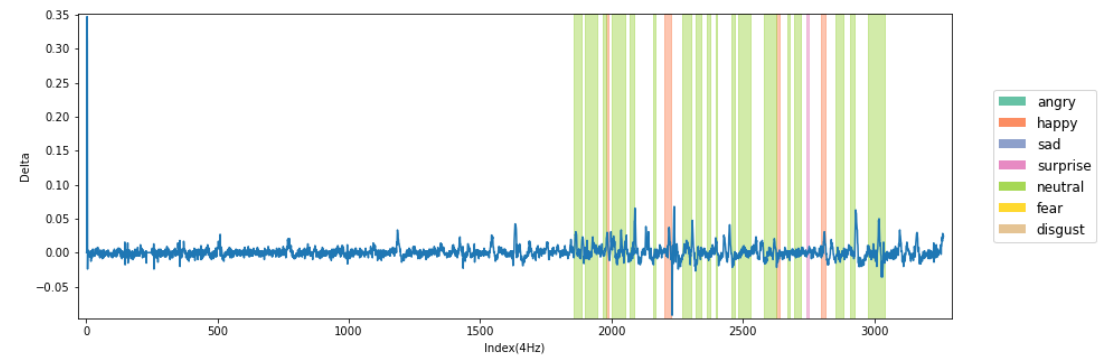


바이오 데이터 - EDA(피부전도도)

EDA of Session 01 - Script 05 - User 002M



EDA of Session 29 - Script 01 - User 057F

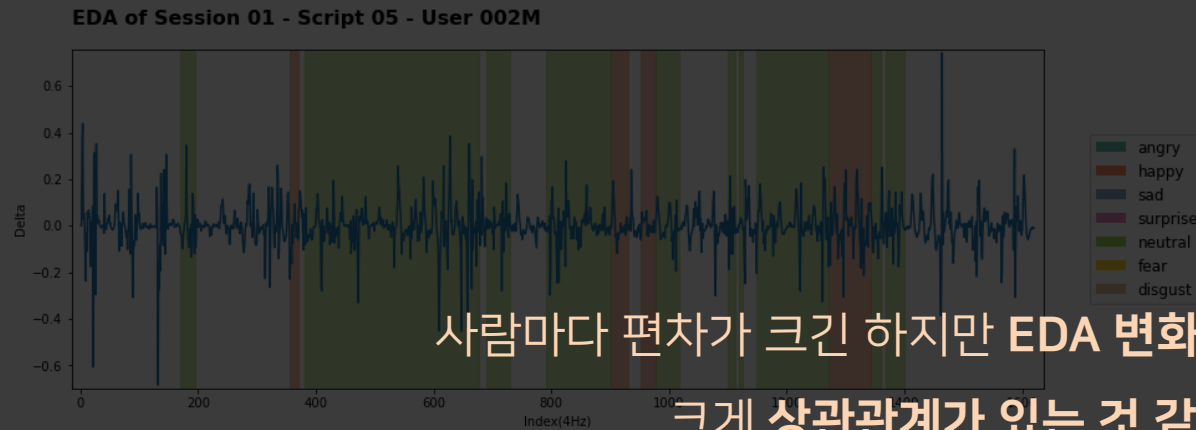




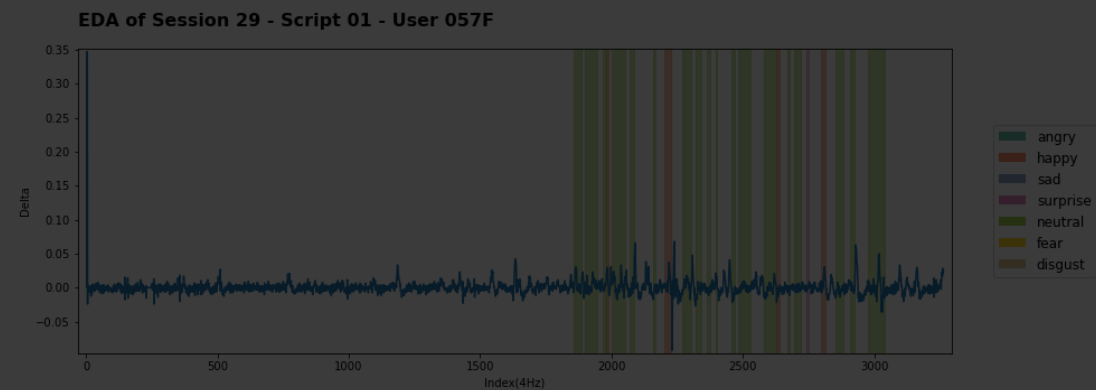
5. 바이오 데이터 분석



바이오 데이터 - EDA(피부전도도)



사람마다 편차가 크긴 하지만 EDA 변화량과 감정 변화량이
크게 상관관계가 있는 것 같지 않음



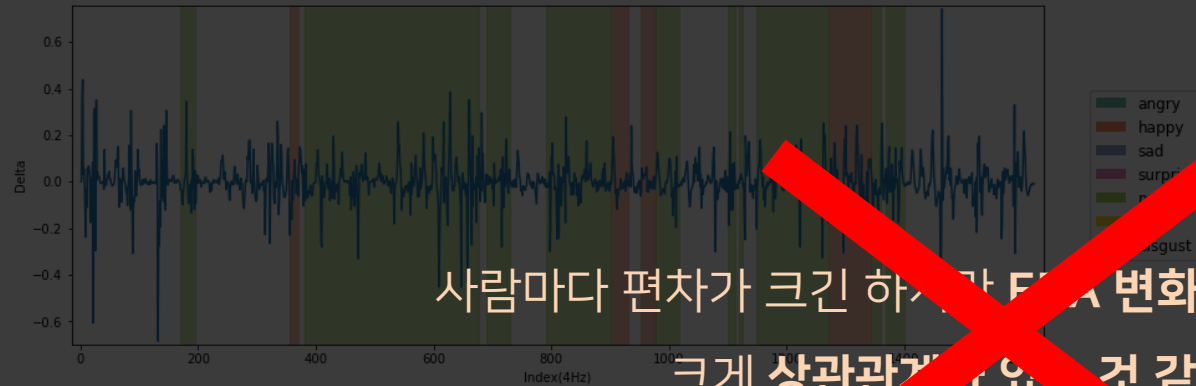


5. 바이오 데이터 분석



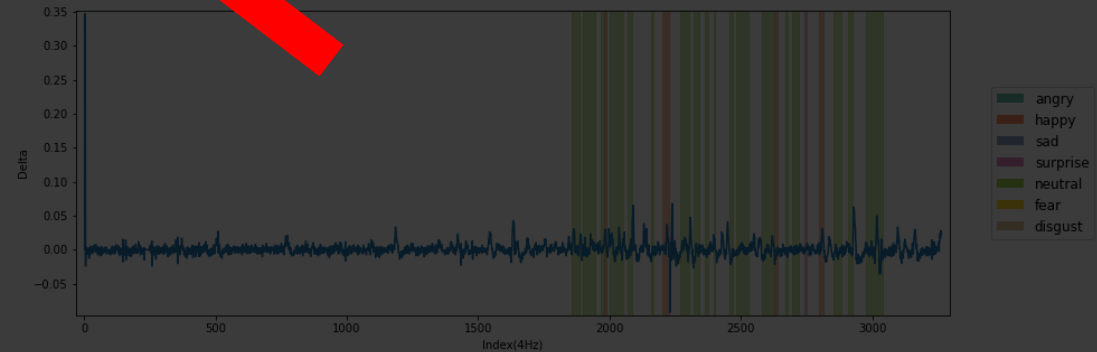
바이오 데이터 - EDA(피부전도도)

EDA of Session 01 - Script 05 - User 002M



사람마다 편차가 크긴 하지만 EDA 변화량과 감정 변화량이
크게 상관관계가 있는 것 같지 않음

EDA of Session 01 - Script 01 - User 057F





5. 바이오 데이터 분석



바이오 데이터 - IBI(심장 박동수)

IBI CSV 파일 형태 - 📁 Sess18_script01_User035M.csv			
14.328125	0.921875	2020-0904-1017-19-328	
40	0.828125	2020-0904-1017-45-000	
395.84375	0.890625	2020-0904-1023-40-843	Sess18_script01_User035M_001
446.9375	0.90625	2020-0904-1024-31-937	Sess18_script01_User035M_005
:			
681.984375	0.828125	2020-0904-1028-26-984	Sess18_script01_User035M_021
726.03125	0.890625	2020-0904-1029-11-031	
샘플링 주기	IBI 값	데이터 측정 시간	해당 IBI값이 속하는 Segment ID





5. 바이오 데이터 분석



바이오 데이터 - IBI(심장 박동수)

IBI CSV 파일 형태 - Sess18_script01_User035M.csv

14.328125	0.821875	2020-0904-1017-19-328	
40	0.828125	2020-0904-1017-45-000	
395.84375	0.890625	2020-0904-1023-40-843	Sess18_script01_User035M_001
446.9375	0.90625	2020-0904-1024-31-937	Sess18_script01_User035M_005
681.984375	0.828125	2020-0904-1028-26-984	Sess18_script01_User035M_021
726.03125	0.890625	2020-0904-1029-11-031	
샘플링 주기	IBI 값	데이터 측정 시간	해당 IBI값이 속하는 Segment ID

EDA 파일과 같은 형태

같은 방식으로 Plot 확인

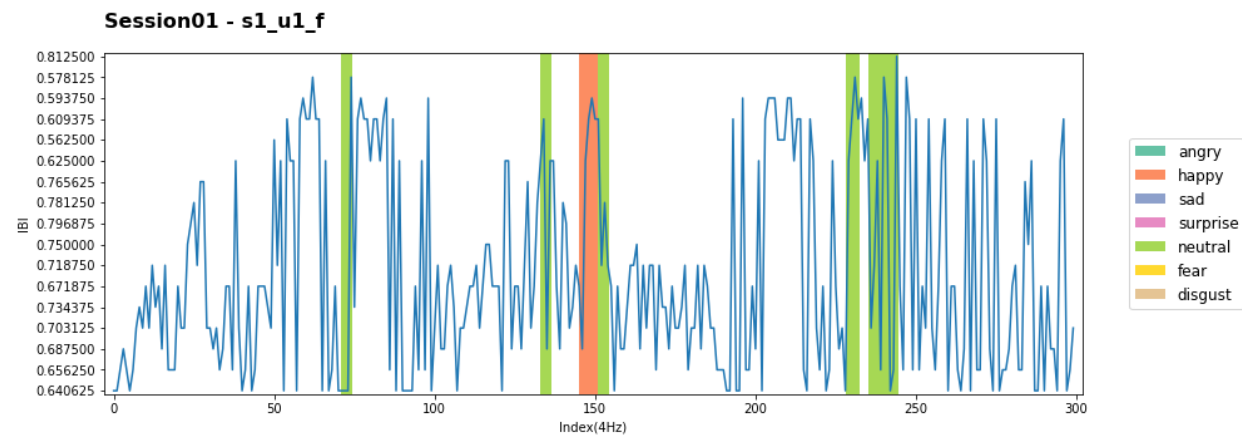
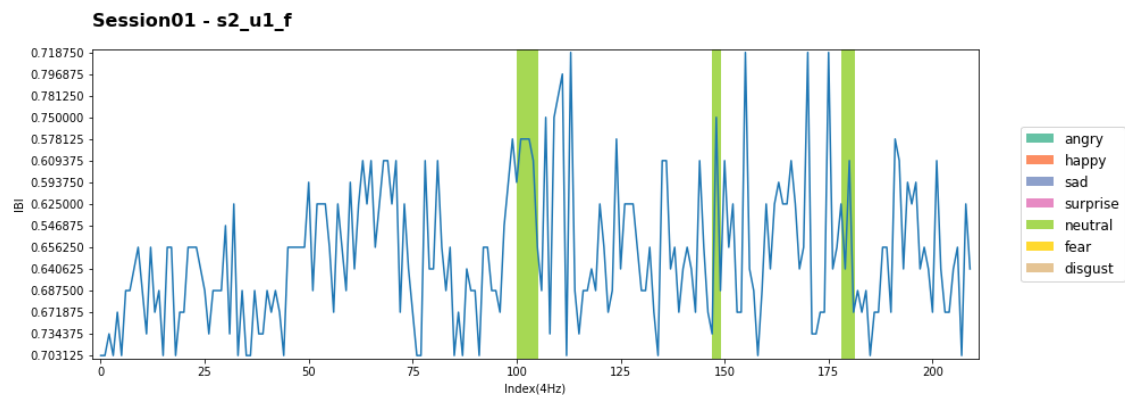




5. 바이오 데이터 분석



바이오 데이터 - IBI(심장 박동수)

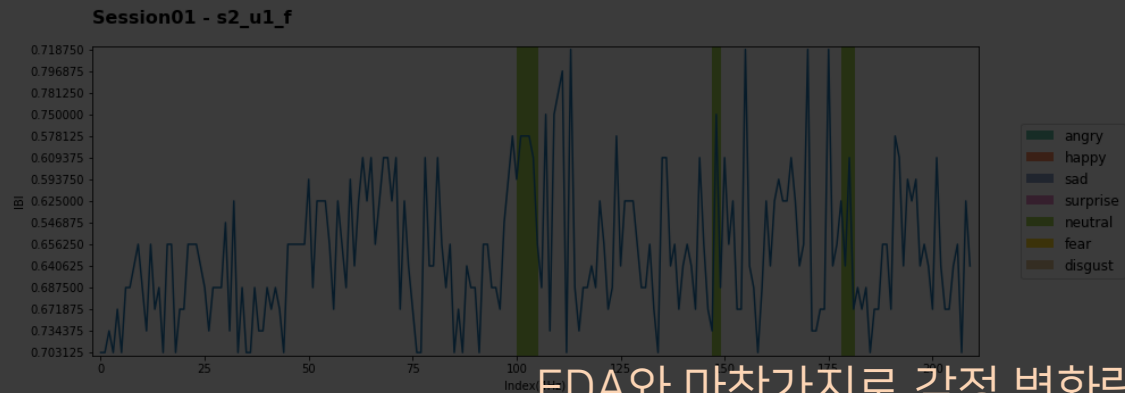




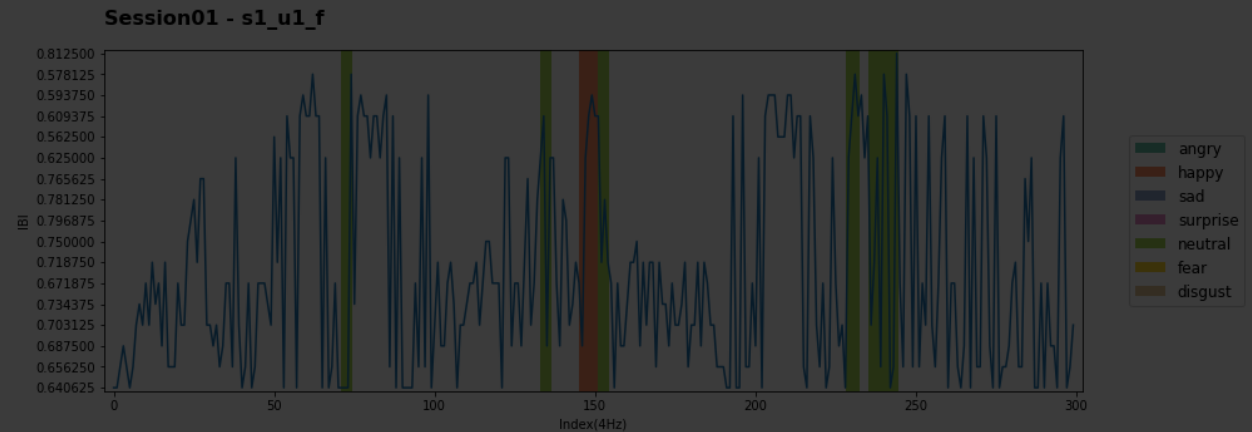
5. 바이오 데이터 분석



바이오 데이터 - IBI(심장 박동수)



EDA와 마찬가지로 감정 변화량과의 상관관계가 없음

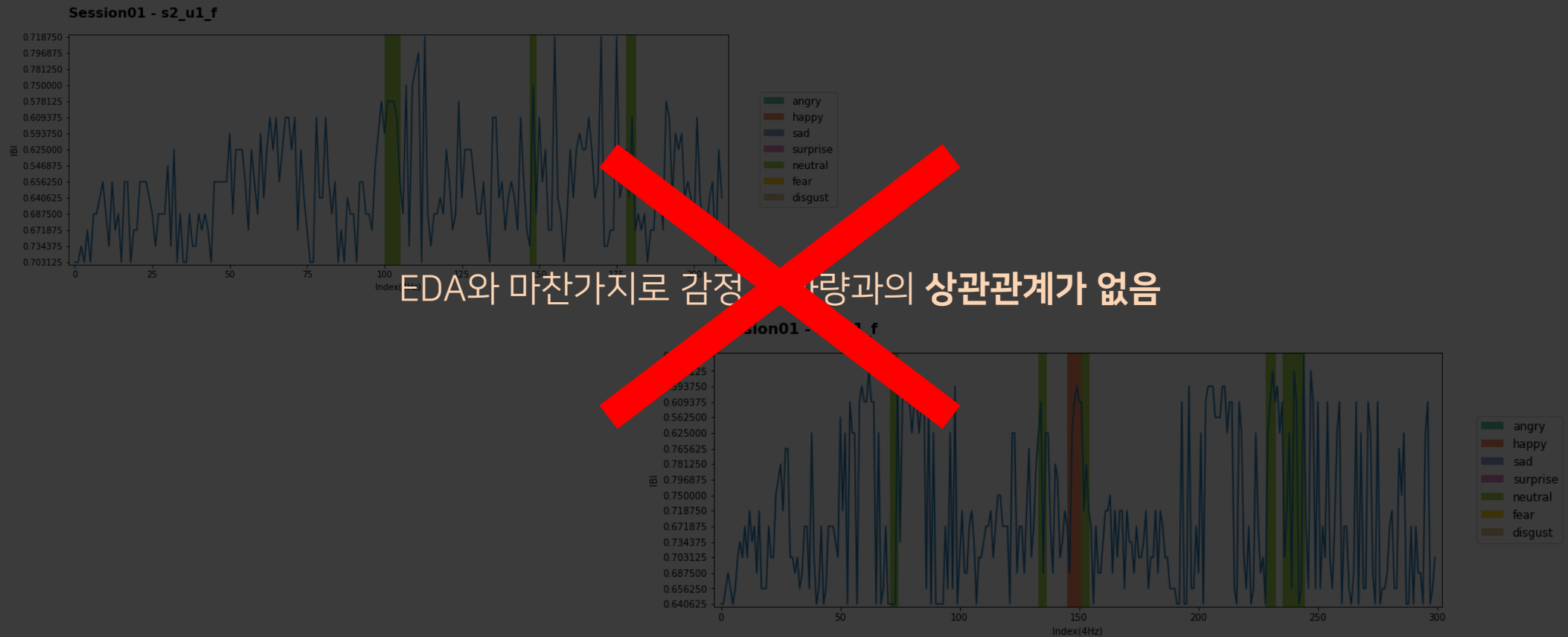




5. 바이오 데이터 분석



바이오 데이터 - IBI(심장 박동수)





5. 바이오 데이터 분석



바이오 데이터 - TEMP(체온)

TEMP CSV 파일 형태 - 📁 Sess20_script05_User040M.csv		
33.42	2020-0906-1134-34-000	
33.39	2020-0906-1134-42-250	
33.13	2020-0906-1139-34-750	Sess20_script05_User040M_001
33.11	2020-0906-1139-36-000	Sess20_script05_User040M_002
:		
33.05	2020-0906-1144-29-500	Sess20_script05_User040M_016
33.07	2020-0906-1145-25-750	
피부온도 값	데이터 측정 시간	해당 TEMP값이 속하는 Segment ID





5. 바이오 데이터 분석



바이오 데이터 - TEMP(체온)

Emotion	TEMP
[0. 0. 0. 1. 0. 0. 0.]	[34.55 34.55 34.55 34.55 34.55 34.55 34.55 34....
[1. 0. 0. 0. 0. 0. 0.]	[34.59 34.59 34.59 34.59 34.59 34.59 34.59 34....
[1. 0. 0. 1. 0. 0. 0.]	[34.61 34.61 34.61 34.61 34.61 34.61 34.61 34....
[0. 1. 0. 1. 0. 0. 0.]	[34.63 34.63 34.63 34.63 34.63 34.63 34.63 34....
[1. 1. 0. 0. 0. 0. 0.]	[35.16 35.16 35.18 35.18 35.18 35.18 35.18 35....
:	
[0. 0. 1. 1. 0. 0. 0.]	[35.21 35.21 35.18 35.18 35.18 35.18 35.21 35....
[0. 0. 0. 0. 0. 1. 0.]	[35.18 35.18 35.18 35.18 35.18 35.18 35.18 35....
[0. 0. 0. 1. 0. 1. 0.]	[35.18 35.18 35.16 35.16 35.16 35.16 35.18 35....





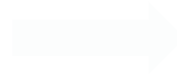
5. 바이오 데이터 분석



바이오 데이터 - TEMP(체온)

24개

Emotion
[0. 0. 0. 1. 0. 0. 0.]
[1. 0. 0. 0. 0. 0. 0.]
[1. 0. 0. 1. 0. 0. 0.]
[0. 1. 0. 1. 0. 0. 0.]
[1. 1. 0. 0. 0. 0. 0.]
[0. 0. 1. 1. 0. 0. 0.]
[0. 1. 0. 0. 0. 0. 0.]
:
[0. 0. 0. 1. 0. 0. 1.]
[0. 0. 1. 0. 0. 0. 0.]
[0. 0. 0. 0. 1. 0. 0.]
[0. 0. 0. 1. 1. 0. 0.]
[0. 0. 0. 0. 0. 0. 1.]



Emotion	
[0. 0. 0. 1. 0. 0. 0.]	Neutral
[0. 0. 0. 0. 0. 0. 1.]	Sad
[1. 0. 0. 0. 0. 0. 0.]	Happy
[0. 0. 1. 0. 0. 0. 0.]	angry
[0. 1. 0. 0. 0. 0. 0.]	Surprise





5. 바이오 데이터 분석



바이오 데이터 - TEMP(체온)

24개

Emotion
[0. 0. 0. 1. 0. 0. 0.]
[1. 0. 0. 0. 0. 0. 0.]
[1. 0. 0. 1. 0. 0. 0.]
[0. 1. 0. 1. 0. 0. 0.]
[1. 1. 0. 0. 0. 0. 0.]
[0. 0. 1. 1. 0. 0. 0.]
[0. 1. 0. 0. 0. 0. 0.]
:
[0. 0. 0. 1. 0. 0. 1.]
[0. 0. 1. 0. 0. 0. 0.]
[0. 0. 0. 0. 1. 0. 0.]
[0. 0. 0. 1. 1. 0. 0.]
[0. 0. 0. 0. 0. 0. 1.]



Emotion	
[0. 0. 0. 1. 0. 0. 0.]	Neutral
[0. 0. 0. 0. 0. 0. 1.]	Sad
[1. 0. 0. 0. 0. 0. 0.]	Happy
[0. 0. 1. 0. 0. 0. 0.]	angry
[0. 1. 0. 0. 0. 0. 0.]	Surprise



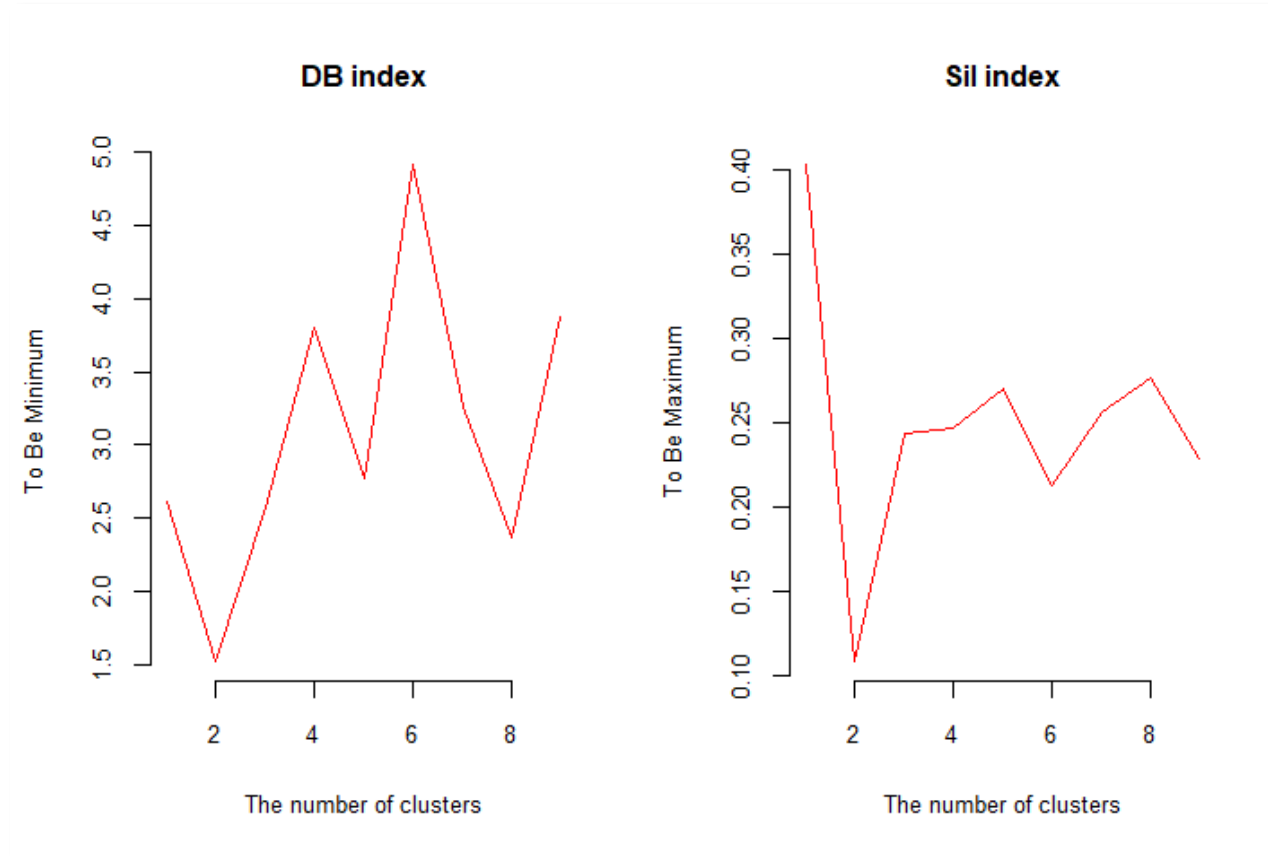


5. 바이오 데이터 분석



바이오 데이터 - DTW

극댓값을
최적의 클러스터로 인식



극소값을
최적의 클러스터로 인식

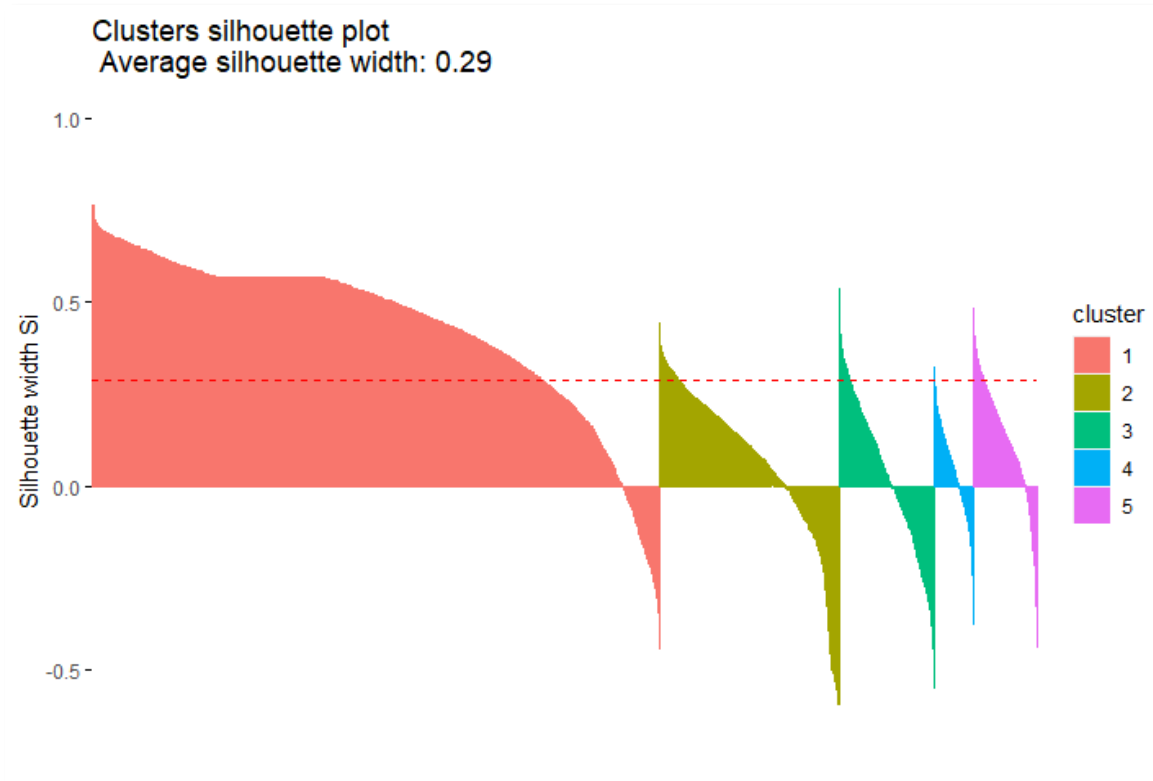




5. 바이오 데이터 분석



바이오 데이터 - Silhouette



색깔 너비
=
군집에 속한 데이터 수

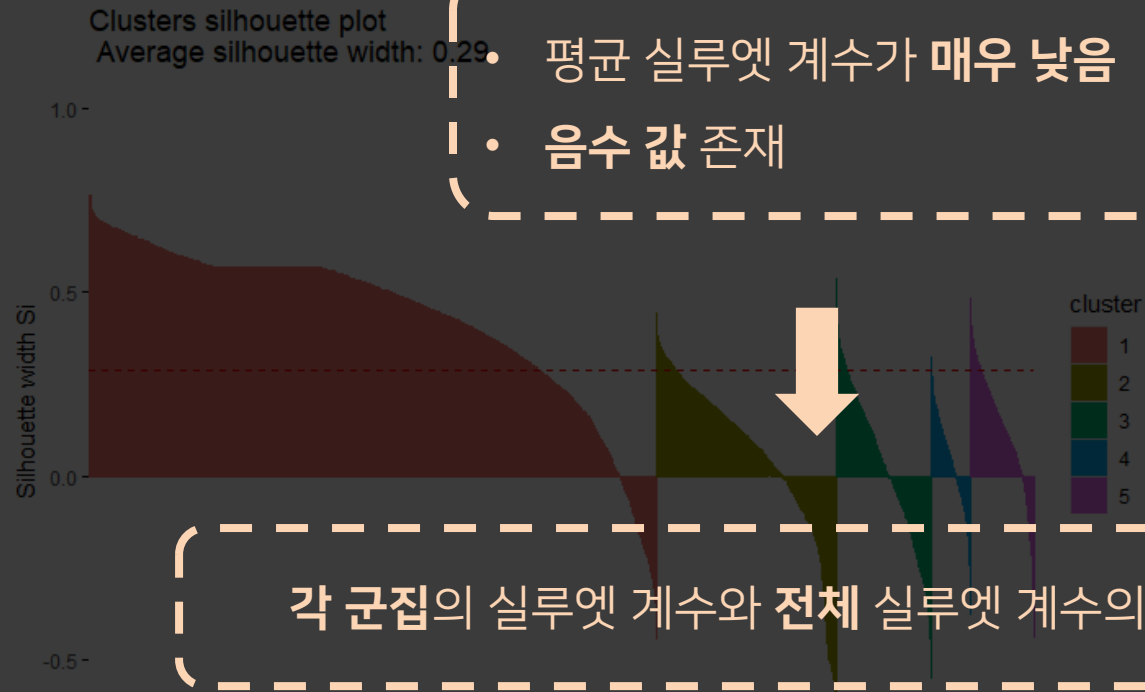




5. 바이오 데이터 분석



바이오 데이터 - Silhouette



색깔 너비

=

군집에 속한 데이터 수

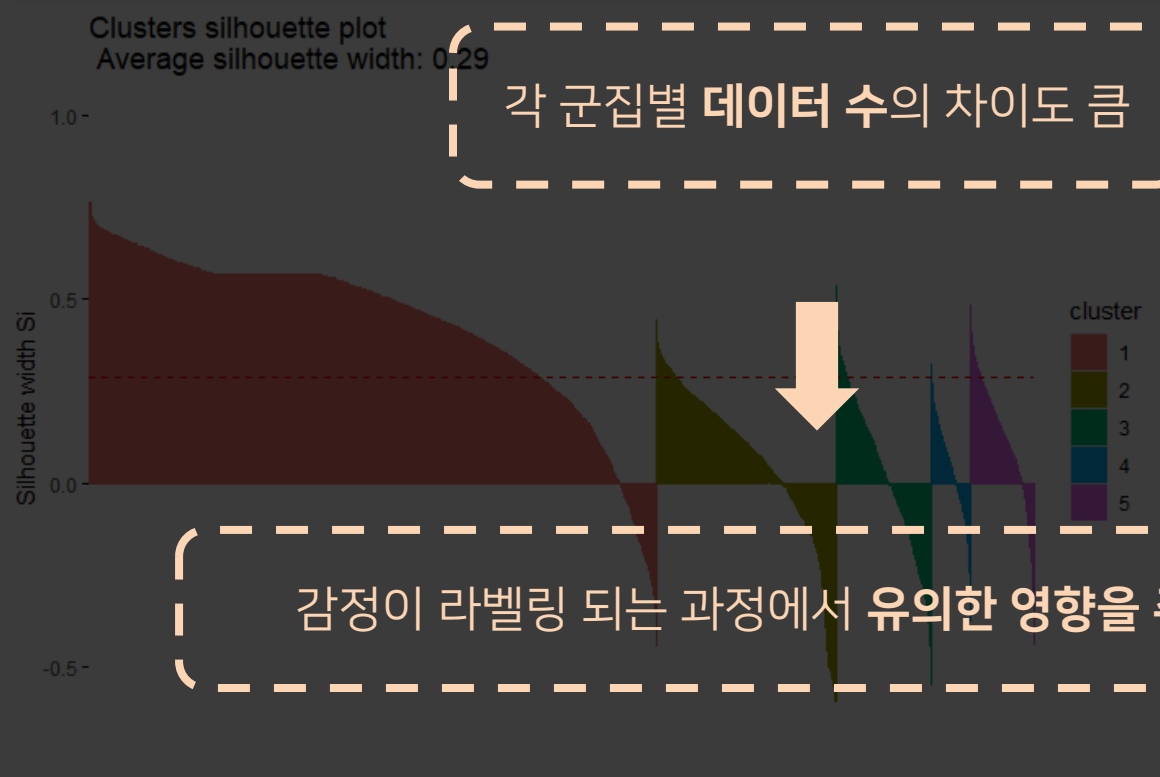




5. 바이오 데이터 분석



바이오 데이터 - Silhouette



색깔 너비

=

군집에 속한 데이터 수

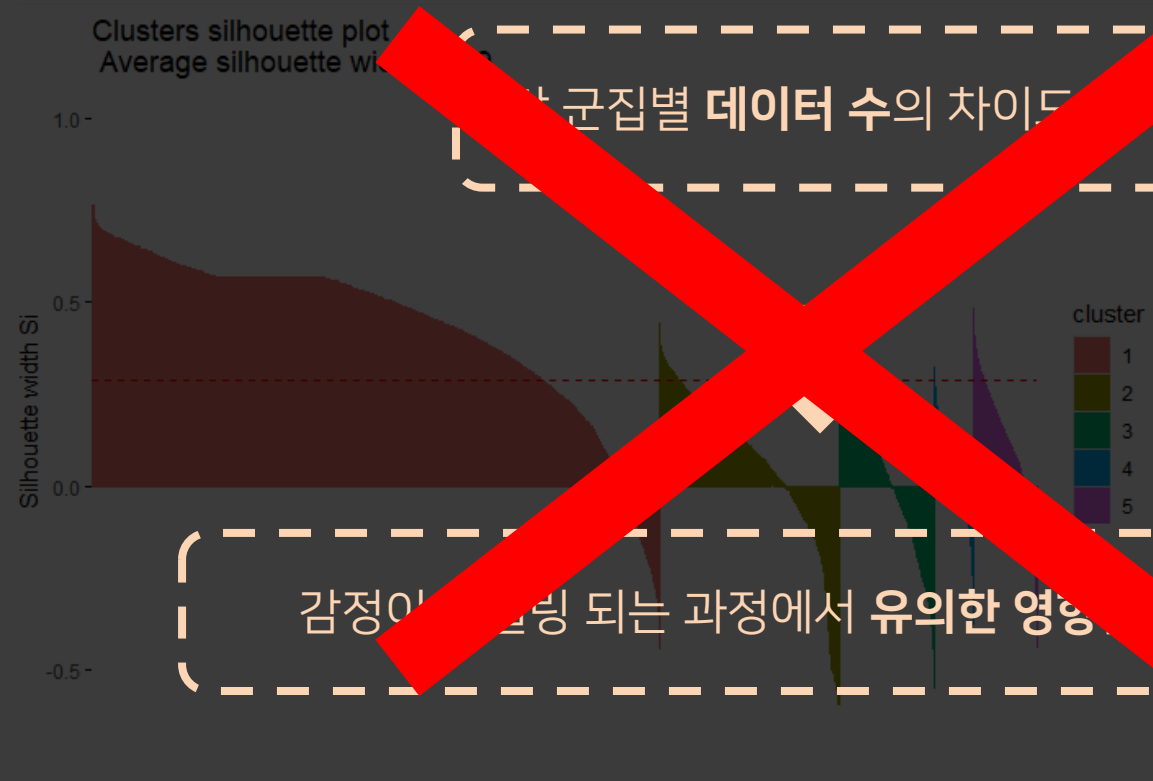




5. 바이오 데이터 분석



바이오 데이터 - Silhouette



색깔 너비
=
군집에 속한 데이터 수





5. 바이오 데이터 분석



최종 데이터

Audio	Text	Emotion	Arousal	Valence	Segment
[[[-46.271706 - 56.406227	n/ 아 친구들도? l/	[0.0.0.1.0.0.0.]	2.8	3.8	Sess01_script01_User001F_001
[[[-54.416702 ... -75...	l/ 나는 생일?	[0.0.0.1.0.0.0.]	3.1	3.3	Sess01_script01_User001F_002
[[[-5.5659237e+01 - 5.0546646e+0...	생일날이면은 b/ 내가 고기를 되...	[0.0.0.1.0.0.0.]	3.1	3.3	Sess01_script01_User001F_003
[[[-63.03308 - 58.875835 - 57.511215	b/ 선물이라 이 보통 돈으로 주...	[1.0.0.0.0.0.0.]	3.1	3.6	Sess01_script01_User001F_004
:					
[[[-5.3404884e+01 - 5.4611393e+0...	u/ 그걸 점점 발전을 시키잖아, 사람처럼 느껴지게.	[0.0.0.1.0.0.0.]	3.7	2.4	Sess40_script06_User080F_024
[[[-5.7383400e+01 -5.5686134e+0...	뭐 진짜 막 어떤 욕심이 가득 찬 사람은 내가...	[0.0.0.1.0.0.0.]	3.5	2.7	Sess40_script06_User080F_025

바이오 데이터(TEMP, IBI, EDA) 제거





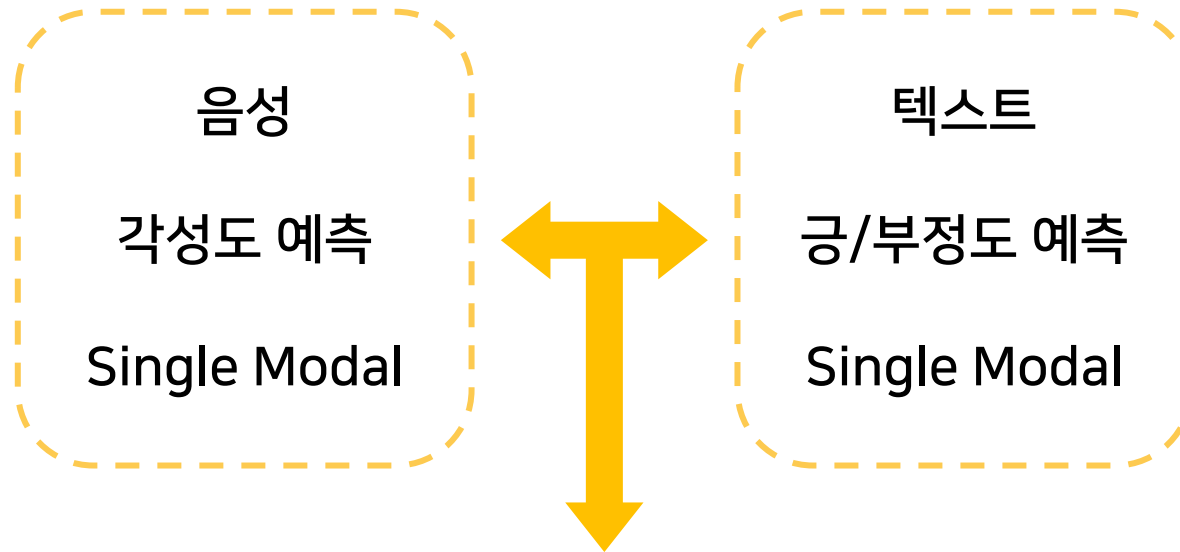
6. 모델 설명



6. 모델 설명



모델 흐름



각각 사전학습 후 Multi-Modal 병합 → 감정 예측 레이어 추가 학습

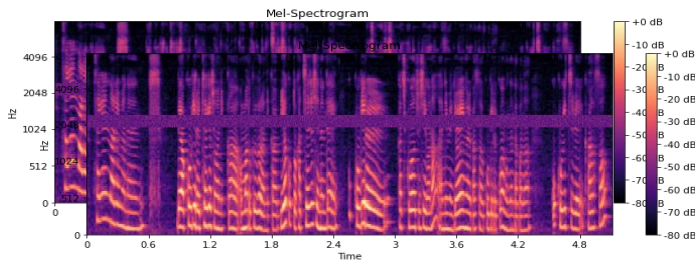




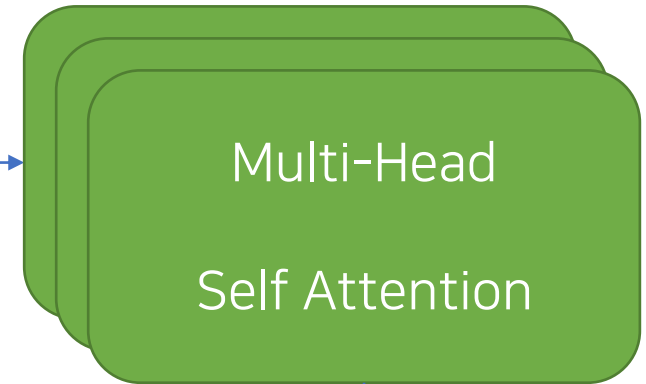
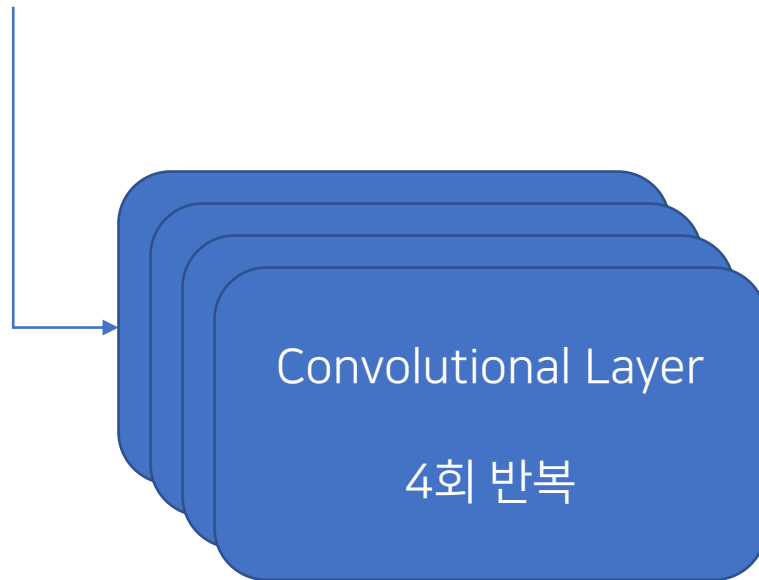
6. 모델 설명



음성



`torch.Size([2, 512, 128])`

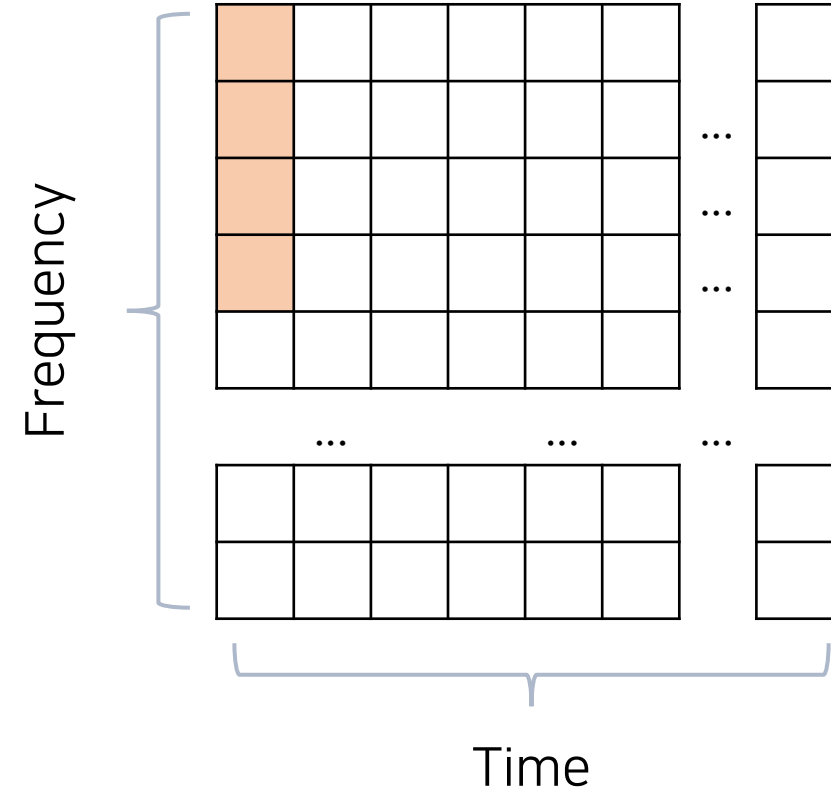
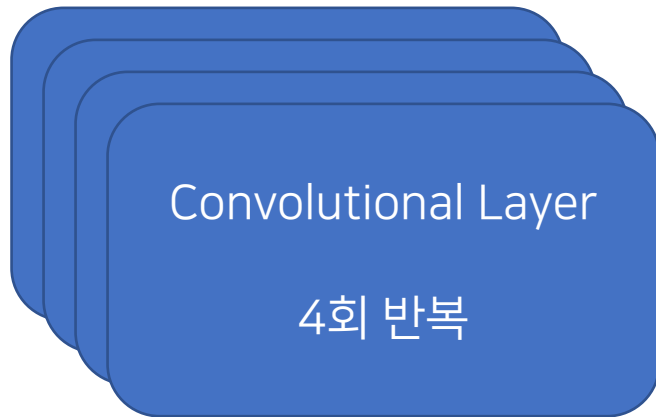




6. 모델 설명



음성 - Feature Extraction



Frequency 방향으로 Kernel 조정 = Frequency 간의 관계 추출 + 시간대별 정보 보존





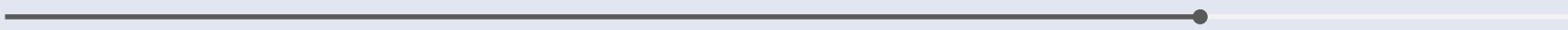
6. 모델 설명



음성 - Masking



- 실제 데이터와 Padding 구간을 구분해주는 Boolean Tensor
- 데이터와 곱하여 Padding 구간은 학습에서 제외
- Pad token에 Attention이 부여되지 않도록 Pad-masking 진행

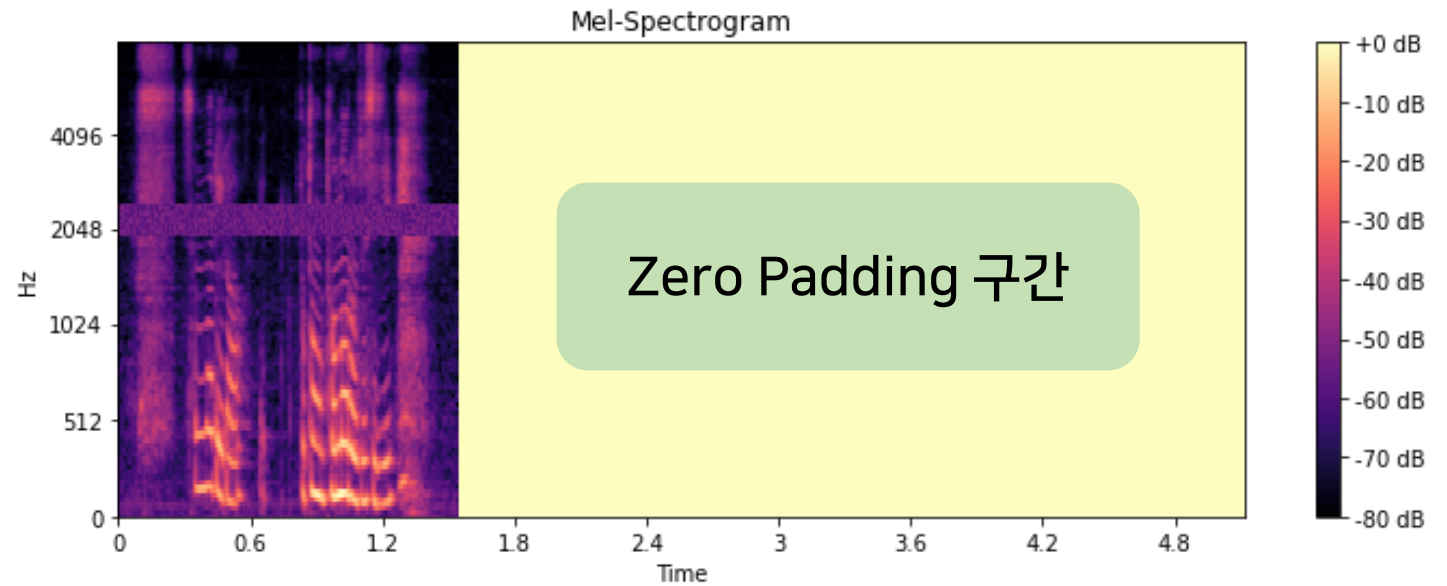
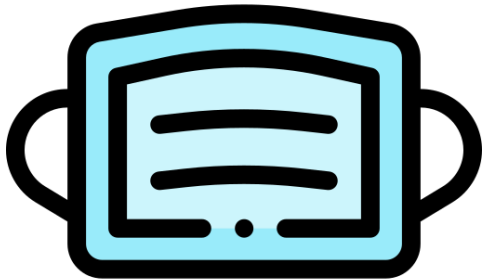




6. 모델 설명



음성 - Masking



Zero Padding 구간은 Attention 학습에서 배제

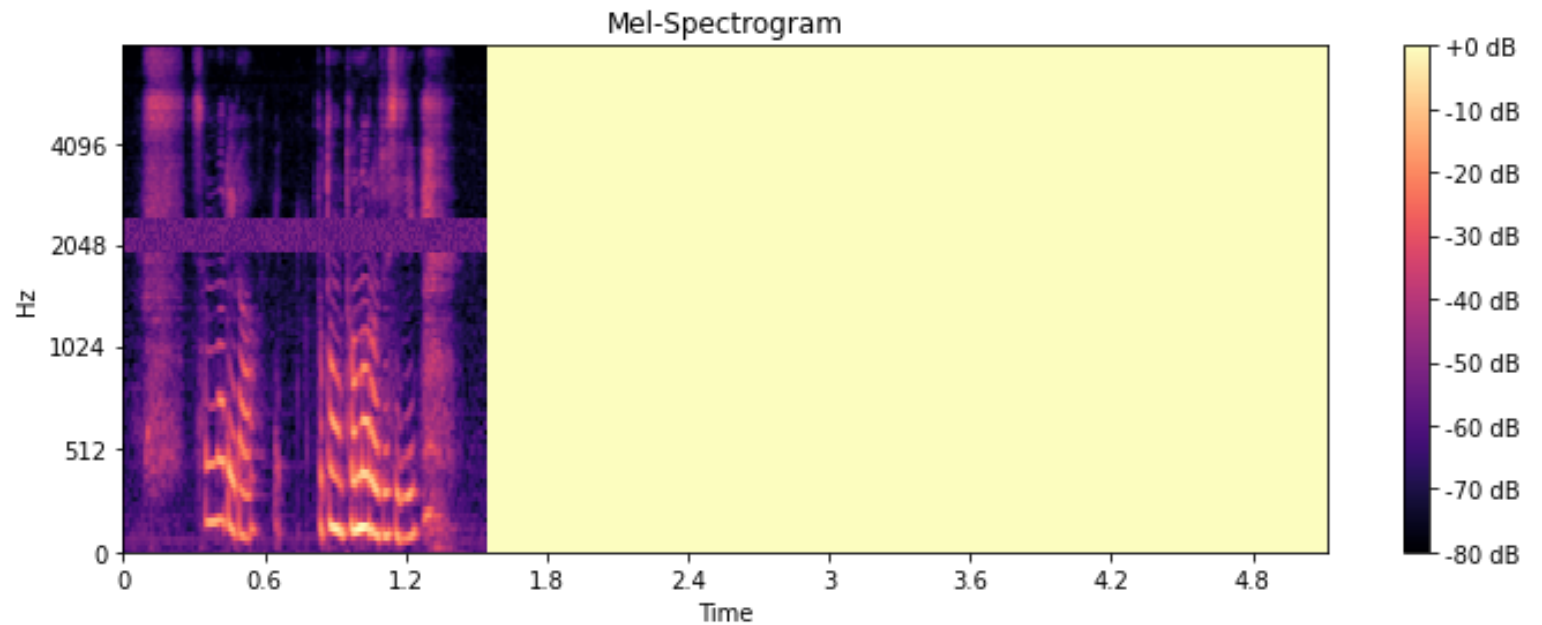
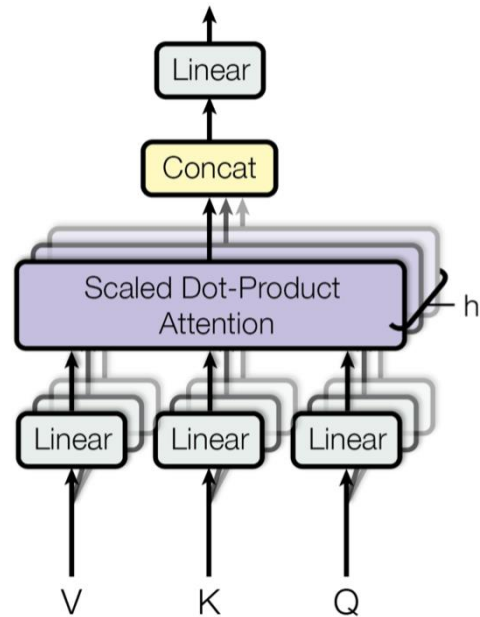




6. 모델 설명



음성 - Multi Head Attention



각성도를 예측하기 위해 필요한 구간에 대한 Attention값 계산

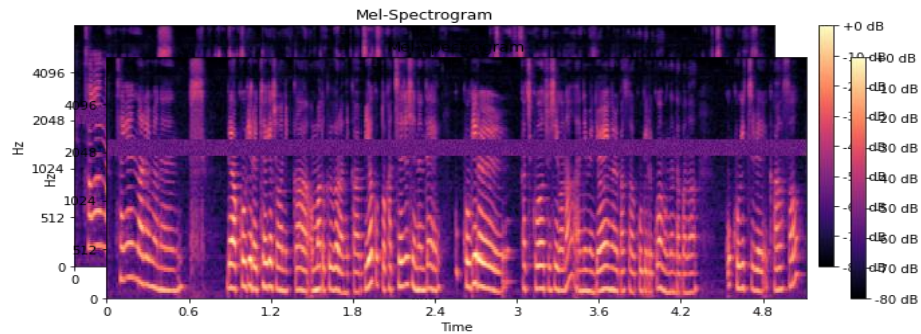




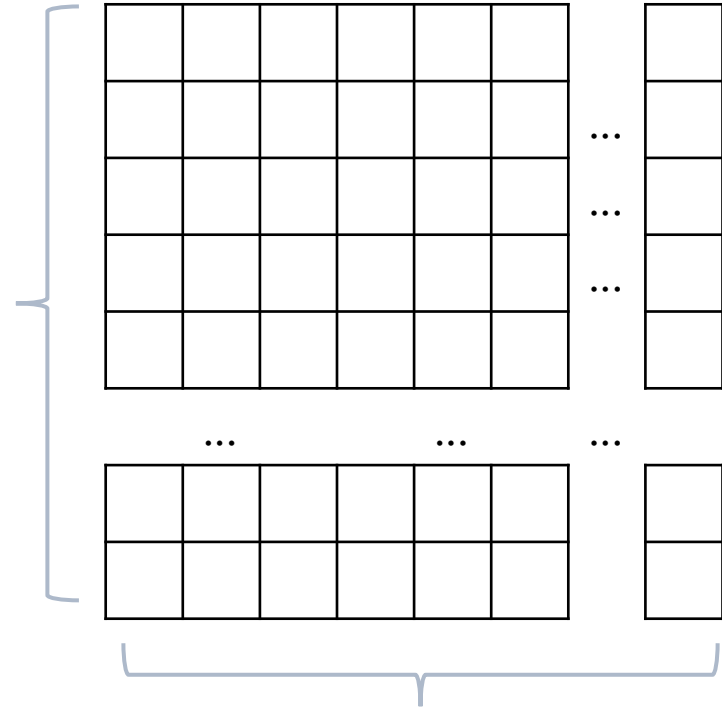
6. 모델 설명



음성 - Multi Head Attention



768(Frequency 간 관계)



512(Time)

Mel Spectrogram

128차원 → 768차원

Self Attention

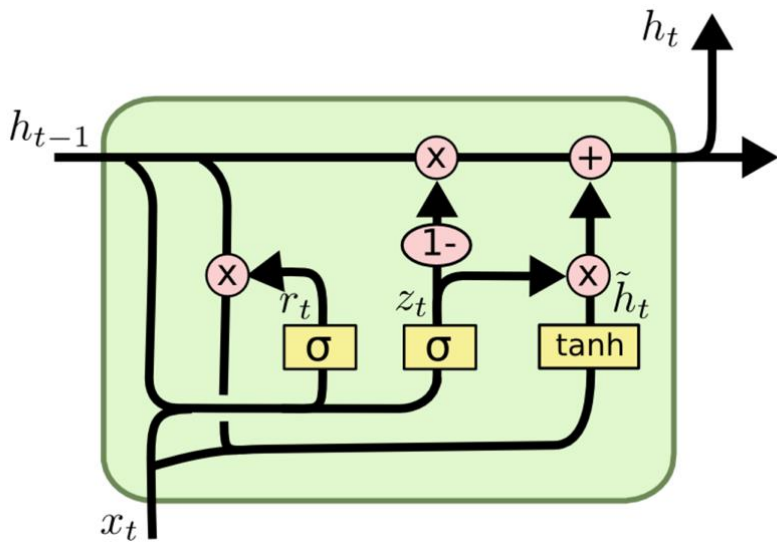




6. 모델 설명



음성 - Arousal 예측



- LSTM의 장기 의존성 문제에 대한 해결책을 유지하며 은닉 상태를 업데이트 하는 계산을 줄임
- 업데이트 게이트와 리셋 게이트로 구성
- LSTM보다 학습 속도가 빠르다고 알려져있음





6. 모델 설명



텍스트 - Valence 예측

_너	는	_내년	_대선	_때	_투표	할	_수	_있어	?
----	---	-----	-----	----	-----	---	----	-----	---



BERT용 input으로 인코딩

64

21	34	2456	3894	...	1	1	1	1	1
0	0	0	0	...	0	0	0	1	1
0	0	0	0	...	0	0	0	0	0

Input_ids

Attention_mask

Token_type_ids



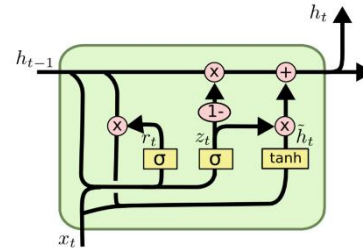


6. 모델 설명



텍스트 - Valence 예측

21	34	2456	3894	...	1	1	1	1	1
0	0	0	0	...	0	0	0	1	1
0	0	0	0	...	0	0	0	0	0



최종 임베딩 벡터 (64, 768)

GRU 모델로 예측

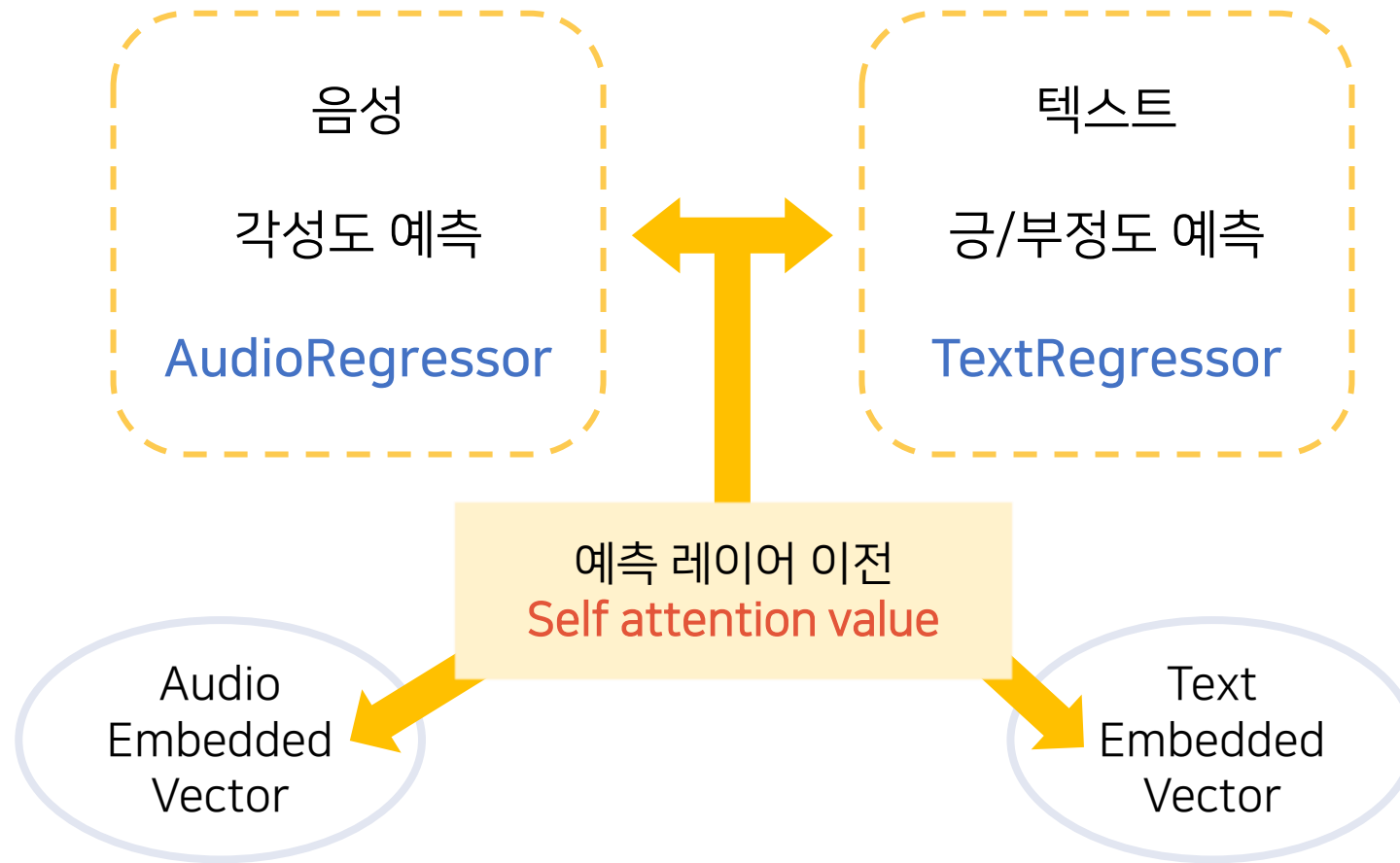




6. 모델 설명



Multimodal Emotion Classifier

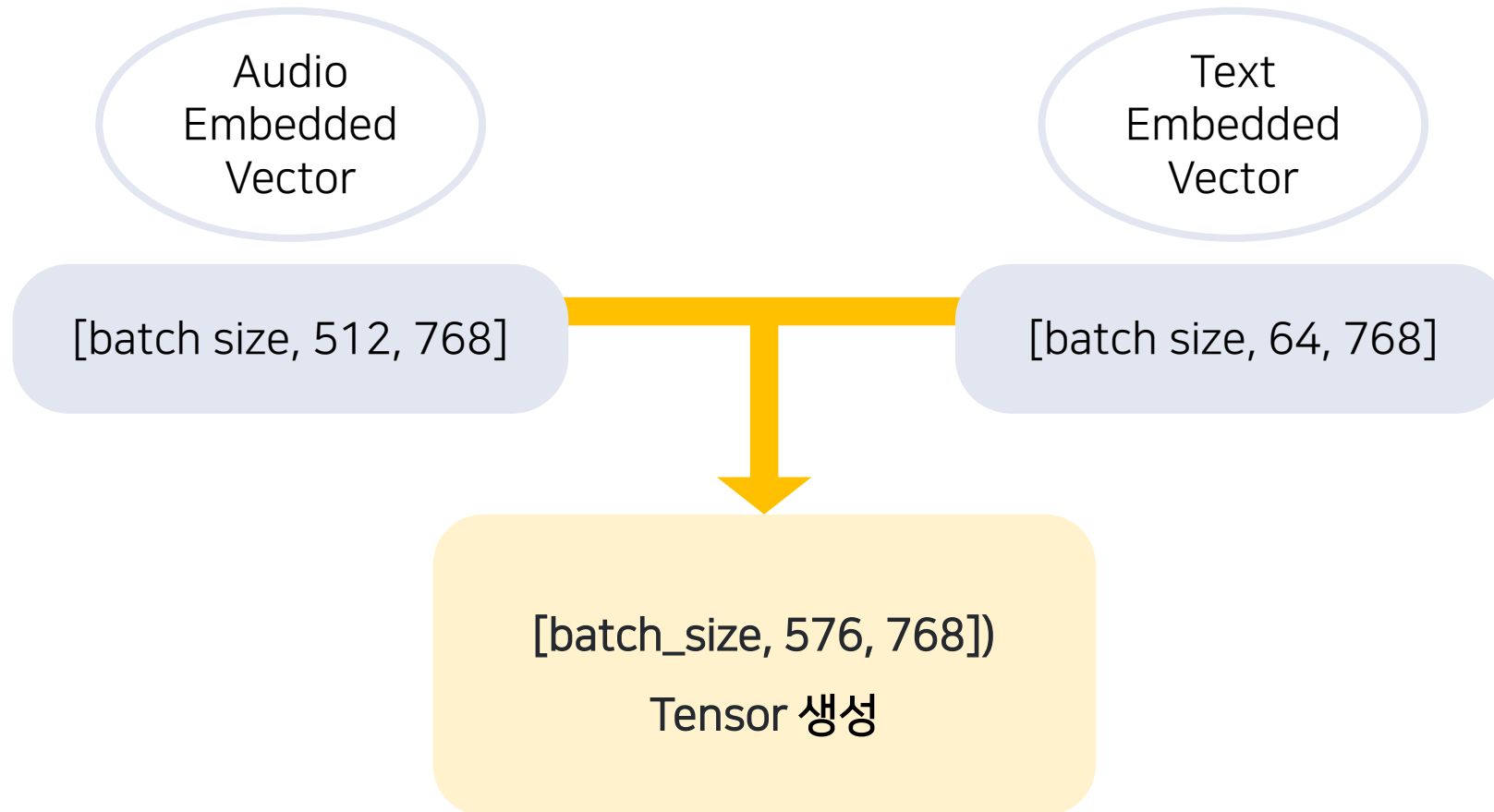




6. 모델 설명



Multimodal Emotion Classifier

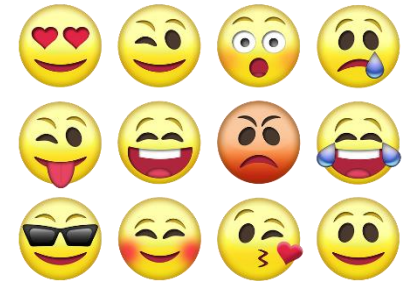
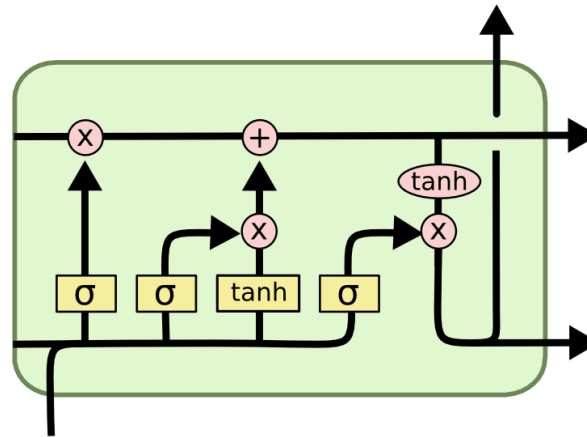
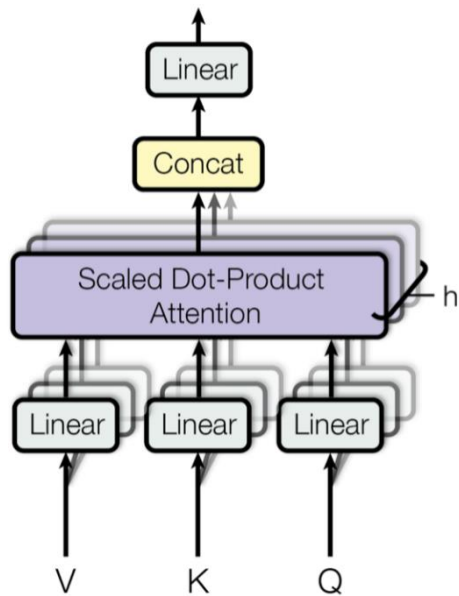




6. 모델 설명



Multimodal Emotion Classifier



Multihead Attention
Self attention 학습

Attention Value

LSTM

감정분류





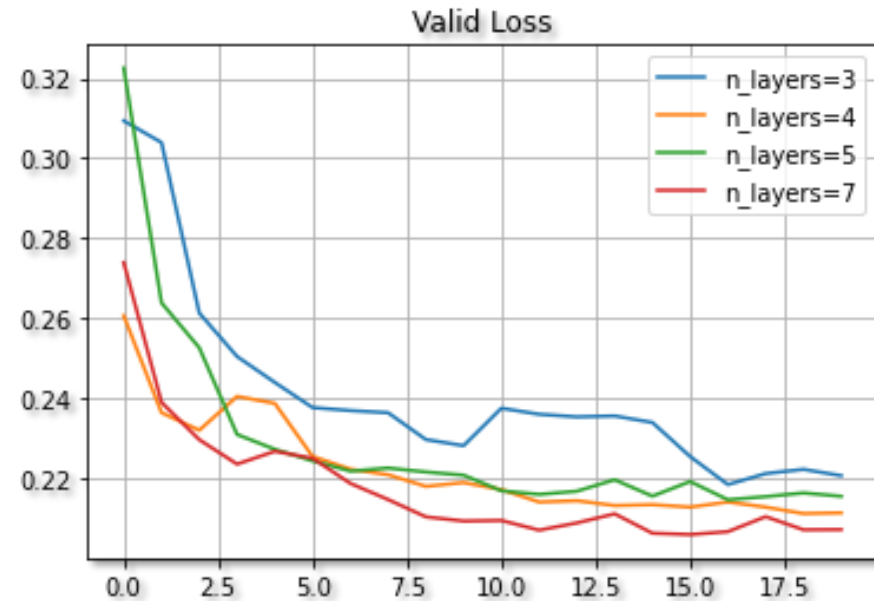
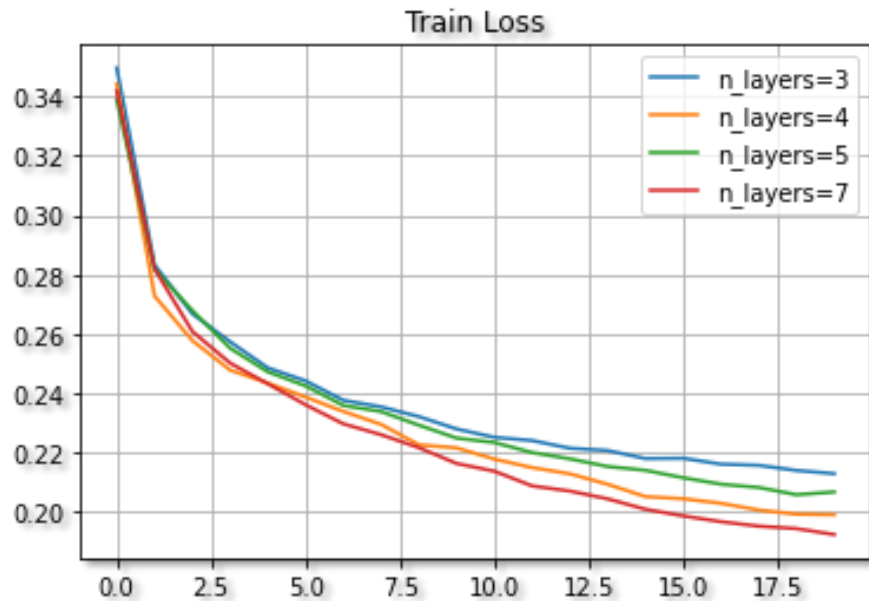
7. 학습 결과



7. 학습 결과



음성 싱글 모달리티- Arousal 예측 결과



Multihead Attention Layer의 반복 횟수에 따른 RMSE변화 → 7회일 때 최저



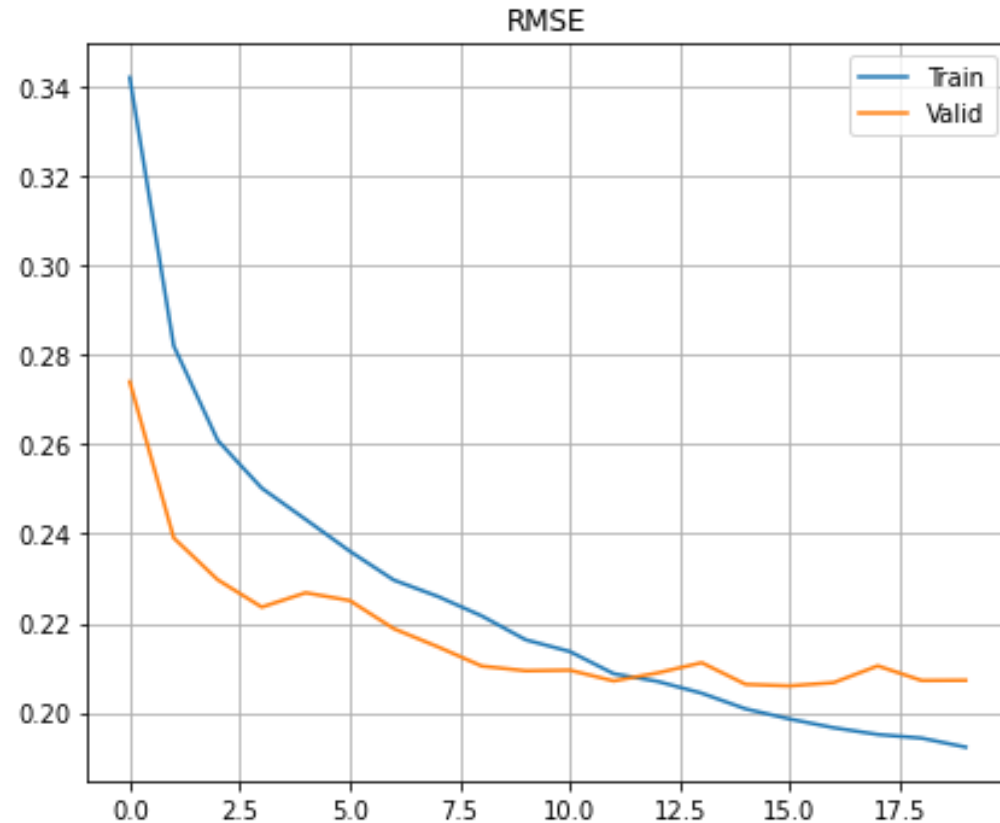


7. 학습 결과



음성 싱글 모달리티- Arousal 예측 결과(n_layer=7)

Train	Valid
0.341884649	0.273954606
0.281943787	0.239045343
0.260837908	0.229711024
0.250174330	0.223605140
.	.
.	.
.	.
0.196662220	0.206742686
0.195157791	0.210485889
0.194336303	0.207206603
0.192330844	0.206007988

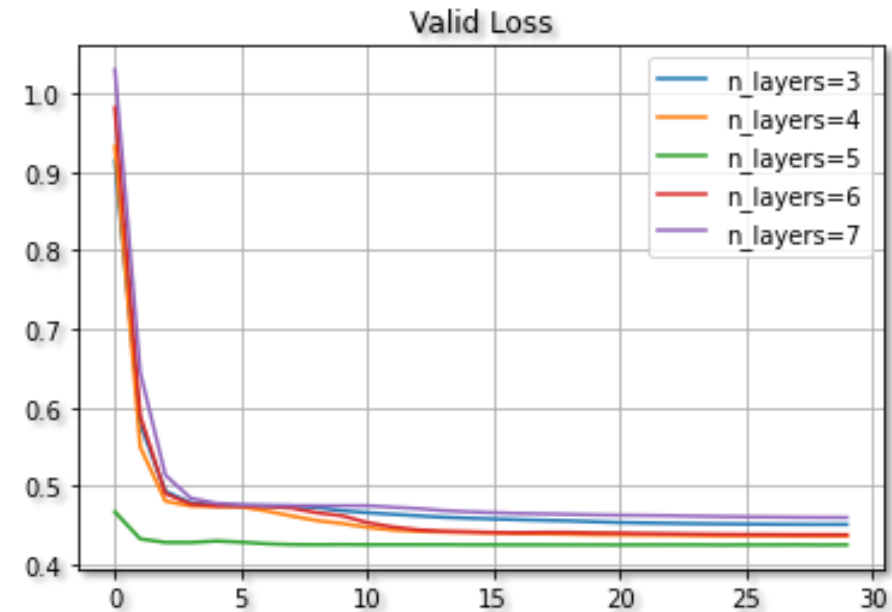
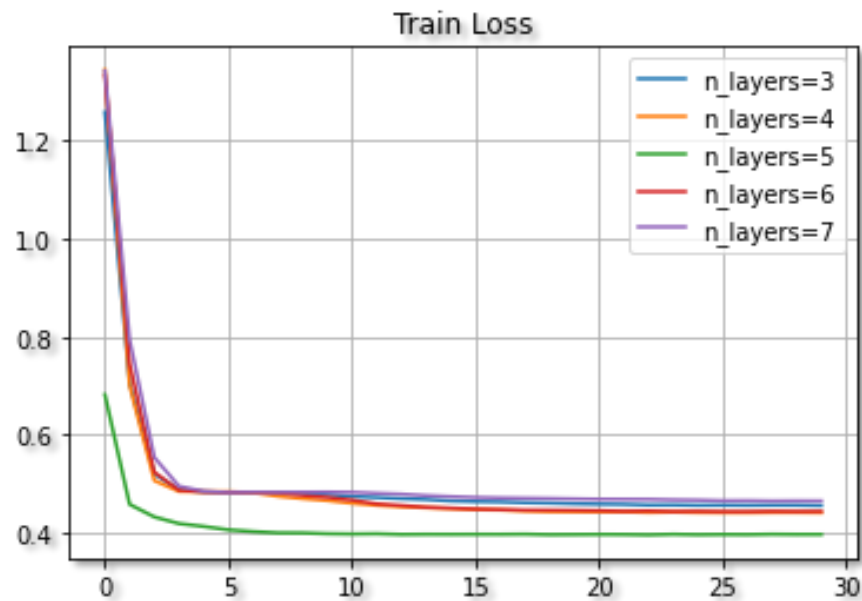




7. 학습 결과



텍스트 싱글 모달리티- Valence 예측 결과



Multihead Attention Layer의 반복 횟수에 따른 RMSE변화 → 5회일 때 최저



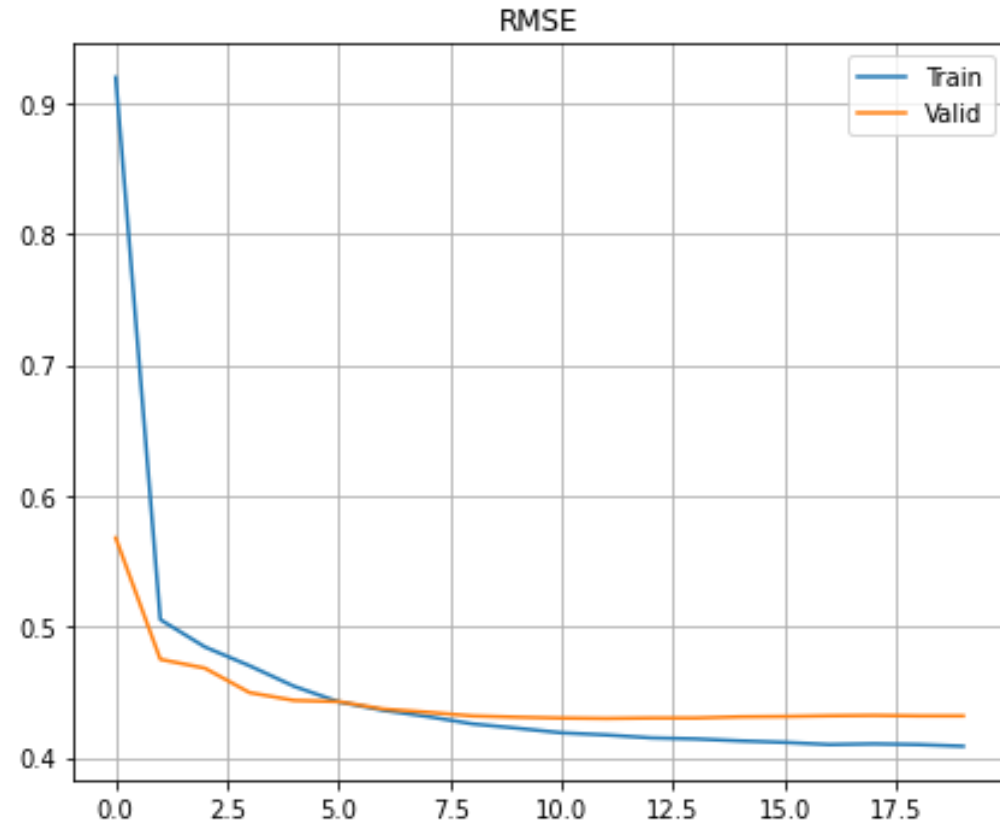


7. 학습 결과



텍스트 싱글 모달리티- Valence 예측 결과(n_layer=5)

Train	Valid
0.91981	0.475089
0.505456	0.468349
0.484568	0.449671
0.470203	0.443702
.	.
.	.
.	.
0.409979	0.431955
0.410524	0.432256
0.409952	0.431876
0.408682	0.431883

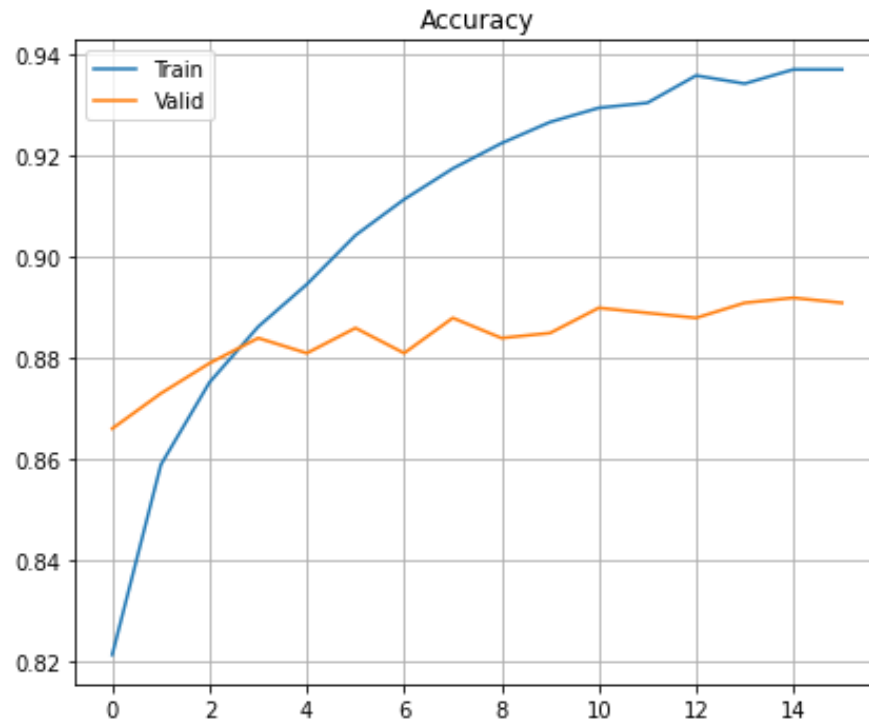




7. 학습 결과



멀티 모달 - 정확도(Accuracy) 평가



정확도가 증가하는 것을
확인할 수 있음

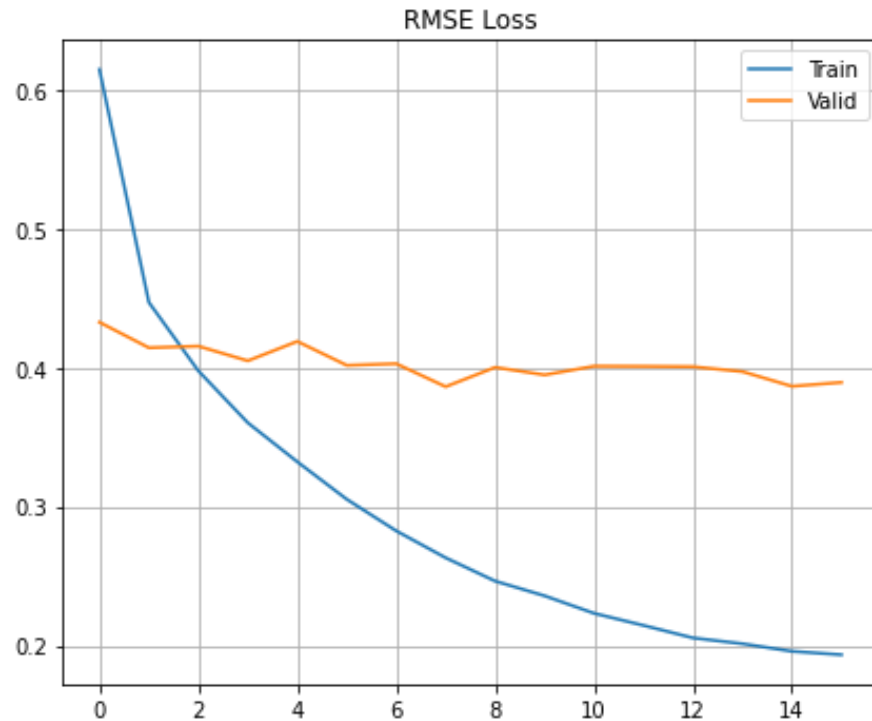




7. 학습 결과



멀티 모달 - RMSE Loss 평가



Loss가 감소하는 것을
확인할 수 있음





7. 학습 결과



멀티 모달 - F1 Score

?? F1 Score란?

- : 주로 분류 클래스 간 데이터가 심각한 불균형을 이루는 경우 사용
- 정밀도와 재현율의 평균

$$F_1 = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

정밀도(precision) : 예측한 것 중에 정답의 비율은?

재현율(recall) : 찾아야 할 것 중에 실제로 찾은 비율은?

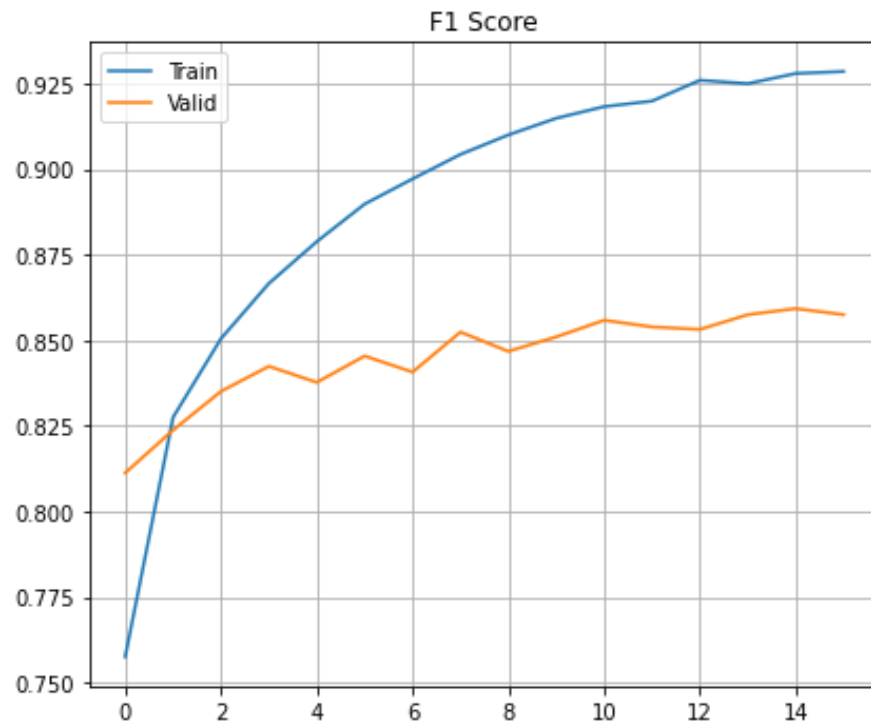




7. 학습 결과



멀티 모달 - F1 Score



증가 폭이 작지만
F1 Score의 값 자체가 높음

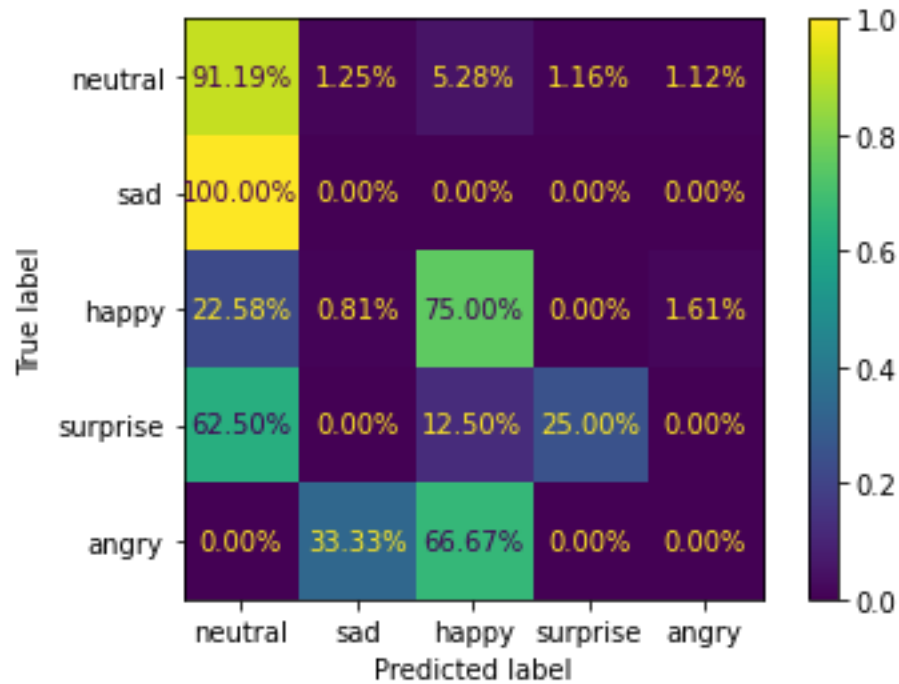




7. 학습 결과



멀티 모달 - Confusion matrix



- F1 Score : 87.54%
- 음성 n_layers = 6(2nd)
- 텍스트 n_layers = 6(2nd)

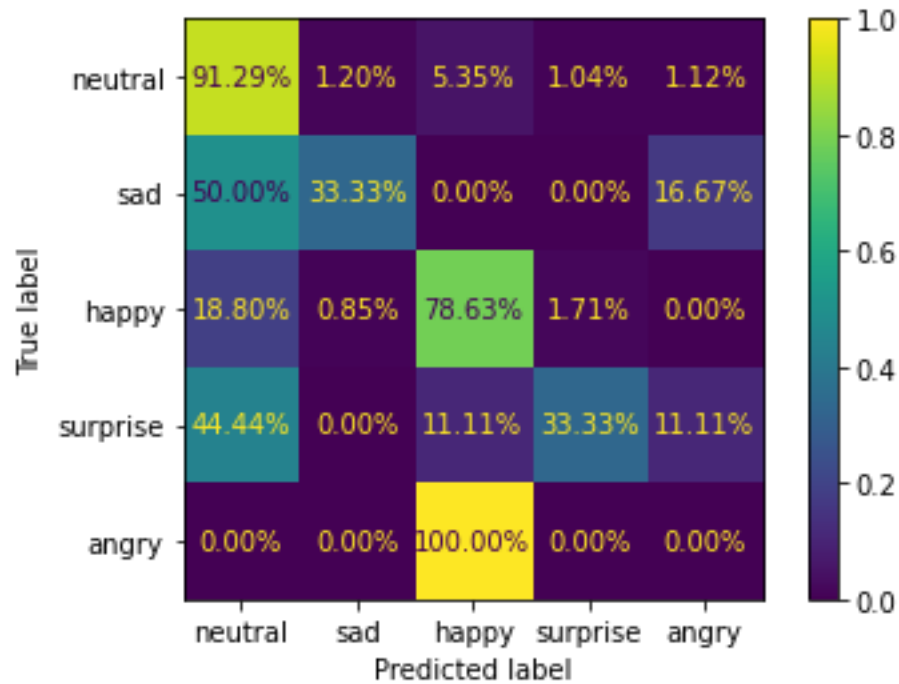




7. 학습 결과



멀티 모달 - Confusion matrix



- F1 Score : 87.60%
- 음성 n_layers = 7(1st)
- 텍스트 n_layers = 6(2nd)

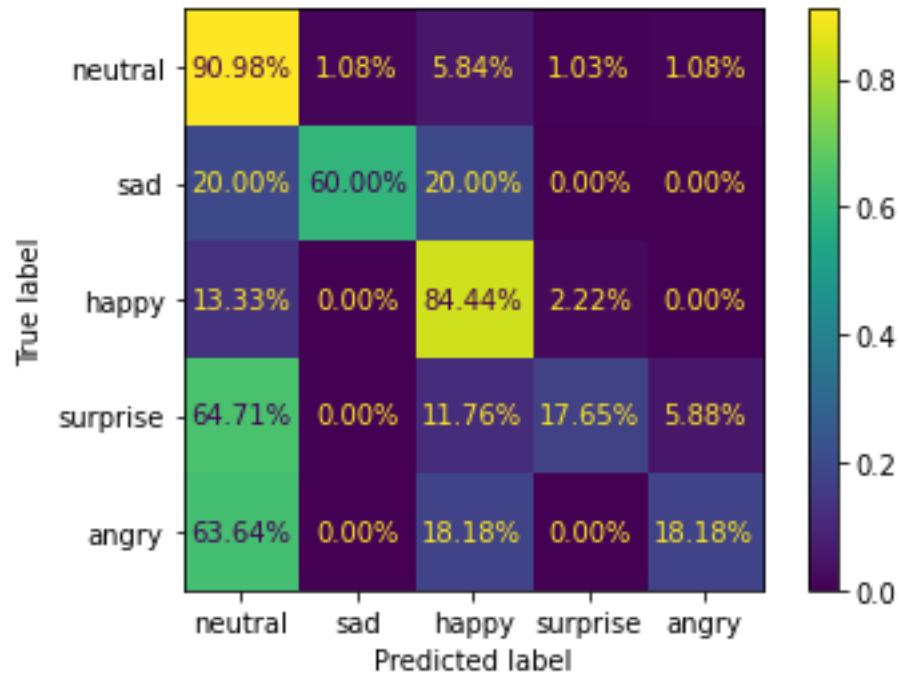




7. 학습 결과



멀티 모달 - Confusion matrix



- F1 Score : 86.93%
- 음성 n_layers = 7(1st)
- 텍스트 n_layers = 5(1st)

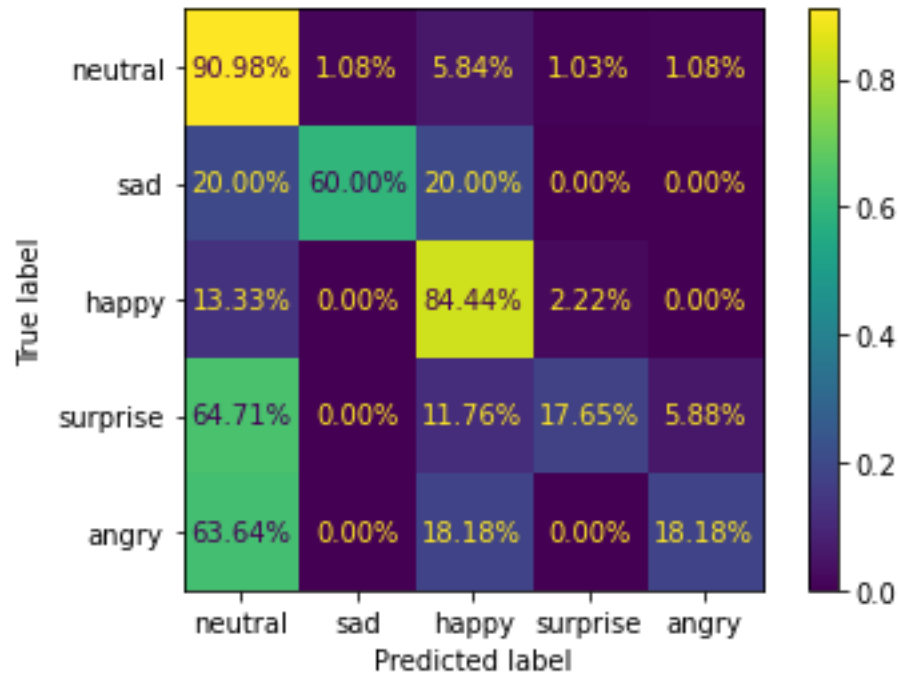




7. 학습 결과

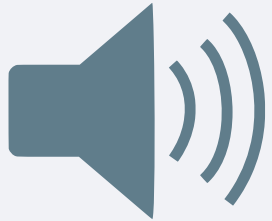


멀티 모달 - Confusion matrix



F1 Score 값은 비교적 낮지만
유일하게 데이터 불균형에
상대적으로 강건함





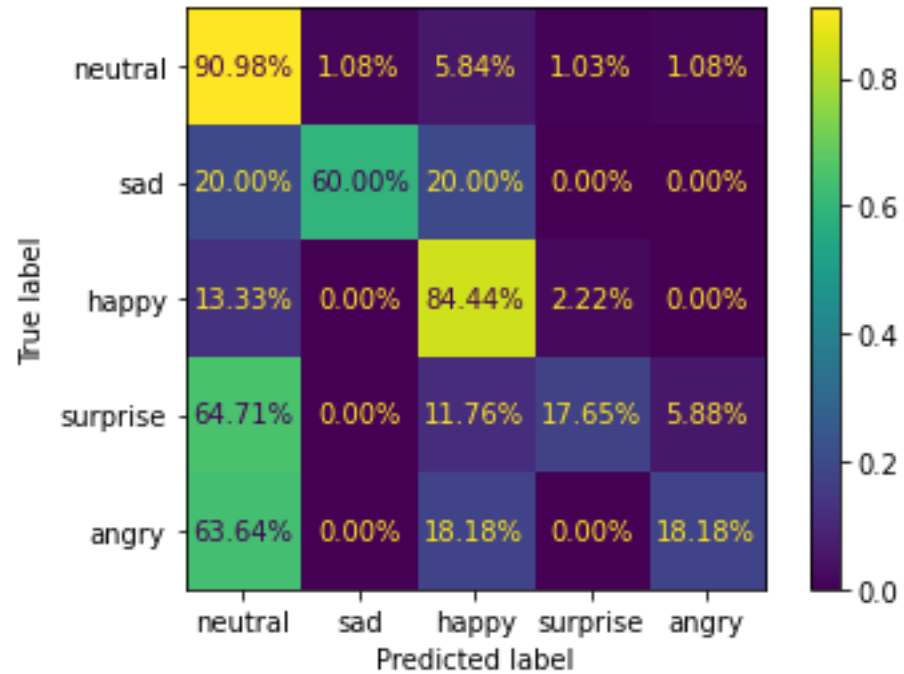
8. 성과 & 한계



8. 성과 & 한계



성과



불균형한 데이터임에도
불구하고 데이터 수가
적은 라벨들도 예측 성공

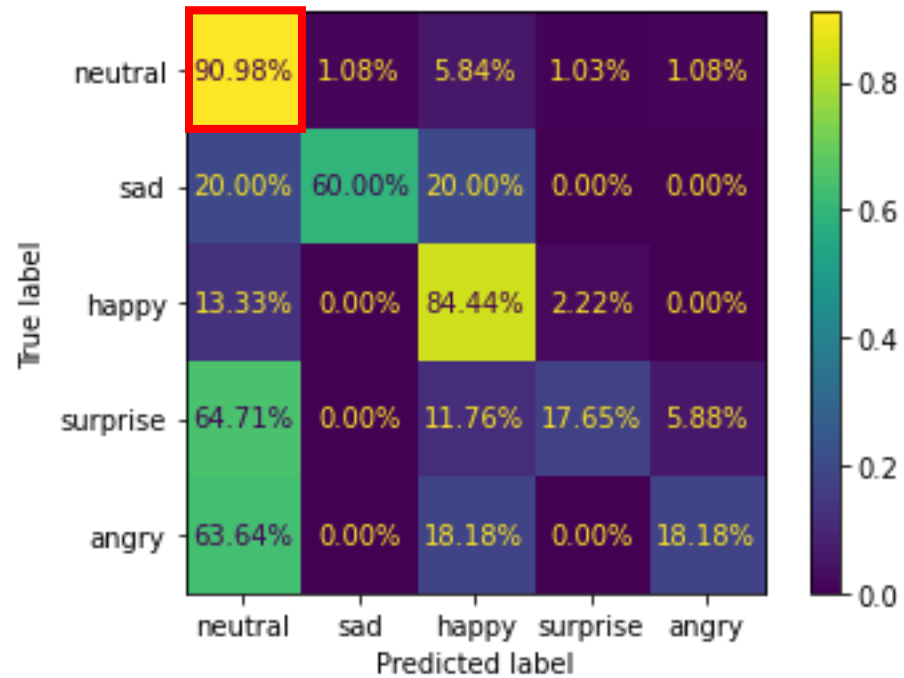




8. 성과 & 한계



성과



데이터 수가 가장 많은
neutral 데이터의 경우
90.98%의 적중률을 보임

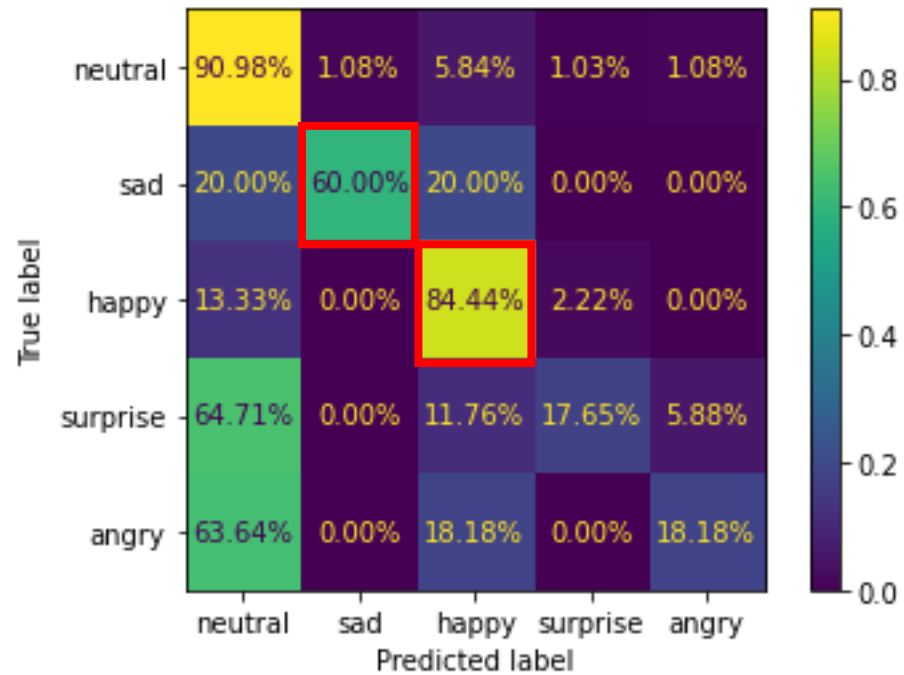




8. 성과 & 한계



성과



데이터의 수가 그나마 많았던
happy는 84.44%,
sad는 60%로
꽤 높은 **적중률**을 보임

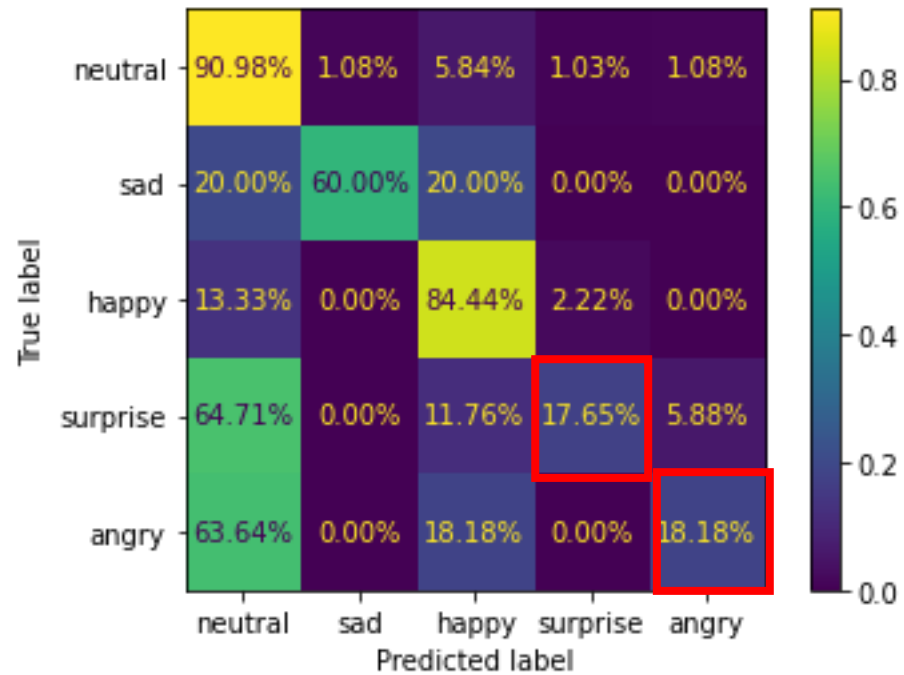




8. 성과 & 한계



한계



하지만 데이터 증강을
했음에도 불구하고
Surprise와 angry는
낮은 적중률을 보임

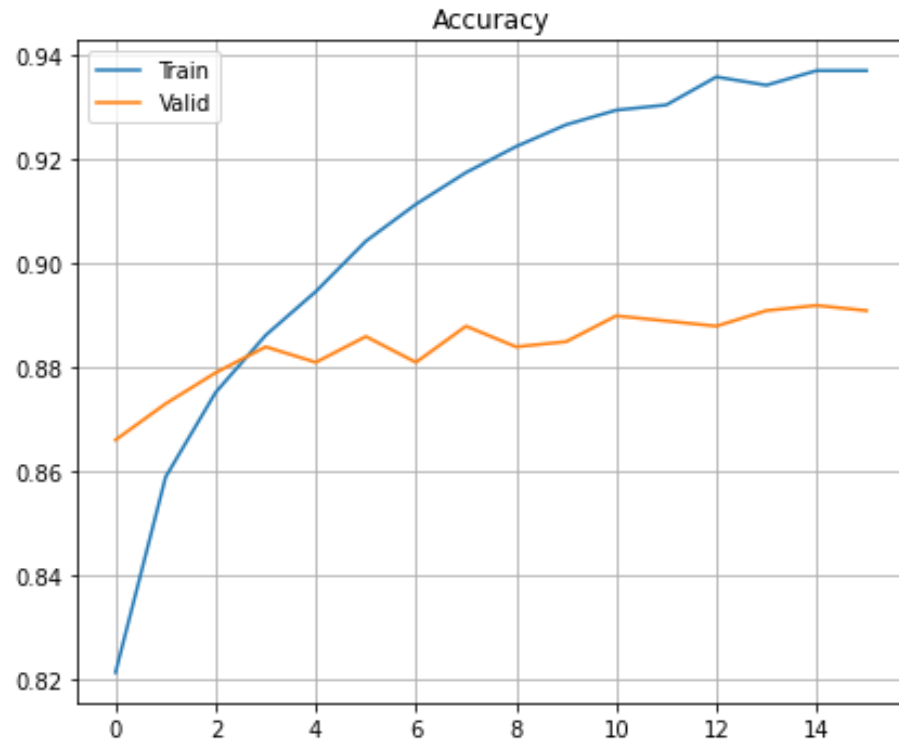




8. 성과 & 한계



의의



실제로 데이터의 수가
많이 불균형한 경우
예측 결과가 높은 accuracy를 보여도
데이터의 수가 많은 라벨에서만
적중하고
나머지 라벨은
예측하지 못하는 경우가 많음

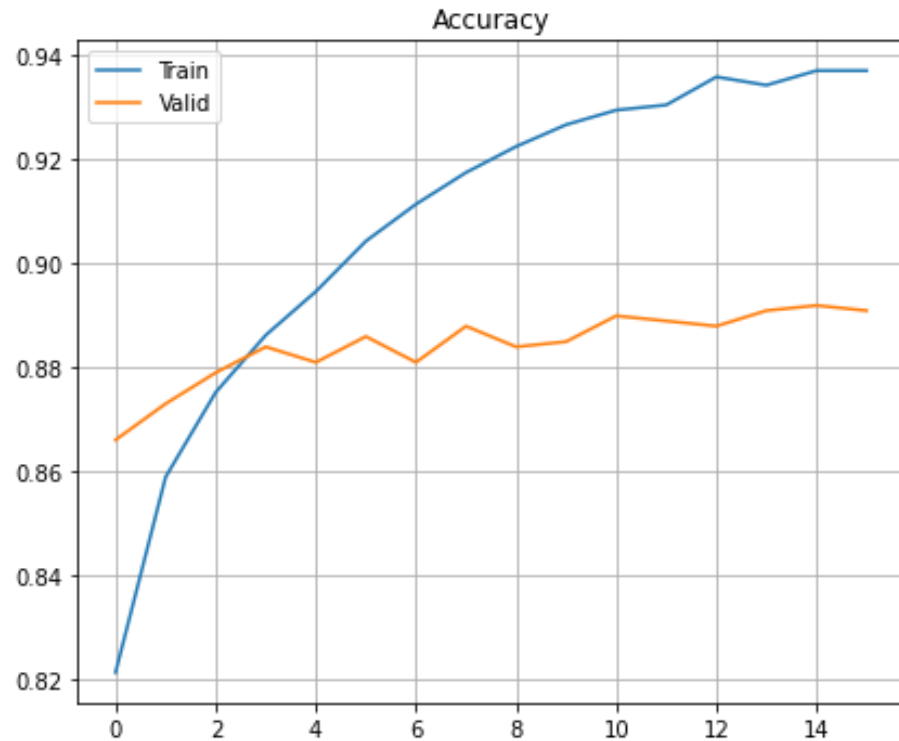




8. 성과 & 한계



의의



멀티모달 기법과 데이터 증강을 이용하여
데이터의 수가 적은 라벨들도
예측을 해냈다는 것이 중요 포인트!!





감사합니다.